Centre de Recherche en Économie et Statistique
CREST

Série des Documents de Travail

n° 2013-51

Asymptotic Inference in Multiple-Threshold Nonlinear Time Series Models

D. LI[1] – S. LING[2]
J.-M. ZAKOIAN[3]

December 2013

1 Tsinghua University, Mathematical Sciences Center, Beijing, China E-mail: li20072010@gmail.com
2 Hong Kong University of Science and Technology. E-mail: maling@usthk
3 CREST, 15 Boulevard Gabriel Péri, 92245 Malakoff Cedex, France. E-mail: zakoian@ensae.fr

Asymptotic inference in multiple-threshold nonlinear time series models *

By Dong Li[†], Shiqing Ling[‡] and Jean-Michel Zakoïan[§]

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 U.S. real GNP data over the period 1947–2013.

Keywords: Compound Poisson process; Ergodicity; Quasi-maximum likelihood estimation; Strict stationarity; MTDAR model, Score test.

JEL Classification: C13 and C22

*The first two authors thank the Hong Kong Research Grants Commission (Grant #HKUST6022/05P and #HKUST4765/03H to SQL) for partial support.

[†]Tsinghua University, Mathematical Sciences Center, Beijing, China E-mail: li20072010@gmail.com

[‡]Hong Kong University of Science and Technology. E-mail: maling@usthk

[§]CREST, 15 boulevard Gabriel Péri, 92245 Malakoff Cedex, France. E-mail: zakoian@ensae.fr

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 probability 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

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

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, $-\infty = r_0 < r_1 < \ldots < r_{m-1} < r_m = +\infty$, $I\{B\}$ 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 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, Ling and Li (2013). 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 *nonlinear*[1] 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 Zakoïan (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. Section 4 develops a score-based test to determine the number of thresholds in MTDAR models. Some special cases of MTDAR models are analyzed in Section 5. 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 First-Order MTDAR Model

In this section we focus on the MTDAR$(m;1)$ model

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

Without loss of generality, we assume 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 solution[2] 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

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

[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 $\{\eta_j : j \leq t\}$.

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*

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

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 $\mu_m < 0$. It can be noted 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 Several particular cases are worth considering.

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

$$\begin{aligned} &(0 < \phi_1 < 1,\ \ 0 < \phi_m < 1) \quad \text{or} \quad (\phi_1 < 0,\ \ \phi_m < 0,\ \ \phi_1\phi_m < 1) \\ \text{or} \quad &(-1 < \phi_1\phi_m < 0,\ \ \phi_1 < 1,\ \ \phi_m < 1), \end{aligned}$$

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 $|\phi_1| < 1$.

ii) 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

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

In particular, if $\alpha_1 = \alpha_m$, we retrieve the standard ARCH(1) condition: $\alpha_1 < \exp\{-E\log\eta_t^2\}$.

iii) 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 2.3 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, their condition is obtained under the assumption that $\sup_{x\in\mathbb{R}}\{(1+|x|)f(x)\} < \infty$ which we do not require.

Remark 2.4 In the proof, we show that, for some constants $u, K > 0$ and $\rho \in (0,1)$, we have

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

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

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

(see Meyn and Tweedie, 1996, Theorem 16.0.1.)

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

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

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

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

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

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: $(\alpha_1 + \alpha_m)\max(\alpha_1, \alpha_m) < 2$.

3 QML Estimation of MTDAR$(m; p)$ Models

We now turn to the estimation of the MDTAR$(m; p)$ model by the QML method, assuming that the orders m and p are known positive integers. The parameter, consisting of the AR and volatility coefficients and the thresholds, is denoted $\boldsymbol{\vartheta} = (\boldsymbol{\lambda}', \mathbf{r}')' \equiv (\boldsymbol{\phi}_1', \boldsymbol{\alpha}_1', \cdots, \boldsymbol{\phi}_m', \boldsymbol{\alpha}_m', \mathbf{r}')'$, with $\boldsymbol{\phi}_i = (c_i, \phi_{i1}, ..., \phi_{ip})'$ and $\boldsymbol{\alpha}_i = (\omega_i, \alpha_{i1}, ..., \alpha_{ip})'$, $i = 1, \cdots, m$, and $\mathbf{r} = (r_1, \cdots, r_{m-1})'$. The true parameter value is denoted $\boldsymbol{\vartheta}_0 = (\boldsymbol{\lambda}_0', \mathbf{r}_0')'$.

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

$$\begin{aligned} L_n(\boldsymbol{\vartheta}) &= \sum_{t=1}^n l_t(\boldsymbol{\vartheta}) \quad \text{with} \\ l_t(\boldsymbol{\vartheta}) &= -\frac{1}{2}\sum_{i=1}^m \left\{\log(\boldsymbol{\alpha}_i'\mathbf{X}_{t-1}) + \frac{(y_t - \boldsymbol{\phi}_i'\mathbf{Y}_{t-1})^2}{\boldsymbol{\alpha}_i'\mathbf{X}_{t-1}}\right\} I\{r_{i-1} < y_{t-d} \leq r_i\}, \end{aligned}$$

where $\mathbf{Y}_{t-1} = (1, y_{t-1}, ..., y_{t-p})'$, $\mathbf{X}_{t-1} = (1, y_{t-1}^2, ..., y_{t-p}^2)'$, $r_0 = -\infty$, $r_m = \infty$. 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 omited (see Berkes, Horváth et Kokoszka (2003), Francq and Zakoïan (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 $\Theta_r = \{\mathbf{r} = (r_1, \cdots, r_{m-1})' \in [-\Gamma, \Gamma]^{m-1} : \ r_{i+1} - r_i \geq \delta\}$ 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 $L_n(\boldsymbol{\vartheta})$ is denoted by $\hat{\boldsymbol{\vartheta}}_n$, i.e.,

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

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

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

Let $\{y_{(1)}, ..., y_{(n)}\}$ denote the order statistics of the sample $\{y_1, ..., y_n\}$. If $(y_{(j_1)}, ..., y_{(j_{m-1})})'$ is an estimate of $\mathbf{r}_0$ for some $j_1 < \cdots < j_{m-1}$, then $L_n^*(\mathbf{r})$ is a constant over the $(m-1)$-dimensional cube $\mathcal{A}$ defined by $\mathcal{A} = \{\mathbf{r} = (r_1, ..., r_{m-1})' : r_i \in [y_{(j_i)}, y_{(j_i+1)}), i = 1, ..., m-1\}$. 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 $\hat{\mathbf{r}}_n = (y_{(j_1)}, ..., y_{(j_{m-1})})'$ as a QMLE of $\mathbf{r}_0$.

The strong consistency of the QMLE $\hat{\boldsymbol{\vartheta}}_n$ of $\boldsymbol{\vartheta}_0$ relies on the following assumptions.

A1: The true value $\boldsymbol{\vartheta}_0$ belongs to Θ and $(\boldsymbol{\phi}_{i0}', \boldsymbol{\alpha}_{i0}')' \neq (\boldsymbol{\phi}_{i+1,0}', \boldsymbol{\alpha}_{i+1,0}')'$ for $i = 1, ..., m-1$.

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, non anticipative 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* $\hat{\boldsymbol{\vartheta}}_n \to \boldsymbol{\vartheta}_0$ *a.s..*

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

A4: $\boldsymbol{\vartheta}_0$ 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, that is, (2.3) is satisfied with $V(y) = 1 + |y|^u$ for some $u > 0$.

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

$$\{(\boldsymbol{\phi}_{i0} - \boldsymbol{\phi}_{i+1,0})'\mathbf{w}_i^*\}^2 + \{(\boldsymbol{\alpha}_{i0} - \boldsymbol{\alpha}_{i+1,0})'\mathbf{W}_i^*\}^2 > 0.$$

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

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

where $\mathbf{J} = \mathrm{diag}\{J_1, \cdots, J_m\}$ *and* $J_i = \boldsymbol{\Sigma}_i^{-1}\boldsymbol{\Omega}_i\boldsymbol{\Sigma}_i^{-1}$ *with*

$$\boldsymbol{\Omega}_i = E\begin{pmatrix} \frac{\mathbf{Y}_{t-1}\mathbf{Y}_{t-1}'I_{it}}{\boldsymbol{\alpha}_{i0}'\mathbf{X}_{t-1}} & \frac{\kappa_3}{2}\frac{\mathbf{Y}_{t-1}\mathbf{X}_{t-1}'I_{it}}{(\boldsymbol{\alpha}_{i0}'\mathbf{X}_{t-1})^{3/2}} \\ \frac{\kappa_3}{2}\frac{\mathbf{X}_{t-1}\mathbf{Y}_{t-1}'I_{it}}{(\boldsymbol{\alpha}_{i0}'\mathbf{X}_{t-1})^{3/2}} & \frac{\kappa_4-1}{4}\frac{\mathbf{X}_{t-1}\mathbf{X}_{t-1}'I_{it}}{(\boldsymbol{\alpha}_{i0}'\mathbf{X}_{t-1})^{2}} \end{pmatrix}, \ \boldsymbol{\Sigma}_i = E\left\{\mathrm{diag}\left(\frac{\mathbf{Y}_{t-1}\mathbf{Y}_{t-1}'I_{it}}{\boldsymbol{\alpha}_{i0}'\mathbf{X}_{t-1}}, \ \frac{1}{2}\frac{\mathbf{X}_{t-1}\mathbf{X}_{t-1}'I_{it}}{(\boldsymbol{\alpha}_{i0}'\mathbf{X}_{t-1})^2}\right)\right\},$$

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

From Theorem 3.2 (i), we know that the convergence rate of $\hat{\mathbf{r}}_n$ is n. To study the limiting distribution of $n(\hat{\mathbf{r}}_n - \mathbf{r}_0)$, we consider the following profile sum of squares errors function:

$$\widetilde{L}_n(\mathbf{s}) = -2\left\{L_n\left(\hat{\boldsymbol{\lambda}}_n(\mathbf{r}_0 + \frac{\mathbf{s}}{n}), \mathbf{r}_0 + \frac{\mathbf{s}}{n}\right) - L_n\left(\hat{\boldsymbol{\lambda}}_n(\mathbf{r}_0), \mathbf{r}_0\right)\right\}, \quad \mathbf{s} = (s_1, ..., s_{m-1})' \in \mathbb{R}^{m-1}. \quad (3.1)$$

By Theorem 3.2 and Taylor's expansion, it is not difficult to show that $\widetilde{L}_n(\mathbf{s})$ can be approximated in $\mathbb{D}(\mathbb{R}^{m-1})$, which is the function space consisting of uniform limits of sequences of simple functions defined on $\mathbb{R}^{m-1}$ that is equipped with the Skorokhod metric (see Li and Ling, 2012), by

$$\wp_n(\mathbf{s}) = -2\left\{L_n\left(\boldsymbol{\lambda}_0, \mathbf{r}_0 + \frac{\mathbf{s}}{n}\right) - L_n\left(\boldsymbol{\lambda}_0, \mathbf{r}_0\right)\right\} = \sum_{i=1}^{m-1}\wp_n^{(i)}(s_i),$$

where

$$\wp_n^{(i)}(s_i) = \sum_{t=1}^{n}\left[\xi_t^{(i,\,i+1)}I\{r_{i0} + \tfrac{s_i}{n} < y_{t-d} \le r_{i0}\}I\{s_i < 0\} + \xi_t^{(i+1,\,i)}I\{r_{i0} < y_{t-d} \le r_{i0} + \tfrac{s_i}{n}\}I\{s_i \ge 0\}\right],$$

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

Denote $F_{(i,\,j)}(\cdot|r_{k0})$ as the conditional distribution function of $\xi_{d+1}^{(i,\,j)}$ given $y_1 = r_{k0}$. Define $(m-1)$ independent 1-dimensional two-sided compound Poisson processes $\{\mathcal{P}_i(z),\ z \in \mathbb{R}\}$ as

$$\mathcal{P}_i(z) = I\{z < 0\} \sum_{k=1}^{N_1^{(i)}(|z|)} Y_k^{(i,\,i+1)} + I\{z \geq 0\} \sum_{k=1}^{N_2^{(i)}(z)} Z_k^{(i+1,\,i)}, \quad z \in \mathbb{R}, \tag{3.2}$$

for $i = 1, ..., m-1$, 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 distributed as $F_{(i,\,i+1)}(\cdot|r_{i0})$ and $F_{(i+1,\,i)}(\cdot|r_{i0})$, respectively. Here, we work with the left continuous version for $N_1^{(i)}(\cdot)$ and the right continuous version for $N_2^{(i)}(\cdot)$ for $i = 1, ..., m-1$.

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

$$\wp(\mathbf{s}) = \sum_{i=1}^{m-1} \mathcal{P}_i(s_i), \quad \mathbf{s} = (s_1, ..., s_{m-1})' \in \mathbb{R}^{m-1}. \tag{3.3}$$

Clearly, $\wp(\mathbf{s})$ goes to ∞ a.s. when $\|\mathbf{s}\| \to \infty$ since $\mathbb{E}Y_1^{(i,\,i+1)} > 0$ and $\mathbb{E}Z_1^{(i+1,\,i)} > 0$ by a conditional argument for $i = 1, ..., m-1$. Therefore, there exists a unique random $(m-1)$-dimensional cube $[\mathbf{M}_-,\ \mathbf{M}_+) \equiv [M_-^{(1)},\ M_+^{(1)}) \times \cdots \times [M_-^{(m-1)},\ M_+^{(m-1)})$ 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

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

Note that, the processes $\{\mathcal{P}_i(z) : i = 1, ..., m-1\}$ being independent, so are $\{M_-^{(i)} : i = 1, ..., m-1\}$. 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 $n \to \infty$, 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* $n \to \infty$. *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. Fortunately, we can use either the subsampling or the resampling method to obtain confidence intervals of the threshold parameter. For more details, see Gonzalo and Wolf (2005) and Li and Ling (2012).

4 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, Ling and Zhang (2013), 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:

$$y_t = \sum_{i=1}^{m} \Big(\boldsymbol{\phi}_i'\mathbf{Y}_{t-1} + \eta_t\sqrt{\boldsymbol{\alpha}_i'\mathbf{X}_{t-1}}\Big) I\{r_{i-1} < y_{t-d} \le r_i\}. \tag{4.1}$$

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 $\boldsymbol{\Sigma} = \mathrm{diag}\{\boldsymbol{\Sigma}_1, \cdots, \boldsymbol{\Sigma}_m\}$, $D_t(\boldsymbol{\vartheta}) = (D_{1t}'(\boldsymbol{\vartheta}), ..., D_{mt}'(\boldsymbol{\vartheta}))'$ with

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

Our test statistic depends on the score-marked empirical process

$$T_{in}(x, \boldsymbol{\vartheta}) = n^{-1/2} \sum_{t=1}^{n} \mathbf{U}^{-1} D_{it}(\boldsymbol{\vartheta}) I\{r_{i-1} < y_{t-d} \le x\}, \quad x \in \mathbb{R}, \tag{4.2}$$

for $i = 1, ..., m$, where $\mathbf{U} = \mathrm{diag}\{\mathbf{I}_{p+1},\ \sqrt{(\kappa_4 - 1)/2}\,\mathbf{I}_{p+1}\}$ with $\mathbf{I}_{p+1}$ being the identity matrix and $I(r_{i-1} < y_{t-d} \le x) = 0$ if $x \le r$ by convention. By Theorem A.1 in Li, Ling and Zhang (2013), Theorem 3.2 and Taylor's expansion, we can get the asymptotic property of $\{T_{in}(x, \hat{\boldsymbol{\vartheta}}_n)\}$.

Theorem 4.1 *Under the null H_0 that $\{y_t\}$ follows model (4.1) with the true value $\boldsymbol{\vartheta}_0$, if the assumptions* **A1-A7** *hold and η_t is symmetrically distributed, then, for $i = 1, ..., m$,*

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

and $T_{in}(x, \hat{\boldsymbol{\vartheta}}_n)$ converges weakly to $G_i(x)$ in $\mathbb{D}(\mathbb{R})$, 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

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

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

Let for $i = 1, ..., m$

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

It is not hard to see that $\widehat{\mathbf{\Sigma}}_{ix}$ is a consistent estimator of $\mathbf{\Sigma}_{ix}$ uniformly in x.

Now, we define the test statistic

$$S_{in} = \max_{x\in[a_i,\ \hat{r}_{in}]} \frac{\left[\boldsymbol{\beta}'\widehat{\mathbf{\Sigma}}_{ix}^{-1}T_{in}(x, \hat{\boldsymbol{\vartheta}}_n)\right]^2}{\boldsymbol{\beta}'(\widehat{\mathbf{\Sigma}}_{i\,a_i}^{-1} - \widehat{\mathbf{\Sigma}}_{i\,\hat{r}_{in}}^{-1})\boldsymbol{\beta}}, \tag{4.3}$$

where $a_i \in (\hat{r}_{i-1,n},\ \hat{r}_{in})$, $\boldsymbol{\beta}$ is a nonzero $2(p+1)\times 1$ constant vector. By Theorem 4.1 and the continuous mapping theorem, we have the following result.

Theorem 4.2 *If the assumptions in Theorem 4.1 hold, then, for any $2(p+1)\times 1$ nonzero constant vector $\boldsymbol{\beta}$, it follows that*

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

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

From the formula in Shorack and Wellner (1986, p.34)

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

we can choose an approximate critical value C_α such that $P(\max_{\tau\in[0,1]} B^2(\tau) \geq C_\alpha) = \alpha$ for rejecting the null H_0 at different significance levels α.

The limiting distribution of the test statistic S_{in} in (4.3), 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). However, it should guarantee that $\widehat{\mathbf{\Sigma}}_{ia_i}^{-1}$ exists. As for the choice of $\boldsymbol{\beta}$, there is no universal criterion. In practice, numerous simulation studies show that $\boldsymbol{\beta} = (1, ..., 1)'$, with a_i around the 15% quantile of data $\{y_t : \hat{r}_{i-1,n} < y_{t-d} \leq \hat{r}_{in}\}$, produces good power.

Since each score test S_{in} only uses the partial information of the whole sample, e.g., $\{y_t : \hat{r}_{i-1,n} < y_{t-d} \leq \hat{r}_{in}\}$, S_{in} may not be always powerful. Alternatively, we can consider some function of all S_{in}'s, for example,

$$S_n = \max_{1\leq i\leq m}\{S_{in}\}. \tag{4.4}$$

Note that under H_0,

$$E\Big\{\Big[D_{it}(\boldsymbol{\vartheta}_0)I\{r_{i-1,0} < y_{t-d} \leq x\} - \mathbf{\Sigma}_{ix}\mathbf{\Sigma}_{i,\infty}^{-1}D_{it}(\boldsymbol{\vartheta}_0)\Big]$$
$$\times\Big[D_{kt}(\boldsymbol{\vartheta}_0)I\{r_{k-1,0} < y_{t-d} \leq y\} - \mathbf{\Sigma}_{ky}\mathbf{\Sigma}_{k,\infty}^{-1}D_{kt}(\boldsymbol{\vartheta}_0)\Big]'\Big\} = 0, \quad i \neq k,$$

for any $x \in (r_{i-1,0}, r_{i0}]$ and any $y \in (r_{k-1,0}, r_{k0}]$, where

$$D_{jt}(\boldsymbol{\vartheta}_0) = \left(\frac{\mathbf{Y}'_{t-1}\eta_t}{(\boldsymbol{\alpha}'_{j0}\mathbf{X}_{t-1})^{1/2}}, \frac{\mathbf{X}'_{t-1}(\eta_t^2 - 1)}{2(\boldsymbol{\alpha}'_{j0}\mathbf{X}_{t-1})} \right)' I\{r_{j-1,0} < y_{t-d} \le r_{j0}\}.$$

Thus, by Theorem 4.1, $\{T_{in}(x_1, \hat{\boldsymbol{\vartheta}}_n), ..., T_{in}(x_j, \hat{\boldsymbol{\vartheta}}_n)\}$ and $\{T_{kn}(y_1, \hat{\boldsymbol{\vartheta}}_n), ..., T_{kn}(y_v, \hat{\boldsymbol{\vartheta}}_n)\}$ are asymptotically independent for any $x_1, ..., x_j \in (r_{i-1,0},\ r_{i0}]$ and $y_1, ..., y_v \in (r_{k-1,0},\ r_{k0}]$, $i \neq k$, so are $\{S_{1n}, ..., S_{mn}\}$. Then, by Theorem 4.2, it follows that

$$\begin{aligned} \lim_{n\to\infty} P\left(S_n \le x\right) &= P\Big(\max_{1\le i\le m} \max_{\tau\in[0,1]} B_i^2(\tau) \le x \Big) \\ &= \left\{ \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} \exp\left(-\frac{(2k+1)^2\pi^2}{8x} \right) \right\}^m, \quad x \ge 0, \end{aligned}$$

where $\{B_i(\tau)\}$ is a sequence of independent standard Brownian motions.

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

Table 1: Approximate critical values of S_n at the significance level α.

$m\backslash\alpha$	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

5 Special cases

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

5.1 Multiple Threshold ARCH Models

The multiple threshold ARCH (MTARCH) model is defined as

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

When threshold parameters are known, variants of this class of MTARCH models were studied by Gouriéroux and Monfort (1992), Rabemananjara and Zakoïan (1993), Zakoïan (1994), Li and Li (1996) among others. Recently, Chan et al. (2013) 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 $\boldsymbol{\lambda} \equiv (\boldsymbol{\alpha}'_1, \cdots, \boldsymbol{\alpha}'_m)'$. By Theorem 3.2, we have

Theorem 5.1 *If the assumptions* **A1-A7** *hold, then*

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

where $\mathbf{J} = (\kappa_4 - 1)\mathrm{diag}\{J_1, \cdots, J_m\}$ *with*

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

The jump distributions used in the corresponding two-sided compound Poisson processes are induced by

$$\xi_t^{(i,\,j)} = \log \frac{\boldsymbol{\alpha}_{j0}'\mathbf{X}_{t-1}}{\boldsymbol{\alpha}_{i0}'\mathbf{X}_{t-1}} + \frac{\eta_t^2(\boldsymbol{\alpha}_{i0} - \boldsymbol{\alpha}_{j0})'\,\mathbf{X}_{t-1}}{\boldsymbol{\alpha}_{j0}'\,\mathbf{X}_{t-1}}.$$

given y_{t-d}.

5.2 Multiple Threshold AR Models with Conditionally Heteroskedastic Errors

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

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

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, asymptotic properties of the QMLE are not available in literature. Let $\boldsymbol{\lambda} \equiv (\boldsymbol{\phi}_1', \cdots, \boldsymbol{\phi}_m', \boldsymbol{\alpha}')'$ with $\boldsymbol{\alpha} = (\omega, \alpha_1, \ldots, \alpha_p)'$.

Theorem 5.2 *If the assumptions* **A1-A7** *hold, then*

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

with

$$\begin{aligned} \boldsymbol{\Sigma} &= E\left\{ \mathrm{diag}\left(\frac{\mathbf{Y}_{t-1}\mathbf{Y}_{t-1}' I_{1t}}{\boldsymbol{\alpha}_0'\mathbf{X}_{t-1}},\ \ldots,\ \frac{\mathbf{Y}_{t-1}\mathbf{Y}_{t-1}' I_{mt}}{\boldsymbol{\alpha}_0'\mathbf{X}_{t-1}},\ \frac{1}{2}\frac{\mathbf{X}_{t-1}\mathbf{X}_{t-1}'}{(\boldsymbol{\alpha}_0'\mathbf{X}_{t-1})^2} \right) \right\}, \\ \boldsymbol{\Omega} &= E\left(\begin{array}{c|c} \widetilde{\boldsymbol{\Omega}} & \begin{array}{c} \frac{\kappa_3}{2}\frac{\mathbf{Y}_{t-1}\mathbf{X}_{t-1}' I_{1t}}{(\boldsymbol{\alpha}_0'\mathbf{X}_{t-1})^{3/2}} \\ \vdots \\ \frac{\kappa_3}{2}\frac{\mathbf{Y}_{t-1}\mathbf{X}_{t-1}' I_{mt}}{(\boldsymbol{\alpha}_0'\mathbf{X}_{t-1})^{3/2}} \end{array} \\ \hline \frac{\kappa_3}{2}\frac{\mathbf{X}_{t-1}\mathbf{Y}_{t-1}' I_{1t}}{(\boldsymbol{\alpha}_0'\mathbf{X}_{t-1})^{3/2}} \quad \cdots \quad \frac{\kappa_3}{2}\frac{\mathbf{X}_{t-1}\mathbf{Y}_{t-1}' I_{mt}}{(\boldsymbol{\alpha}_0'\mathbf{X}_{t-1})^{3/2}} & \frac{\kappa_4 - 1}{4}\frac{\mathbf{X}_{t-1}\mathbf{X}_{t-1}'}{(\boldsymbol{\alpha}_0'\mathbf{X}_{t-1})^2} \end{array} \right), \\ \widetilde{\boldsymbol{\Omega}} &= \mathrm{diag}(I_{1t}, \ldots, I_{mt}) \otimes \left(\frac{\mathbf{Y}_{t-1}\mathbf{Y}_{t-1}'}{\boldsymbol{\alpha}_0'\mathbf{X}_{t-1}} \right) \end{aligned}$$

where $\otimes$ *is the Kronecker product.*

The jump distributions used in the corresponding two-sided compound Poisson processes are induced by

$$\xi_t^{(i,\,j)} = \frac{\left\{(\boldsymbol{\phi}_{i0} - \boldsymbol{\phi}_{j0})'\mathbf{Y}_{t-1}\right\}^2}{\boldsymbol{\alpha}_0'\mathbf{X}_{t-1}} + \frac{2\eta_t(\boldsymbol{\phi}_{i0} - \boldsymbol{\phi}_{j0})'\mathbf{Y}_{t-1}}{\sqrt{\boldsymbol{\alpha}_0'\mathbf{X}_{t-1}}}$$

given y_{t-d}.

5.3 AR Models with Multiple-Threshold Conditionally Heteroskedastic 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, Roy and Zakoïan (2005)). To further specify the errors dependence structure, various heteroscedasticity models were introduced, see Li, Ling and McAleer (2002), Francq and Zakoïan (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:

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

Let $\boldsymbol{\lambda} \equiv (\boldsymbol{\phi}', \boldsymbol{\alpha}_1', \cdots, \boldsymbol{\alpha}_m')'$.

Theorem 5.3 *If the assumptions* **A1-A7** *hold, then*

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

where

$$\boldsymbol{\Sigma} = E\left\{\operatorname{diag}\left(\sum_{i=1}^{m}\frac{\mathbf{Y}_{t-1}\mathbf{Y}_{t-1}' I_{it}}{\boldsymbol{\alpha}_{i0}'\mathbf{X}_{t-1}},\ \frac{1}{2}\frac{\mathbf{X}_{t-1}\mathbf{X}_{t-1}' I_{1t}}{(\boldsymbol{\alpha}_{10}'\mathbf{X}_{t-1})^2},\ \ldots,\ \frac{1}{2}\frac{\mathbf{X}_{t-1}\mathbf{X}_{t-1}' I_{mt}}{(\boldsymbol{\alpha}_{m0}'\mathbf{X}_{t-1})^2}\right)\right\},$$

$$\boldsymbol{\Omega} = E\left(\begin{array}{c|ccc} \sum_{i=1}^{m}\frac{\mathbf{Y}_{t-1}\mathbf{Y}_{t-1}' I_{it}}{\boldsymbol{\alpha}_{i0}'\mathbf{X}_{t-1}} & \frac{\kappa_3}{2}\frac{\mathbf{Y}_{t-1}\mathbf{X}_{t-1}' I_{1t}}{(\boldsymbol{\alpha}_{10}'\mathbf{X}_{t-1})^{3/2}} & \cdots & \frac{\kappa_3}{2}\frac{\mathbf{Y}_{t-1}\mathbf{X}_{t-1}' I_{mt}}{(\boldsymbol{\alpha}_{m0}'\mathbf{X}_{t-1})^{3/2}} \\ \hline \frac{\kappa_3}{2}\frac{\mathbf{X}_{t-1}\mathbf{Y}_{t-1}' I_{1t}}{(\boldsymbol{\alpha}_{10}'\mathbf{X}_{t-1})^{3/2}} & & & \\ \vdots & & \widetilde{\boldsymbol{\Omega}} & \\ \frac{\kappa_3}{2}\frac{\mathbf{X}_{t-1}\mathbf{Y}_{t-1}' I_{mt}}{(\boldsymbol{\alpha}_{m0}'\mathbf{X}_{t-1})^{3/2}} & & & \end{array}\right),$$

$$\widetilde{\boldsymbol{\Omega}} = \frac{\kappa_4 - 1}{4}\ \operatorname{diag}\left\{(\boldsymbol{\alpha}_{10}'\mathbf{X}_{t-1})^{-2} I_{1t},\ \ldots,\ (\boldsymbol{\alpha}_{m0}'\mathbf{X}_{t-1})^{-2} I_{mt}\right\} \otimes (\mathbf{X}_{t-1}\mathbf{X}_{t-1}').$$

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 1,000 replications of the following model

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

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 $\eta_t \overset{i.i.d.}{\sim} N(0, 1)$.

For model (6.1), by direct calculation, we find $E \log |\phi_1 + \eta_t \sqrt{\alpha_1}| \approx -0.87_{(0.01)} < 0$ and $E \log |\phi_3 + \eta_t \sqrt{\alpha_3}| \approx -0.56_{(0.01)} < 0$, where the number 0.01 in parentheses is the margin of error. Thus, model (6.1) is strictly stationary, although $E \log |\phi_2 + \eta_t \sqrt{\alpha_2}| \approx 0.07_{(0.01)} > 0$.

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 $\hat{\boldsymbol{\lambda}}_n$ and $\hat{\mathbf{r}}_n$ are computed by using $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$ in Theorem 3.2 and by simulating the compound Poisson processes defined in (3.2), respectively. From Table 2, we see that the consistency of the estimators is confirmed from the EM and the closeness of the ESD to the ASD. We also see that the values of the ESD for $\hat{\mathbf{r}}_n$ are approximately halved whenever the value of n is doubled. This partially illustrates the n-consistency of $\hat{\mathbf{r}}_n$.

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
	EM	0.4915	0.9476	0.2977	1.0125	0.4534	3.1094	-0.7081	1.0302	0.4540	-0.9884	0.0158
300	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
	EM	0.4942	0.9778	0.2980	1.0144	0.4675	3.0777	-0.7010	1.0320	0.4631	-0.9933	0.0092
600	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
	EM	0.4978	0.9778	0.3023	1.0028	0.4798	3.0872	-0.6977	1.0162	0.4802	-0.9949	0.0044
900	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

To see the overall feature of the estimated threshold, Figure 1 displays the histograms of $n(\hat{r}_{1n} - r_{10})$ and $n(\hat{r}_{2n} - r_{20})$, respectively, when $n = 600$.

For Model (6.1), by Theorem 2.1, we know that $\phi_2 = 1$ does not entail a unit-root behaviour. This fact can also be seen from Figure 2(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 asymptotically normal. This fact can be seen from Figure 2 (b).

To examine the performance of our score-based tests in finite samples, we take $\beta = (1, ..., 1)'$

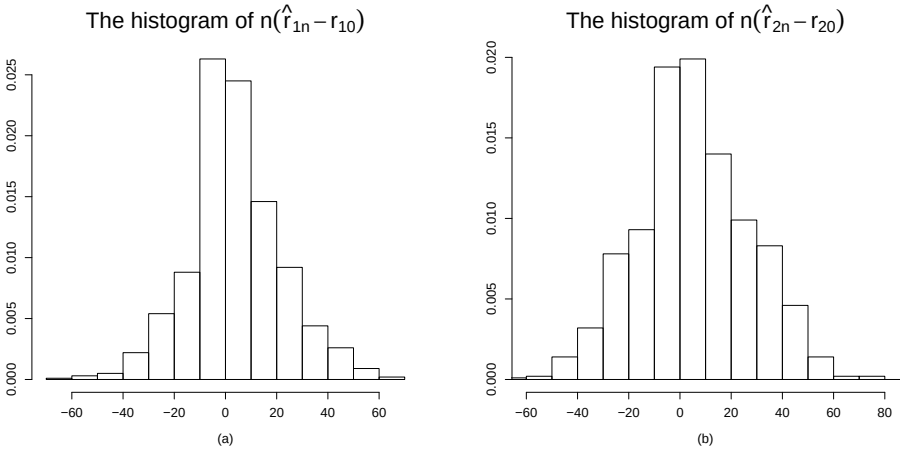

Figure 1: The histograms of $n(\hat{r}_{1n} - r_{10})$ and $n(\hat{r}_{2n} - r_{20})$ when $n = 600$.

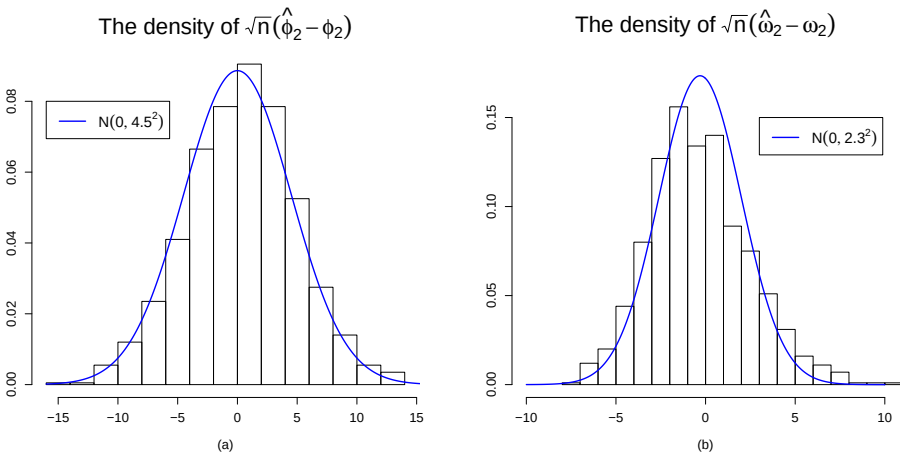

Figure 2: Empirical and asymptotic densities of $\sqrt{n}(\hat{\phi}_2 - \phi_2)$ and $\sqrt{n}(\hat{\omega}_2 - \omega_2)$ when $n = 600$. The asymptotic variances are computed by Theorem 3.2.

and use 1,000 replications. Under the null, H_0, $\{y_t\}$ follows a MTDAR(2;1) model, i.e.,

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

where $\eta_t \overset{i.i.d.}{\sim} N(0,1)$. The alternative H_1 is a MTDAR(3;1) model, i.e.,

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

At the significance level 0.05, the empirical sizes of the test statistics S_{1n}, S_{2n} and S_n for H_0 against H_1, 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 the test based on S_n gets closer to its nominal value than the two other tests as the sample size increases. Figure 3 illustrates the power of the tests S_{1n} and S_{2n} defined in (4.3) when λ varies. It can be seen that such tests can have little power, depending on the value of λ. This can be explained by the fact that these tests only use partial information contained in the data: S_{1n} mainly uses the subsample $\{y_t : y_{t-1} \le \hat{r}_n\}$ while S_{2n} uses the remainder. From Figure 4, we can see that S_n, which uses the whole sample, is more powerful than S_{1n} and S_{2n}.

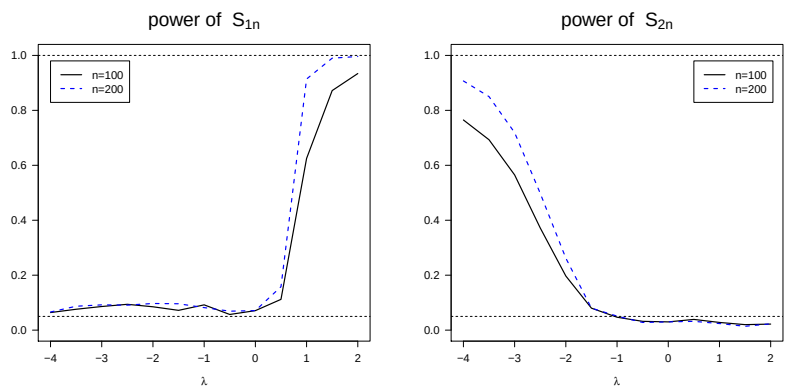

Figure 3: Powers of the test statistics S_{1n} and S_{2n} at the significance level 5%, based on 1,000 simulations of the MTDAR(3;1) model.

7 An empirical example

It is well known that gross national product (GNP) is the market value of all the products and services produced in one year by labor and property supplied by the residents of a country. It measures the economic condition or total economic activity of a country. In modern macroeconomics, the U.S. GNP is perhaps the most examined univariate time series, see Potter (1995) and references therein. Many researchers pointed out that the U.S. GNP sequence contains nonlinearity and asymmetric effects between the contraction and expansion regimes. To characterize such nonlinearity

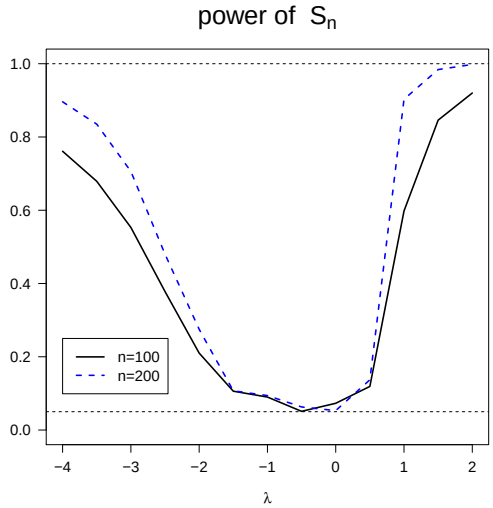

Figure 4: Power of the test statistic S_n at the significance level 5%, based on 1,000 simulations of the MTDAR(3;1) model.

and asymmetric effects, Tiao and Tsay (1994) suggested that two-regime threshold models may be appropriate for recession 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 principle, Li and Ling (2012) used a three-regime TAR model to fit the growth rate of the quarterly U.S. real GNP data over the period 1947–2009.

In this section, we will use the score test in Section 4 to reanalyze the quarterly U.S. real GNP data over the period 1947 — 06/2013 with a total of 266 observations. Let $x_1, ..., x_{266}$ denote the original data. We define the growth rate series as

$$y_t = 100(\log y_t - \log y_{t-1}), \quad t = 2, ..., 266.$$

The data $\{x_t\}$ and the growth rate series $\{y_t\}$ are plotted in Figue 5.

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

$$y_t = \begin{cases} 0.177 + 0.424y_{t-1} - 0.306y_{t-2} + \eta_t\sqrt{0.690 + 0.631y_{t-3}^2}, & \text{if } y_{t-2} \le 0.244, \\ 0.421 + 0.331y_{t-1} + 0.213y_{t-2} - 0.105y_{t-3} + \eta_t\sqrt{0.601 + 0.022y_{t-1}^2}, & \text{if } y_{t-2} > 0.244, \end{cases} \tag{7.1}$$

where the estimated value of d is 2, $\widehat{r}_n = 0.244$, the AIC and BIC's values are 207.246 and 257.362, respectively. 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 Figure 6. Moreover, the value of the score test S_n in (4.4) is 0.767, which indicates that the two-regime TDAR model (7.1) is sufficient for $\{y_t\}$. The number of observations in the recession and expansion regimes are 58 and 207, respectively, meaning that the U.S. total economic activity was most of the time in expansion since the World War II.

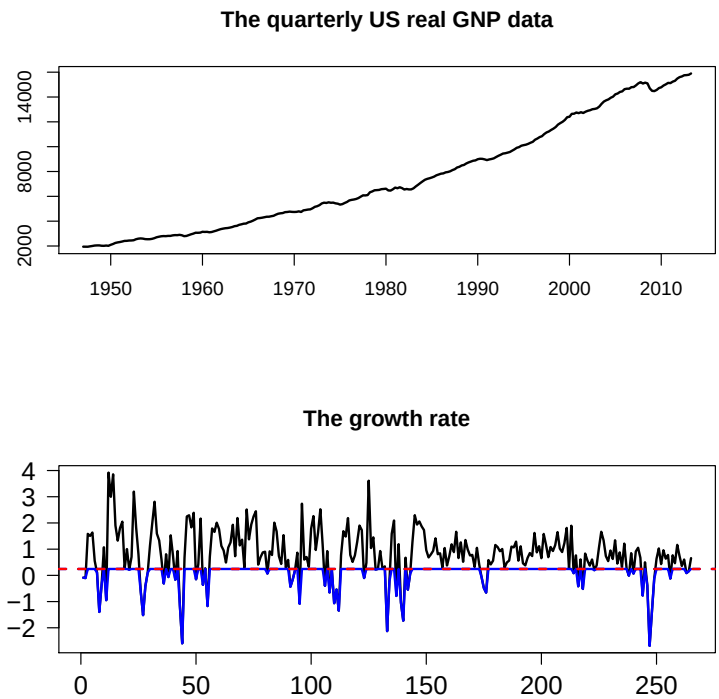

Figure 5: The original data and the growth rate. The dash line below is the frontier between the recession regime and the expansion regime.

8 Appendix

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

8.1 Proof of Theorem 2.1

Proof. We will verify the following criterion, which is straightforwardly deduced from Meyn and Tweedie (Theorem 19.1.3, 1996): 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 $V : \mathbb{R} \to \mathbb{R}^+$ such that

$$V(y) \geq 1, \qquad \forall y \in C \tag{8.1}$$

and for some $\delta > 0$

$$E[V(y_t)|Y_{t-s} = y] \leq (1-\delta)V(y), \qquad \forall y \notin C \tag{8.2}$$

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

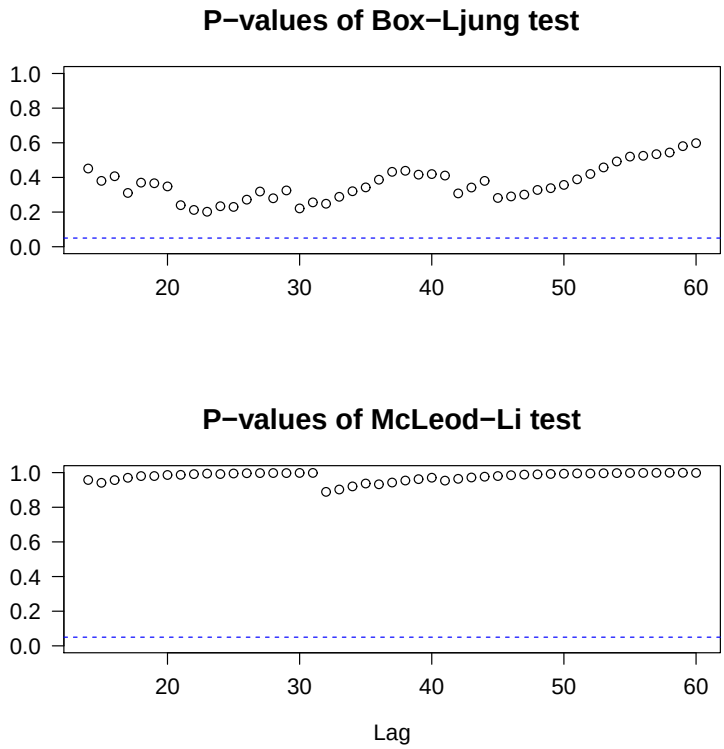

Figure 6: The p-values of the Ljung-Box test statistic and the McLeod-Li test statistic.

transition probabilities of $\{y_t\}$ are given by

$$P(y,B) = P(y_t \in B | y_{t-1} = y) = \sum_{i=1}^{m} \left(\int_{\frac{B-\phi_i y - c_i}{\sqrt{\omega_i + \alpha_i y^2}}} f(x)dx \right) I\{y \in \mathcal{R}_i\},$$

for $y \in \mathbb{R},\ B \in \mathcal{B}(\mathbb{R})$. 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

$$\begin{aligned} y_t &= [a_1(\eta_t) I\{y_{t-1} < r_1\} + a_m(\eta_t) I\{y_{t-1} > r_{m-1}\}]\, y_{t-1} \\ &\quad + \{u_1(y_{t-1}) + u_m(y_{t-1})\}\, \eta_t + z(y_{t-1}, \eta_t), \end{aligned} \tag{8.3}$$

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 (8.3) we find that if $y_{t-2} = y$,

$$y_t = b(\eta_t, y_{t-1})b(\eta_{t-1}, y)y + v(y_{t-1}, \eta_t, y), \tag{8.4}$$

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

$$I\{y_{t-1} < r_1\} = I\{\eta_{t-1} < \frac{r_1 - \phi_m y}{\sqrt{\omega_m + \alpha_m y^2}}\} = I\{\eta_{t-1} < -\frac{\phi_m}{\sqrt{\alpha_m}}\} + \varepsilon_1 I\{\eta_{t-1} \in A_1(y)\}, \tag{8.5}$$

and

$$I\{y_{t-1} > r_{m-1}\} = I\{\eta_{t-1} > \frac{r_{m-1} - \phi_m y}{\sqrt{\omega_m + \alpha_m y^2}}\} = I\{\eta_{t-1} > -\frac{\phi_m}{\sqrt{\alpha_m}}\} + \varepsilon_2 I\{\eta_{t-1} \in A_2(y)\}, \tag{8.6}$$

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 $y < r_1$,

$$I\{y_{t-1} < r_1\} = I\{\eta_{t-1} < \frac{\phi_1}{\sqrt{\alpha_1}}\} + \varepsilon_3 I\{\eta_{t-1} \in A_3(y)\}, \tag{8.7}$$

and

$$I\{y_{t-1} > r_{m-1}\} = I\{\eta_{t-1} > \frac{\phi_1}{\sqrt{\alpha_1}}\} + \varepsilon_4 I\{\eta_{t-1} \in A_4(y)\}. \tag{8.8}$$

Let

$$H_1(\eta_0, \eta_1) = \left\{a_1(\eta_0)I\{\eta_1 < \frac{\phi_1}{\sqrt{\alpha_1}}\} + a_m(\eta_0)I\{\eta_1 > \frac{\phi_1}{\sqrt{\alpha_1}}\}\right\} a_1(\eta_1), \tag{8.9}$$

$$H_m(\eta_0, \eta_1) = \left\{a_1(\eta_0)I\{\eta_1 < -\frac{\phi_m}{\sqrt{\alpha_m}}\} + a_m(\eta_0)I\{\eta_1 > -\frac{\phi_m}{\sqrt{\alpha_m}}\}\right\} a_m(\eta_1), \tag{8.10}$$

We thus have, in view of (8.4)-(8.8), for $y > r_{m-1}$

$$\begin{aligned} y_t &= \left\{a_1(\eta_t)I\{\eta_{t-1} < -\frac{\phi_m}{\sqrt{\alpha_m}}\} + a_m(\eta_t)I\{\eta_{t-1} > -\frac{\phi_m}{\sqrt{\alpha_m}}\}\right\} a_m(\eta_{t-1})y + v(y_{t-1}, \eta_t, y) \\ &\quad + [a_1(\eta_t)\varepsilon_1 I\{\eta_{t-1} \in A_1(y)\} + a_m(\eta_t)\varepsilon_2 I\{\eta_{t-1} \in A_2(y)\}]\, a_m(\eta_{t-1})y \\ &:= H_m(\eta_t, \eta_{t-1})y + R_m(\eta_t, \eta_{t-1}, y), \end{aligned} \tag{8.11}$$

and for $y < r_1$,

$$y_t = H_1(\eta_t, \eta_{t-1})y + R_1(\eta_t, \eta_{t-1}, y). \tag{8.12}$$

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

$$\gamma = \max\{E\log|H_1(\eta_0,\eta_1)|,\ E\log|H_m(\eta_0,\eta_1)|\} < 0. \tag{8.13}$$

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 Zakoïan (2010), Lemma 2.2). Let $V : y \to V(y) = 1 + |y|^u$. Using the elementary equality $(a+b)^u \le a^u + b^u$ for $a, b \in \mathbb{R}$, we have, for $y > r_{m-1}$, by (8.11)

$$E[|y_t|^u|y_{t-2} = y] - |y|^u \quad \le \quad \{\rho - 1\}\,|y|^u + E\{R_m(\eta_t, \eta_{t-1}, y)\}^u. \tag{8.14}$$

Now

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

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

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

The same inequalities hold with a_1 replaced by a_m. Thus, by (8.14), we have, for $y > r_{m-1}$,

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

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

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

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

The set C is a non-empty compact set such that $\lambda(C) > 0$. Moreover (8.1) and (8.2) 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. □

8.2 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 (8.9)-(8.10). Starting from (8.11) we have, for $y > r_{m-1}$,

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

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 $y > r_{m-1}$

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

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

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

we conclude that (8.1) and (8.2) hold with $s = 2$, and $V(y) = 1 + y^2$. The proof is complete. □

8.3 Proof of Theorem 3.1

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

Lemma 8.1 *Under the assumptions of Theorem 3.1, $El_t(\boldsymbol{\vartheta})$ has a unique maximum at $\boldsymbol{\vartheta}_0$.*

Proof. Let $\beta(\boldsymbol{\vartheta}) = -2E\{l_t(\boldsymbol{\vartheta}) - l_t(\boldsymbol{\vartheta}_0)\}$, $A_i = I\{r_{i-1} < y_{t-d} \leq r_i\}$, $A_{i0} = I\{r_{i-1,0} < y_{t-d} \leq r_{i0}\}$ and

$$\Gamma_{i,\,j0} = \log\Big(\frac{\boldsymbol{\alpha}_i'\mathbf{X}_{t-1}}{\boldsymbol{\alpha}_{j0}'\mathbf{X}_{t-1}}\Big) + \frac{\boldsymbol{\alpha}_{j0}'\mathbf{X}_{t-1}}{\boldsymbol{\alpha}_i'\mathbf{X}_{t-1}} - 1 + \frac{\{(\boldsymbol{\phi}_{j0} - \boldsymbol{\phi}_i)'\mathbf{Y}_{t-1}\}^2}{\boldsymbol{\alpha}_i'\mathbf{X}_{t-1}}.$$

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, $\beta(\boldsymbol{\vartheta}) = \sum_{i=1}^m \sum_{j=1}^m E\{\Gamma_{i,j0}A_iA_{j0}\} \geq 0$.

Next, we show that $\beta(\boldsymbol{\vartheta}) = 0$ holds if and only if $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0$. Assume $\beta(\boldsymbol{\vartheta}) = 0$. We first prove $r_1 = r_{10}$. If $r_1 > r_{10}$, then $\Gamma_{1,10}A_1A_{10} = \Gamma_{1,10}A_{10} = 0$ and $\Gamma_{1,20}A_1A_{20} = \Gamma_{1,20}I\{r_{10} < y_{t-d} \leq r_1 \wedge r_{20}\} = 0$. 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 $(\boldsymbol{\phi}_{10}', \boldsymbol{\alpha}_{10}')' = (\boldsymbol{\phi}_1', \boldsymbol{\alpha}_1')' = (\boldsymbol{\phi}_{20}', \boldsymbol{\alpha}_{20}')'$. This contradicts **A1**. Hence, $r_1 \leq r_{10}$. If $r_1 < r_{10}$, using the same technique, we can get $(\boldsymbol{\phi}_{10}', \boldsymbol{\alpha}_{10}')' = (\boldsymbol{\phi}_2', \boldsymbol{\alpha}_2')' = (\boldsymbol{\phi}_{20}', \boldsymbol{\alpha}_{20}')'$ if $r_2 \geq r_{10}$, which results in $r_1 < r_2 < r_{10}$. Repeating the above procedure, we can get $r_1 < ... < r_m < r_{10}$, 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 $(\boldsymbol{\phi}_1', \boldsymbol{\alpha}_1')' = (\boldsymbol{\phi}_{10}', \boldsymbol{\alpha}_{10}')'$ implied by $\Gamma_{1,10}A_{10} = 0$. Similarly, we have $r_i = r_{i0}$ for $i = 2, ..., m-1$ and then in turn $(\boldsymbol{\phi}_j', \boldsymbol{\alpha}_j')' = (\boldsymbol{\phi}_{j0}', \boldsymbol{\alpha}_{j0}')'$ for $j = 2, ..., m$. Thus, $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0$ and $El_t(\boldsymbol{\vartheta})$ is uniquely maximized at $\boldsymbol{\vartheta}_0$. □

Lemma 8.2 *Under the assumptions of Theorem 3.1, for any $\eta > 0$,*

$$\lim_{l\to\infty} P\Big(\max_{l\leq n<\infty} \sup_{\|\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_0\|>\eta} \sum_{t=1}^n [l_t(\boldsymbol{\vartheta}) - l_t(\boldsymbol{\vartheta}_0)] \geq 0\Big) = 0.$$

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

$$EX_t(\tilde{\eta}) \to 0 \tag{8.15}$$

as $\tilde{\eta} \to 0$. 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

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

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

$$P\Big(\max_{l\le n<\infty} \frac{1}{n}\sum_{t=1}^{n} X_t(\tilde{\eta}) \ge \epsilon\Big) \le P\Big(\max_{l\le n<\infty} \frac{1}{n}\Big|\sum_{t=1}^{n}[X_t(\tilde{\eta}) - EX_t(\tilde{\eta})]\Big| \ge \frac{\epsilon}{2}\Big) < \epsilon_1. \tag{8.16}$$

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

$$\lim_{l\to\infty} P\Big(\max_{l\le n<\infty} \Big|\frac{1}{n}\sum_{t=1}^{n}[l_t(\boldsymbol{\vartheta}) - El_t(\boldsymbol{\vartheta})]\Big| \ge \epsilon\Big) = 0. \tag{8.17}$$

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 ith ball, we take a point ξ_i and denote this ball by $V(\xi_i)$. By (8.15)-(8.17), for any $\epsilon > 0$, we have

$$\begin{aligned}
&P\Big(\max_{l\le n<\infty} \frac{1}{n}\sup_{\Theta}\Big|\sum_{t=1}^{n}[l_t(\boldsymbol{\vartheta}) - El_t(\boldsymbol{\vartheta})]\Big| \ge \epsilon\Big)\\
&\le P\Big(\max_{1\le j\le N}\sup_{\boldsymbol{\vartheta}\in V(\xi_j)}\max_{l\le n<\infty}\Big|\frac{1}{n}\sum_{t=1}^{n}[l_t(\boldsymbol{\vartheta}) - l_t(\xi_j)]\Big| \ge \frac{\epsilon}{3}\Big)\\
&\quad + P\Big(\max_{1\le j\le N}\sup_{\boldsymbol{\vartheta}\in V(\xi_j)}\Big|E[l_t(\boldsymbol{\vartheta}) - l_t(\xi_j)]\Big| \ge \frac{\epsilon}{3}\Big) + P\Big(\max_{1\le j\le N}\max_{l\le n<\infty}\Big|\frac{1}{n}\sum_{t=1}^{n}[l_t(\xi_j) - El_t(\xi_j)]\Big| \ge \frac{\epsilon}{3}\Big)\\
&\le P\Big(\sup_{\xi_i\in\Theta}\sup_{\boldsymbol{\vartheta}\in V(\xi_j)}\max_{l\le n<\infty}\Big|\frac{1}{n}\sum_{t=1}^{n}[l_t(\boldsymbol{\vartheta}) - l_t(\xi_j)]\Big| \ge \frac{\epsilon}{3}\Big) + \sum_{j=1}^{N} P\Big(\sup_{\boldsymbol{\vartheta}\in V(\xi_j)}\Big|E[l_t(\boldsymbol{\vartheta}) - l_t(\xi_j)]\Big| \ge \frac{\epsilon}{3}\Big)\\
&\quad + \sum_{j=1}^{N} P\Big(\max_{l\le n<\infty}\Big|\frac{1}{n}\sum_{t=1}^{n}[l_t(\xi_j) - El_t(\xi_j)]\Big| \ge \frac{\epsilon}{3}\Big) < \epsilon,
\end{aligned} \tag{8.18}$$

as l is large enough and Δ is small enough.

Since $E[l_t(\boldsymbol{\vartheta})]$ has a unique maximum at $\boldsymbol{\vartheta}_0$, Θ is compact, and $El_t(\boldsymbol{\vartheta})$ is continuous, there exists a constant $c > 0$, such that

$$\max_{\|\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_0\|>\eta} E[l_t(\boldsymbol{\vartheta}) - l_t(\boldsymbol{\vartheta}_0)] \le -c, \tag{8.19}$$

for any $\eta > 0$. By (8.18)-(8.19), it follows that

$$
\begin{aligned}
&P\Big(\max_{l\le n<\infty}\sup_{\|\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_0\|>\eta}\Big\{\sum_{t=1}^{n}[l_t(\boldsymbol{\vartheta})-l_t(\boldsymbol{\vartheta}_0)]+\frac{cn}{2}\Big\}>0\Big)\\
&=P\Big(\max_{l\le n<\infty}\sup_{\|\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_0\|>\eta}\Big\{\sum_{t=1}^{n}[l_t(\boldsymbol{\vartheta})-El_t(\boldsymbol{\vartheta})]\\
&\quad-\sum_{t=1}^{n}[l_t(\boldsymbol{\vartheta}_0)-El_t(\boldsymbol{\vartheta}_0)]+n[El_t(\boldsymbol{\vartheta})-El_t(\boldsymbol{\vartheta}_0)]+\frac{cn}{2}\Big\}>0\Big)\\
&\le P\Big(\max_{l\le n<\infty}\sup_{\Theta}\Big\{2\Big|\sum_{t=1}^{n}[l_t(\boldsymbol{\vartheta})-El_t(\boldsymbol{\vartheta})]\Big|-cn+\frac{cn}{2}\Big\}>0\Big)\\
&\le P\Big(\max_{l\le n<\infty}\sup_{\Theta}\Big|\frac{1}{n}\sum_{t=1}^{n}[l_t(\boldsymbol{\vartheta})-El_t(\boldsymbol{\vartheta})]\Big|>\frac{c}{4}\Big)\to 0,\quad \text{as } l\to\infty. \qquad (8.20)
\end{aligned}
$$

□

Proof of Theorem 3.1. By Lemma 8.2, for any $\epsilon > 0$, we have

$$
\begin{aligned}
\lim_{l\to\infty}P(\max_{l\le n<\infty}\|\hat{\boldsymbol{\vartheta}}_n-\boldsymbol{\vartheta}_0\|>\epsilon)&=\lim_{l\to\infty}P\Big\{\max_{l\le n<\infty}\|\hat{\boldsymbol{\vartheta}}_n-\boldsymbol{\vartheta}_0\|>\epsilon,\max_{l\le n<\infty}\sum_{t=1}^{n}\Big[l_t(\hat{\boldsymbol{\vartheta}}_n)-l_t(\boldsymbol{\vartheta}_0)\Big]\ge 0\Big\}\\
&\le\lim_{l\to\infty}P\Big\{\max_{l\le n<\infty}\sup_{\|\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_0\|>\epsilon}\sum_{t=1}^{n}\Big[l_t(\boldsymbol{\vartheta})-l_t(\boldsymbol{\vartheta}_0)\Big]\ge 0\Big\}=0.
\end{aligned}
$$

Thus, the conclusion holds. □

8.4 Proof of Theorem 3.2

Since $\hat{\boldsymbol{\vartheta}}_n$ is consistent by Theorem 3.1, we restrict the parameter space to an open neighborhood of $\boldsymbol{\vartheta}_0$. To this end, define $V_\delta = \{\boldsymbol{\vartheta}\in\Theta : \|\boldsymbol{\lambda}-\boldsymbol{\lambda}_0\|<\delta,\ |r_i-r_{i0}|<\delta, i=1,...,m-1\}$ for some $0<\delta<1$ to be determined later. Choose δ small enough so that $\{r : |r-r_{i-1,0}|<\delta\}\cap\{r : |r-r_{i0}|<\delta\}=\varnothing$ for $i=2,...,m-1$. Note that

$$
L_n(\boldsymbol{\lambda},\mathbf{r})-L_n(\boldsymbol{\lambda},\mathbf{r}_0)=\sum_{i=1}^{m-1}L_n^{(i)}(\boldsymbol{\lambda},r_i),
$$

where

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

For (i), it is equivalent to prove $n|\hat{r}_{in}-r_{i0}|=O_p(1)$ for each $i=1,...,m$. 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 $\varepsilon>0$,

$$
P\left(\sup_{\substack{B/n<|r_i-r_{i0}|\le\delta\\ \boldsymbol{\vartheta}\in V_\delta}}\frac{L_n^{(i)}(\boldsymbol{\lambda},r_i)-L_n^{(i)}(\boldsymbol{\lambda},r_{i0})}{nG_i(|r_i-r_{i0}|)}<-\gamma\right)>1-\varepsilon \qquad (8.21)
$$

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

We now show (8.21) holds for the case $p = 1$ and $i = 1$. The proof of the general case is similar. Here, we only treat the case $r_1 > r_{10}$. The proof for the case $r_1 \le r_{10}$ is similar. Writing $r_1 = r_{10} + u$ for some $u \ge 0$. By a simple calculation, it follows that

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

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

$$K_1 = \log \frac{\boldsymbol{\alpha}_{10}'\mathbf{X}}{\boldsymbol{\alpha}_{20}'\mathbf{X}} + \frac{\boldsymbol{\alpha}_{20}'\mathbf{X}}{\boldsymbol{\alpha}_{10}'\mathbf{X}} - 1 + \frac{\{(\phi_{20} - \phi_{10}) r_{10}\}^2}{\boldsymbol{\alpha}_{10}'\mathbf{X}},$$
$$K_2 = \frac{2\{(\phi_{10} - \phi_{20}) r_{10}\}\sqrt{\boldsymbol{\alpha}_{20}'\mathbf{X}}}{\boldsymbol{\alpha}_{10}'\mathbf{X}} \quad \text{and} \quad K_3 = \frac{(\boldsymbol{\alpha}_{10} - \boldsymbol{\alpha}_{20})'\mathbf{X}}{\alpha_{10}'\mathbf{X}}$$

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

$$P\Big(\sup_{B/n < u \le \delta} \Big| \frac{G_{1n}(u)}{G_1(u)} - 1 \Big| < \epsilon \Big) > 1 - \varepsilon,$$
$$P\Big(\sup_{B/n < u \le \delta} \Big| \frac{\sum_{t=1}^n \eta_t I(r_{10} < y_{t-1} \le r_{10} + u)}{nG_1(u)} \Big| < \epsilon \Big) > 1 - \varepsilon,$$
$$P\Big(\sup_{B/n < u \le \delta} \Big| \frac{\sum_{t=1}^n (\eta_t^2 - 1) I(r_{10} < y_{t-1} \le r_{10} + u)}{nG_1(u)} \Big| < \epsilon \Big) > 1 - \varepsilon.$$

Note that $K_1 > 0$ by **A7**. Choosing δ small enough and $\gamma = K_1/4$, (8.21) holds and so does (i). □

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

$$\begin{aligned}
&\sup_{\|\mathbf{r} - \mathbf{r}_0\| \le B/n} \Big\| \frac{1}{n} \frac{\partial L_n(\boldsymbol{\lambda}_0, \mathbf{r})}{\partial \boldsymbol{\lambda}} - \frac{1}{n} \frac{\partial L_n(\boldsymbol{\lambda}_0, \mathbf{r}_0)}{\partial \boldsymbol{\lambda}} \Big\| = O_p(n^{-1}), \\
&\sup_{\|\boldsymbol{\lambda} - \boldsymbol{\lambda}_0\| < \epsilon} \sup_{\mathbf{r} \in \Theta_r} \Big\| \frac{1}{n} \frac{\partial^2 L_n(\boldsymbol{\lambda}, \mathbf{r})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}'} - \frac{1}{n} \frac{\partial^2 L_n(\boldsymbol{\lambda}_0, \mathbf{r})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}'} \Big\| = O_p(\epsilon), \\
&\sup_{\|\mathbf{r} - \mathbf{r}_0\| \le B/n} \Big\| \frac{1}{n} \frac{\partial^2 L_n(\boldsymbol{\lambda}_0, \mathbf{r})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}'} - \frac{1}{n} \frac{\partial^2 L_n(\boldsymbol{\lambda}_0, \mathbf{r}_0)}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}'} \Big\| = O_p(n^{-1}).
\end{aligned} \tag{8.22}$$

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

$$0 = \frac{1}{n} \frac{\partial L_n(\hat{\boldsymbol{\lambda}}_n(\mathbf{r}), \mathbf{r})}{\partial \boldsymbol{\lambda}} = \frac{1}{n} \frac{\partial L_n(\boldsymbol{\lambda}_0, \mathbf{r})}{\partial \boldsymbol{\lambda}} + \frac{1}{n} \frac{\partial^2 L_n(\bar{\boldsymbol{\lambda}}, \mathbf{r})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}'} \Big[\hat{\boldsymbol{\lambda}}_n(\mathbf{r}) - \boldsymbol{\lambda}_0 \Big], \tag{8.23}$$

where $\bar{\boldsymbol{\lambda}}$ lies in the ball $\mathcal{B}(\boldsymbol{\lambda}_0, \|\hat{\boldsymbol{\lambda}}_n(\mathbf{r}) - \boldsymbol{\lambda}_0\|)$. By the ergodic theorem, we have

$$-\frac{1}{n} \frac{\partial^2 L_n(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}'} \longrightarrow \boldsymbol{\Sigma} := \mathrm{diag}(\boldsymbol{\Sigma}_1, ..., \boldsymbol{\Sigma}_m), \quad \text{a.s.}$$

as $n \to \infty$. Thus, by (8.22) and (8.23),

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

which implies that

$$\begin{aligned} \sqrt{n} \sup_{\|\mathbf{r}-\mathbf{r}_0\| \le B/n} \left\| \hat{\boldsymbol{\lambda}}_n(\mathbf{r}) - \hat{\boldsymbol{\lambda}}_n(\mathbf{r}_0) \right\| \le & \sup_{\|\mathbf{r}-\mathbf{r}_0\| \le B/n} \left\| \sqrt{n} \left[\hat{\boldsymbol{\lambda}}_n(\mathbf{r}) - \boldsymbol{\lambda}_0 \right] - \boldsymbol{\Sigma}^{-1} \frac{1}{\sqrt{n}} \frac{\partial L_n(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\lambda}} \right\| \\ & + \left\| \sqrt{n} \left[\hat{\boldsymbol{\lambda}}_n(\mathbf{r}_0) - \boldsymbol{\lambda}_0 \right] - \Sigma^{-1} \frac{1}{\sqrt{n}} \frac{\partial L_n(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\lambda}} \right\| \\ = & \, o_p(1). \end{aligned}$$

From the martingale central limit theorem it follows that

$$\frac{1}{\sqrt{n}} \frac{\partial L_n(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\lambda}} \xrightarrow{d} \mathcal{N}(0, \ \boldsymbol{\Omega}) \quad \text{with} \quad \Omega = \mathrm{diag}(\Omega_1, ..., \Omega_m).$$

Thus, the result holds. □

REFERENCES

Bardet, J-M. and Wintenberger, O. (2009) Asymptotic normality of the quasi-maximum likelihood estimator for multidimensional causal processes. *The Annals of Statistics* **37**, 2730–2759.

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

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. Roy. Statist. Soc. Ser. B* **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. and Tong, H. (2013) On conditional heteroscedastic AR models with thresholds. To appear in *Statist. Sinica.*

Chan, K.S., Petruccelli, J.D., Tong, H. and Woolford, S.W. (1985) A multiple-threshold AR(1) model. *J. Appl. Probab.* **22**, 267-279.

Chan, K.S. and Tong, H. (1985) On the use of the deterministic Lyapunov function for the ergodicity of stochastic difference equations. *Adv. in Appl. Probab.* **17**, 666-678.

Chan, K.S. and Tong, H. (1990) On Likelihood Ratio tests for threshold autoregression. *J. Roy. Statist. Soc. Ser. B*, **52**, 469–476

Chow, Y.S. and Teicher, H. (1978) *Probability theory: independence, interchangeability, martingales.* Springer-Verlag.

Cline, D.B.H. and Pu, H-M. (2004) Stability and the Lyapounov 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. and Zakoian, J-M. (2004) Maximum Likelihood Estimation of Pure GARCH and ARMA-GARCH Processes. *Bernoulli* **10**, 605–637.

Francq, C., Roy, R. and Zakoïan, J-M. (2005) Diagnostic checking in ARMA models with uncorrelated errors. *J. Amer. Statist. Assoc.* **100**, 532–544.

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

Gonzalo, J. and Wolf, M. (2005) Subsampling inference in threshold autoregressive models. *J. Econom.* **127**, 201–224.

Gouriéroux, C. and Monfort, A. (1992) Qualitative threshold ARCH models. *J. Econom.* **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 U.S. unemployment. J. Bus. Econ. Stat. 17, 298–312.

Li, C.W. and Li, W.K. (1996) On a double-threshold autoregressive heteroscedastic time series model. *J. Appl. Econom.* 11, 253–274.

Li, D. and Ling, S. (2012) On the least squares estimation of multiple-regime threshold autoregressive models. *J. Econom.* **167**, 240–253.

Li, D., Ling, S. and Li, W.K.(2013) Asymptotic theory on the least squares estimation of threshold moving-average models. *Economet. Theor.* **29**, 482–516.

Li, D., Ling, S. and Zhang, R.M. (2013) On a threshold double autoregressive model. *J. Bus. Econom. Statist.* (Revised and resubmitted)

Li, W.K., Ling, S. and McAleer, M. (2002) Recent theoretical results for time series models with GARCH errors. *J. Econ. Surveys* **16**, 245–269.

Ling, S. (2004) Estimation and testing stationarity for double-autoregressive models. *J. Roy. Statist. Soc. Ser. B* **66**, 63–78.

Ling, S. (2007) A double AR(p) model: structure and estimation. *Statist. Sinica* **17**, 161–175.

Ling, S. and Li, D. (2008) Asymptotic inference for a non-stationary double AR(1) model. *Biometrika* **95**, 257–263.

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

McLeod, A. I. and Li, W. K. (1983) Diagnostic checking ARMA time series models using squared-residual autocorrelations. *J. Time Ser. Anal.* **4**, 269–273.

Meitz, M. and Saikkonen, P. (2011) Parameter estimation in nonlinear AR-GARCH models. *Economet. Theor.* **27**, 1236–1278.

Meyn, S.P. and 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. Econom. 10, 109–125.

Rabemananjara, R., and Zakoïan, J.M. (1993), Threshold ARCH models and asymmetries in volatility. *J. Appl. Econom.*, 8, 31–49.

Shorack, G.R. and 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. Forecasting* **13**, 109–131.

Tong, H. (1978) On a threshold model. *Pattern recognition and signal processing.* (ed. C.H.Chen). Sijthoff and Noordhoff, Amsterdam.

Tong, H. (1990) *Nonlinear time series. A dynamical system approach.* Oxford University Press, New York.

Weiss, A.A. (1984) ARMA models with ARCH errors. *J. Time Ser. Anal.* **5**, 129-143.

Wong, C.S. and Li, W.K. (1997), Testing for threshold autoregression with conditional heteroscedasticity. *Biometrika*, 84, 407–418.

Wong, C.S. and Li, W.K. (2000), Testing for double threshold autoregressive conditional heteroscedastic model. *Statist. Sinica*, 10, 173–189.

Zakoïan, J-M. (1994) Threshold heteroskedastic models. *J. Econ. Dyn. Control* **18**, 931–955.

Zhang, X., Wong, H., Li, Y. and Ip, W.-C. (2011). A class of threshold autoregressive conditional heteroscedastic models. *Stat. Interface* **4**, 149–157.