REVSTAT – Statistical Journal
Volume 14, Number 4, October 2016, 395–413

ON THE IDENTIFIABILITY CONDITIONS IN SOME NONLINEAR TIME SERIES MODELS

Authors: Jungsik Noh
– Department of Bioinformatics,
University of Texas Southwestern Medical Center, Dallas, USA
jungsik.noh@utsouthwestern.edu

Sangyeol Lee
– Department of Statistics, Seoul National University,
Seoul, Korea
sylee@stats.snu.ac.kr

Received: March 2014 Revised: February 2015 Accepted: March 2015

Abstract:

- In this study, we consider the identifiability problem for nonlinear time series models. Special attention is paid to smooth transition GARCH, nonlinear Poisson autoregressive, and multiple regime smooth transition autoregressive models. Some sufficient conditions are obtained to establish the identifiability of these models.

Key-Words:

- *identifiability, nonlinear time series models; GARCH-type models; smooth transition GARCH models; Poisson autoregressive models; smooth transition autoregressive models.*

AMS Subject Classification:

- 62F10, 62M10.

1. INTRODUCTION

Verifying the identifiability conditions for time series models is a fundamental task in constructing the consistent estimators of model parameters and ensuring the positive definiteness of their asymptotic covariance matrices. Although time series models are assumed to be identifiable in many situations, its verification is often nontrivial and even troublesome, especially in handling nonlinear generalized autoregressive conditional heteroscedasticity (GARCH) models. This issue has a long history and there exist a vast amount of relevant studies in the literature. For instance, Rothenberg [28] introduced the global and local identification concept and verified that local identifiability is equivalent to the nonsingularity of the information matrix. Phillips [26] derived asymptotic theories in partially identified models. Hansen [12] and Francq *et al.* [7] proposed a test for the hypothesis wherein nuisance parameters are unidentifiable. Komunjer [16] provided the primitive conditions for global identification in moment restriction models. In most cases, the identifiability condition is inherent to given statistical models; for example, the multiple linear regression model is unidentifiable when exact multicollinearity exists. Thus, in nature, the verification of identifiability is more complicated in nonlinear time series models with volatilities, such as threshold autoregressive and smooth transition GARCH models (see, for instance, Chan [3] and Meitz and Saikkonen [24]). Thus, there is a need to develop a more refined approach than the existing ones to cope with the problem more adequately.

In this study, we deal with the identifiability problem within a framework similar to that of the M-estimation. To elucidate, let us consider the nonlinear least squares (NLS) estimation from a strictly stationary ergodic process $\{(Y_t, Z_t)\}$, with $E(Y_t|Z_t) = f(Z_t, \beta^\circ)$ for some known function f. Then, the limit of the random objective functions for parameter estimation is uniquely minimized at β° when the following identifiability condition holds:

$$f(Z_1, \beta) = f(Z_1, \beta^\circ) \qquad \text{a.s. implies} \quad \beta = \beta^\circ\,. \tag{1.1}$$

In most M-estimation procedures, the identifiability conditions are given in the form of (1.1), where f can be a conditional mean, variance, or quantile function (see Hayashi [13, p. 463], Berkes *et al.* [2], and Lee and Noh [19]). Moreover, as seen in Wu [33], to ensure the positive definiteness of asymptotic covariance matrices of the NLS estimator, one needs to verify that $\lambda^T \partial f(Z_1, \beta^\circ)/\partial\beta = 0$ a.s. implies $\lambda = 0$. The method described in this study is also useful to verify the positive definiteness of asymptotic covariance matrices of parameter estimators.

As a representative study on the issue with nonlinear time series, we can refer to Chan and Tong [4], who studied the asymptotic theory of NLS estimators for the smooth transition AR (STAR) models and verified the positive definiteness of asymptotic variance matrices. Later, many authors handled this problem using various GARCH-type models because it is crucial when verifying the asymptotic properties of quasi-maximum likelihood estimators (QMLEs). For example, Straumann–Mikosch [29], Medeiros and Veiga [22], Kristensen and Rahbek [17], Meitz and Saikkonen [24], and Lee and Lee [18] consider the identifiability problem in exponential and asymmetric GARCH(p, q) models, flexible coefficient GARCH$(1, 1)$ models nesting a smooth transition GARCH$(1, 1)$ (STGARCH) model, nonlinear ARCH models, nonlinear AR(p) models with nonlinear GARCH$(1, 1)$ errors, STAR(p)–STGARCH$(1, 1)$ models, and Box–Cox transformed threshold GARCH(p, q) models. To ensure (1.1), these authors developed their own methods that reflect the nonlinear structure of underlying models.

In this study, we develop a method that refines existing ones to deduce the identifiability conditions for various nonlinear time series models, tribute to STGARCH(p, q), Poisson autoregressive, and multiple regime STAR(p) models. The remainder of this paper is organized as follows. In Section 2, we describe our method using some examples. In Section 3, we investigate the identifiability conditions in the aforementioned models. The proofs are provided in Section 4.

2. EXAMPLES AND MOTIVATION

In this section, we explore some existing methods that verify the identifiability of STAR models and asymmetric GARCH (AGARCH) models. In what follows, $\{X_t\}$ and $\mathcal{F}_t$ denote the data-generating process and the σ-field generated by $\{X_s : s \le t\}$.

First, we consider the STAR model with two regimes as follows:

$$X_t = m(X_{t-1}, ..., X_{t-p}; \theta^\circ) + \varepsilon_t ,$$
$$m(X_{t-1}, ..., X_{t-p}; \theta^\circ) = {\beta_0^\circ}^T \mathbf{X}_{t-1} + {\beta_1^\circ}^T \mathbf{X}_{t-1} F\left(\frac{X_{t-d} - c^\circ}{z^\circ}\right),$$

where $\{\varepsilon_t\}$ are iid random variables, $\theta^{\circ T} = ({\beta_0^\circ}^T, {\beta_1^\circ}^T, c^\circ, r^\circ)$ and $\mathbf{X}_{t-1} = (1, X_{t-1}, ..., X_{t-p})^T$, and $F(\cdot)$ is a smooth distribution function. Chan and Tong [4] verified the positive definiteness of $E[\dot{m}_t(\theta^\circ)\, \dot{m}_t(\theta^\circ)^T]$, where $\dot{m}_t(\theta^\circ) = \dot{m}(X_{t-1}, ..., X_{t-p}; \theta^\circ)$ denotes the gradient of $m(x; \theta)$ at θ°, by showing that for a given $\lambda \ne 0$, there exists $S \subset \mathbb{R}^p$, such that $\{\lambda^T \dot{m}(x; \theta^\circ)\}^2$ is positive for any $x \in S$ and $P(\{(X_{t-1}, ..., X_{t-p}) \in S\}) > 0$. On the other hand, Meitz and Saikkonen [24] also considered the above model and verified that $m(X_{t-1}, ..., X_{t-p}; \theta) = m(X_{t-1}, ..., X_{t-p}; \theta^\circ)$ a.s. implies $\theta = \theta^\circ$. In both cases, the main step is commonly to show that the function $x \mapsto g(x; \theta, \theta^\circ)$, which equals $(\theta - \theta^\circ)^T \dot{m}(x; \theta^\circ)$ in Chan and Tong [4] and $m(x; \theta) - m(x; \theta^\circ)$ in Meitz and Saikkonen [24], satisfies $g(x; \theta, \theta^\circ) = 0$ for all $x \in \text{supp}(X_{t-1}, ..., X_{t-p})$, where $\text{supp}(Y)$ denotes the distribution support of the random vector Y. With this equation, they could deduce certain conditions to guarantee $\theta = \theta^\circ$. Motivated by these studies, we take a similar approach to deduce the identifiability conditions for nonlinear time series models. In fact, our method is handier than those in the existing studies, such as Kristensen–Rahbek [17], Meitz and Saikkonen [24], and Lee and Lee [18]. For example, our method no longer requires the condition that either the observations or their conditional volatilities should take all values of an open interval with a positive probability.

Next, we consider the case of an AGARCH(1,1) model with power 2:

$$X_t = \sigma_t \eta_t \,, \qquad \sigma_t^2 \;=\; \omega^\circ + \alpha^\circ \left(|X_{t-1}| - \gamma^\circ X_{t-1}\right)^2 + \beta^\circ \sigma_{t-1}^2 \,, \tag{2.1}$$

where $\{\eta_t\}$ is a sequence of iid random variables with $E\eta_t = 0$ and $E\eta_t^2 = 1$. Kristensen and Rahbek [17] and Straumann and Mikosch [29] derived identifiability conditions for asymmetric power ARCH and AGARCH models. We denote $\theta^\circ = (\omega^\circ, \alpha^\circ, \beta^\circ, \gamma^\circ)^T$ and $\Theta = (0,\infty) \times [0,\infty) \times [0,1) \times [-1,1]$, where $\alpha^\circ > 0$. Assuming that Model (2.1) has a strictly stationary solution $\{X_t\}$, for $\theta \in \Theta$, we define a strictly stationary process $\{\sigma_t^2(\theta)\}$ as the solution of

$$\sigma_t^2(\theta) \;=\; \omega + \alpha \left(|X_{t-1}| - \gamma X_{t-1}\right)^2 + \beta \sigma_{t-1}^2(\theta) \,, \qquad \forall t \in \mathbb{Z} \,, \tag{2.2}$$

where $\sigma_t^2(\theta^\circ)$ is equal to σ_t^2.

In this case, the identifiability condition is that $\sigma_t^2 = \sigma_t^2(\theta)$ a.s. for some $t \in \mathbb{Z}$ and $\theta \in \Theta$ implies $\theta = \theta^\circ$, which is crucial to verify the strong consistency of QMLE. Below, we demonstrate the approach of Straumann and Mikosch [29]. Note that $\sigma_t^2 = \sigma_t^2(\theta)$ a.s. for all t because $\{\sigma_t^2 - \sigma_t^2(\theta)\}$ is stationary. Then, one can obtain

$$\omega^\circ - \omega + \sigma_{t-1}^2 Y_{t-1} = \; 0 \quad \text{a.s.}\,, \tag{2.3}$$

where $Y_{t-1} = \alpha^\circ \left(|\eta_{t-1}| - \gamma^\circ \eta_{t-1}\right)^2 - \alpha \left(|\eta_{t-1}| - \gamma \eta_{t-1}\right)^2 + \beta^\circ - \beta$. As shown in Lemma 5.3 of Straumann and Mikosch [29], Y_{t-1} is $\mathcal{F}_{t-2}$-measurable due to (2.3), but at the same time, it is independent of $\mathcal{F}_{t-2}$. Then, $\theta = \theta^\circ$ can be easily deduced from the degeneracy of Y_{t-1} and certain mild conditions on the distribution of η_{t-1}. This approach, however, cannot be extended straightforwardly to more complicated models. Thus, in our study, we take a different approach.

Our idea is to interpret the left-hand side of equation (2.3) as a function of η_{t-1}. Considering that σ_{t-1} is given, for example, as constant σ, we introduce the continuous function:

$$g(x,\sigma) \;=\; \omega^\circ - \omega + \sigma^2 \left\{ \alpha^\circ \left(|x| - \gamma^\circ x\right)^2 - \alpha \left(|x| - \gamma x\right)^2 + \beta^\circ - \beta \right\}.$$

Since (2.3) implies $g(\eta_{t-1}, \sigma_{t-1}) = 0$ a.s., it follows that $g(x,\sigma) = 0$ for all $(x,\sigma) \in \text{supp}(\eta_{t-1}, \sigma_{t-1})$. Further, owing to the independence of η_{t-1} and σ_{t-1}, we have

$g(x,\sigma) = 0$ for all $(x,\sigma) \in \mathrm{supp}(\eta_{t-1}) \times \mathrm{supp}(\sigma_{t-1})$. This, in turn, implies

$$P\Big\{g(x,\sigma_{t-1}) = 0 \ \text{ for all } \ x \in \mathrm{supp}(\eta_{t-1})\Big\} \;=\; 1 \,. \tag{2.4}$$

Assume that $\mathrm{supp}(\eta_{t-1}) = \mathbb{R}$; in fact, it is sufficient to assume that $\mathrm{supp}(\eta_{t-1})$ comprises three distinct (one positive and one negative) real numbers. Then, $g(x,\sigma_{t-1}) = 0$ a.s. for all $x \in \mathbb{R}$ and, particularly $g(0,\sigma_{t-1}) = \omega^\circ - \omega + \sigma_{t-1}^2\,(\beta^\circ - \beta) = 0$ a.s., which leads to $\beta = \beta^\circ$ and $\omega = \omega^\circ$ owing to the nondegeneracy of σ_{t-1}^2. Henceforth, the equation $g(x,\sigma_{t-1}) = 0$ a.s. $\forall x \in \mathbb{R}$ is now reduced to

$$\alpha^\circ\,(|x| - \gamma^\circ x)^2 - \alpha\,(|x| - \gamma x)^2 \;=\; 0 \;, \qquad \forall x \in \mathbb{R} \,, \tag{2.5}$$

and thus, $\theta = \theta^\circ$ is derived. This AGARCH$(1,1)$ example demonstrates that equation (2.4) plays a crucial role in obtaining the conditions to guarantee the identifiability of a time series model. Later, to obtain the desired results for general nonlinear time series models, such as STGARCH, nonlinear Poisson autoregressive, and multiple regime STAR models, we will often apply the equations analogous to (2.4) and results such as $P\{\lim_{x\to\infty} g(x,\sigma_{t-1}) = 0\} = 1$ or $P\big\{\lim_{x\to-\infty} x^{-2} g(x,\sigma_{t-1}) = 0\big\} = 1$, as seen in the proof of Theorem 3.1.

3. IDENTIFIABILITY IN NONLINEAR TIME SERIES

3.1. Smooth transition GARCH models

González-Rivera [11] introduced the STGARCH(p,q,d) model:

$$\begin{aligned} X_t \;&=\; \sigma_t \eta_t \,, \\ \sigma_t^2 \;&=\; \omega^\circ + \sum_{i=1}^{q} \alpha_{1i}^\circ X_{t-i}^2 + \left(\sum_{i=1}^{q} \alpha_{2i}^\circ X_{t-i}^2\right) F(X_{t-d},\gamma^\circ) + \sum_{j=1}^{p} \beta_j^\circ \sigma_{t-j}^2 \,, \end{aligned} \tag{3.1}$$

where $\{\eta_t\}$ is the same as that in Model (2.1),

$$F(X_{t-d},\gamma^\circ) \;=\; \frac{1}{1+e^{\gamma^\circ X_{t-d}}} - \frac{1}{2} \,,$$

$d \in \{1,...,q\}$ is pre-specified, and $\gamma^\circ > 0$ is the smoothness parameter that determines the speed of transition. It is noteworthy that when $\gamma^\circ \to \infty$, the STGARCH$(1,1,1)$ model becomes a GJR-GARCH$(1,1)$ model proposed by Glosten *et al.* [10], which is identical to Model (2.1).

We denote the true parameter vector by $\theta^\circ = (\gamma^\circ, \omega^\circ, \alpha_{11}^\circ, ..., \alpha_{1q}^\circ, \alpha_{21}^\circ, ..., \alpha_{2q}^\circ, \beta_1^\circ, ..., \beta_p^\circ)^T$. Let $\Theta = [0,\infty) \times (0,\infty) \times A \times B$ be the parameter space, where

$$A = \left\{(\alpha_{11}, ..., \alpha_{1q}, \alpha_{21}, ..., \alpha_{2q}) \in \mathbb{R}^{2q} : \alpha_{1i} \geq 0,\ |\alpha_{2i}| \leq 2\alpha_{1i}, \forall i\right\},$$

$$B = \left\{(\beta_1, ..., \beta_p) \in [0,1)^p : \sum_{j=1}^{p} \beta_j < 1\right\}, \tag{3.2}$$

and assume that $\theta^\circ \in \Theta$ for the conditional variance to be positive.

Sufficient conditions to ensure the existence of a stationary solution for Model (3.1) are not specified in the literature. For instance, Straumann–Mikosch [29] and Meitz and Saikkonen [23] derived such conditions only for general GARCH-type models. However, for example, it can be seen that the STGARCH$(1,1,1)$ model is stationary when $E\left[\log\left\{\beta_1^\circ + \left(\alpha_{11}^\circ + \frac{1}{2}|\alpha_{21}^\circ|\right)\eta_{t-1}^2\right\}\right] < 0$ (cf. Example 4 and Table 1 of Meitz and Saikkonen [23]).

Given the stationary solution $\{X_t\}$ and a parameter vector $\theta \in \Theta$, we define

$$c_t(\alpha) = \omega + \sum_{i=1}^{q} \alpha_{1i} X_{t-i}^2 + \left(\sum_{i=1}^{q} \alpha_{2i} X_{t-i}^2\right) F(X_{t-d}, \gamma),$$

where $\alpha = (\gamma, \omega, \alpha_{11}, ..., \alpha_{1q}, \alpha_{21}, ..., \alpha_{2q})$. Note that the polynomial $\beta(z) = 1 - \sum_{j=1}^{p} \beta_j z^j$ has all its zeros outside the unit disc because of (3.2). Define $\sigma_t^2(\theta) = \beta(B)^{-1} c_t(\alpha)$, where B is the backshift operator. Then, we have the following.

Theorem 3.1. *Let $\{X_t\}$ be a stationary process satisfying (3.1) and suppose that*

(**a**) $\alpha_{2i}^\circ \neq 0$ *for some* $1 \leq i \leq q$ *and* $\gamma^\circ > 0$.

(**b**) *The support of the distribution of* η_1 *is* $\mathbb{R}$.

Then, if $\sigma_t^2 = \sigma_t^2(\theta)$ *a.s. for some* $t \in \mathbb{Z}$ *and* $\theta \in \Theta$, *we have* $\theta = \theta^\circ$.

Remark 3.1. It is remarkable that the identifiability in the STGARCH models needs no restriction concerning orders p and q. The above theorem shows that the STGARCH(p,q,d) models can be consistently estimated by fitting any STGARCH(p^*,q^*,d) models with $p^* \geq p$ and $q^* \geq q$. However, this is not true for GARCH and AGARCH models, wherein conditions such as (c) in Theorem 3.2 below are necessary. See Francq and Zakoïan [8] and Straumann and Mikosch [29].

Remark 3.2. As pointed out by a referee, the common root condition for the STGARCH models is not required owing to the reasons described below. Consider a STGARCH$(0,1,d)$ model and let σ_t^2 be the conditional variance. Multiplying $(1-\beta B)$ to both sides of the volatility equation, we get $(1-\beta B)\sigma_t^2 = (1-\beta)\omega + \alpha_{11} X_{t-1}^2 - \beta\alpha_{11} X_{t-2}^2 + \alpha_{21} X_{t-1}^2 F(X_{t-d}, \gamma) - \beta\alpha_{21} X_{t-2}^2 F(X_{t-d-1}, \gamma)$.

This, however, is not expressible as a form of STGARCH$(1,2,d)$ models, unlike we see in GARCH and AGARCH models.

Remark 3.3. As in the case of the AGARCH model in Section 2, the support needs not be $\mathbb{R}$. For example, $\mathrm{supp}(\eta_1) = \mathbb{Z}$ is sufficient.

Condition (a) in Theorem 3.1 suggests that there exists a smooth transition mechanism, that is, conditional variances asymmetrically respond to positive and negative news. When it fails, the STGARCH model becomes a standard GARCH model. The following theorem demonstrates that model parameters in (3.1) are only partially identified when no such transition mechanism exists.

Theorem 3.2. *Let $\{X_t\}$ be a stationary process satisfying (3.1) with $\gamma^\circ = 0$ or $\alpha^\circ_{2i} = 0$, $i = 1, ..., q$. Suppose that condition (b) in Theorem 3.1 and the following condition hold:*

(**c**) *$\alpha^\circ_{1i} > 0$ for some $1 \le i \le q$, $(\alpha^\circ_{1q}, \beta^\circ_p) \neq (0,0)$, and the polynomials $\alpha^\circ_1(z) = \sum_{i=1}^q \alpha^\circ_{1i} z^i$ and $\beta^\circ(z) = 1 - \sum_{j=1}^p \beta^\circ_j z^j$ have no common zeros.*

If $\sigma_t^2 = \sigma_t^2(\theta)$ a.s. for some $t \in \mathbb{Z}$ and $\theta \in \Theta$, then $\omega = \omega^\circ$, $\alpha_{1i} = \alpha^\circ_{1i}$, $\beta_j = \beta^\circ_j$ for $1 \le i \le q$, $1 \le j \le p$, and either $\gamma = 0$ or $\alpha_{2i} = 0$, $1 \le i \le q$ holds.

Remark 3.4. The hypothesis testing of whether the smoothness mechanism exists has been studied by González-Rivera [11]. This is a type of testing problem wherein nuisance parameters are unidentifiable under the null hypothesis. In addition, inference in a similar situation has been studied by Hansen [12] and Francq *et al.* [7].

3.2. Threshold Poisson autoregressive models

Poisson autoregressive models (or integer-valued GARCH models) are used to model time series of counts with over-dispersion and have been widely applied in fields ranging from finance to epidemiology to estimate, for example, the number of transactions per minute of certain stocks and the daily epileptic seizure counts of patients. See Fokianos *et al.* [5], Kang and Lee [15], and the references therein.

Let $\{X_t : t \ge 0\}$ be a time series of counts and $\{\lambda_t : t \ge 0\}$ its intensity process. Let $\mathcal{F}_{0,t}$ denote the σ-field generated $\{\lambda_0, X_0, ..., X_t\}$. An integer-valued threshold GARCH (INTGARCH) model is then defined by

$$
\begin{aligned}
&X_t | \mathcal{F}_{0,t-1} \sim \mathrm{Poisson}(\lambda_t)\,, \\
&\lambda_t = \omega^\circ + \alpha^\circ_1 X_{t-1} + (\alpha^\circ_2 - \alpha^\circ_1)(X_{t-1} - l^\circ)^+ + \beta^\circ \lambda_{t-1}\,,
\end{aligned}
\tag{3.3}
$$

for $t \geq 1$, where a^+ denotes $\max\{0, a\}$. We assume that the true parameter vector $\theta^\circ = (\omega^\circ, \alpha_1^\circ, \alpha_2^\circ, \beta^\circ, l^\circ)$ belongs to a parameter space $\Theta = (0, \infty) \times [0, 1)^3 \times \mathbb{N}$. Theorem 2.1 of Neumann [25] indicates that if $\beta^\circ + \max\{\alpha_1^\circ, \alpha_2^\circ\} < 1$, there exists a unique stationary bivariate process $\{(X_t, \lambda_t) : t \geq 0\}$ satisfying (3.3). Then, the time domain can be extended from $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ to $\mathbb{Z}$. Franke *et al.* [9] considered the conditional LS estimation in these models.

Given the stationary process $\{X_t : t \in \mathbb{Z}\}$ and a parameter vector $\theta \in \Theta$, we define a stationary process $\{\lambda_t(\theta)\}$ as the solution of

$$\lambda_t(\theta) \;=\; \omega + \alpha_1 X_{t-1} + (\alpha_2 - \alpha_1)(X_{t-1} - l)^+ + \beta \lambda_{t-1}(\theta)\;, \qquad t \in \mathbb{Z}\;.$$

Then, we have the following.

Theorem 3.3. *Suppose that $\{X_t : t \in \mathbb{Z}\}$ is a stationary process satisfying (3.3) and $\alpha_1^\circ \neq \alpha_2^\circ$. Then, if $\lambda_t = \lambda_t(\theta)$ a.s. for some $t \in \mathbb{Z}$ and $\theta \in \Theta$, we have $\theta = \theta^\circ$.*

Remark 3.5. When $\alpha_1^\circ = \alpha_2^\circ > 0$, Model (3.3) becomes an integer-valued GARCH(1,1) model. In this case, it can be seen that parameters, except the threshold parameter l, are identifiable.

3.3. General Poisson autoregressive models

Neumann [25] considered a class of nonlinear Poisson autoregressive models $\{X_t : t \in \mathbb{Z}\}$ of counts with intensity process $\{\lambda_t : t \in \mathbb{Z}\}$ such as

$$X_t | \mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t)\;, \qquad \lambda_t = f(\lambda_{t-1}, X_{t-1}, \theta^\circ)\;, \tag{3.4}$$

for some known function $f : [0, \infty) \times \mathbb{N}_0 \times \Theta \to [0, \infty)$. According to Theorems 2.1 and 3.1 of Neumann [25], when $f(\cdot, \theta^\circ)$ satisfies the following contractive condition:

$$|f(\lambda, y, \theta^\circ) - f(\lambda', y', \theta^\circ)| \;\leq\; \kappa_1 |\lambda - \lambda'| + \kappa_2 |y - y'|\;, \quad \forall \lambda, \lambda' \geq 0\;,\;\; \forall y, y' \in \mathbb{N}_0\;,$$

where $\kappa_1, \kappa_2 \geq 0$ and $\kappa_1 + \kappa_2 < 1$, there exists a stationary process $\{(X_t, \lambda_t)\}$ with $\lambda_t \in \mathcal{F}_{t-1}$ satisfying (3.4). Further, in view of Theorem 3.1 in Neumann [25], one can define a stationary process $\{\lambda_t(\theta)\}$ satisfying

$$\lambda_t(\theta) = f(\lambda_{t-1}(\theta), X_{t-1}, \theta)\;, \qquad \forall t \in \mathbb{Z}\;,$$

for the stationary process $\{X_t\}$ and parameter vector $\theta \in \Theta$. Fokianos and Tjøstheim [6] studied ML estimation in these models.

The following theorem presents the mild requirements of f for their identifiability assumptions. Its proof is straightforward in view of the proof of Theorem 3.3.

Theorem 3.4. *Let $\{(X_t, \lambda_t)\}$ be a stationary process satisfying (3.4) and suppose that*

(**a**) *For each $\theta \in \Theta$, $f(\cdot, \theta)$ is continuous on $\mathrm{supp}(\lambda_1) \times \mathbb{N}_0$.*

(**b**) *$f(\lambda, y, \theta) = f(\lambda, y, \theta^\circ)$, $\forall \lambda \in \mathrm{supp}(\lambda_1)$, $\forall y \in \mathbb{N}_0$ implies $\theta = \theta^\circ$.*

Then, if $\lambda_t = \lambda_t(\theta)$ a.s. for some $t \in \mathbb{Z}$ and $\theta \in \Theta$, $\theta = \theta^\circ$.

3.4. Multiple regime smooth transition autoregressive models

Regime switching models for financial data have received considerable attention. For example, Teräsvirta [32] studied inference for two-regime STAR models and McAleer and Medeiros [21] and Li and Ling [20] considered multiple-regime smooth transition and threshold AR models. In this subsection, we consider the nonlinear LS estimation in a multiple-regime STAR model with heteroscedastic errors proposed by McAleer and Medeiros [21].

Suppose that $\{X_t\}$ follows a multiple-regime STAR model of order p with $M+1$ (limiting) regimes, that is,

$$X_t \;=\; \beta_0^{\circ T} \mathbf{X}_{t-1} + \sum_{i=1}^{M} \beta_i^{\circ T} \mathbf{X}_{t-1} G(X_{t-d^\circ}; \gamma_i^\circ, c_i^\circ) + \varepsilon_t \,, \tag{3.5}$$

where $\{\varepsilon_t\}$ is white noise, $\beta_i^\circ = (\phi_{i0}^\circ, \phi_{i1}^\circ, ..., \phi_{ip}^\circ)^T$ for $0 \le i \le M$, $\mathbf{X}_{t-1} = (1, X_{t-1}, ..., X_{t-p})^T$, and $G(X_{t-d^\circ}; \gamma_i^\circ, c_i^\circ)$ is a logistic transition function given by

$$G(X_{t-d^\circ}; \gamma_i^\circ, c_i^\circ) \;=\; \frac{1}{1 + e^{-\gamma_i^\circ (X_{t-d^\circ} - c_i^\circ)}} \,, \tag{3.6}$$

wherein the regime switches according to the value of transition variable X_{t-d°: $d^\circ \in \{1, ..., p\}$ is a delay parameter, $-\infty < c_1^\circ < \cdots < c_M^\circ < \infty$ are threshold parameters, and $\gamma_i^\circ > 0$, $i = 1, ..., M$, are smoothing parameters. When γ_i° is quite large, Model (3.5) is barely distinguishable from the threshold model studied by Li and Ling [20].

In the literature, one can find sufficient conditions under which Model (3.5) is stationary when the error terms are iid. For example, Theorem 2 of McAleer and Medeiros [21] ensures the stationarity of Model (3.5) of order 1. Using the same reasoning and Lemma 2.1 of Berkes *et al.* [2], we can see that Model (3.5) has a stationary solution if

$$\sum_{j=1}^{p} \sup_{x\in\mathbb{R}} \left| \phi^\circ_{0j} + \sum_{i=1}^{M} \phi^\circ_{ij} G(x; \gamma^\circ_i, c^\circ_i) \right| < 1 .$$

It is also true if $\max_{0\le i\le M} \sum_{j=1}^{p} \left|\sum_{k=0}^{i} \phi^\circ_{kj}\right| < 1$, which can be deduced from Theorem 3.2 and Example 3.6 in An and Huang [1].

We denote by $\theta = (\beta_0^T, \beta_1^T, ..., \beta_M^T, \gamma_1, ..., \gamma_M, c_1, ..., c_M, d)^T$ a parameter vector belonging to a parameter space $\Theta \subset \mathbb{R}^{(M+1)(p+1)+2M} \times \{1, ..., p\}$ and set

$$m(X_{t-1}, ..., X_{t-p}, \theta) \;=\; {\beta_0}^T \mathbf{X}_{t-1} + \sum_{i=1}^{M} {\beta_i}^T \mathbf{X}_{t-1} G(X_{t-d}; \gamma_i, c_i) .$$

Then, we have the following.

Theorem 3.5. *Let $\{X_t\}$ be a stationary process satisfying (3.5). Assume that*

- (**a**) *For each $i = 1, ..., M$, $\beta^\circ_i \neq (0, ..., 0)^T \in \mathbb{R}^{p+1}$.*
- (**b**) *The support of the stationary distribution of $(X_p, ..., X_1)$ is $\mathbb{R}^p$.*
- (**c**) *The parameter space Θ satisfies that $\gamma_i > 0$, $i = 1, ..., M$, and $-\infty < c_1 < \cdots < c_M < \infty$.*

Then, if $m(X_{t-1}, ..., X_{t-p}, \theta^\circ) = m(X_{t-1}, ..., X_{t-p}, \theta)$ a.s. for some $t \in \mathbb{Z}$ and $\theta \in \Theta$, we have $\theta = \theta^\circ$.

Remark 3.6. Theorem 3.5 is closely related to the identifiability of the finite mixture of logistic distributions (see Lemma 4.1 in Section 4). Although the restriction on threshold parameters has a natural interpretation, it is not necessarily required. In fact, if we only assume that $(\gamma^\circ_i, c^\circ_i), i = 1, ..., M$, are distinct, instead of the condition $c^\circ_i < c^\circ_{i+1}$, then Model (3.5) is weakly identifiable in the sense of Redner and Walker [27].

4. PROOFS

Proof of Theorem 3.1: We only prove the theorem when $d = 1$ since the other cases can be handled similarly. Owing to the stationarity, we have $\sigma_t^2 = \sigma_t^2(\theta)$ a.s. for any $t \in \mathbb{Z}$. Since $\beta^\circ(z) \neq 0$ for $|z| \le 1$ and $\sigma_t^2 = \beta^\circ(B)^{-1} c_t(\alpha^\circ)$, we can express

$$(4.1) \qquad c_t(\alpha) \;=\; \beta(B)\sigma_t^2(\theta) \;=\; \beta(B)\beta^\circ(B)^{-1} c_t(\alpha^\circ) \;=\; c_t(\alpha^\circ) + \sum_{j=1}^{\infty} b_j c_{t-j}(\alpha^\circ) ,$$

where $1+\sum_{j=1}^{\infty} b_j z^j = \beta(z)/\beta^{\circ}(z)$ for $|z| \le 1$. As discussed in Section 2, we can express (4.1) as a function of η_{t-1} and $\mathcal{F}_{t-2}$-measurable random variables:

$$\begin{aligned} g_1(\eta_{t-1}, \sigma_{t-1}, A_{t,2}, B_{t,2}, A^{\circ}_{t,2}, B^{\circ}_{t,2}, D_{t,2}) \ &:= \\ := \ & (\alpha_{11}-\alpha^{\circ}_{11})\sigma^2_{t-1}\eta^2_{t-1} + A_{t,2} - A^{\circ}_{t,2} + \big(\alpha_{21}\sigma^2_{t-1}\eta^2_{t-1} + B_{t,2}\big) F(\sigma_{t-1}\eta_{t-1}, \gamma) \\ & - \big(\alpha^{\circ}_{21}\sigma^2_{t-1}\eta^2_{t-1} + B^{\circ}_{t,2}\big) F(\sigma_{t-1}\eta_{t-1}, \gamma^{\circ}) - D_{t,2} \\ = \ & 0 \quad \text{a.}s., \end{aligned}$$

where for $2 \le i^* \le q$ and $2 \le k$,

$$A_{t,i^*} = \omega + \sum_{i=i^*}^{q} \alpha_{1i} X^2_{t-i}\,, \qquad B_{t,i^*} = \sum_{i=i^*}^{q} \alpha_{2i} X^2_{t-i}\,, \qquad D_{t,k} = \sum_{j=k-1}^{\infty} b_j c_{t-j}(\alpha^{\circ})\,,$$
$$A^{\circ}_{t,i^*} = \omega^{\circ} + \sum_{i=i^*}^{q} \alpha^{\circ}_{1i} X^2_{t-i}\,, \quad B^{\circ}_{t,i^*} = \sum_{i=i^*}^{q} \alpha^{\circ}_{2i} X^2_{t-i}\,.$$

Using the arguments that obtain (2.4) and condition (b), we can see that with probability 1, $g_1(x, \sigma_{t-1}, A_{t,2}, B_{t,2}, A^{\circ}_{t,2}, B^{\circ}_{t,2}, D_{t,2}) = 0$ for all $x \in \mathbb{R}$. Particularly, this implies

$$(4.2) \qquad g_1(0, \sigma_{t-1}, A_{t,2}, B_{t,2}, A^{\circ}_{t,2}, B^{\circ}_{t,2}, D_{t,2}) \ = \ A_{t,2} - A^{\circ}_{t,2} - D_{t,2} \ = \ 0 \quad \text{a.s..}$$

Then, viewing (4.2) as a function of η_{t-2} and $\mathcal{F}_{t-3}$-measurable random variables, we can express

$$\begin{aligned} g_2(\eta_{t-2}, \sigma_{t-2}, A_{t,3}, A^{\circ}_{t,3}, A^{\circ}_{t-1,2}, B^{\circ}_{t-1,2}, D_{t,3}) \ &:= \\ := \ & (\alpha_{12}-\alpha^{\circ}_{12})\sigma^2_{t-2}\eta^2_{t-2} + A_{t,3} - A^{\circ}_{t,3} - b_1 c_{t-1}(\alpha^{\circ}) - D_{t,3} \\ (4.3) \qquad = \ & (\alpha_{12}-\alpha^{\circ}_{12})\sigma^2_{t-2}\eta^2_{t-2} + A_{t,3} - A^{\circ}_{t,3} - D_{t,3} \\ & - b_1\big\{\alpha^{\circ}_{11}\sigma^2_{t-2}\eta^2_{t-2} + A^{\circ}_{t-1,2} + \big(\alpha^{\circ}_{21}\sigma^2_{t-2}\eta^2_{t-2} + B^{\circ}_{t-1,2}\big) F(\sigma_{t-2}\eta_{t-2}, \gamma^{\circ})\big\} \\ = \ & 0 \quad \text{a.s.}\,, \end{aligned}$$

which entails

$$(4.4) \qquad P\Big(g_2(x, \sigma_{t-2}, A_{t,3}, A^{\circ}_{t,3}, A^{\circ}_{t-1,2}, B^{\circ}_{t-1,2}, D_{t,3}) = 0,\ \forall x \in \mathbb{R}\Big) \ = \ 1\,.$$

Note that if

$$(4.5) \qquad f(x) \ := \ ax^2 + b + (cx^2 + d)F(\sigma x, \gamma^{\circ}) \ = \ 0$$

for all $x \in \mathbb{R}$, where a, b, c, d, $\sigma > 0$, $\gamma^{\circ} > 0$ are real numbers, because $\lim_{x\to\pm\infty} x^{-2} f(x) = 0$ and $\lim_{x\to\pm\infty} f(x) = 0$, it must hold that $a = c = 0$ and $b = d = 0$. Then, combining this and (4.4), we get $b_1\alpha^{\circ}_{21} = 0$ and $b_1 B^{\circ}_{t-1,2} = 0$ a.s.. Further, $B^{\circ}_{t-1,2} = 0$ a.s. if and only if $\alpha^{\circ}_{22} = \cdots = \alpha^{\circ}_{2q} = 0$. Due to condition (a) and (4.3), we have $b_1 = 0$ and $A_{t,3} - A^{\circ}_{t,3} - D_{t,3} = 0$ a.s., and similarly, it can be seen that $b_k = 0$, $k \ge 2$, $A_{t,k+2} - A^{\circ}_{t,k+2} - D_{t,k+2} = 0$ a.s., $2 \le k \le q-2$, and $\omega - \omega^{\circ} - D_{t,k+2} = 0$ a.s., $k \ge q-1$. This implies $\beta(\cdot) = \beta^{\circ}(\cdot)$, $\omega = \omega^{\circ}$, and

$A_{t,2} = A^\circ_{t,2}, ..., A_{t,q} = A^\circ_{t,q}$ a.s., and subsequently, $\alpha_{1q} = \alpha^\circ_{1q}, ..., \alpha_{12} = \alpha^\circ_{12}$. From this and (4.1), we can obtain

$$
\begin{aligned}
h_1(\eta_{t-1}, \sigma_{t-1}, B_{t,2}, B^\circ_{t,2}) \;&:= \\
:=\; &(\alpha_{11} - \alpha^\circ_{11})\sigma^2_{t-1}\eta^2_{t-1} + \big(\alpha_{21}\sigma^2_{t-1}\eta^2_{t-1} + B_{t,2}\big)\, F(\sigma_{t-1}\eta_{t-1}, \gamma) \\
&- \big(\alpha^\circ_{21}\sigma^2_{t-1}\eta^2_{t-1} + B^\circ_{t,2}\big)\, F(\sigma_{t-1}\eta_{t-1}, \gamma^\circ) \\
=\; &0 \quad \text{a.}s.
\end{aligned}
\tag{4.6}
$$

Suppose that $\gamma = 0$. Then, $F(X_{t-1}, \gamma) \equiv 0$, and using (4.5) and (4.6), we get $\alpha^\circ_{21} = 0$ and $B^\circ_{t,2} = 0$ a.s. Since this is a contradiction to condition (a), γ must be positive. Thus, from (4.6), we have

$$\lim_{x\to\infty} x^{-2} h_1(x, \sigma_{t-1}, B_{t,2}, B^\circ_{t,2}) \;=\; \sigma^2_{t-1}\left\{\alpha_{11} - \alpha^\circ_{11} - 2^{-1}(\alpha_{21} - \alpha^\circ_{21})\right\} \;=\; 0 \quad \text{a.s.}.$$

Further, taking the limit $x \to -\infty$, we obtain $\alpha_{11} = \alpha^\circ_{11}$ and $\alpha_{21} = \alpha^\circ_{21}$, so that

$$\lim_{x\to\infty} h_1(x, \sigma_{t-1}, B_{t,2}, B^\circ_{t,2}) \;=\; -2^{-1} B_{t,2} + 2^{-1} B^\circ_{t,2} = 0 \quad \text{a.s.}\,,$$

which results in $\alpha_{2i} = \alpha^\circ_{2i},\ 2 \le i \le q$. Then, in view of (4.6), we obtain

$$
\begin{aligned}
h_2(\eta_{t-1}, \sigma_{t-1}, B^\circ_{t,2}) \;&:=\; \big(\alpha^\circ_{21}\sigma^2_{t-1}\eta^2_{t-1} + B^\circ_{t,2}\big)\left(\frac{1}{1+e^{\gamma\sigma_{t-1}\eta_{t-1}}} - \frac{1}{1+e^{\gamma^\circ\sigma_{t-1}\eta_{t-1}}}\right) \\
&=\; 0 \quad \text{a.s.}\,.
\end{aligned}
$$

If $\gamma < \gamma^\circ$ and additionally if $\alpha^\circ_{21} \neq 0$, we should have

$$\lim_{x\to\infty} x^{-2} e^{\gamma\sigma_{t-1}x} h_2(x, \sigma_{t-1}, B^\circ_{t,2}) \;=\; \alpha^\circ_{21}\sigma^2_{t-1} \;=\; 0 \quad \text{a.s.}\,,$$

which leads to a contradiction. However, if $\alpha^\circ_{21} = 0$, we have $\lim_{x\to\infty} e^{\gamma\sigma_{t-1}x} h_2(x, \sigma_{t-1}, B^\circ_{t,2}) = B^\circ_{t,2} = 0$ a.s., which also leads to a contradiction to condition (a). Hence, we must have $\gamma \ge \gamma^\circ$. Since $\gamma > \gamma^\circ$ is also impossible, we conclude that $\gamma = \gamma^\circ$, which completes the proof. □

Proof of Theorem 3.2: As in handling (4.3), we follow the same lines in the proof of Theorem 3.1 to obtain

$$
\begin{aligned}
g'_2(\eta_{t-2}, \sigma_{t-2}, A_{t,3}, A^\circ_{t,3}, A^\circ_{t-1,2}, D_{t,3}) \;&:= \\
:=\; &(\alpha_{12} - \alpha^\circ_{12} - b_1\alpha^\circ_{11})\sigma^2_{t-2}\eta^2_{t-2} + A_{t,3} - A^\circ_{t,3} - b_1 A^\circ_{t-1,2} - D_{t,3} \\
=\; &0 \quad \text{a.s.}\,.
\end{aligned}
$$

Then, as in handling (4.2), we get $A_{t,3} - A^\circ_{t,3} - b_1 A^\circ_{t-1,2} - D_{t,3} = 0$ a.s. Similarly, it can be seen that $\omega - \omega^\circ - b_1 A^\circ_{t-1,q} - b_2 A^\circ_{t-2,q-1} - \cdots - b_{q-1} A^\circ_{t-q+1,2} - D_{t,q+1} = 0$ a.s. Then, with probability 1, for all $x \in \mathbb{R}$,

$$
\begin{aligned}
g(x) \;:=\; &(\omega - \omega^\circ) - b_1\big(\alpha^\circ_{1q}\sigma^2_{t-q-1}x^2 + \omega^\circ\big) - b_2\big(\alpha^\circ_{1,q-1}\sigma^2_{t-q-1}x^2 + A^\circ_{t-2,q}\big) - \cdots \\
&- b_{q-1}\big(\alpha^\circ_{12}\sigma^2_{t-q-1}x^2 + A^\circ_{t-q+1,3}\big) - b_q\big(\alpha^\circ_{11}\sigma^2_{t-q-1}x^2 + A^\circ_{t-q,2}\big) - D_{t,q+2} \\
=\; &0\,,
\end{aligned}
\tag{4.7}
$$

which, in turn, implies

$$P\left(\lim_{x\to\infty}\frac{-g(x)}{\sigma_{t-q-1}^2 x^2} = b_1\alpha_{1q}^\circ + \cdots + b_q\alpha_{11}^\circ = 0\right) = 1\ .$$

In fact, we can obtain an analogous relationship between η_{t-q-k} and $\mathcal{F}_{t-q-k-1}$-measurable random variables, $k \geq 2$, and as such, $b_k\alpha_{1q}^\circ + \cdots + b_{k+q-1}\alpha_{11}^\circ = 0$ for all $k \geq 1$, which implies that $\beta(z)\beta^\circ(z)^{-1}\alpha_1^\circ(z)$ is a polynomial of at most q orders. Then, using condition (c) and the arguments similar to those in Straumann and Mikosch (2006), p. 2481, we can see that $\beta(\cdot) = \beta^\circ(\cdot)$, and thus, $b_j = 0$ for $j \geq 1$. Combining this, (4.2) and (4.7), we get $A_{t,2} = A_{t,2}^\circ, ..., A_{t,q} = A_{t,q}^\circ$ a.s. and $\omega = \omega^\circ$, which, in turn, implies $\alpha_{1q} = \alpha_{1q}^\circ, ..., \alpha_{12} = \alpha_{12}^\circ$. Hence, (4.1) can be reexpressed as

$$\begin{aligned} h_1'(\eta_{t-1}, \sigma_{t-1}, B_{t,2}) &:= (\alpha_{11} - \alpha_{11}^\circ)\sigma_{t-1}^2\eta_{t-1}^2 + \left(\alpha_{21}\sigma_{t-1}^2\eta_{t-1}^2 + B_{t,2}\right) F(\sigma_{t-1}\eta_{t-1}, \gamma) \\ &= 0 \quad \text{a.}s.. \end{aligned}$$

From this, we can easily obtain $\alpha_{11} = \alpha_{11}^\circ$ and the same equation as in (4.5), which finally leads to $\alpha_{21} = \cdots = \alpha_{2q} = 0$. This completes the proof. □

Proof of Theorem 3.3: First, we conjecture that the support of the stationary distribution of (X_1, λ_1) is a Cartesian product of $\mathbb{N}_0$ and $\text{supp}(\lambda_1)$. If it is not true, there exists $(m', \lambda') \in \mathbb{N}_0 \times \text{supp}(\lambda_1)$ such that $(m', \lambda') \notin \text{supp}(X_1, \lambda_1)$, and for some positive real number r,

$$0 = P\left(X_1 = m', \lambda_1 \in (\lambda' - r, \lambda' + r)\right) = \int_{\lambda'-r}^{\lambda'+r} (m'!)^{-1} e^{-u} u^{m'} \mathrm{d}F_{\lambda_1}(u)\ ,$$

where F_{λ_1} is the distribution function of λ_1. Since the integrand is positive, it must hold that $P(\lambda_1 \in (\lambda' - r, \lambda' + r)) = 0$, which, however, contradicts to the fact that $\lambda' \in \text{supp}(\lambda_1)$. Thus, our conjecture is validated.

Note that owing to the stationarity, for all $t \in \mathbb{Z}$,

$$\begin{aligned} g(X_{t-1}, \lambda_{t-1}) &:= (\omega - \omega^\circ) + (\alpha_1 - \alpha_1^\circ)X_{t-1} + (\alpha_2 - \alpha_1)(X_{t-1} - l)^+ \\ &\quad - (\alpha_2^\circ - \alpha_1^\circ)(X_{t-1} - l^\circ)^+ + (\beta - \beta^\circ)\lambda_{t-1} \\ &= 0 \quad \text{a.s.}\ , \end{aligned}$$

and therefore,

$$(4.8)\qquad g(m, \lambda) = 0 \qquad \text{for all} \quad m \in \mathbb{N}_0 \quad \text{and} \quad \lambda \in \text{supp}(\lambda_1)\ ,$$

since $g(\cdot)$ is continuous and $\text{supp}(X_1, \lambda_1) = \mathbb{N}_0 \times \text{supp}(\lambda_1)$. In particular, $g(0, \lambda) = (\omega - \omega^\circ) + (\beta - \beta^\circ)\lambda = 0$ for any $\lambda \in \text{supp}(\lambda_1)$. Note that λ_t is not degenerate when $\alpha_1^\circ \neq \alpha_2^\circ$, since otherwise, X_{t-1} should be degenerate. Thus, we have $\omega = \omega^\circ$ and $\beta = \beta^\circ$, so that $g(1, \lambda) = \alpha_1 - \alpha_1^\circ = 0$. Further, it follows from (4.8) that $\lim_{m\to\infty} m^{-1} g(m, \lambda) = \alpha_2 - \alpha_2^\circ = 0$. Then, using the fact that $g(l, \lambda) = g(l^\circ, \lambda) = 0$ and $\alpha_1^\circ \neq \alpha_2^\circ$, we obtain $l = l^\circ$, which completes the proof. □

Proof of Theorem 3.5: For simplicity, we assume that $d^\circ = 1$: the other cases can be handled similarly. From condition (b) and the continuity of $m(\cdot, \theta)$, we can see that

$$m(x_1, ..., x_p, \theta^\circ) = m(x_1, ..., x_p, \theta) \;, \qquad \forall x_j \in \mathbb{R}\,, \quad 1 \le j \le p\,. \tag{4.9}$$

Suppose that $d \neq 1$. From (4.9), we can express

$$\begin{aligned} m(x_1, ..., x_p, \theta^\circ) - m(x_1, ..., x_p, \theta) \;=& \\ =\;& \left\{ f_0^\circ(\mathbf{x}_2) - f_0(\mathbf{x}_2) - \sum_{i=1}^{M} f_i(\mathbf{x}_2) G(x_d; \gamma_i, c_i) \right\} \\ &+ \left\{ \phi_{01}^\circ - \phi_{01} - \sum_{i=1}^{M} \phi_{i1} G(x_d; \gamma_i, c_i) \right\} x_1 \\ &+ \sum_{i=1}^{M} \left(f_i^\circ(\mathbf{x}_2) + \phi_{i1}^\circ x_1\right) G(x_1; \gamma_i^\circ, c_i^\circ) \\ =\;& 0\,, \end{aligned} \tag{4.10}$$

where $G(\cdot)$ is the one in (3.6), $\mathbf{x}_2 = (x_2, ..., x_p)^T$, and

$$f_i^\circ(\mathbf{x}_2) = \phi_{i0}^\circ + \sum_{2 \le j \le p} \phi_{ij}^\circ x_j\,, \quad f_i(\mathbf{x}_2) = \phi_{i0} + \sum_{2 \le j \le p} \phi_{ij} x_j\,, \qquad \text{for} \quad i = 0, 1, ..., M\,.$$

Then, applying Lemma 4.1 below to (4.10), we have $\phi_{11}^\circ = 0$ and $f_1^\circ(\mathbf{x}_2) = 0$ for each $\mathbf{x}_2 \in \mathbb{R}^{p-1}$, which, however, contradicts to condition (a). Thus, it must hold that $d = d^\circ = 1$. Owing to the above, we can reexpress (4.9) as

$$\begin{aligned} \left(f_0^\circ(\mathbf{x}_2) + \phi_{01}^\circ x_1\right) + \sum_{i=1}^{M} \left(f_i^\circ(\mathbf{x}_2) + \phi_{i1}^\circ x_1\right) G(x_1; \gamma_i^\circ, c_i^\circ) \;=& \\ =\; \left(f_0(\mathbf{x}_2) + \phi_{01} x_1\right) + \sum_{i=1}^{M} \left(f_i(\mathbf{x}_2) + \phi_{i1} x_1\right) G(x_1; \gamma_i, c_i)\,,& \\ \forall x_j \in \mathbb{R}\,, \quad 1 \le j \le p\,.& \end{aligned} \tag{4.11}$$

Lemma 4.1 ensures that a family of real-valued functions $\mathcal{G} = \{1, i(\cdot)\} \cup \{G(\cdot; \gamma, c) : \gamma > 0, c \in \mathbb{R}\} \cup \{i(\cdot) G(\cdot; \gamma, c) : \gamma > 0, c \in \mathbb{R}\}$, where $i(\cdot)$ is an identity function, i.e., $i(y) = y$, are linearly independent. Thus, any element of the linear span of $\mathcal{G}$ is uniquely represented as a linear combination of the elements of $\mathcal{G}$: see Yakowitz and Spragins [34]. Further, there exists a vector $\mathbf{x}_2' \in \mathbb{R}^{p-1}$ such that $(f_i^\circ(\mathbf{x}_2'), \phi_{i1}^\circ) \neq (0, 0)$ for all $i = 1, ..., M$; unless otherwise, $\phi_{i0}^\circ = \cdots = \phi_{ip}^\circ = 0$ for some i, which contradicts condition (a). Then, viewing (4.11) with $\mathbf{x}_2$ substituted by $\mathbf{x}_2'$ as a function of x_1 and using condition (c), we obtain $\phi_{01}^\circ = \phi_{01}$ and $\phi_{i1}^\circ = \phi_{i1}$, $\gamma_i^\circ = \gamma_i$, $c_i^\circ = c_i$ for $i = 1, ..., M$. Subsequently, owing to (4.11), for all $x_1 \in \mathbb{R}$ and $\mathbf{x}_2 \in \mathbb{R}^{p-1}$, we get

$$\left(f_0^\circ(\mathbf{x}_2) - f_0(\mathbf{x}_2)\right) + \sum_{i=1}^{M} \left(f_i^\circ(\mathbf{x}_2) - f_i(\mathbf{x}_2)\right) G(x_1; \gamma_i^\circ, c_i^\circ) \;=\; 0\,.$$

Then, applying Lemma 4.1 again, we conclude that $\phi_{i0}^{\circ} = \phi_{i0}$ and $\phi_{ij}^{\circ} = \phi_{ij}$, $j = 2, ..., p$, $i = 0, 1, ..., M$. This completes the proof. $\square$

Lemma 4.1. *Let $(\gamma_1, c_1), ..., (\gamma_k, c_k)$ be distinct real vectors with $\gamma_i > 0$, $i = 1, ..., k$. Suppose that for all $y \in \mathbb{R}$,*

$$d_{00} + d_{01}y + \sum_{i=1}^{k}(d_{i0} + d_{i1}y)\frac{1}{1 + e^{-\gamma_i(y-c_i)}} \;=\; 0\,. \tag{4.12}$$

Then, $d_{i0} = d_{i1} = 0$ for $i = 0, 1, ..., k$.

Proof: Denote by $g(y)$ the left-hand side of (4.12). Then, $\lim_{y\to-\infty} y^{-1}g(y) = d_{01} = 0$, and thus, $\lim_{y\to-\infty} g(y) = d_{00} = 0$. In what follows, for function $f : \mathbb{R} \to \mathbb{R}$, we denote by $\mathcal{L}\{f\}$ its two-sided Laplace transform, that is, $\mathcal{L}\{f(\cdot)\}(s) = \int_{-\infty}^{\infty} e^{-sy} f(y)\mathrm{d}y$. Note that the transform of the logistic distribution function is as follows:

$$F_0(s; \gamma, c) \;:=\; \mathcal{L}\{G(\cdot; \gamma, c)\}(s) \;=\; \frac{\pi\gamma^{-1}e^{-cs}}{\sin \pi\gamma^{-1}s}\,, \qquad 0 < s < \gamma\,.$$

Further,

$$\begin{aligned} F_1(s; \gamma, c) \;&:=\; \mathcal{L}\{i(\cdot)G(\cdot; \gamma, c)\}(s) \\ &=\; \frac{\pi\gamma^{-1}ce^{-cs}}{\sin \pi\gamma^{-1}s} + \frac{\pi^2\gamma^{-2}e^{-cs}\cos \pi\gamma^{-1}s}{\sin^2 \pi\gamma^{-1}s}\,, \qquad 0 < s < \gamma\,. \end{aligned}$$

Without loss of generality, assume that (γ_i, c_i), $i = 1, ..., k$, satisfy a lexicographical ordering, that is, $\gamma_i \le \gamma_{i+1}$ and $c_i < c_{i+1}$ when $\gamma_i = \gamma_{i+1}$. Suppose that $\gamma_1 = \cdots = \gamma_l < \gamma_{l+1} \le \cdots \le \gamma_k$ and $c_1 < \cdots < c_l$. Then, applying the two-sided Laplace transformation to (4.12), we have that for all $0 < s < \gamma_1$,

$$\sum_{i=1}^{k} d_{i0}F_0(s; \gamma_i, c_i) + \sum_{i=1}^{k} d_{i1}F_1(s; \gamma_i, c_i) \;=\; 0\,. \tag{4.13}$$

Since the numerator of the left-hand side of (4.13) is an analytic function on $\mathbb{R}$, (4.13) is still valid for all $s \in \mathbb{R}\backslash D$, where $D = \{s : s = \gamma_i m, 1 \le i \le k, m \in \mathbb{Z}\}$. Multiplying $\sin^2 \pi\gamma_1^{-1}s$ to both the sides of (4.13), we attain

$$\begin{aligned} &\sin \pi\gamma_1^{-1}s \sum_{i=1}^{l} \left\{d_{i0}\pi\gamma_1^{-1}e^{-c_i s} + d_{i1}\pi\gamma_1^{-1}c_i e^{-c_i s}\right\} \;+ \\ &+\; \sin^2 \pi\gamma_1^{-1}s \sum_{i=l+1}^{k} \left\{d_{i0}\frac{\pi\gamma_i^{-1}e^{-c_i s}}{\sin \pi\gamma_i^{-1}s} + d_{i1}\frac{\pi\gamma_i^{-1}c_i e^{-c_i s}}{\sin \pi\gamma_i^{-1}s}\right\} \\ &+\; \cos \pi\gamma_1^{-1}s \sum_{i=1}^{l} d_{i1}\pi^2\gamma_1^{-2}e^{-c_i s} + \sin^2 \pi\gamma_1^{-1}s \sum_{i=l+1}^{k} d_{i1}\frac{\pi^2\gamma_i^{-2}e^{-c_i s}\cos \pi\gamma_i^{-1}s}{\sin^2 \pi\gamma_i^{-1}s} \;=\; 0\,. \end{aligned}$$

Then, if we set $\mathbb{N}_1 = \{n \in \mathbb{N} : \gamma_1 n \neq \gamma_i m \text{ for all } l < i \leq k, m \in \mathbb{N}\}$, for any fixed $n \in \mathbb{N}_1$, letting $s \to \gamma_1 n$ through the values in $\mathbb{R} \backslash D$, we can have

$$\sum_{i=1}^{l} d_{i1} e^{-c_i \gamma_1 n} = 0 . \tag{4.14}$$

Since (4.14) holds for all $n \in \mathbb{N}_1$, multiplying $e^{c_1 \gamma_1 n}$ to both the sides of (4.14) and letting $n \to \infty$ through the values in $\mathbb{N}_1$, we get $d_{11} = 0$. Similarly, it can be seen that $d_{21} = \cdots = d_{l1} = 0$. Meanwhile, multiplying $\sin \pi \gamma_1^{-1} s$ to both the sides to (4.13) and letting $s \to \gamma_1 n$, we can have $\sum_{i=1}^{l} d_{i0} e^{-c_i \gamma_1 n} = 0$ for any $n \in \mathbb{N}_1$, and henceforth, $d_{10} = \cdots = d_{l0} = 0$. Continuing the above process, one can finally establish the lemma. □

Remark 4.1. Lemma 4.1 actually entails the identifiability of logistic mixture distributions (cf. Yakowitz and Spragins [34] and Sussmann [30]). Hwang and Ding [14] also proved the linear independence of logistic distributions and their density functions to deal with the identifiability problem in artificial neural networks. However, their results do not directly imply Lemma 4.1. Our proof is simpler and is based on Theorem 2 of Teicher [31].

ACKNOWLEDGMENTS

The authors thank the referees for their careful reading and valuable comments. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2012R1A2A2A01046092).

REFERENCES

[1] An, H.Z. and Huang, F.C. (1996). The geometrical ergodicity of nonlinear autoregressive models, *Statist. Sinica*, **6**, 943–956.

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

[3] Chan, K.S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model, *Ann. Statist.*, **21**, 520–533.

[4] Chan, K.S. and Tong, H. (1986). On estimating thresholds in autoregressive models, *J. Time Ser. Anal.*, **7**, 179–190.

[5] Fokianos, K.; Rahbek, A. and Tjøstheim, D. (2009). Poisson autoregression, *J. Amer. Statist. Assoc.*, **104**, 1430–1439. With electronic supplementary materials available online.

[6] Fokianos, K. and Tjøstheim, D. (2012). Nonlinear Poisson autoregression, *Ann. Inst. Statist. Math.*, **64**, 1205–1225.

[7] Francq, C.; Horvath, L. and Zakoïan, J.-M. (2010). Sup-tests for linearity in a general nonlinear AR(1) model, *Economet. Theor.*, **26**, 965–993.

[8] Francq, C. and Zakoïan, J.-M. (2004). Maximum likelihood estimation of pure GARCH and ARMA–GARCH processes, *Bernoulli*, **10**, 605–637.

[9] Franke, J.; Kirch, C. and Kamgaing, J. (2012). Changepoints in times series of counts, *J. Time Series Anal.*, **33**, 757–770.

[10] Glosten, L.; Jagannathan, R. and Runkle, D. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks, *J. Finance*, **48**, 1779–1801.

[11] González-Rivera, G. (1998). Smooth-transition GARCH models, *Stud. Nonlinear Dyn. Econom.*, **3**, 61–78.

[12] Hansen, B.E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis, *Econometrica*, **64**, 413–430.

[13] Hayashi, F. (2000). *Econometrics*, Princeton University Press, Princeton, NJ.

[14] Hwang, J.T.G. and Ding, A.A. (1997). Prediction intervals for artificial neural networks, *J. Amer. Statist. Assoc.*, **92**, 748–757.

[15] Kang, J. and Lee, S. (2014). Parameter change test for Poisson autoregressive models, *Scand. J. Statist.* (to appear).

[16] Komunjer, I. (2012). Global identification in nonlinear models with moment restrictions, *Economet. Theor.*, **28**, 719–729.

[17] Kristensen, D. and Rahbek, A. (2009). Asymptotics of the QMLE for nonlinear ARCH models, *J. Time Ser. Econom.*, **1**, 38.

[18] Lee, S. and Lee, T. (2012). Inference for Box–Cox transformed threshold GARCH models with nuisance parameters, *Scand. J. Statist.*, **39**, 568–589.

[19] Lee, S. and Noh, J. (2013). Quantile regression estimator for GARCH models, *Scand. J. Stat.*, **40**, 2–20.

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

[21] McAleer, M. and Medeiros, M.C. (2008). A multiple regime smooth transition heterogeneous autoregressive model for long memory and asymmetries, *J. Econometrics*, **147**, 104–119.

[22] Medeiros, M.C. and Veiga, A. (2009). Modeling multiple regimes in financial volatility with a flexible coefficient GARCH(1,1) model, *Economet. Theor.*, **25**, 117–161.

[23] Meitz, M. and Saikkonen, P. (2008). Ergodicity, mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD models, *Economet. Theor.*, **24**, 1291–1320.

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

[25] Neumann, M.H. (2011). Absolute regularity and ergodicity of Poisson count processes, *Bernoulli*, **17**, 1268–1284.

[26] Phillips, P.C.B. (1989). Partially identified econometric models, *Economet. Theor.*, **5**, 181–240.

[27] Redner, R.A. and Walker, H.F. (1984). Mixture densities, maximum likelihood and the EM algorithm, *SIAM Rev.*, **26**, 195–239.

[28] Rothenberg, T.J. (1971). Identification in parametric models, *Econometrica*, **39**, 577–591.

[29] Straumann, D. and Mikosch, T. (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach, *Ann. Statist.*, **34**, 2449–2495.

[30] Sussmann, H. (1992). Uniqueness of the weights for minimal feedforward nets with a given input-output map, *Neural Networks*, **5**, 589–593.

[31] Teicher, H. (1963). Identifiability of finite mixtures, *Ann. Math. Statist.*, **34**, 1265–1269.

[32] Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models, *J. Amer. Statist. Assoc.*, **89**, 208–218.

[33] Wu, C.-F. (1981). Asymptotic theory of nonlinear least squares estimation, *Ann. Statist.*, **9**, 501–513.

[34] Yakowitz, S.J. and Spragins, J.D. (1968). On the identifiability of finite mixtures, *Ann. Math. Statist.*, **39**, 209–214.