On Mixture Double Autoregressive Time Series Models

Guodong Li, Qianqian Zhu, Zhao Liu, and Wai Keung Li
Department of Statistics and Actuarial Science, University of Hong Kong, Hong Kong (gdli@hku.hk; qianqzhu@hku.hk; zhaoliu@hku.hk; hrntlwk@hku.hk)

This article proposes a mixture double autoregressive model by introducing the flexibility of mixture models to the double autoregressive model, a novel conditional heteroscedastic model recently proposed in the literature. To make it more flexible, the mixing proportions are further assumed to be time varying, and probabilistic properties including strict stationarity and higher order moments are derived. Inference tools including the maximum likelihood estimation, an expectation–maximization (EM) algorithm for searching the estimator and an information criterion for model selection are carefully studied for the logistic mixture double autoregressive model, which has two components and is encountered more frequently in practice. Monte Carlo experiments give further support to the new models, and the analysis of an empirical example is also reported.

KEYWORDS: Double autoregressive model; EM algorithm; Mixture model; Stationarity.

1. INTRODUCTION

The conditional heteroscedastic models have become a standard family of nonlinear time series models since the appearances of the autoregressive conditional heteroscedastic (ARCH) model (Engle 1982) and the generalized autoregressive conditional heteroscedastic (GARCH) model (Bollerslev 1986). Among hundreds of members in this family, the double autoregressive (AR) model recently has attracted more and more attentions; see Ling and Li (2008), Zhu and Ling (2013), and references therein. This model has the form of

$$y_t = \theta_0 + \sum_{i=1}^{p} \theta_i y_{t-i} + \varepsilon_t \sqrt{\beta_0 + \sum_{j=1}^{p} \beta_j y_{t-j}^2},$$

where $\beta_0 > 0$, $\beta_j \geq 0$, and $\{\varepsilon_t\}$ are identically and independently distributed (iid) random variables with mean zero and variance one. It is a special case of the AR-ARCH models in Weiss (1986), and will reduce to the ARCH model when θ_i's are zero. Comparing with the AR-ARCH or ARMA-GARCH model, the double AR model has two novel properties. First, it has an even larger parameter space than that of the commonly used AR model. For example, when $p = 1$, the double AR model may still be stationary even as $|\theta_1| \geq 1$ (Ling 2004). Second as we know, the finite fourth moment of the ARMA-GARCH process is unavoidable in deriving the asymptotic distribution of the Gaussian quasi-maximum likelihood estimation (Francq and Zakoian 2004), and this makes the available parameter space much narrower (Li and Li 2009). However, for the double AR model, we usually do not need to assume the moment condition on y_t to derive the asymptotic normality of its parameter estimators; see Ling (2007) and Zhu and Ling (2013).

In the meanwhile, many time series have exhibited a multimodal marginal or conditional distribution. For example, the classical Canadian lynx data was shown by Tong (1990) to have a bimodal marginal distribution, and Wong and Li (2000) showed that a bimodal conditional distribution is more suitable for certain stock prices since they may rise or decline sharply when markets become volatile. The mixture AR model was first proposed by Wong and Li (2000) to capture the phenomenon of multimodal conditional distributions, and Wong and Li (2001b) extended it to the mixture AR-ARCH model. The heavy tail is another important phenomenon in financial time series (Li and Li 2005, 2008), and the mixture processes can explain it to some extent; see Zhang, Li, and Yuen (2006) for a discussion of the mixture GARCH model. Another promising property is that a mixture of a stationary component and a nonstationary component may result in a stationary process.

Section 2 proposes a mixture double autoregressive (MDAR) model by introducing the flexibility of mixture models to the double AR time series model, a novel conditional heteroscedastic model proposed by Ling (2004). Moreover, it is natural to expect that some exogenous variables may affect the prediction and description of the time series via the mixing proportions (Wong and Li 2001a; Cheng, Yu, and Li 2009). To make our model more flexible, we further assume that the mixing proportions are time varying with a logit link function. The strict stationarity and ergodicity of the new model are derived, and we also discuss its existence of higher order moments. In practice, the most commonly used mixture model is that with two components, and the proposed model then reduces to the logistic MDAR model. For this special MDAR model, Section 3 derives the asymptotic normality of its maximum likelihood estimation (MLE). An expectation–maximization (EM) algorithm is employed to search for the MLE, and the observed information matrix as well as an information matrix based on the first-order derivatives are presented. The Bayesian information criterion (BIC) is also discussed for model selection in Section 3. Section 4 conducts several simulation experiments to study the finite-sample performance of these proposed inference tools in

Journal of Business & Economic Statistics
April 2017, Vol. 35, No. 2
DOI: 10.1080/07350015.2015.1102735
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/r/jbes.

Section 3, and an empirical example is analyzed to demonstrate the usefulness of the proposed models. All technical proofs are relegated to the Appendix.

2. MIXTURE DOUBLE AUTOREGRESSIVE MODELS

2.1 Mixture Double Autoregressive Models

Consider a time series $\{y_t\}$ with some exogenous variables $\{\mathbf{x}_t\}$, where $\mathbf{x}_t$ may include the lagged values of y_t. Let $\mathcal{F}_t = \sigma(y_t, y_{t-1}, \ldots)$ and $\Omega_t = \sigma(\mathbf{x}_t, \mathbf{x}_{t-1}, \ldots)$ be the σ-fields generated by $\{y_t, y_{t-1}, \ldots\}$ and $\{\mathbf{x}_t, \mathbf{x}_{t-1}, \ldots\}$, respectively. Suppose that the conditional distribution of y_t has the form of

$$F(y|\mathcal{F}_{t-1}, \Omega_t) = \sum_{k=1}^{K} \alpha_{kt} \Phi\left(\frac{y - \theta_{k0} - \theta_{k1} y_{t-1} - \cdots - \theta_{kp_k} y_{t-p_k}}{\sqrt{\beta_{k0} + \beta_{k1} y_{t-1}^2 + \cdots + \beta_{kp_k} y_{t-p_k}^2}}\right), \quad \text{for} \quad y \in \mathbb{R}, \tag{1}$$

and the mixing proportions follow the logit model (Agresti 2002),

$$\ln(\alpha_{jt}/\alpha_{Kt}) = \mathbf{x}_t' \boldsymbol{\varphi}_j, \qquad j = 1, \ldots, K-1, \tag{2}$$

where α_{Kt} is the baseline proportion, $\alpha_{kt} > 0$, $\sum_{k=1}^{K} \alpha_{kt} = 1$, $\beta_{k0} > 0$, $\beta_{kj} \geq 0$ for all j, and Φ is the cumulative distribution function of the standard normal distribution. Note that the logit model at (2) can be replaced by other link functions. In this article, model (1) is called the K-component mixture double autoregressive (MDAR) model.

As for other mixture time series models (Wong and Li 2000, 2001b), the shape of the conditional distribution at (1) changes over time. Especially, we have that

$$E(y_t|\mathcal{F}_{t-1}, \Omega_t) = \sum_{k=1}^{K} \alpha_{kt}(\theta_{k0} + \theta_{k1} y_{t-1} + \cdots + \theta_{kp_k} y_{t-p_k}) := \sum_{k=1}^{K} \alpha_{kt} \mu_{kt},$$

and

$$\text{var}(y_t|\mathcal{F}_{t-1}, \Omega_t) = \sum_{k=1}^{K} \alpha_{kt} h_{kt} + \sum_{k=1}^{K} \alpha_{kt} \mu_{kt}^2 - \left(\sum_{k=1}^{K} \alpha_{kt} \mu_{kt}\right)^2,$$

where $h_{kt} = \beta_{k0} + \beta_{k1} y_{t-1}^2 + \cdots + \beta_{kp_k} y_{t-p_k}^2$ and the difference in the conditional means also contributes to the conditional variance (Wong and Li 2001b).

Note that the conditional distribution at (1) can be multimodal, and the quantity $E(y_{n+1}|\mathcal{F}_n, \Omega_{n+1})$ may not be the best predictor of y_{n+1}. We therefore suggest the mode of the distribution of y_{n+1} conditional on $\mathcal{F}_n$ and Ω_{n+1} to be the predicted value of y_{n+1}, which is denoted by $\widehat{y}_n(1)$. The prediction interval with the shortest length can be constructed by following the highest density region method in Hyndman (1996), and it will always include $\widehat{y}_n(1)$ as an interior point. The corresponding mean squared prediction error has the value of $E\{[y_{n+1} - \widehat{y}_n(1)]^2|\mathcal{F}_n, \Omega_{n+1}\} \geq \text{var}(y_{n+1}|\mathcal{F}_n, \Omega_{n+1})$, which implies that the prediction of $E(y_{n+1}|\mathcal{F}_n, \Omega_{n+1})$ has the smallest mean squared prediction error although it may not be suitable here. For general r-step-ahead predictions, we refer to Wong and Li (2000, 2001b) for more discussions.

2.2 Probabilistic Properties

We first consider the case with constant mixing proportions, and then α_{kt} can be denoted by α_k for simplicity, where $k = 1, \ldots, K$. Moreover, we assume that $p_1 = \cdots = p_K = p$ without loss of generality.

Let $\{\varepsilon_t\}$ be iid random variables with the standard normal distribution, and $\{\mathbf{z}_t\}$ be iid random vectors with one element equal to one and the others equal to zero. Denote $\mathbf{z}_t = (z_{1t}, \ldots, z_{Kt})'$. We further assume that $P(z_{kt} = 1) = \alpha_k$ for $1 \leq k \leq K$, and time series $\{\varepsilon_t\}$ and $\{\mathbf{z}_t\}$ are independent. Then model (1) can be represented as

$$y_t = \sum_{k=1}^{K} z_{kt} \left(\theta_{k0} + \theta_{k1} y_{t-1} + \cdots + \theta_{kp} y_{t-p} + \varepsilon_t \sqrt{\beta_{k0} + \beta_{k1} y_{t-1}^2 + \cdots + \beta_{kp} y_{t-p}^2}\right), \tag{3}$$

where $\{\varepsilon_t\}$ are innovations, and $\{z_{kt}\}$ are latent variables. Moreover, let $\{\boldsymbol{\xi}_t\}$ be iid p-dimensional standard normal random vectors independent of $\{y_t\}$, $\{\varepsilon_t\}$, and $\{\mathbf{z}_t\}$, where $\boldsymbol{\xi}_t = (\xi_{1t}, \ldots, \xi_{pt})'$. We consider a random coefficient autoregressive (RCAR) model,

$$y_t^* = \sum_{k=1}^{K} z_{kt}(\theta_{k1} + \xi_{1t}\sqrt{\beta_{k1}}) y_{t-1}^* + \cdots + \sum_{k=1}^{K} z_{kt}(\theta_{kp} + \xi_{pt}\sqrt{\beta_{kp}}) y_{t-p}^* + \sum_{k=1}^{K} z_{kt}\theta_{k0} + \sum_{k=1}^{K} (z_{kt}\sqrt{\beta_{k0}})\varepsilon_t. \tag{4}$$

For models (3) and (4), it can be verified that, for any $(a_1, \ldots, a_p)' \in \mathbb{R}^p$,

$$F(y_t|y_{t-1} = a_1, \ldots, y_{t-p} = a_p) = F(y_t^*|y_{t-1}^* = a_1, \ldots, y_{t-p}^* = a_p).$$

As a result, their corresponding Markov chains will have the same transition probability, and then it is equivalent to discuss the ergodicity of the RCAR model at (4).

Denote $A_t = \sum_{k=1}^{K} z_{kt} A_{kt}$, where

$$A_{kt} = \begin{pmatrix} \theta_{k1} + \xi_{1t}\sqrt{\beta_{k1}} & \ldots & \theta_{k,p-1} + \xi_{p-1,t}\sqrt{\beta_{k,p-1}} & \theta_{kp} + \xi_{pt}\sqrt{\beta_{kp}} \\ & \mathbf{I}_{p-1} & & \mathbf{0} \end{pmatrix},$$

$\mathbf{I}_m$ is the $m \times m$ identity matrix, and $\mathbf{0}$ is a zero vector. The top Lyapounov exponent can be defined as

$$\gamma = \inf\left\{\frac{1}{n} E \ln \|A_1 \ldots A_n\|, \quad n \geq 1\right\},$$

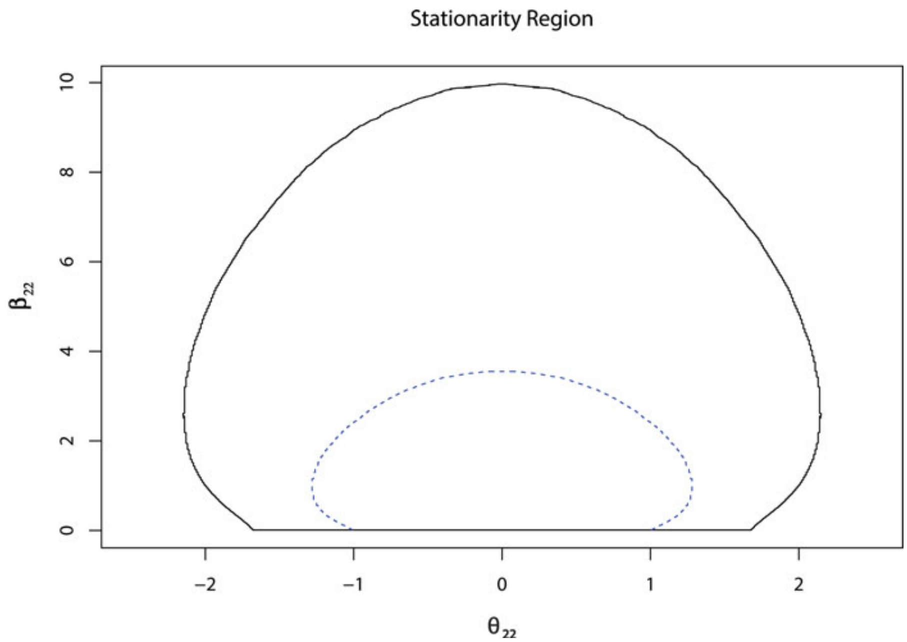

Figure 1. Stationarity region for the two-component MDAR model with $p = 1$ and $(\alpha_{1t}, \alpha_{2t}, \theta_{11}, \beta_{11}) = (0.5, 0.5, 0.5, 1)$. The dashed line is the stationarity region of the corresponding double AR model.

where $\|M\| = \sqrt{\mathrm{tr}(MM')}$ for a vector or matrix M, and $\mathrm{tr}(M)$ is the trace of the matrix M.

Theorem 1. Suppose that the mixing proportions $\{\alpha_{kt}\}$ are constants and independent of t. Then there exists a strictly stationary solution $\{y_t\}$ to model (1) if and only if $\gamma < 0$, and $\{y_t\}$ is unique and geometrically ergodic with $E|y_t|^\delta < \infty$ for some $\delta > 0$.

Note that $E\ln^+\|A_1\| = \sum_{k=1}^K \alpha_k E\ln^+\|A_{k1}\| < \infty$, where $\ln^+(x) = \max\{\ln(x), 0\}$. Then, by the subadditive ergodic theorem (Hall and Heyde 1980, Theorem 7.5), we can show that

$$\gamma = \lim_{n\to\infty} \frac{1}{n} \ln \|A_1 \dots A_n\|$$

with probability one. For the case with $p = 1$, it holds that $\gamma = \sum_{k=1}^K \alpha_k E \ln|\theta_{k1} + \sqrt{\beta_{k1}}\xi_{1t}|$, while the stationarity condition of the kth component is that $E\ln|\theta_{k1} + \sqrt{\beta_{k1}}\xi_{1t}| < 0$ (Ling 2007). As a result, it is not necessary for each component of the MDAR processes to be stationary (Wong and Li 2000). As an illustration, we consider a two-component MDAR model with $p = 1$, $\alpha_1 = \alpha_2 = 0.5$, and $(\theta_{11}, \beta_{11}) = (0.5, 1)$, and it can be verified that the first component is stationary. Figure 1 gives the stationarity region of the MDAR model with respect to $(\theta_{22}, \beta_{22})$, and it can be seen that the region is larger than that of the corresponding double AR model.

Let $A_t^{\otimes m}$ be the Kronecker product of m matrices, and $\rho(M)$ be the modulus of matrix M, which is defined as the maximum of the absolute eigenvalues of matrix M. We give a sufficient condition of some higher order moments of the MDAR process as follows.

Theorem 2. Under the assumptions of Theorem 1, if $\rho(E[A_t^{\otimes m}]) < 1$, then the mth-order moment of y_t is finite with $m = 2$ and 4.

Note that $E[A_t^{\otimes m}] = \sum_{k=1}^K \alpha_k E[A_{kt}^{\otimes m}]$. When $m = 2$, the condition $\rho(E[A_{kt}^{\otimes 2}]) < 1$ is necessary and sufficient for the kth component. Ling (1999) discussed the moment conditions for the GARCH process, which are similar to that in the above theorem.

We next consider the general case of model (1), and assume that $\mathbf{x}_t = (1 y_{t-1}, \dots, y_{t-p})'$ without loss of generality. Define

$$\gamma_k = \inf\left\{\frac{1}{n} E\ln\|A_{k1}\dots A_{kn}\|,\quad n \geq 1\right\} \quad \text{for} \quad k = 1, \dots, K.$$

Corollary 1. There exists a strictly stationary solution $\{y_t\}$ to model (1) if $\gamma_k < 0$ for all k, and $\{y_t\}$ is unique and geometrically ergodic with $E|y_t|^\delta < \infty$ for some $\delta > 0$.

Unlike the results of Theorem 1, this corollary requires all components to be stationary, and the condition is no longer necessary. Note that each α_{kt} can arbitrarily approach zero and one. As a result, it may not be easy to further relax the restriction of the above corollary. Moreover, if $\max_{1\leq k\leq K} \rho(E[A_{kt}^{\otimes m}]) < 1$, then $E|y_t|^m < \infty$ with $m = 2$ and 4.

3. STATISTICAL INFERENCE FOR LOGISTIC MDAR MODELS

3.1 Logistic MDAR Models

In practice, the most commonly used mixture model is that with two components (Wong and Li 2001a), and this section will focus on the inference of model (1) with two components only, which can be called the logistic MDAR model (Agresti 2002).

Let $\mathbf{y}_{k1t} = (1, y_{t-1}, \dots, y_{t-p_k})'$, $\mathbf{y}_{k2t} = (1, y_{t-1}^2, \dots, y_{t-p_k}^2)'$, $\boldsymbol{\theta}_k = (\theta_{k0}, \theta_{k1}, \dots, \theta_{kp_k})'$, and $\boldsymbol{\beta}_k = (\beta_{k0}, \beta_{k1}, \dots, \beta_{kp_k})'$ with $k = 1$ and 2. We can rewrite model (1) with $K = 2$ into the following form,

$$F(y|\mathcal{F}_{t-1},\Omega_t)=\sum_{k=1}^{2}\alpha_{kt}\Phi\left(\frac{y-\mathbf{y}_{k1t}'\boldsymbol{\theta}_k}{\sqrt{\mathbf{y}_{k2t}'\boldsymbol{\beta}_k}}\right),\quad \text{for}\quad y\in\mathbb{R},\quad (5)$$

and the mixing proportions satisfying $0<\alpha_{1t}<1$, $\alpha_{2t}=1-\alpha_{1t}$ and

$$\ln(\alpha_{1t}/\alpha_{2t})=\mathbf{x}_t'\boldsymbol{\varphi}=\varphi_0+\varphi_1x_{1t}+\cdots+\varphi_lx_{lt},\qquad (6)$$

with $\mathbf{x}_t=(1,x_{1t},\ldots,x_{lt})'$ and $\boldsymbol{\varphi}=(\varphi_0,\varphi_1,\ldots,\varphi_l)'$.

When $\mathbf{x}_t$ includes a lagged value of y_t only, say y_{t-d}, the logistic MDAR model defined by (5) and (6) has the following form:

$$y_t=\begin{cases}\mathbf{y}_{11t}'\boldsymbol{\theta}_1+\varepsilon_t\sqrt{\mathbf{y}_{12t}'\boldsymbol{\beta}_1} & \text{with probability } \alpha_{1t},\\ \mathbf{y}_{21t}'\boldsymbol{\theta}_2+\varepsilon_t\sqrt{\mathbf{y}_{22t}'\boldsymbol{\beta}_2} & \text{with probability } 1-\alpha_{1t},\end{cases}$$

where

$$\alpha_{1t}=[\exp\{-\varphi_0-\varphi_1y_{t-d}\}+1]^{-1},$$

and it tends to either zero or one as φ_1y_{t-d} tends to $-\infty$ or $+\infty$, respectively. It is noteworthy to point out that the above model will become the two-regime smooth transition threshold model (Chan and Tong 1986) if α_{1t} and $1-\alpha_{1t}$ are the weights for the summation of these two components instead. As a result, the logistic MDAR model may have a performance similar to that of the smooth transition threshold model, while keeping the piecewise structures; see also Li et al. (2015).

3.2 Maximum Likelihood Estimation

Denote the parameter vector by $\boldsymbol{\lambda}=(\boldsymbol{\varphi}',\boldsymbol{\theta}_1',\boldsymbol{\beta}_1',\boldsymbol{\theta}_2',\boldsymbol{\beta}_2')'$, and its true value by $\boldsymbol{\lambda}_0=(\boldsymbol{\varphi}_0',\boldsymbol{\theta}_{01}',\boldsymbol{\beta}_{01}',\boldsymbol{\theta}_{02}',\boldsymbol{\beta}_{02}')'$. The parameter space $\Lambda\subset\mathbb{R}^{2(p_1+p_2)+l+5}$ is a compact set. We further assume that $\beta_{kj}>0$ for $j=0,1,\ldots,p_k$ and $k=1$ and 2, and the identification condition is given below.

Assumption 1. It holds that $(\boldsymbol{\theta}_{01}',\boldsymbol{\beta}_{01}')\neq(\boldsymbol{\theta}_{02}',\boldsymbol{\beta}_{02}')$ if $p_1=p_2$, and the first nonzero element of $\boldsymbol{\varphi}_0$ is negative. Moreover, $E(\mathbf{x}_t\mathbf{x}_t')$ is a finite and positive definite matrix.

The condition that $(\boldsymbol{\theta}_{01}',\boldsymbol{\beta}_{01}')\neq(\boldsymbol{\theta}_{02}',\boldsymbol{\beta}_{02}')$ if $p_1=p_2$ makes sure that the two components have different structures. For the logistic MDAR model given by (5) and (6), if we exchange the two components and replace $\boldsymbol{\varphi}_0$ by $-\boldsymbol{\varphi}_0$, then the resulting model is still the same. As a result, it is necessary to restrict the first nonzero element of $\boldsymbol{\varphi}_0$ to be either negative or positive.

Let $p=\max\{p_1,p_2\}$. Define the functions $\alpha_{1t}(\boldsymbol{\varphi})=1/[1+\exp(-\mathbf{x}_t'\boldsymbol{\varphi})]$, $\alpha_{2t}(\boldsymbol{\varphi})=1-\alpha_{1t}(\boldsymbol{\varphi})$, $\mu_{kt}(\boldsymbol{\theta}_k)=\mathbf{y}_{k1t}'\boldsymbol{\theta}_k$, and $h_{kt}(\boldsymbol{\beta}_k)=\mathbf{y}_{k2t}'\boldsymbol{\beta}_k$ with $k=1$ and 2. Then, up to a constant, the log-likelihood function of the logistic MDAR model is $L_n(\boldsymbol{\lambda})=\sum_{t=p+1}^{n}l_t(\boldsymbol{\lambda})$, where

$$l_t(\boldsymbol{\lambda})=\ln\left\{\sum_{k=1}^{2}\frac{\alpha_{kt}(\boldsymbol{\varphi})}{\sqrt{h_{kt}(\boldsymbol{\beta}_k)}}\exp\left[-\frac{(y_t-\mu_{kt}(\boldsymbol{\theta}_k))^2}{2h_{kt}(\boldsymbol{\beta}_k)}\right]\right\}.$$

Thus, the MLE can be defined as

$$\widehat{\boldsymbol{\lambda}}_n=\operatorname*{argmax}_{\boldsymbol{\lambda}\in\Lambda}L_n(\boldsymbol{\lambda}).$$

Denote

$$\tau_{1t}(\boldsymbol{\lambda})=\frac{\alpha_{1t}(\boldsymbol{\varphi})\exp\{-0.5[y_t-\mu_{1t}(\boldsymbol{\theta}_1)]^2/h_{1t}(\boldsymbol{\beta}_1)\}/\sqrt{h_{1t}(\boldsymbol{\beta}_1)}}{\sum_{k=1}^{2}\alpha_{kt}(\boldsymbol{\varphi})\exp\{-0.5[y_t-\mu_{kt}(\boldsymbol{\theta}_k)]^2/h_{kt}(\boldsymbol{\beta}_k)\}/\sqrt{h_{kt}(\boldsymbol{\beta}_k)}},\qquad (7)$$

and $\tau_{2t}(\boldsymbol{\lambda})=1-\tau_{1t}(\boldsymbol{\lambda})$. The derivative functions of $l_t(\boldsymbol{\lambda})$ have the form,

$$\frac{\partial l_t(\boldsymbol{\lambda})}{\partial\boldsymbol{\varphi}}=[\tau_{1t}(\boldsymbol{\lambda})-\alpha_{1t}(\boldsymbol{\lambda})]\mathbf{x}_t,$$

$$\frac{\partial l_t(\boldsymbol{\lambda})}{\partial\boldsymbol{\theta}_k}=\frac{\tau_{kt}(\boldsymbol{\lambda})[y_t-\mu_{kt}(\boldsymbol{\theta}_k)]}{h_{kt}(\boldsymbol{\beta}_k)}\mathbf{y}_{k1t},$$

and

$$\frac{\partial l_t(\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}_k}=\frac{\tau_{kt}(\boldsymbol{\lambda})}{2h_{kt}(\boldsymbol{\beta}_k)}\left\{\frac{[y_t-\mu_{kt}(\boldsymbol{\theta}_k)]^2}{h_{kt}(\boldsymbol{\beta}_k)}-1\right\}\mathbf{y}_{k2t}$$

with $k=1$ and 2.

Theorem 3. Suppose that the strictly stationary and ergodic time series $\{y_t\}$ is generated by Equations (5) and (6). If Assumption 1 holds, then $\widehat{\boldsymbol{\lambda}}_n$ converges to $\boldsymbol{\lambda}_0$ in the almost surely sense as $n\to\infty$. Moreover, $\sqrt{n}(\widehat{\boldsymbol{\lambda}}_n-\boldsymbol{\lambda}_0)\to N(0,\Sigma^{-1})$ as $n\to\infty$, where $\Sigma=E\{[\partial l_t(\boldsymbol{\lambda}_0)/\partial\boldsymbol{\lambda}][\partial l_t(\boldsymbol{\lambda}_0)/\partial\boldsymbol{\lambda}']\}$.

The multimodal marginal or conditional distribution of a mixture model provides more flexibility in fitting real data. However, it also leads to some difficulty in directly maximizing the corresponding likelihood function, for example, some algorithms such as the Newton–Raphson may not work well (McLachlan and Krishnan 1997). Moreover, for the logistic MDAR model, the number of parameters $2(p_1+p_2)+l+5$ usually is large, and the information matrix $E[\partial^2l_t(\boldsymbol{\lambda}_0)/\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}']=-\Sigma$ is not block diagonal. The EM algorithm is commonly used in the literature of mixture models to search for the MLE, and it is based on the idea of replacing the difficult likelihood maximization with a sequence of easier maximizations; see Dempster, Laird, and Rubin (1977), McLachlan and Krishnan (1997), McLachlan and Peel (2000), and references therein. Hence, we will introduce the EM algorithm to search for the MLE $\widehat{\boldsymbol{\lambda}}_n$ in the next subsection.

3.3 EM Algorithm and Information Matrices

By taking into account latent variables $\{z_{kt}\}$, we treat $\{y_t,\mathbf{z}_t,t=1,\ldots,n\}$ as the complete data and, up to a constant, the complete log-likelihood function is

$$L_{cn}(\boldsymbol{\lambda})=\sum_{t=p+1}^{n}\sum_{k=1}^{2}z_{kt}\left\{\ln[\alpha_{kt}(\boldsymbol{\varphi})]-0.5\ln[h_{kt}(\boldsymbol{\beta}_k)]-0.5\frac{[y_t-\mu_{kt}(\boldsymbol{\theta}_k)]^2}{h_{kt}(\boldsymbol{\beta}_k)}\right\}.$$

The iterative EM procedure was demonstrated by Dempster, Laird, and Rubin (1977) to be flexible for estimating the parameters in the mixture-type models including the mixture time series models. We describe this procedure as follows.

- *E-step*: It holds that $E(z_{kt}|\boldsymbol{\lambda},\mathcal{F}_n,\Omega_n)=\tau_{kt}(\boldsymbol{\lambda})$ with $k=1$ and 2, where $\tau_{1t}(\boldsymbol{\lambda})$ and $\tau_{2t}(\boldsymbol{\lambda})$ are defined as in (7). Then we can replace the missing data $\{z_{kt}\}$ by their conditional expectations $\tau_{kt}(\boldsymbol{\lambda})$'s, that is, the E-equation is the same as that in (7).

- *M-step*: Suppose the missing data $\{z_{kt}\}$ are known by replacing them with $\{\tau_{kt}(\widetilde{\boldsymbol{\lambda}}_n^i)\}$, where $\widetilde{\boldsymbol{\lambda}}_n^i$ is from the last *M*-step. We then can estimate the parameter vector $\boldsymbol{\lambda}$ by maximizing the likelihood function

$$\widetilde{\boldsymbol{\lambda}}_n^{i+1} = \operatorname*{argmax}_{\boldsymbol{\lambda}\in\Lambda} L_{cn}(\boldsymbol{\lambda}).$$

By Theorems 2 and 3 of Wu (1983), the above EM algorithm will monotonically converge to a stationary point of the log-likelihood function $L_n(\boldsymbol{\lambda})$, that is, $\widetilde{\boldsymbol{\lambda}}_n^i \to \widetilde{\boldsymbol{\lambda}}_n$ as $i \to \infty$, $L_n(\widetilde{\boldsymbol{\lambda}}_n^i) \le L_n(\widetilde{\boldsymbol{\lambda}}_n^{i+1})$, and $\partial L_n(\widetilde{\boldsymbol{\lambda}}_n)/\partial\boldsymbol{\lambda} = 0$; see also McLachlan and Krishnan (1997). Note that $\widetilde{\boldsymbol{\lambda}}_n$ may not be the global maximum of $L_n(\boldsymbol{\lambda})$ (i.e., the MLE $\widehat{\boldsymbol{\lambda}}_n$). As recommended by Wu (1983) for real applications, we may try several EM iterations with different starting points, which are representatives of the parameter space.

In the M-step, it is equivalent to do the following optimizations,

$$\widetilde{\boldsymbol{\varphi}}_n^{i+1} = \operatorname*{argmax}_{\boldsymbol{\varphi}} \sum_{t=p+1}^{n} \{\widetilde{\tau}_{1t}\ln[\alpha_{1t}(\boldsymbol{\varphi})] + \widetilde{\tau}_{2t}\ln[\alpha_{2t}(\boldsymbol{\varphi})]\},$$

and

$$(\widetilde{\boldsymbol{\theta}}_{kn}^{i+1}, \widetilde{\boldsymbol{\beta}}_{kn}^{i+1}) = \operatorname*{argmax}_{\boldsymbol{\theta}_k,\boldsymbol{\beta}_k} \sum_{t=p+1}^{n} \widetilde{\tau}_{kt}\left\{-\ln[h_{kt}(\boldsymbol{\beta}_k)] - \frac{[y_t - \mu_{kt}(\boldsymbol{\theta}_k)]^2}{h_{kt}(\boldsymbol{\beta}_k)}\right\}, \quad k = 1 \text{ and } 2, \tag{8}$$

where $\widetilde{\tau}_{kt} = \tau_{kt}(\widetilde{\boldsymbol{\lambda}}_n^i)$ with $k = 1$ and 2. Moreover,

$$\frac{\partial L_{cn}(\boldsymbol{\lambda})}{\partial\boldsymbol{\theta}_k} = \sum_{t=p+1}^{n} \frac{z_{kt}y_t}{h_{kt}(\boldsymbol{\beta}_k)}\mathbf{y}_{k1t} - \sum_{t=p+1}^{n} \frac{z_{kt}}{h_{kt}(\boldsymbol{\beta}_k)}\mathbf{y}_{k1t}\mathbf{y}_{k1t}'\boldsymbol{\theta}_k.$$

As a result, we may further simplify the optimization in (8) as follows,

$$\widetilde{\boldsymbol{\beta}}_{kn}^{i+1} = \operatorname*{argmax}_{\boldsymbol{\beta}_k} \sum_{t=p+1}^{n} \widetilde{\tau}_{kt}\left\{-\ln[h_{kt}(\boldsymbol{\beta}_k)] - \frac{[y_t - \widetilde{\mu}_{kt}]^2}{h_{kt}(\boldsymbol{\beta}_k)}\right\},$$

and

$$\widetilde{\boldsymbol{\theta}}_{kn}^{i+1} = \left(\sum_{t=p+1}^{n} \frac{\widetilde{\tau}_{kt}}{h_{kt}(\widetilde{\boldsymbol{\beta}}_{kn}^{i+1})}\mathbf{y}_{k1t}\mathbf{y}_{k1t}'\right)^{-1} \sum_{t=p+1}^{n} \frac{\widetilde{\tau}_{kt}y_t}{h_{kt}(\widetilde{\boldsymbol{\beta}}_{kn}^{i+1})}\mathbf{y}_{k1t},$$

where $\widetilde{\mu}_{kt} = \mu_{kt}(\widetilde{\boldsymbol{\lambda}}_n^i)$ with $k = 1$ and 2. Note that we do not need to perform the time-consuming optimizations to obtain $\widetilde{\boldsymbol{\theta}}_{kn}^{i+1}$, and some optimizing algorithms such as the Newton–Raphson can be employed to calculate $\widetilde{\boldsymbol{\varphi}}_n^{i+1}$ and $\widetilde{\boldsymbol{\beta}}_{kn}^{i+1}$.

Denote the complete information matrix by I_c, and the missing information matrix by I_m. As in Louis (1982) and Wong and Li (2001a), the observed information matrix can be calculated by

$$I_1 = I_c - I_m = E\left(-\frac{\partial^2 L_{cn}(\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}'}\Big|\boldsymbol{\lambda}, \mathcal{F}_n, \Omega_n\right)_{\widehat{\boldsymbol{\lambda}}_n} - \mathrm{var}\left(\frac{\partial L_{cn}(\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}}\Big|\boldsymbol{\lambda}, \mathcal{F}_n, \Omega_n\right)_{\widehat{\boldsymbol{\lambda}}_n}, \tag{9}$$

where $I_c = \mathrm{diag}\{I_{0c}, I_{1c}, I_{2c}\}$, $I_{0c} = \sum_{t=p+1}^n \alpha_{1t}(\widehat{\boldsymbol{\lambda}}_n)\alpha_{2t}(\widehat{\boldsymbol{\lambda}}_n)\mathbf{x}_t\mathbf{x}_t'$, $I_{1c} = \sum_{t=p+1}^n \tau_{1t}(\widehat{\boldsymbol{\lambda}}_n)\iota_{1t}(\widehat{\boldsymbol{\lambda}}_n)$, $I_{2c} = \sum_{t=p+1}^n \tau_{2t}(\widehat{\boldsymbol{\lambda}}_n)\iota_{2t}(\widehat{\boldsymbol{\lambda}}_n)$, the symmetric matrices

$$\iota_{kt}(\boldsymbol{\lambda}) = \begin{pmatrix} h_{kt}^{-1}(\boldsymbol{\beta}_k)\mathbf{y}_{k1t}\mathbf{y}_{k1t}' & h_{kt}^{-2}(\boldsymbol{\beta}_k)[y_t - \mu_{kt}(\boldsymbol{\theta}_k)]\mathbf{y}_{k1t}\mathbf{y}_{k2t}' \\ * & h_{kt}^{-2}(\boldsymbol{\beta}_k)\{[y_t - \mu_{kt}(\boldsymbol{\theta}_k)]^2/h_{kt}(\boldsymbol{\beta}_k) - 0.5\}\mathbf{y}_{k2t}\mathbf{y}_{k2t}' \end{pmatrix}$$

with $k = 1$ and 2, $I_m = \sum_{t=p+1}^n \tau_{1t}(\widehat{\boldsymbol{\lambda}}_n)\tau_{2t}(\widehat{\boldsymbol{\lambda}}_n)\iota_{3t}(\widehat{\boldsymbol{\lambda}}_n)\iota_{3t}'(\widehat{\boldsymbol{\lambda}}_n)$, and the vector

$$\iota_{3t}(\boldsymbol{\lambda}) = \left(\mathbf{x}_t', \frac{y_t - \mu_{1t}(\boldsymbol{\theta}_1)}{h_{1t}(\boldsymbol{\beta}_1)}\mathbf{y}_{11t}', \left\{\frac{[y_t - \mu_{1t}(\boldsymbol{\theta}_1)]^2}{h_{1t}(\boldsymbol{\beta}_1)} - 1\right\}\frac{\mathbf{y}_{12t}'}{2h_{1t}(\boldsymbol{\beta}_1)}, -\frac{y_t - \mu_{2t}(\boldsymbol{\theta}_2)}{h_{2t}(\boldsymbol{\beta}_2)}\mathbf{y}_{21t}', -\left\{\frac{[y_t - \mu_{2t}(\boldsymbol{\theta}_2)]^2}{h_{2t}(\boldsymbol{\beta}_2)} - 1\right\}\frac{\mathbf{y}_{22t}'}{2h_{2t}(\boldsymbol{\beta}_2)}\right)'.$$

Note that, from the proof of Theorem 3, $I_1 = -\partial^2 L_n(\widehat{\boldsymbol{\lambda}}_n)/\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}' = n\Sigma + o_p(n)$. This information matrix may not be positive definite in practice, and we can alternatively consider

$$I_2 = \sum_{t=p+1}^{n} \frac{\partial l_t(\widehat{\boldsymbol{\lambda}}_n)}{\partial\boldsymbol{\lambda}}\frac{\partial l_t(\widehat{\boldsymbol{\lambda}}_n)}{\partial\boldsymbol{\lambda}'} = n\Sigma + o_p(n).$$

The derivation of Equation (9) is given in the Appendix.

3.4 Model Selection

For the logistic MDAR model defined as in (5) and (6), we introduce the Bayesian information criterion (BIC) to select its orders,

$$\mathrm{BIC}(\mathbf{p}) = -2L_n(\widehat{\boldsymbol{\lambda}}_n) + \ln(n-p)[2(p_1 + p_2) + l + 5], \tag{10}$$

where $\mathbf{p} = (l, p_1, p_2)$. Let $\widehat{\mathbf{p}}_n = \mathrm{argmin}_{0\le l,p_1,p_2\le p_{\max}} \mathrm{BIC}(\mathbf{p})$, where $p_{\max}$ is a predetermined positive number, and it can be different for l, p_1, and p_2.

Theorem 4. Under the assumptions of Theorem 3, if $p_{\max} \ge \max\{l_0, p_{10}, p_{20}\}$, then $P(\widehat{\mathbf{p}}_n = \mathbf{p}_0) \to 1$ as $n \to \infty$, where $\mathbf{p}_0 = (l_0, p_{10}, p_{20})$ are the true orders, that is, $|\varphi_{0\ell_0}| > 0$ and $|\theta_{0kp_{k0}}| + |\beta_{0kp_{k0}}| > 0$ with $k = 1$ and 2.

Note that $L_{cn}(\boldsymbol{\lambda})$ with z_{kt} replaced by $\tau_{kt}(\widehat{\boldsymbol{\lambda}}_n)$ is also a likelihood function. We can similarly design an information criterion based on $L_{cn}(\boldsymbol{\lambda})$, however, Wong and Li (2000) showed by simulations that its performance is poorer than that given by (10). Some other information criteria such as Akaike information criterion (AIC) can be similarly discussed.

For logistic MDAR models given by (5) and (6), as in Wong and Li (2001a), we may be more interested in whether the mixing proportions are time varying, that is, we would like to test the following null hypothesis,

$$H_0: \quad \varphi_1 = \cdots = \varphi_l = 0.$$

Table 1. Estimating results for the MDAR model with mixing proportions being constant and both components being stationary

		1st component				2nd component				
n		θ_{10}	θ_{11}	β_{10}	β_{11}	θ_{20}	θ_{21}	β_{20}	β_{21}	φ_0
300	Bias	0.0020	0.1462	−0.0347	−0.0196	−0.0010	−0.0073	−0.0045	−0.0292	−0.0777
	MSE	0.1677	0.5182	0.0852	0.4609	0.0587	0.1588	0.0338	0.1110	0.9125
	ASE1	0.0946	0.3857	0.0618	0.3877	0.0429	0.1187	0.0243	0.0955	0.7271
	ASE2	0.1031	0.6025	0.0846	0.6758	0.0448	0.1227	0.0269	0.1102	1.0067
500	Bias	−0.0018	0.1263	−0.0163	−0.0466	0.0029	−0.0033	−0.0021	−0.0296	−0.0705
	MSE	0.1088	0.4451	0.0679	0.3739	0.0413	0.1213	0.0276	0.0958	0.8115
	ASE1	0.0741	0.3371	0.0540	0.3094	0.0327	0.0977	0.0206	0.0793	0.6501
	ASE2	0.0763	0.5074	0.0673	0.4894	0.0333	0.1018	0.0222	0.0890	0.8708
1000	Bias	0.0029	0.1226	−0.0011	−0.0778	−0.0007	0.0008	−0.0016	−0.0207	−0.0716
	MSE	0.0589	0.3954	0.0468	0.2741	0.0257	0.0831	0.0188	0.0660	0.6825
	ASE1	0.0507	0.2742	0.0404	0.2220	0.0228	0.0751	0.0158	0.0613	0.5352
	ASE2	0.0516	0.3907	0.0465	0.3189	0.0231	0.0779	0.0171	0.0664	0.6876

The results in Theorem 3 make sure that the likelihood ratio test can be employed for this purpose and, under the null hypothesis, the test statistic converges to χ^2_l, the chi-squared distribution with l degrees of freedom, in distribution.

4. NUMERICAL STUDIES

4.1 Simulation Experiments

This subsection constructs three simulation experiments to study the finite-sample performance of the methodology in the previous sections: the EM algorithm and information matrices, the BIC, and the one-step-ahead prediction.

In the first experiment, we consider four data-generating processes including two MDAR models with constant mixing proportions and two logistic MDAR models. The MDAR models with constant mixing proportions are

$$y_t = \begin{cases} 0.45y_{t-1} + \varepsilon_t\sqrt{0.2 + 0.6y_{t-1}^2} & \text{with probability } \alpha_{1t}, \\ -0.5y_{t-1} + \varepsilon_t\sqrt{0.1 + 0.2y_{t-1}^2} & \text{with probability } \alpha_{2t}, \end{cases} \quad (11)$$

and

$$y_t = \begin{cases} 1.2y_{t-1} + \varepsilon_t\sqrt{0.2 + 1.8y_{t-1}^2} & \text{with probability } \alpha_{1t}, \\ -0.5y_{t-1} + \varepsilon_t\sqrt{0.1 + 0.2y_{t-1}^2} & \text{with probability } \alpha_{2t}, \end{cases} \quad (12)$$

where $\{\varepsilon_t\}$ are iid standard normal random variables, $\alpha_{2t} = 1 - \alpha_{1t}$ and

$$\ln(\alpha_{1t}/\alpha_{2t}) = -0.7.$$

Note that model (12) is stationary while its first component is nonstationary since $E \ln |1.2 + \sqrt{1.8}\xi_t| > 0$ with ξ_t being a standard normal random variable. For the logistic MDAR models, we use the same settings as in (11) and (12), and the time varying mixing proportions are generated by

$$\ln(\alpha_{1t}/\alpha_{2t}) = -0.7 + 0.3y_{t-1} - 0.5x_t, \quad (13)$$

where $\{x_t\}$ are from an AR(2) model, $x_t = 0.6x_{t-1} - 0.2x_{t-2} + \eta_t$, with $\{\eta_t\}$ being iid standard normal random variables and independent of $\{\varepsilon_t\}$.

We consider three sample sizes, $n = 300, 500$, and 1000, and the number of replications is 1000. The EM algorithm in Section 3.3 is employed to search for the MLEs, and the corresponding standard errors are calculated based on information matrices I_1 and I_2 (ASE1 and ASE2). Tables 1–4 give their biases (Bias), mean squared errors (MSE), and two types of standard errors, ASE1 and ASE2. It can be seen that the bias decreases as sample size n increases. These two types of standard errors are close to each other, and we occasionally encounter the problem that the information matrix I_1 is not positive definite. Note that the two components in model (12) are more separated from each other. The MSEs in Tables 2 and 4 are close to their two theoretical versions when the sample size is as small as $n = 300$, and the corresponding EM algorithms also converge quickly.

The data-generating process in the second experiment is

$$y_t = \begin{cases} 0.4 + 0.5y_{t-1} + \varepsilon_t\sqrt{0.2 + 0.6y_{t-1}^2} & \text{with probability } \alpha_{1t}, \\ -0.4 - 0.8y_{t-1} + \varepsilon_t\sqrt{0.1 + 0.5y_{t-1}^2} & \text{with probability } \alpha_{2t}, \end{cases} \quad (14)$$

with the mixing proportions satisfying

$$\ln(\alpha_{1t}/\alpha_{2t}) = -0.8 - 0.5x_t + 0.7y_{t-1},$$

where $\{\varepsilon_t\}$ and $\{x_t\}$ are generated as in the first experiment. It is noteworthy that this is a logistic MDAR model with orders $(l, p_1, p_2) = (2, 1, 1)$, and then we have three orders to select in total. To save the computing time, the selection in the logistic function is limited to that of the lagged values of y_t, that is, the exogenous variable x_t is kept in the model. We employ the BIC in Section 3.4 to select p_1, p_2, and l with $0 \le p_1, p_2 \le 2$, and $1 \le l \le 3$, and there are 27 candidate models in total. The sample size is set to 1000, and there are 100 replications generated. As a result, the BIC correctly identifies all true orders at a rate of 98%, and wrongly selects larger orders for the remaining two replications.

The third experiment is for the comparison of two different one-step-ahead prediction methods, and the first one is introduced in Section 2.1. The other one is to use the conditional mean $E(y_{n+1}|\mathcal{F}_n, \Omega_{n+1})$ to be the predicted value, while the lower and upper limits of prediction intervals with confidence level $1 - \alpha$ take the values of the $\alpha/2$- and $(1 - \alpha/2)$-quantiles of the conditional distribution, respectively. The data-generating process is as in (11) with time-varying mixing proportions (13), and all other settings are the same as in the first experiment. We first estimate the model as in the first experiment, and then perform predictions based on estimated models. Six confidence lev-

Table 2. Estimating results for the MDAR model with constant mixing proportions, one stationary component and one nonstationary component

		1st component				2nd component				
n		θ_{10}	θ_{11}	β_{10}	β_{11}	θ_{20}	θ_{21}	β_{20}	β_{21}	φ_0
300	Bias	−0.0009	0.2682	−0.0098	−0.3566	0.0004	0.0198	0.0046	0.0062	−0.2025
	MSE	0.1133	0.5048	0.0890	0.7106	0.0382	0.0870	0.0251	0.0879	0.4674
	ASE1	0.1004	0.4111	0.0754	0.5858	0.0369	0.0782	0.0230	0.0737	0.3923
	ASE2	0.1019	0.5291	0.0903	0.8270	0.0369	0.0766	0.0237	0.0783	0.4626
500	Bias	−0.0019	0.1886	−0.0085	−0.2317	0.0004	0.0123	0.0035	0.0028	−0.1459
	MSE	0.0863	0.4167	0.0653	0.5759	0.0304	0.0626	0.0197	0.0647	0.3750
	ASE1	0.0764	0.3407	0.0583	0.4877	0.0287	0.0603	0.0181	0.0584	0.3173
	ASE2	0.0776	0.4154	0.0652	0.6232	0.0286	0.0601	0.0183	0.0610	0.3634
1000	Bias	−0.0011	0.1121	−0.0049	−0.1527	−0.0003	0.0069	0.0023	0.0021	−0.0876
	MSE	0.0542	0.3162	0.0431	0.4349	0.0203	0.0427	0.0136	0.0462	0.2817
	ASE1	0.0527	0.2521	0.0407	0.3518	0.0203	0.0430	0.0131	0.0426	0.2346
	ASE2	0.0532	0.2895	0.0431	0.4126	0.0202	0.0429	0.0132	0.0436	0.2581

Table 3. Estimating results for the logistic MDAR model with both components being stationary

		1st component				2nd component				Mixing proportions		
n		θ_{10}	θ_{11}	β_{10}	β_{11}	θ_{20}	θ_{21}	β_{20}	β_{21}	φ_0	φ_1	φ_2
300	Bias	0.0002	0.0869	−0.0283	−0.0123	−0.0006	−0.0068	−0.0021	−0.0315	−0.1363	0.3460	−0.4305
	MSE	0.1596	0.4597	0.0779	0.4191	0.0587	0.1556	0.0332	0.1076	1.8606	1.9164	1.5227
	ASE1	0.0905	0.3135	0.0586	0.3279	0.0401	0.1092	0.0229	0.0906	0.9268	0.8978	0.6837
	ASE2	0.0937	0.3913	0.0731	0.4796	0.0393	0.1060	0.0234	0.0989	1.1314	1.2360	0.9557
500	Bias	−0.0041	0.0846	−0.0101	−0.0452	−0.0012	−0.0005	−0.0001	−0.0255	−0.1054	0.0803	−0.1605
	MSE	0.1024	0.3814	0.0627	0.3231	0.0364	0.1105	0.0253	0.0874	0.9793	0.6862	0.5745
	ASE1	0.0803	0.3165	0.0575	0.3026	0.0321	0.0942	0.0202	0.0798	0.8002	0.5608	0.5463
	ASE2	0.0938	0.4370	0.0703	0.4621	0.0313	0.0913	0.0200	0.0833	0.9080	0.5753	0.6054
1000	Bias	−0.0013	0.0847	−0.0021	−0.0697	0.0001	−0.0011	0.0003	−0.0169	−0.0802	0.0280	−0.0558
	MSE	0.0562	0.3136	0.0412	0.2368	0.0245	0.0781	0.0178	0.0614	0.6526	0.3461	0.2147
	ASE1	0.0509	0.2562	0.0389	0.2175	0.0219	0.0696	0.0148	0.0595	0.5231	0.2842	0.1993
	ASE2	0.0511	0.3215	0.0426	0.2833	0.0215	0.0685	0.0147	0.0609	0.5803	0.2793	0.1932

els are considered for prediction intervals, $1 - \alpha = 0.95$, 0.90, 0.80, 0.70, 0.60, and 0.50. Their empirical coverage rates (ECR) and lengths of prediction intervals (LPI) are listed in Table 5. The mean squared prediction errors (MSPE) are also provided for each predicted value. It can be seen that MSPEs based on conditional means are all smaller than those based on conditional modes, which is consistent with the fact that the predicted value based on the conditional mean has the smallest MSPE. Moreover, the MSPE decreases as the sample size increases, and it is because the estimated model is more accurate for larger sample size. Second, as expected, the LPIs based on the high density region method have smaller values. Finally, the ECRs

Table 4. Estimating results for the logistic MDAR model with one stationary component and one nonstationary component

		1st component				2nd component				Mixing proportions		
n		θ_{10}	θ_{11}	β_{10}	β_{11}	θ_{20}	θ_{21}	β_{20}	β_{21}	φ_0	φ_1	φ_2
300	Bias	0.0017	0.0327	−0.0129	−0.0969	0.0003	0.0041	0.0012	−0.0116	−0.0504	0.0317	−0.0543
	MSE	0.1116	0.3207	0.0795	0.5026	0.0428	0.0798	0.0253	0.0714	0.4259	0.2490	0.2714
	ASE1	0.0964	0.2790	0.0698	0.4432	0.0392	0.0749	0.0236	0.0661	0.3703	0.1929	0.2438
	ASE2	0.0991	0.3202	0.0827	0.5522	0.0386	0.0738	0.0240	0.0698	0.3914	0.2031	0.2435
500	Bias	0.0025	0.0232	−0.0090	−0.0781	−0.0004	−0.0019	0.0017	−0.0054	−0.0514	0.0166	−0.0250
	MSE	0.0773	0.2433	0.0602	0.3694	0.0307	0.0605	0.0198	0.0560	0.3083	0.1476	0.1856
	ASE1	0.0740	0.2225	0.0540	0.3455	0.0298	0.0581	0.0184	0.0532	0.2821	0.1375	0.1756
	ASE2	0.0751	0.2475	0.0599	0.4072	0.0295	0.0579	0.0185	0.0551	0.2975	0.1403	0.1765
1000	Bias	−0.0015	−0.0092	−0.0051	−0.0290	0.0008	−0.0014	0.0006	−0.0045	−0.0037	0.0070	−0.0151
	MSE	0.0523	0.1478	0.0401	0.2385	0.0218	0.0420	0.0135	0.0384	0.1927	0.0963	0.1253
	ASE1	0.0514	0.1531	0.0379	0.2386	0.0211	0.0411	0.0131	0.0382	0.1928	0.0940	0.1216
	ASE2	0.0521	0.1632	0.0403	0.2612	0.0210	0.0410	0.0132	0.0389	0.1983	0.0952	0.1217

Table 5. One-step-ahead predicting results including empirical coverage rates (ECR), lengthes of prediction intervals (LPI), and mean squared prediction errors (MSPE)

		Confidence level					
n		0.95	0.90	0.80	0.70	0.60	0.50
		Prediction with the conditional mode					
300	ECR	0.955	0.911	0.820	0.721	0.619	0.504
	LPI	2.049	1.683	1.287	1.027	0.823	0.640
	MSPE	0.393					
500	ECR	0.954	0.905	0.797	0.691	0.600	0.498
	LPI	2.009	1.660	1.267	1.010	0.809	0.629
	MSPE	0.369					
1000	ECR	0.950	0.901	0.800	0.695	0.601	0.490
	LPI	1.938	1.599	1.216	0.974	0.777	0.616
	MSPE	0.326					
		Prediction with the conditional mean					
300	ECR	0.955	0.916	0.818	0.730	0.620	0.502
	LPI	2.075	1.714	1.309	1.044	0.838	0.667
	MSPE	0.369					
500	ECR	0.951	0.902	0.798	0.709	0.599	0.502
	LPI	2.020	1.673	1.282	1.026	0.827	0.659
	MSPE	0.305					
1000	ECR	0.948	0.898	0.802	0.695	0.591	0.495
	LPI	1.948	1.611	1.232	0.983	0.791	0.630
	MSPE	0.291					

are close to their corresponding nominal values when the sample size is as small as $n = 300$.

4.2 An Empirical Example

This subsection considers the weekly exchange rates of U.S. Dollar (USD) to Japanese Yen (JPY) from February 16, 2008 to February 9, 2013, and there are 262 observations in total. We focus on the sequence of log returns, and its time plot is given in Figure 3. The fitted kernel density and the Hill estimator (Hill 1975, 2010) of these log returns are presented in Figure 2. We can observe a bimodal marginal distribution, and the estimated tail index $\widehat{\alpha}$ is clearly not greater than one. As a result, a mixture model is preferred, and it seems not suitable to interpret the volatility as well as the conditional mean structure by an AR-ARCH model since its inference usually needs a finite fourth-order moment (Francq and Zakoian 2004).

We consider the logistic MDAR model to the sequence of log returns $\{y_t\}$ to further take into account the interest rates since they are closely related to exchange rates. The interest rate spread is defined as the difference between the federal funds rate in U.S. and the interest rate in Japan, and they can be downloaded from the websites of the Federal Reserve Bank of New York (*http://www.federalreserve.gov/econresdata/*) and the Bank of Japan (*http://www.stat-search.boj.or.jp/ssi/mtshtml/w1_en.html*), respectively. There is an obvious trend in the sequence of interest rate spreads, especially during the period of the financial crisis at 2008, and its differenced sequence $\{x_t\}$ is then introduced to the mixing proportions of the logistic MDAR model; see Figure 2 for its time plot.

The BIC is employed to select the orders with $1 \le l, p_1, p_2 \le 3$, and there are 3^3 candidate models in total. The best-fitted model has the form of

$$y_t = \begin{cases} \varepsilon_t\{1.5101 \times 10^{-6}_{1.7434\times 10^{-7}} + 0.3677_{0.0088}y_{t-1}^2 \\ \quad +0.3168_{0.0069}y_{t-2}^2\}^{1/2} \\ -0.0108_{0.0001} + 0.1777_{0.0106}y_{t-1} + 0.0504_{0.0081}y_{t-2} \\ \qquad\qquad\qquad\qquad \text{with probability } \alpha_{1t} \\ \varepsilon_t\{8.7764 \times 10^{-5}_{1.2331\times 10^{-6}} + 0.1725_{0.0057}y_{t-1}^2\}^{1/2} \\ +0.0029_{0.0001} + 0.2101_{0.0055}y_{t-1} \\ \qquad\qquad\qquad\qquad \text{with probability } \alpha_{2t} \end{cases}$$

with the mixing proportions satisfying $\alpha_{1t} = 1 - \alpha_{2t}$ and

$$\ln(\alpha_{1t}/\alpha_{2t}) = -1.2858_{0.0388} - 27.7333_{1.0512}x_{t-1},$$

where standard errors of parameter estimates are given in subscripts, and are calculated based on the information matrix I_1 at Section 3.3. The estimated coefficients in the above model are all significant at the level of 5%.

The fitted values of the probability α_{1t} for the first component are presented in Figure 3, and they roughly have values around 0.2 after the financial crisis. Note that the volatility of the first component has a longer persistence than that of the second component. For the first 50 time points at the period of the financial crisis, both exchange rates and interest rates have much higher volatility, and the values of $\widehat{\alpha}_{1t}$ also vary dramatically. This may be due to the unstable financial environment during the financial crisis. Figure 3 also gives the density of the predictive conditional distribution $F(y_{t+1}|\mathcal{F}_t, \Omega_{t+1})$ at $t = 34$, and a bimodal feature can be very clearly observed. Moreover, it is skewed to the right, and this implies that the US Dollar is relatively strong compared with the Japanese Yen at that time point. Finally, Figure 3 lists the estimated 95% one-step ahead prediction intervals based on the estimated model. In sum, we may conclude the usefulness of the proposed MDAR models.

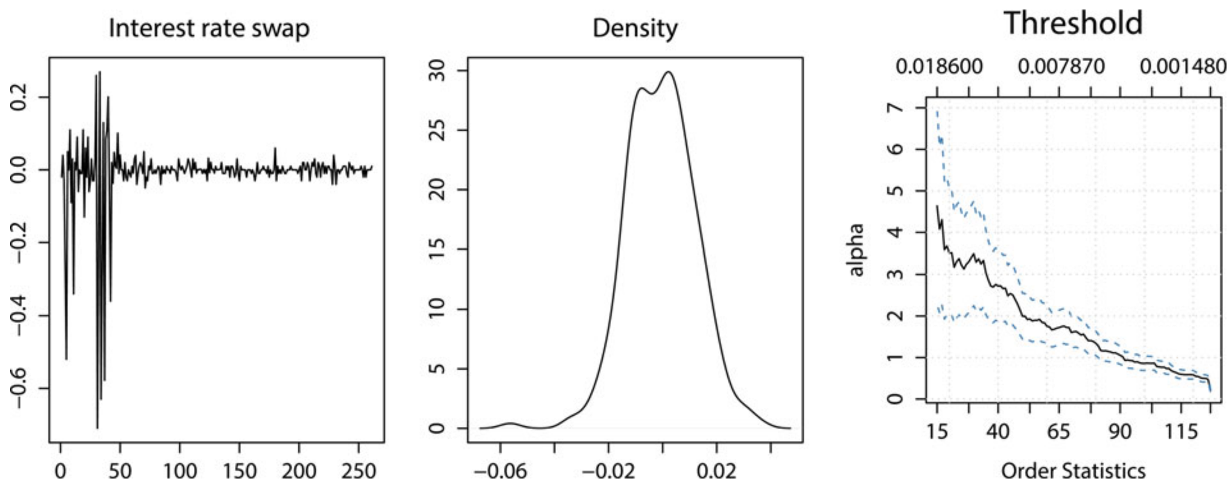

Figure 2. Differenced sequence of interest rate spreads (left panel), fitted kernel density (middle panel) and Hill estimator (right panel), of log returns for the weekly exchange rates of U.S. Dollar (USD) to Japanese Yen (JPY) from February 16, 2008, to February 9, 2013.

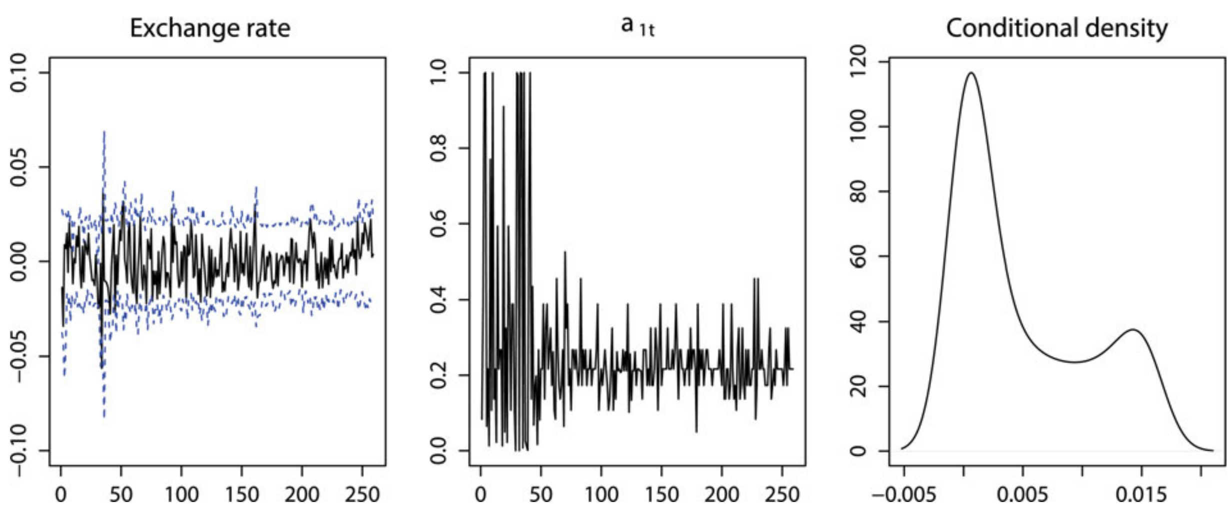

Figure 3. Time plot of log returns with estimated 95% one-step ahead prediction intervals (left panel), the probability of the first component $\widehat{\alpha}_{1t}$ (middle panel) and the conditional density of $F(y_{t+1}|\mathcal{F}_t, \Omega_{t+1})$ at time point $t = 34$ (right panel).

5. CONCLUSIONS

The double AR model (Ling 2004) has been shown to have a better performance in interpreting the heavy-tailed time series by allowing a wider parameter space, while these sequences may have a multimodal marginal or conditional distribution. The proposed MDAR model, especially the logistic MDAR model, is then suggested to interpret such type of time series, and their probabilistic properties and statistical inference are also derived. These models greatly extend the usefulness of the double AR model. Their potential in application is illustrated by the USD/JPY exchange rates returns.

APPENDIX: TECHNICAL DETAILS

This appendix gives proofs of Theorems 1–4, Corollary 1, and the derivation of Equation (9).

Proof of Theorem 1. Let $\mathbf{y}_t = (y_t, y_{t-1}, \ldots, y_{t-p+1})'$, and denote by $(\mathbb{R}^p, \mathcal{B}^p, \nu_p)$ the state space of process $\{\mathbf{y}_t\}$, where $\mathcal{B}^p$ is the class of Borel sets of $\mathbb{R}^p$, and ν_p is the Lebesgue measure on $(\mathbb{R}^p, \mathcal{B}^p)$. Then the process $\{\mathbf{y}_t\}$ is a homogeneous Markov chain with the transition probability

$$P(\mathbf{a}, \mathbf{A}) = \int_{m(\mathbf{A})} \sum_{k=1}^{K} \frac{\alpha_k}{\sqrt{\boldsymbol{\beta}_k' \mathbf{a}_2}} \phi\left(\frac{y - \boldsymbol{\theta}_k' \mathbf{a}_1}{\sqrt{\boldsymbol{\beta}_k' \mathbf{a}_2}}\right) dy,$$

$$\text{for} \quad \mathbf{a} \in \mathbb{R}^p \quad \text{and} \quad \mathbf{A} \in \mathcal{B}^p, \tag{A.1}$$

where $\mathbf{a} = (a_p, \ldots, a_1)'$, $\mathbf{a}_1 = (1, \mathbf{a}')'$, $\mathbf{a}_2 = (1, a_p^2, \ldots, a_1^2)'$, $\boldsymbol{\theta}_k = (\theta_{k0}, \ldots, \theta_{kp})'$, $\boldsymbol{\beta}_k = (\beta_{k0}, \ldots, \beta_{kp})'$, $\phi(\cdot)$ is the density function of the standard normal distribution, and $m(\cdot)$ is the projection map onto the first coordinate, that is, $m(\mathbf{a}) = a_p$. Note that the density function $\phi(\cdot)$ is positive everywhere and then, from (A.1), $P(\mathbf{a}, \mathbf{A}) > 0$ for each $\mathbf{a} \in \mathbb{R}^p$ whenever $\nu_p(\mathbf{A}) > 0$. It is implied that $\{\mathbf{y}_t\}$ is ν_p-irreducible.

We first prove the sufficiency. Since $\gamma < 0$, there exists an integer s such that $E \ln \|A_1 \ldots A_s\| < 0$. Let $\tilde{A}_t = A_t \ldots A_{t-s+1}$. It can be shown that the function $q(x) = E\|\tilde{A}_t\|^x$ is continuous and differentiable on $[0, 2)$ with $\lim_{x\to 0} q'(x) = E \ln \|\tilde{A}_t\| < 0$, where $q'(x)$ is the derivative function of $q(x)$. Then there exists a constant $0 < \delta < 1$ such that $E\|\tilde{A}_t\|^\delta < q(0) = 1$. Let $g(x) = 1 + \|x\|^\delta$. Consider a random coefficient autoregressive (RCAR) model,

$$\mathbf{y}_t^* = A_t \mathbf{y}_{t-1}^* + \boldsymbol{\varepsilon}_t, \tag{A.2}$$

where $\mathbf{y}_t^* = (y_t^*, y_{t-1}^*, \ldots, y_{t-p+1}^*)'$ and $\boldsymbol{\varepsilon}_t = (\sum_{k=1}^K z_{kt}[\theta_{k0} + \sqrt{\beta_{k0}}\varepsilon_t], 0, \ldots, 0)'$. For a compact set G with $\nu_p(G) > 0$, by a method similar to the proof of Theorem 2.1 in Ling (2007), we can show that

$$E[g(\mathbf{y}_{st}^*)|\mathbf{y}_{s(t-1)}^* = x] \le (1-\epsilon)g(x), \quad x \in G^c, \tag{A.3}$$

$$E[g(\mathbf{y}_{st}^*)|\mathbf{y}_{s(t-1)}^* = x] \le M, \quad x \in G \tag{A.4}$$

for some $0 < \epsilon < 1$ and $M > 0$, where $\{\mathbf{y}_{st}^*\}$ is the s-step Markov chain of $\{\mathbf{y}_t^*\}$ in (A.2). Note that $\{\mathbf{y}_t^*\}$ is a homogenous Markov chain with the same transition probability as that in (A.1), and then Equations (A.3) and (A.4) still hold for $\{\mathbf{y}_{st}\}$, the s-step Markov chain of $\{\mathbf{y}_t\}$. By Theorem 4 (ii) of Tweedie (1983) and Theorems 1 and 2 of Feigin and Tweedie (1985), we can show that process $\{\mathbf{y}_{st}\}$ is geometrically ergodic with a unique stationary distribution $\pi(\cdot)$ and $\int \|\mathbf{y}_{st}\|^\delta d\pi \le \int_{\mathbb{R}^p} g(x)\pi(dx) < \infty$. Finally, by Lemma 3.1 of Tjostheim (1990), it can be shown that $\mathbf{y}_t$ or y_t is geometrically ergodic with $E|y_t|^\delta < \infty$.

We next consider the necessity. Suppose that $\{y_t\}$ is the strictly stationary solution to the MDAR model (1) with a stationary distribution $\pi(\cdot)$. Then there also exists a strictly stationary solution to the RCAR model (A.2) since their transition probabilities are the same. Consider the p-step Markov chain $\{\mathbf{y}_{pt}^*\}$,

$$\mathbf{y}_{tp}^* = \tilde{A}_{tp}\mathbf{y}_{(t-1)p}^* + \boldsymbol{\varepsilon}_t^*, \tag{A.5}$$

where $s = p$ and $\boldsymbol{\varepsilon}_t^* = \boldsymbol{\varepsilon}_{tp} + \sum_{i=1}^{p-1}\prod_{j=0}^{i-1} A_{tp-j}\boldsymbol{\varepsilon}_{tp-i}$. By a method similar to the proof of Theorem 2.1 in Ling (2007), we can verify that model (A.5) is irreducible, and the corresponding process $\{\mathbf{y}_{pt}^*\}$ is its nonanticipative strictly stationary solution. By Theorem 2.5 of Bougerol and Picard (1992), it holds that

$$\tilde{\gamma} = \inf\{t^{-1}E\ln\|\tilde{A}_p\tilde{A}_{2p}\ldots\tilde{A}_{tp}\|, t \ge 1\} < 0.$$

Note that $\tilde{A}_p\tilde{A}_{2p}\ldots\tilde{A}_{tp} = A_1\ldots A_{tp}$, and then $\gamma < p^{-1}\tilde{\gamma} < 0$. We accomplish the proof of Theorem 1. □

Proof of Theorem 2. From the proof of Theorem 1, we know that the MDAR process $\{\mathbf{y}_t\}$ and the RCAR process $\{\mathbf{y}_t^*\}$ at (A.2) have the same transition probability. Then it is sufficient to show that $E\|\mathbf{y}_t^*\|^m < \infty$ with $m = 2$ and 4, and this can be accomplished by a method similar to Ling (1999). □

Proof of Theorem 3. For these functions in $l_t(\boldsymbol{\lambda})$ and $\partial l_t(\boldsymbol{\lambda})/\partial\boldsymbol{\lambda}$, we have their derivatives as follows,

$$\frac{\partial\mu_{kt}(\boldsymbol{\theta}_k)}{\partial\boldsymbol{\theta}_k} = \mathbf{y}_{k1t}, \qquad \frac{\partial h_{kt}(\boldsymbol{\beta}_k)}{\partial\boldsymbol{\beta}_k} = \mathbf{y}_{k2t},$$

$$\frac{\partial\alpha_{1t}(\boldsymbol{\varphi})}{\partial\boldsymbol{\varphi}} = \alpha_{1t}(\boldsymbol{\varphi})\alpha_{2t}(\boldsymbol{\varphi})\mathbf{x}_t, \qquad \frac{\partial\alpha_{2t}(\boldsymbol{\varphi})}{\partial\boldsymbol{\varphi}} = -\alpha_{1t}(\boldsymbol{\varphi})\alpha_{2t}(\boldsymbol{\varphi})\mathbf{x}_t,$$

$$\frac{\partial\ln[\alpha_{1t}(\boldsymbol{\varphi})]}{\partial\boldsymbol{\varphi}} = \alpha_{2t}(\boldsymbol{\varphi})\mathbf{x}_t, \qquad \frac{\partial\ln[\alpha_{2t}(\boldsymbol{\varphi})]}{\partial\boldsymbol{\varphi}} = -\alpha_{1t}(\boldsymbol{\varphi})\mathbf{x}_t,$$

$$\frac{\partial\tau_{1t}(\boldsymbol{\lambda})}{\partial\boldsymbol{\varphi}} = \tau_{1t}(\boldsymbol{\lambda})\tau_{2t}(\boldsymbol{\lambda})\mathbf{x}_t,$$

$$\frac{\partial\tau_{1t}(\boldsymbol{\lambda})}{\partial\boldsymbol{\theta}_k} = (-1)^{k+1}\tau_{1t}(\boldsymbol{\lambda})\tau_{2t}(\boldsymbol{\lambda})\frac{[y_t-\mu_{kt}(\boldsymbol{\theta}_k)]}{h_{kt}(\boldsymbol{\beta}_k)}\mathbf{y}_{k1t},$$

$$\frac{\partial\tau_{1t}(\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}_k} = (-1)^{k+1}\tau_{1t}(\boldsymbol{\lambda})\tau_{2t}(\boldsymbol{\lambda})\frac{1}{2h_{kt}(\boldsymbol{\beta}_k)} \times \left\{\frac{[y_t-\mu_{kt}(\boldsymbol{\theta}_k)]^2}{h_{kt}(\boldsymbol{\beta}_k)} - 1\right\}\mathbf{y}_{k2t},$$

and $\partial\tau_{2t}(\boldsymbol{\lambda})/\partial\boldsymbol{\lambda} = -\partial\tau_{1t}(\boldsymbol{\lambda})/\partial\boldsymbol{\lambda}$, where $k = 1$ and 2. Note that $E\|\mathbf{x}_t\|^2 < \infty$. By a method similar to Ling (2007), we can verify that

$$E\sup_{\boldsymbol{\lambda}\in\Lambda}\left\|\frac{\partial l_t(\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}}\right\|^2 < \infty \quad \text{and} \quad E\sup_{\boldsymbol{\lambda}\in\Lambda}\left\|\frac{\partial^2 l_t(\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}'}\right\|^2 < \infty.$$

We first prove the consistency and, by following the standard arguments in Huber (1967), it is sufficient to verify the following two claims:

(i) $E[l_t(\boldsymbol{\lambda})] \le E[l_t(\boldsymbol{\lambda}_0)]$ for all $\boldsymbol{\lambda} \in \Lambda$, and the equality holds if and only if $\boldsymbol{\lambda} = \boldsymbol{\lambda}_0$.

(ii) For any $\boldsymbol{\lambda} \in \Lambda$, define an open neighborhood of $\boldsymbol{\lambda}$ with radius $0 < \eta < 1$ as $U_\lambda(\eta) = \{\boldsymbol{\lambda}^* \in \Lambda : \|\boldsymbol{\lambda}^* - \boldsymbol{\lambda}\| < \eta\}$. Then it holds that $E\sup_{\boldsymbol{\lambda}^*\in U_\lambda(\eta)}|l_t(\boldsymbol{\lambda}^*) - l_t(\boldsymbol{\lambda})| \to 0$ as $\eta \to 0$.

Denote

$$f_t(y,\boldsymbol{\lambda}) = \sum_{k=1}^2 \frac{\alpha_{kt}(\boldsymbol{\varphi})}{\sqrt{h_{kt}(\boldsymbol{\beta}_k)}}\exp\left[-\frac{(y-\mu_{kt}(\boldsymbol{\theta}_k))^2}{2h_{kt}(\boldsymbol{\beta}_k)}\right].$$

Note that $l_t(\boldsymbol{\lambda}) = \ln[f_t(y_t,\boldsymbol{\lambda})]$, and $f_t(y,\boldsymbol{\lambda})$ is the conditional density of the logistic MDAR process at (5). By Jensen's inequality, it holds that

$$E[l_t(\boldsymbol{\lambda})] - E[l_t(\boldsymbol{\lambda}_0)] = E\left\{E\left[\ln\frac{f_t(y_t,\boldsymbol{\lambda})}{f_t(y_t,\boldsymbol{\lambda}_0)}|\mathcal{F}_{t-1},\Omega_t\right]\right\} \le E\left\{\ln E\left[\frac{f_t(y_t,\boldsymbol{\lambda})}{f_t(y_t,\boldsymbol{\lambda}_0)}|\mathcal{F}_{t-1},\Omega_t\right]\right\} = 0,$$

where the equality holds if and only if

$$f_t(y,\boldsymbol{\lambda}) = f_t(y,\boldsymbol{\lambda}_0) \quad \text{with probability one.} \tag{A.6}$$

By conditioning $f_t(y,\boldsymbol{\lambda})$ and $f_t(y,\boldsymbol{\lambda}_0)$ on σ-fields $\mathcal{F}_{t-1}$ and Ω_t, together with the assumption of $\varphi_0 \le 0$, we can obtain from (A.6) that, with probability one,

$$(\alpha_{1t}(\boldsymbol{\varphi}),\mu_{1t}(\boldsymbol{\theta}_1),h_{1t}(\boldsymbol{\beta}_1),\alpha_{2t}(\boldsymbol{\varphi}),\mu_{2t}(\boldsymbol{\theta}_2),h_{2t}(\boldsymbol{\beta}_2)) = (\alpha_{1t}(\boldsymbol{\varphi}_0),\mu_{1t}(\boldsymbol{\theta}_{01}),h_{1t}(\boldsymbol{\beta}_{01}),\alpha_{2t}(\boldsymbol{\varphi}_0),\mu_{2t}(\boldsymbol{\theta}_{02}),h_{2t}(\boldsymbol{\beta}_{02})),$$

or equivalently

$$\mathbf{x}_t'(\boldsymbol{\varphi}-\boldsymbol{\varphi}_0) = \mathbf{y}_{k1t}'(\boldsymbol{\theta}_k-\boldsymbol{\theta}_{0k}) = \mathbf{y}_{k2t}'(\boldsymbol{\beta}_k-\boldsymbol{\beta}_{0k}) = 0, \quad k = 1 \text{ and } 2.$$

This implies that $\boldsymbol{\lambda} = \boldsymbol{\lambda}_0$, and hence we complete the proof of Claim (i).

Due to the compactness of the parameter space Λ, there exists a constant $c_\beta > 0$ such that $\beta_{kj} \ge c_\beta$ with $j = 0, 1, \ldots, p_k$ and $k = 1$ and 2. Note that $0 < \tau_{kt}(\boldsymbol{\lambda}) < 1$, $0 < \alpha_{kt}(\boldsymbol{\varphi}) < 1$, and $y_{t-j}^2/h_{kt}(\boldsymbol{\beta}_k) < \beta_{kj}^{-1} \le c_\beta^{-1}$ with $1 \le j \le p_k$, where $k = 1$ and 2. It can be shown that $E\sup_{\boldsymbol{\lambda}\in\Lambda}\|\partial l_t(\boldsymbol{\lambda})/\partial\boldsymbol{\lambda}\| < \infty$, and

$$E\sup_{\boldsymbol{\lambda}^*\in U_\lambda(\eta)}|l_t(\boldsymbol{\lambda}^*) - l_t(\boldsymbol{\lambda})| \le E\sup_{\boldsymbol{\lambda}\in\Lambda}\left\|\frac{\partial l_t(\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}}\right\|\cdot\eta.$$

Hence, Claim (ii) holds, and we accomplish the proof of consistency.

We now consider the asymptotic normality. The second-order derivative function of $l_t(\boldsymbol{\lambda})$ has the form of

$$\frac{\partial^2 l_t(\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}'} = \tau_{1t}(\boldsymbol{\lambda})\tau_{2t}(\boldsymbol{\lambda})\iota_{3t}(\boldsymbol{\lambda})\iota_{3t}'(\boldsymbol{\lambda}) - \text{diag}\{\alpha_{1t}(\boldsymbol{\lambda})\alpha_{2t}(\boldsymbol{\lambda})\mathbf{x}_t\mathbf{x}_t', \tau_{1t}(\boldsymbol{\lambda})\iota_{1t}(\boldsymbol{\lambda}), \tau_{2t}(\boldsymbol{\lambda})\iota_{2t}(\boldsymbol{\lambda})\},$$

where the symmetric matrices

$$\iota_{kt}(\boldsymbol{\lambda}) = \begin{pmatrix} h_{kt}^{-1}(\boldsymbol{\beta}_k)\mathbf{y}_{k1t}\mathbf{y}_{k1t}' & h_{kt}^{-2}(\boldsymbol{\beta}_k)[y_t-\mu_{kt}(\boldsymbol{\theta}_k)]\mathbf{y}_{k1t}\mathbf{y}_{k2t}' \\ * & h_{kt}^{-2}(\boldsymbol{\beta}_k)\{[y_t-\mu_{kt}(\boldsymbol{\theta}_k)]^2/h_{kt}(\boldsymbol{\beta}_k) - 0.5\}\mathbf{y}_{k2t}\mathbf{y}_{k2t}' \end{pmatrix}$$

with $k = 1$ and 2, and

$$\iota_{3t}(\boldsymbol{\lambda}) = \left(\mathbf{x}_t', \frac{y_t-\mu_{1t}(\boldsymbol{\theta}_1)}{h_{1t}(\boldsymbol{\beta}_1)}\mathbf{y}_{11t}', \left\{\frac{[y_t-\mu_{1t}(\boldsymbol{\theta}_1)]^2}{h_{1t}(\boldsymbol{\beta}_1)} - 1\right\}\frac{\mathbf{y}_{12t}'}{2h_{1t}(\boldsymbol{\beta}_1)}, -\frac{y_t-\mu_{2t}(\boldsymbol{\theta}_2)}{h_{2t}(\boldsymbol{\beta}_2)}\mathbf{y}_{21t}', -\left\{\frac{[y_t-\mu_{2t}(\boldsymbol{\theta}_2)]^2}{h_{2t}(\boldsymbol{\beta}_2)} - 1\right\}\frac{\mathbf{y}_{22t}'}{2h_{2t}(\boldsymbol{\beta}_2)}\right)';$$

see also Section 3.3. Note that $\sup_{\boldsymbol{\lambda}\in\Lambda}\|\mathbf{y}_{k2t}/h_{kt}(\boldsymbol{\beta}_k)\| \le c_\beta^{-1}\sqrt{p_k}$, and it then can be verified that

$$E\sup_{\boldsymbol{\lambda}\in\Lambda}\left\|\frac{\partial^2 l_t(\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}'} - E\left[\frac{\partial^2 l_t(\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}'}\right]\right\| \le 2E\sup_{\boldsymbol{\lambda}\in\Lambda}\left\|\frac{\partial^2 l_t(\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}'}\right\| < \infty.$$

Moreover, the process $\{y_t\}$ is strictly stationary and ergodic. Applying Theorem 3.1 of Ling and McAleer (2003) and Lemma B.1 of Ling (2007), we can obtain that

$$\sup_{\boldsymbol{\lambda}\in\Lambda}\left\| \frac{1}{n}\sum_{t=p+1}^{n}\frac{\partial^2 l_t(\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}'} - E\left[\frac{\partial^2 l_t(\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}'}\right]\right\| = o_p(1). \quad \text{(A.7)}$$

Suppose that there exists a constant vector $\mathbf{a} = (\mathbf{c}', \mathbf{a}_1', \mathbf{b}_1', \mathbf{a}_2', \mathbf{b}_2')' \in R^{l+2p_1+2p_2+5}$ such that

$$0 = \mathbf{a}'E\left[\frac{\partial^2 l_t(\boldsymbol{\lambda}_0)}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}'}\right]\mathbf{a} = \mathbf{a}'E\left[\frac{\partial l_t(\boldsymbol{\lambda}_0)}{\partial\boldsymbol{\lambda}}\frac{\partial l_t(\boldsymbol{\lambda}_0)}{\partial\boldsymbol{\lambda}'}\right]\mathbf{a} = E\left[\mathbf{a}'\frac{\partial l_t(\boldsymbol{\lambda}_0)}{\partial\boldsymbol{\lambda}}\right]^2.$$

For simplicity, we denote $\tau_{kt} = \tau_{kt}(\boldsymbol{\lambda}_0)$, $\alpha_{kt} = \alpha_{kt}(\boldsymbol{\lambda}_0)$, $\mu_{kt} = \mu_{kt}(\boldsymbol{\theta}_{0k})$, and $h_{kt} = h_{kt}(\boldsymbol{\beta}_{0k})$ with $k = 1$ and 2. Note that $\tau_{1t} > 0$ and

$$\frac{\tau_{2t}}{\tau_{1t}} = \frac{\alpha_{2t}}{\alpha_{1t}}\left(\frac{h_{1t}}{h_{2t}}\right)^{1/2}\exp\left\{-\frac{(y_t-\mu_{2t})^2}{2h_{2t}} + \frac{(y_t-\mu_{1t})^2}{2h_{1t}}\right\}. \quad \text{(A.8)}$$

It then holds that, with probability one,

$$\begin{aligned} 0 &= \tau_{1t}^{-1}\cdot\mathbf{a}'\frac{\partial l_t(\boldsymbol{\lambda}_0)}{\partial\boldsymbol{\lambda}} \\ &= \frac{\tau_{2t}}{\tau_{1t}}\left\{\frac{\mathbf{b}_2'\mathbf{y}_{22t}}{2h_{2t}^2}(y_t-\mu_{2t})^2 + \frac{\mathbf{a}_2'\mathbf{y}_{21t}}{h_{2t}}(y_t-\mu_{2t})\alpha_{1t}\mathbf{c}'\mathbf{x}_t - \frac{\mathbf{b}_2'\mathbf{y}_{22t}}{2h_{2t}}\right\} \\ &\quad + \frac{\mathbf{b}_1'\mathbf{y}_{12t}}{2h_{1t}^2}(y_t-\mu_{1t})^2 + \frac{\mathbf{a}_1'\mathbf{y}_{11t}}{h_{1t}}(y_t-\mu_{1t}) + (1-\alpha_{1t})\mathbf{c}'\mathbf{x}_t - \frac{\mathbf{b}_1'\mathbf{y}_{12t}}{2h_{1t}}. \end{aligned}$$

By conditioning the above equation on σ-fields $\mathcal{F}_{t-1}$ and Ω_t, together with (A.8) and the fact that the conditional distribution of y_t is a mixture Gaussian, we can obtain that $\mathbf{c}'\mathbf{x}_t = \mathbf{a}_k'\mathbf{y}_{k1t}h_{kt}^{-1/2} = \mathbf{b}_k'\mathbf{y}_{k2t}h_{kt}^{-1} = 0$ with probability one. Note that matrices $E(\mathbf{x}_t\mathbf{x}_t')$, $E(h_{kt}^{-1}\mathbf{y}_{k1t}\mathbf{y}_{k1t}')$, and $E(h_{kt}^{-2}\mathbf{y}_{k2t}\mathbf{y}_{k2t}')$ are all positive definite. Hence, $\mathbf{a}$ is a zero vector and $E[\partial^2 l_t(\boldsymbol{\lambda}_0)/(\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}')]$ is then a positive definite matrix. Moreover, by Taylor expansion,

$$0 = \frac{\partial L_n(\widehat{\boldsymbol{\lambda}}_n)}{\partial\boldsymbol{\lambda}} = \frac{\partial L_n(\boldsymbol{\lambda}_0)}{\partial\boldsymbol{\lambda}} + \frac{\partial^2 L_n(\boldsymbol{\lambda}_n^*)}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}'}(\widehat{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}_0),$$

where $\boldsymbol{\lambda}_n^*$ is between $\widehat{\boldsymbol{\lambda}}_n$ and $\boldsymbol{\lambda}_0$. Together with (A.7), the consistency of $\widehat{\boldsymbol{\lambda}}_n$, the positive definiteness of $E[\partial^2 l_t(\boldsymbol{\lambda}_0)/(\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}')]$ and the central limit theorem, we finish the proof for asymptotic normality. □

Proof of Theorem 4. We first consider the case with $\mathbf{p} > \mathbf{p}_0$, that is, $l \geq l_0$, $p_1 \geq p_{10}$, $p_2 \geq p_{20}$ and at least one inequality holds, where $\mathbf{p} = (l, p_1, p_2)$ and $\mathbf{p}_0 = (l_0, p_{10}, p_{20})$. Notations $\boldsymbol{\lambda}^{\mathbf{p}}$, $\boldsymbol{\lambda}_0^{\mathbf{p}}$, and $\widehat{\boldsymbol{\lambda}}_n^{\mathbf{p}}$ are employed to emphasize their dependence on the orders $\mathbf{p}$. Note that the model with orders $\mathbf{p}$ corresponds to a bigger model, and it holds that

$$l_t(\boldsymbol{\lambda}_0^{\mathbf{p}}) = l_t(\boldsymbol{\lambda}_0^{\mathbf{p}_0}) \quad \text{and} \quad L_n(\boldsymbol{\lambda}_0^{\mathbf{p}}) = L_n(\boldsymbol{\lambda}_0^{\mathbf{p}_0}). \quad \text{(A.9)}$$

Let $\mathbf{v} = n^{1/2}(\widehat{\boldsymbol{\lambda}}_n^{\mathbf{p}_0} - \boldsymbol{\lambda}_0^{\mathbf{p}_0})$ and, by a method similar to Li and Li (2008), we can show that

$$\begin{aligned} L_n(\widehat{\boldsymbol{\lambda}}_n^{\mathbf{p}_0}) - L_n(\boldsymbol{\lambda}_0^{\mathbf{p}_0}) &= L_n(\boldsymbol{\lambda}_0^{\mathbf{p}_0} + n^{-1/2}\mathbf{v}) - L_n(\boldsymbol{\lambda}_0^{\mathbf{p}_0}) \\ &= \mathbf{v}'n^{-1/2}\partial L_n(\boldsymbol{\lambda}_0^{\mathbf{p}_0})/\partial\boldsymbol{\lambda}^{\mathbf{p}} + \mathbf{v}'\Sigma\mathbf{v} \\ &\quad + o_p(1) = O_p(1) \end{aligned} \quad \text{(A.10)}$$

since $n^{-1/2}\partial L_n(\boldsymbol{\lambda}_0^{\mathbf{p}_0})/\partial\boldsymbol{\lambda}^{\mathbf{p}} = O_p(1)$. Similarly, $L_n(\widehat{\boldsymbol{\lambda}}_n^{\mathbf{p}}) - L_n(\boldsymbol{\lambda}_0^{\mathbf{p}}) = O_p(1)$. As a result,

$$\begin{aligned} L_n(\widehat{\boldsymbol{\lambda}}_n^{\mathbf{p}}) - L_n(\widehat{\boldsymbol{\lambda}}_n^{\mathbf{p}_0}) &= [L_n(\widehat{\boldsymbol{\lambda}}_n^{\mathbf{p}}) - L_n(\boldsymbol{\lambda}_0^{\mathbf{p}})] - [L_n(\widehat{\boldsymbol{\lambda}}_n^{\mathbf{p}_0}) - L_n(\boldsymbol{\lambda}_0^{\mathbf{p}_0})] \\ &\quad + [L_n(\boldsymbol{\lambda}_0^{\mathbf{p}}) - L_n(\boldsymbol{\lambda}_0^{\mathbf{p}_0})] = O_p(1), \end{aligned}$$

and

$$\begin{aligned} \text{BIC}(\mathbf{p}) - \text{BIC}(\mathbf{p}_0) &= -2[L_n(\widehat{\boldsymbol{\lambda}}_n^{\mathbf{p}}) - L_n(\widehat{\boldsymbol{\lambda}}_n^{\mathbf{p}_0})] \\ &\quad + (p^* - p_0^*)\ln(n-p) \\ &= O_p(1) + (p^* - p_0^*)\ln(n-p) \to +\infty \end{aligned}$$

as $n \to +\infty$, where $p^* = 2(p_1 + p_2) + l + 5$, $p_0^* = 2(p_{10} + p_{20}) + l_0 + 5$ and $p^* - p_0^* > 0$.

We next consider the case with $\mathbf{p} < \mathbf{p}_0$, that is, $l < l_0$, $p_1 < p_{10}$, or $p_2 < p_{20}$. Let $\boldsymbol{\lambda}_0^{\mathbf{p}} = \text{argmax}\, E[l_t(\boldsymbol{\lambda}^{\mathbf{p}})]$ and, by a method similar to the proof of Theorem 3 and together with (A.10), we can show that $\sqrt{n}(\widehat{\boldsymbol{\lambda}}_n^{\mathbf{p}} - \boldsymbol{\lambda}_0^{\mathbf{p}}) = O_p(1)$ and

$$L_n(\widehat{\boldsymbol{\lambda}}_n^{\mathbf{p}}) - L_n(\boldsymbol{\lambda}_0^{\mathbf{p}}) = O_p(1). \quad \text{(A.11)}$$

Let $\mathbf{p}^{\max} = \max(\mathbf{p}, \mathbf{p}_0) = [\max(l, l_0), \max(p_1, p_{10}), \max(p_2, p_{20})]$. Denote by $\boldsymbol{\lambda}_0^{\mathbf{p}^{\max}}$ and $\boldsymbol{\lambda}_0^{\mathbf{p}_0^{\max}}$ the parameter vectors of $\boldsymbol{\lambda}_0^{\mathbf{p}}$ and $\boldsymbol{\lambda}_0^{\mathbf{p}_0}$ with extra parameters being zero, respectively. From the conclusion of Claim (i) in the proof of Theorem 3 and by a method similar to (A.9) and the ergodic theorem, we have that $c = E[l_t(\boldsymbol{\lambda}_0^{\mathbf{p}^{\max}})] - E[l_t(\boldsymbol{\lambda}_0^{\mathbf{p}_0^{\max}})] < 0$, and

$$L_n(\boldsymbol{\lambda}_0^{\mathbf{p}}) - L_n(\boldsymbol{\lambda}_0^{\mathbf{p}_0}) = L_n(\boldsymbol{\lambda}_0^{\mathbf{p}^{\max}}) - L_n(\boldsymbol{\lambda}_0^{\mathbf{p}_0^{\max}}) = cn + o_p(n).$$

As a result, by (A.10)–(A.11),

$$\begin{aligned} L_n(\widehat{\boldsymbol{\lambda}}_n^{\mathbf{p}}) - L_n(\widehat{\boldsymbol{\lambda}}_n^{\mathbf{p}_0}) &= [L_n(\widehat{\boldsymbol{\lambda}}_n^{\mathbf{p}}) - L_n(\boldsymbol{\lambda}_0^{\mathbf{p}})] - [L_n(\widehat{\boldsymbol{\lambda}}_n^{\mathbf{p}_0}) - L_n(\boldsymbol{\lambda}_0^{\mathbf{p}_0})] \\ &\quad + [L_n(\boldsymbol{\lambda}_0^{\mathbf{p}}) - L_n(\boldsymbol{\lambda}_0^{\mathbf{p}_0})] \\ &= O_p(1) + o_p(n) + cn \end{aligned}$$

and

$$\begin{aligned} \text{BIC}(\mathbf{p}) - \text{BIC}(\mathbf{p}_0) &= -2[L_n(\widehat{\boldsymbol{\lambda}}_n^{\mathbf{p}}) - L_n(\widehat{\boldsymbol{\lambda}}_n^{\mathbf{p}_0})] + (p^* - p_0^*)\ln(n-p) \\ &= -2cn + O_p(1) + o_p(n) + O(\ln n) \to +\infty \end{aligned}$$

as $n \to +\infty$. This completes the proof. □

Proof of Corollary 1. We consider an autoregressive model,

$$\mathbf{y}_t^* = \sum_{k=1}^{K} z_{kt}A_{kt}\mathbf{y}_{t-1}^* + \boldsymbol{\varepsilon}_t, \quad \text{(A.12)}$$

where $P(z_{kt} = 1) = \alpha_{kt}$ for $1 \leq k \leq K$, $\ln(\alpha_{jt}/\alpha_{Kt}) = \mathbf{x}_t^{*\prime}\boldsymbol{\varphi}_j$ for $1 \leq j \leq K-1$, $\mathbf{x}_t^* = (1, y_{t-1}^*, \ldots, y_{t-p}^*)'$, $\mathbf{y}_t^* = (y_t^*, y_{t-1}^*, \ldots, y_{t-p+1}^*)'$, and $\boldsymbol{\varepsilon}_t = (\sum_{k=1}^{K} z_{kt}[\theta_{k0} + \sqrt{\beta_{k0}}\varepsilon_t], 0, \ldots, 0)'$.

For each k, since $\gamma_k < 0$, there exists an integer s_k such that $E\ln\|A_{k1}\ldots A_{ks_k}\| < 0$. Let $s = \prod_{k=1}^{K} s_k$ and $g(x) = 1 + \|x\|^\delta$ with $0 < \delta < 1$. By the proof of Theorem 1, we can verify the inequalities at (A.3) and (A.4), and the remaining proof can be accomplished similarly. □

Derivations of Equation (9). Let

$$l_{ct}(\boldsymbol{\lambda}) = \sum_{k=1}^{2} z_{kt}\left\{\ln[\alpha_{kt}(\boldsymbol{\varphi})] - 0.5\ln[h_{kt}(\boldsymbol{\beta}_k)] - 0.5\frac{[y_t - \mu_{kt}(\boldsymbol{\theta}_k)]^2}{h_{kt}(\boldsymbol{\beta}_k)}\right\},$$

and then $L_{cn}(\boldsymbol{\lambda}) = \sum_{t=p+1}^{n} l_{ct}(\boldsymbol{\lambda})$. The derivative functions of $l_{ct}(\boldsymbol{\lambda})$ have the form of

$$\frac{\partial l_{ct}(\boldsymbol{\lambda})}{\partial\boldsymbol{\varphi}} = [z_{1t} - \alpha_{1t}(\boldsymbol{\varphi})]\mathbf{x}_t, \qquad \frac{\partial l_{ct}(\boldsymbol{\lambda})}{\partial\boldsymbol{\theta}_k} = z_{kt}\frac{y_t - \mu_{kt}(\boldsymbol{\theta}_k)}{h_{kt}(\boldsymbol{\beta}_k)}\mathbf{y}_{k1t},$$

$$\frac{\partial l_{ct}(\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}_k} = z_{kt}\left\{\frac{[y_t - \mu_{kt}(\boldsymbol{\theta}_k)]^2}{h_{kt}(\boldsymbol{\beta}_k)} - 1\right\}\frac{\mathbf{y}_{k2t}}{2h_{kt}(\boldsymbol{\beta}_k)},$$

and

$$\frac{\partial^2 l_{ct}(\boldsymbol{\lambda})}{\partial\boldsymbol{\varphi}\partial\boldsymbol{\varphi}'} = -\alpha_{1t}(\boldsymbol{\varphi})\alpha_{2t}(\boldsymbol{\varphi})\mathbf{x}_t\mathbf{x}_t',$$

$$\frac{\partial^2 l_{ct}(\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}_k\partial\boldsymbol{\theta}_k'} = -z_{kt}\frac{y_t-\mu_{kt}(\boldsymbol{\theta}_k)}{h_{kt}^2(\boldsymbol{\beta}_k)}\mathbf{y}_{k2t}\mathbf{y}_{k1t}',$$

$$\frac{\partial^2 l_{ct}(\boldsymbol{\lambda})}{\partial\boldsymbol{\theta}_k\partial\boldsymbol{\theta}_k'} = -\frac{z_{kt}}{h_{kt}(\boldsymbol{\beta}_k)}\mathbf{y}_{k1t}\mathbf{y}_{k1t}',$$

$$\frac{\partial^2 l_{ct}(\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}_k\partial\boldsymbol{\beta}_k'} = -\frac{z_{kt}}{h_{kt}^2(\boldsymbol{\beta}_k)}\left\{\frac{[y_t-\mu_{kt}(\boldsymbol{\theta}_k)]^2}{h_{kt}(\boldsymbol{\beta}_k)}-\frac{1}{2}\right\}\mathbf{y}_{k2t}\mathbf{y}_{k2t}',$$

$$\frac{\partial^2 l_{ct}(\boldsymbol{\lambda})}{\partial\boldsymbol{\varphi}\partial\boldsymbol{\theta}_k'} = \frac{\partial^2 l_{ct}(\boldsymbol{\lambda})}{\partial\boldsymbol{\varphi}\partial\boldsymbol{\beta}_k'} = \frac{\partial^2 l_{ct}(\boldsymbol{\lambda})}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_2'} = \frac{\partial^2 l_{ct}(\boldsymbol{\lambda})}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\beta}_2'}$$

$$= \frac{\partial^2 l_{ct}(\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}_1\partial\boldsymbol{\theta}_2'} = \frac{\partial^2 l_{ct}(\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}_1\partial\boldsymbol{\beta}_2'} = 0$$

with $k = 1$ and 2. Note that $\{\mathbf{z}_t\}$ are independent conditional on the σ-fields $\mathcal{F}_n$ and Ω_n; see Wong and Li (2000). Thus,

$$\text{var}\left(\frac{\partial L_{cn}(\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}}\Big|\boldsymbol{\lambda},\mathcal{F}_n,\Omega_n\right) = \sum_{t=p+1}^{n}\text{var}\left(\frac{\partial l_{ct}(\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}}\Big|\boldsymbol{\lambda},\mathcal{F}_n,\Omega_n\right).$$

Together with the fact that $\text{var}(z_{kt}|\boldsymbol{\lambda},\mathcal{F}_n,\Omega_n) = \tau_{1t}(\boldsymbol{\lambda})\tau_{2t}(\boldsymbol{\lambda})$ and $\text{cov}(z_{1t},z_{2t}|\boldsymbol{\lambda},\mathcal{F}_n,\Omega_n) = -\tau_{1t}(\boldsymbol{\lambda})\tau_{2t}(\boldsymbol{\lambda})$, we can derive the results in (9). □

ACKNOWLEDGMENTS

The authors are grateful to the editor, an associate editor, and two anonymous referees for their valuable comments and constructive suggestions that led to the substantial improvement of this article. The authors also thank the Hong Kong RGC grant HKU702613P for partial support.

[Received February 2015. Revised September 2015.]

REFERENCES

Agresti, A. (2002), *Categorical Data Analysis* (2nd ed.), Hoboken: Wiley. [307,308]

Bollerslev, T. (1986), "Generalized Autoregression Conditional Heteroscedasticity," *Journal of Econometrics*, 31, 307–327. [306]

Bougerol, P., and Picard, N. (1992), "Strict Stationarity of Generalized Autoregressive Processes," *The Annals of Probability*, 20, 1714–1730. [315]

Chan, K. S., and Tong, H. (1986), "On Estimating Thresholds in Autoregressive Models," *Journal of Time Series Analysis*, 7, 178–190. [309]

Cheng, X. X., Yu, P. L. H., and Li, W. K. (2009), "On a Dynamic Mixture GARCH Models," *Journal of Forecasting*, 28, 247–265. [306]

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood From Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society,* Series B, 39, 1–38. [309]

Engle, R. F. (1982), "Autoregression Conditional Heteroscedasticity With Estimates of the Variance of U.K. Inflation," *Econometrica*, 50, 987–1008. [306]

Feigin, P. D., and Tweedie, R. L. (1985), "Random Coefficient Autoregressive Processes: A Markov Chain Analysis of Stationarity and Finiteness of Moments," *Journal of Time Series Analysis*, 6, 1–14. [315]

Francq, C., and Zakoian, J. M. (2004), "Maximum Likelihood Estimation of Pure GARCH and ARMA-GARCH Processes," *Bernoulli*, 10, 605–637. [306,313]

Hall, P., and Heyde, C. C. (1980), *Martingale Limit Theory and Its Application*, New York: Academic Press. [308]

Hill, B. M. (1975), "A Simple General Approach to Inference About the Tail of a Distribution," *The Annals of Statistics*, 3, 1163–1174. [313]

Hill, J. B. (2010), "On Tail Index Estimation for Dependent, Heterogeneous Data," *Econometric Theory*, 26, 1398–1436. [313]

Huber, P. J. (1967), "The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1), Berkeley: University of California Press, pp. 221–233. [315]

Hyndman, R. J. (1996), "Computing and Graphing Highest Density Regions," *The American Statistician*, 50, 120–126. [307]

Li, G., Guan, B., Li, W. K., and Yu, P. L. H. (2015), "Hysteretic Autoregressive Time Series Models," *Biometrika*, 102, 717–723. [309]

Li, G., and Li, W. K. (2005), "Diagnostic Checking for Time Series Models With Conditional Heteroscedasticity Estimated by the Least Absolute Deviation Approach," *Biometrika*, 92, 691–701. [306]

——— (2008), "Least Absolute Deviation Estimation for Fractionally Integrated Autoregressive Moving Average Time Series Models With Conditional Heteroscedasticity," *Biometrika*, 95, 399–414. [306,316]

——— (2009), "Least Absolute Deviation Estimation for Unit Root Processes With GARCH Errors," *Econometric Theory*, 25, 1208–1227. [306]

Ling, S. (1999), "On Probability Properties of a Double Threshold ARMA Conditional Heteroskedasticity Model," *Journal of Applied Probability*, 36, 688–705. [308,315]

——— (2004), "Estimation and Testing Stationarity for Double-autoregressive Models," *Journal of the Royal Statistical Society,* Series B, 66, 63–78. [306,314]

——— (2007), "A Double AR(p) Model: Structure and Estimation," *Statistica Sinica*, 17, 161–175. [306,308,315]

Ling, S., and Li, D. (2008), "Asymptotic Inference for a Nonstationary Double AR(1) Model," *Biometrika*, 95, 257–263. [306]

Ling, S., and McAleer, M. (2003), "Asymptotic Theory for a Vector ARMA-GARCH Model," *Econometric Theory*, 19, 280–310. [316]

Louis, T. A. (1982), "Finding the Observed Information Matrix When Using the EM Algorithm," *Journal of the Royal Statistical Society,* Series B, 91, 1504–1514. [310]

McLachlan, G. J., and Krishnan, T. (1997), *The EM Algorithm and Extensions*, New York: Wiley. [309,310]

McLachlan, G. J., and Peel, D. (2000), *Finite Mixture Models*, New York: Wiley. [309]

Tjostheim, D. (1990), "Nonlinear Time Series and Markov Chains," *Advances in Applied Probability*, 22, 587–611. [315]

Tong, H. (1990), *Nonlinear Time Series: A Dynamical System Approach*, Oxford: Oxford University Press. [306]

Tweedie, R. L. (1983), "Criteria for Rates of Convergence of Markov Chains, With Application to Queueing and Storage Theory," in *Probability, Statistics and Analysis*, eds. J. F. C. Kingman and G. E. H. Reuter, Cambridge: Cambridge University Press, pp. 260–276. [315]

Weiss, A. (1986), "Asymptotic Theory for ARCH Models: Estimation and Testing," *Econometric Theory*, 2, 107–131. [306]

Wong, C. S., and Li, W. K. (2000), "On a Mixture Autoregressive Model," *Journal of the Royal Statistical Society,* Series B, 62, 95–115. [306,307,308,310,317]

——— (2001a), "On a Logistic Mixture Autoregressive Model," *Biometrika*, 88, 833–846. [306,308,310]

——— (2001b), "On a Mixture Autoregressive Conditional Heteroscedastic Model," *Journal of the American Statistical Association*, 96, 982–995. [306,307]

Wu, C. F. J. (1983), "On the Convergence Properties of the EM Algorithm," *The Annals of Statistics*, 11, 95–103. [310]

Zhang, Z., Li, W. K., and Yuen, K. C. (2006), "On a Mixture GARCH Time-Series Model," *Journal of Time Series Analysis*, 27, 577–597. [306]

Zhu, K., and Ling, S. (2013), "Quasi-Maximum Exponential Likelihood Estimators for a Double AR(p) Model," *Statistica Sinica*, 23, 251–270. [306]