A General Autoregressive Model with Markov Switching: Estimation and Consistency

Yingfu Xie, Jun Yu and Bo Ranneby

Research Report
Centre of Biostochastics

Swedish University of Agricultural Sciences

Report 2007:6
ISSN 1651-8543

A general autoregressive model with Markov switching: estimation and consistency

Yingfu Xie[1], Jun Yu and Bo Ranneby

Centre of Biostochastics
Swedish University of Agricultural Sciences
SE-901 83 Umeå, Sweden

Abstract

In this paper, a general autoregressive model with Markov switching is considered, where the autoregression may be of an infinite order. The consistency of the maximum likelihood estimators for this model is obtained under regular assumptions. Examples of finite and infinite order Markov switching AR models are discussed. The simulation study with these examples illustrates the consistency and asymptotic normality of the estimators.

Keywords: General autoregressive model, Markov switching, MLE, consistency, asymptotic normality.

[1]E-mail address to the correspondence author: yingfu.xie@sekon.slu.se

1 Introduction

In economic and financial studies, more and more researchers think it may be more reasonable to consider that there are different economic states behind an economic system, and the outcomes of the system depend on these states. For example, in the analysis of the US annual GNP (gross national product) growth rate series, Hamilton (1989) treated the expansion and recession periods as two states (called regimes in econometric literature), and assumed that the shift between these two regimes was governed by a Markov chain. But the Markov chain is unobservable, the inference has to be based only on the observations, the outcomes of some economic variables. Hamilton (1989) then studied the GNP series with a (linear) autoregressive model with Markov switching. The original model (Hamilton, 1989) may be written as

$$(Y_t - \mu_{X_t}) = \textstyle\sum_{i=1}^{s} \beta_i (Y_{t-i} - \mu_{X_{t-i}}) + \varepsilon_t,$$

where $\{Y_t\}$ are the observations in question, β_i, $i = 1, ..., s$, are coefficients, X_t is the regime at time t, μ_{X_t} is a constant depending on the regime X_t (assuming two regimes in his model) and ε_t is distributed as Gaussian, $N(0, \sigma^2)$. Hamilton (1989) showed that this model better fits the data than other, e.g. one-regime autoregressive models.

During the last two decades, the idea of Markov switching has been widely accepted and applied to many areas. See e.g. McCulloch and Tsay (1994) and Doucet et al. (2000) among others. In this paper, we are interested in a general autoregressive model with Markov switching (GARMS)

$$Y_t = f_\theta(\bar{\mathbf{Y}}_{t-1}^{s}, X_t; \varepsilon_t), \tag{1}$$

where $\{\varepsilon_t\}$ is an independent, identically distributed (i.i.d.) innovation process, $\bar{\mathbf{Y}}_{t-1}^{s}$ denotes observations $\{Y_s, ..., Y_{t-1}\}$ from possibly an infinite past, i.e. s may equal to $-\infty$. By convention, this set is empty when $s > t-1$. The parameter θ may depend on the regimes R_t, and is assumed to be finite-dimensional. An infinite-dimensional setting sounds attractive but is technically formidable. f_θ is a family of measurable functions indexed by θ and has implicit requirements imposed by the (conditional) density of Y_t. The maximum likelihood (ML) estimation of model (1) will be considered.

When $s > t-1$, i.e. the conditional distribution of Y_t does not depend on lagged Y's but only on X_t, model (1) leads to an important class of models, called hidden Markov models (HMMs). HMMs are widely used in many different areas such as engineering, biology, physiology, and econometrics. The readers are referred to the monograph by MacDonald and Zucchini (1997) and references therein for a comprehensive introduction. The asymptotic properties of the ML estimator (MLE) for HMMs are well developed. Baum and Petrie (1966) and Petrie (1969) studied the consistency and asymptotic normality of the MLE for probabilistic functions of a Markov chain where observations and the Markov states both take finite values. Leroux (1992) initially proved the consistency of MLE for general HMMs using ergodic theorem of subadditive processes from Kingman (1973) and the idea of Wald (1949). The asymptotic normality of MLE for general HMMs was obtained by Bickel and Ritov (1996) and Bickel et al. (1998) for Markov chain with finite state space. This result was extended to HMMs with compact state-space regimes by Jensen and Petersen (1999). The proofs of asymptotic normality in these papers follow the classic Cramér method, i.e. involving a central limit theorem of the Fisher score function and a law of large numbers for the observed Fisher information.

When $s = t - 1$, i.e. a first order autoregression in (1), Yao and Attali (2000) studied the stability of this process, including conditions under which there exist a stationary solution and moments of $\{Y_t\}$ and limit theorems can be applied. Francq and Roussignol (1998) considered the stability of the process when $s = t - 1$. In addition, they obtained the consistency of the MLE in this case. Yao (2001) also studied the condition for the existence of a square-integrable stationary solution when (1) is an ordinary autoregressive (AR) process (with coefficients depending on the regimes). In these works, the Markov chain has a finite state space.

A natural and interesting case is that of s less than $t-1$ but finite. The finite order autoregressive model with Markov switching is called ARMS model. In fact, the results for ARMS model are available only rather recently. Following basically Leroux's (1992) idea on HMMs, Krishnamerthy and Rydén (1998) obtained the consistency of the MLE when the Markov chain of this model has a finite state space. Douc et al. (2004) not only extended it to continuous state space but also proved the asymptotic normality of MLE for both stationary and non-stationary observation sequences. It is worth noting that the ARMS model includes the ARCH with Markov switching (MS-ARCH) model as a special case, which is non-linear and incorporates the idea of Markov regimes into the conditional variance equation of the seminal ARCH model (Engle, 1982). Therefore, the results of Douc et al. (2004) also cover the previous findings for MS-ARCH models by, e.g. Cai (1994) and Francq et al. (2001).

In this paper, we will consider the GARMS model (1) where the conditional density of Y_t can depend on all past observations, from possibly an infinite past, and the Markov chain has a finite state space. The ML estimation of this model is discussed and the consistency is proved. The assumptions for consistency are discussed using examples of finite and infinite order Markov switching autoregressive (MS-AR) models. Simulation studies with these examples are also reported. This paper is organized as follows. The notation and assumptions for the model are introduced and discussed in Section 2. In Section 3 the consistency of the MLE is proved. For the examples of finite and infinite order MS-AR models, conditions associated with the consistency are discussed in Section 4. Simulation studies using these two models are conducted to illustrate the consistency. The asymptotic normality of the MLEs is also conjectured by the numerical experiment. In Section 5 some conclusions are drawn.

2 Notation and assumptions

Let $\{X_t\}_{t\in\mathbb{Z}}$ be the Markov chain with finite state $\mathbf{E} = \{1, 2, ..., d\}$, where d is assumed known and fixed. Its transition probability matrix is $\mathbf{A} = (\alpha_{kl})$, where $\alpha_{kl} = P(X_t = l|X_{t-1} = k), k, l \in \mathbf{E}$. We assume that

(**A1**) The Markov chain $\{X_t\}_{t\in\mathbb{Z}}$ is aperiodic and irreducible.

An aperiodic and irreducible Markov chain has a unique stationary distribution. Denote this stationary distribution vector as π and $\pi(k) = P(X_t = k) > 0$ for all $k \in \mathbf{E}$.

Let $\{Y_t\}_{t\in\mathbb{Z}}$ be a real-valued observable sequence governed by (1). Assume that given the distribution of ε_t, the regime $X_t = k$, observations $\bar{\mathbf{y}}_{t-1}^s$ and function f_{θ_k}, the conditional distribution of Y_t has a density $q(y_t|\bar{\mathbf{y}}_{t-1}^s; \theta_k)$ with respect to some Lesbegue measure. Here $\theta_k, k \in \mathbf{E}$, belong to a finite-dimensional parameter space Θ. For $\{Y_t\}$, further assume that

(**A2**) $\{Y_t\}_{t\in\mathbb{Z}}$ is a strictly stationary and ergodic process.

Remark 1. Stationarity and ergodicity are usually assumed so as to be able to apply ergodic theorems. In the HMMs case, these properties carry over from the irreducibility of finite state Markov chain (Leroux, 1992, Lemma 1). For the ARMS model, Krishnamurthy and Rydén (1998) assumed the process $\{(X_t, Y_t, ..., Y_{t-s+1})\}_{t\in\mathbb{Z}}$ (with s finite) to be stationary and ergodic. Douc et al. (2004) first obtained the consistency and asymptotic normality for stationary $\{Y_t\}$ and then extended the results to a non-stationary case.

The conditions for the existence of such a stationary and ergodic solution of (1) have been investigated in some special cases. For the first order autoregressive model, i.e. $s = t-1$ in (1), conditions were obtained by Francq and Roussignol (1998) and Yao (2001). Roughly speaking, they required that the function f_θ should have some property like Lipschitz continuity and that the autogressive system (1) is "contractive on average". These conditions are usually not easy to verify except by simulations, and the details can be found in the two references. Yao and Attali (2000) also considered the conditions for the existence of moments (larger than one) and for limit theorems to be applied when $s = t-1$ in (1). However, the conditions for the general model (1) are still unknown.

Denote the parameters that characterize this model wholly as ϕ, which belongs to a parameter space Φ, a subset of the Euclidean space. That is, formally we have $\mathbf{A}(\phi) = (\alpha_{kl}(\phi))$ and $\theta_k(\phi) \in \Phi$ for $k, l \in \mathbf{E}$. The usual case is just $\phi = \{\alpha_{11}, \alpha_{12}, ..., \alpha_{dd}, \theta_1, ..., \theta_d\}$ (θ_k may be a vector), and $\alpha_{kl}(\cdot)$ and $\theta_k(\cdot)$ equal to coordinate projections. The true parameter is denoted by ϕ_0 and assume $\phi_0 \in \Phi$. With the Euclidean distance $\lambda(\cdot, \cdot)$, it is assumed that

(**A3**) For each l and k in $\mathbf{E}$, $\alpha_{kl}(\cdot)$ and $\theta_k(\cdot)$ are continuous on Φ, and $q(y_t|\bar{\mathbf{y}}_{t-1}^s; \theta_k(\phi))$ is continuous on Φ for all realizations $\bar{\mathbf{y}}_{t-1}^s$.

Suppose that observations $\{y_1, ..., y_n\}$ are given while the Markov chain $\{X_t\}$ is not observed. In this paper, we will study the consistency of the estimator of the parameter ϕ using MLE. The likelihood we work with has the form

$$L_n^*(y_n, ..., y_1; \phi) = \sum_{(x_1, ..., x_n)} \pi(x_1) \left\{ \prod_{t=2}^n \alpha(x_{t-1}, x_t) \right\} \left\{ \prod_{t=1}^n q(y_t|\bar{\mathbf{y}}_{t-1}^1; \theta_{x_t}(\phi)) \right\}. \tag{2}$$

The MLE is defined as any parameter $\hat{\phi}_n$ that maximizes the likelihood L_n^* over a compact subset of Φ. Its existence is guaranteed from the continuity assumption (A3) and the compactness. The likelihood (2) can be numerically maximized rather quickly by observing that it can be rewritten as a product of matrices. Let $\mathbf{M}_t(y_t; \phi)$ be the diagonal matrix $\text{diag}\{q(y_t|\bar{\mathbf{y}}_{t-1}^1; \theta_k(\phi)), k \in \mathbf{E}\}$, yielding

$$L_n^*(y_n, ..., y_1; \phi) = \pi \mathbf{M}_1(y_1; \phi) \left\{ \prod_{t=2}^n \mathbf{M}_t(y_t; \phi) \mathbf{A}(\phi) \right\} \mathbf{1}, \tag{3}$$

where $\mathbf{1}$ is an $d \times 1$ vector of ones.

REMARK 2. The initial state probabilities $\pi(x_1)$ in (2) may be taken as arbitrary positive quantities without changing the asymptotic property of model. However, using the stationary distribution of the Markov chain is the most convenient way to handle the initial state. Other possibilities include conditioning on it and then maximizing the likelihood with or without respect to it. For more discussions on the initial state, see e.g. Douc et al. (2004).

It will facilitate our proofs if we approximate the likelihood (2) by the conditional likelihood conditioning on all observations $\bar{\mathbf{y}}_0^{-\infty}$, which is

$$\begin{aligned} & L_n(y_n, ..., y_1|\bar{\mathbf{y}}_0^{-\infty}; \phi) \\ & \quad = \sum_{(x_1,...,x_n)\in\mathbf{E}^n} \pi(x_1) \left\{ \prod_{t=2}^{n} \alpha(x_{t-1}, x_t) \right\} \left\{ \prod_{t=1}^{n} q(y_t|\bar{\mathbf{y}}_{t-1}^{-\infty}; \theta_{x_t}(\phi)) \right\} \\ & \quad = \pi \mathbf{M}_1^{'}(y_1; \phi) \left\{ \prod_{t=2}^{n} \mathbf{M}_t^{'}(y_t; \phi) \mathbf{A}(\phi) \right\} \mathbf{1}, \end{aligned} \tag{4}$$

where $\mathbf{M}_t^{'}(y_t; \phi) = \text{diag}\{q(y_t|\bar{\mathbf{y}}_{t-1}^{-\infty}; \theta_k(\phi)), k \in \mathbf{E}\}$. Define $p^*(Y_t|\bar{\mathbf{Y}}_{t-1}^{1}; \phi)$ as the conditional density of Y_t given $\bar{\mathbf{Y}}_{t-1}^{1}$ and $g^*(Y_t|\bar{\mathbf{Y}}_{t-1}^{1}; \phi)$ as its logarithm; similarly define $p(Y_t|\bar{\mathbf{Y}}_{t-1}^{-\infty}; \phi)$ as the conditional density of Y_t given $\bar{\mathbf{Y}}_{t-1}^{-\infty}$ and its logarithm $g(Y_t|\bar{\mathbf{Y}}_{t-1}^{-\infty}; \phi)$.

Denote the term "almost surely under ϕ_0" by $a.s. - P_{\phi_0}$ and the following assumption is needed.

(A4) $\forall\ \phi \in \Phi$, $0 < \min_{k\in\mathbf{E}} q(Y_t|\bar{\mathbf{Y}}_{t-1}^{-\infty}; \theta_k(\phi)) \leq \max_{k\in\mathbf{E}} q(Y_t|\bar{\mathbf{Y}}_{t-1}^{-\infty}; \theta_k(\phi)) < \infty$ for all $\bar{\mathbf{Y}}_t^{-\infty}$ $a.s. - P_{\phi_0}$; and there exists a neighborhood of ϕ, $V(\phi) = \{\phi^{'} : \lambda(\phi, \phi^{'}) \leq \delta\}$ such that $E_{\phi_0}\left[\sup_{\phi^{'}\in V(\phi)} |g(Y_t|\bar{\mathbf{Y}}_{t-1}^{-\infty}; \phi^{'})|\right] < \infty$ for some $\delta > 0$.

Assumption (A4) implies that, among others, there exists at least one compact subset of Φ containing ϕ_0, $V(\phi_0)$, over which the expectation under ϕ_0 of $|g(Y_t|\bar{\mathbf{Y}}_{t-1}^{-\infty}; \phi)|$ is uniformly finite. Similar conditions as (A4) are also found in, e.g. Francq and Roussignol (1998) or Douc et al. (2004).

The last assumption concerns the identifiability of the parameters. Because the label of the Markov regimes can be switched without changing the law of the model, the parameters are not strictly identifiable up to permutation. Leroux (1992) henceforth defined an equivalence class which consists of parameters that induce the same law for the observations and stated the consistency result on the equivalence class of the MLE. However, we will use a more convenient and essentially equivalent assumption from Francq and Roussignol (1998).

(**A5**) For any ϕ_1 and $\phi_2 \in \Phi$, if for all $\bar{\mathbf{Y}}_t^{-\infty}$, $p(Y_t|\bar{\mathbf{Y}}_{t-1}^{-\infty};\phi_1) = p(Y_t|\bar{\mathbf{Y}}_{t-1}^{-\infty};\phi_2)$ $a.s. - P_{\phi_0}$, then $\phi_1 = \phi_2$.

3 The consistency of MLE

We prove the consistency of MLE in this section. Our methods are benefitted from the paper of Francq and Roussignol (1998). First, as we noted before, we approximate the likelihood L_n^* by L_n over a compact subset of Φ. In fact, (A2) and (A4) imply that the amount of information from $\bar{\mathbf{Y}}_0^{-\infty}$ will be equivalently revealed by $\bar{\mathbf{Y}}_n^1$ when n tends to ∞. Hence it is expected that the approximation holds for some more general likelihood functions as Lemma 1 shows.

Lemma 1 *Suppose that a process* $\{Y_t\}_{t\in\mathbb{Z}}$ *satisfies Assumptions (A2) and (A4). Let* L_n^**,* L_n*,* $g^*(Y_t|\bar{\mathbf{Y}}_{t-1}^1;\phi)$ *and* $g(Y_t|\bar{\mathbf{Y}}_{t-1}^{-\infty};\phi)$ *be defined as before except that the likelihoods* L_n^* *and* L_n *do not necessarily have the form of (2) and (4), respectively. Then, for all* ϕ *in a compact subset of* Φ*,*

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\log L_n^*(\bar{\mathbf{Y}}_n^1;\phi) &= \lim_{n\to\infty}\frac{1}{n}\log L_n(\bar{\mathbf{Y}}_n^1|\bar{\mathbf{Y}}_0^{-\infty};\phi)\\ &= E_{\phi_0}g(Y_t|\bar{\mathbf{Y}}_{t-1}^{-\infty};\phi)\ \text{a.s.} - P_{\phi_0}. \qquad (5)\end{aligned}$$

Proof. Since we confine us to a compact set, the expectation of $|g(Y_t|\bar{\mathbf{Y}}_{t-1}^{-\infty};\phi)|$ is uniformly finite by (A4). In addition, $g(Y_t|\bar{\mathbf{Y}}_{t-1}^{-\infty};\phi)$ is stationary and ergodic, which is carried over from the stationarity and ergodicity of $\{Y_t\}$ (see, e.g. Stout, 1974, Theorem 3.5.8). Then the second equality of (5) follows from applying the ergodic theorem, e.g. Stout (1974, Theorem 3.5.7), by observing that

$$\frac{1}{n}\log L_n(\bar{\mathbf{Y}}_n^1|\bar{\mathbf{Y}}_0^{-\infty};\phi) = \frac{1}{n}\sum_{t=1}^{n} g(Y_t|\bar{\mathbf{Y}}_{t-1}^{-\infty};\phi). \qquad (6)$$

Similarly,

$$\frac{1}{n}\log L_n^*(\bar{\mathbf{Y}}_n^1;\phi) = \frac{1}{n}\sum_{t=1}^{n} g^*(Y_t|\bar{\mathbf{Y}}_{t-1}^1;\phi). \qquad (7)$$

Now we need to show that (7) asymptotically equals to (6) a.s. as $n \to \infty$. Analogous to Karlin and Taylor (1975, p.502), define

$$Z_t^T = \sup_{l\geq T}|g^*(Y_t|\bar{\mathbf{Y}}_{t-1}^{t-l};\phi) - g(Y_t|\bar{\mathbf{Y}}_{t-1}^{-\infty};\phi)|.$$

Then $\{Z_t^T\}$ is stationary, ergodic from (A2) and $E_{\phi_0}|Z_t^T| < \infty$. We have

$$\begin{aligned}
&\limsup_{n\to\infty}\left|\frac{1}{n}\sum_{t=1}^{n}\{g^*(Y_t|\bar{\mathbf{Y}}_{t-1}^{1};\phi)-g(Y_t|\bar{\mathbf{Y}}_{t-1}^{-\infty};\phi)\}\right| \\
&\le \limsup_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\left|g^*(Y_t|\bar{\mathbf{Y}}_{t-1}^{1};\phi)-g(Y_t|\bar{\mathbf{Y}}_{t-1}^{-\infty};\phi)\right| \\
&\le \limsup_{n\to\infty}\frac{1}{n}\sum_{t=T+1}^{n} Z_t^T = E_{\phi_0}[Z_t^T].
\end{aligned}$$

But as $T\to\infty$, $Z_t^T \to 0$, and the interchange of limit and expectation can be justified to give $\lim_{T\to\infty} E_{\phi_0}[Z_t^T] = 0$. This completes the proof. □

Next, define

$$R_n(\phi) = \frac{1}{n}\log\frac{L_n^*(Y_n,...,Y_1;\phi)}{L_n^*(Y_n,...,Y_1;\phi_0)}$$

and we have the following lemma.

Lemma 2 *Assume (A1)-(A2) and (A4)-(A5). For any ϕ in a compact subset of Φ,*

$$\lim_{n\to\infty} R_n(\phi) = \lim_{n\to\infty}\frac{1}{n}\log\frac{L_n(Y_n,...,Y_1|\bar{\mathbf{Y}}_0^{-\infty};\phi)}{L_n(Y_n,...,Y_1|\bar{\mathbf{Y}}_0^{-\infty};\phi_0)} \le 0, \quad a.s. - P_{\phi_0}.$$

The limit equals to zero a.s. if and only if $\phi = \phi_0$.

Proof. The first equality is from Lemma 1. By Lemma 1 and Jensen's inequality,

$$\begin{aligned}
\lim_{n\to\infty}\frac{1}{n}\log\frac{L_n(Y_n,...,Y_1|\bar{\mathbf{Y}}_0^{-\infty};\phi)}{L_n(Y_n,...,Y_1|\bar{\mathbf{Y}}_0^{-\infty};\phi_0)} &= E_{\phi_0}\log\frac{p(Y_t|\bar{\mathbf{Y}}_{t-1}^{-\infty};\phi)}{p(Y_t|\bar{\mathbf{Y}}_{t-1}^{-\infty};\phi_0)} \\
&\le \log E_{\phi_0}\frac{p(Y_t|\bar{\mathbf{Y}}_{t-1}^{-\infty};\phi)}{p(Y_t|\bar{\mathbf{Y}}_{t-1}^{-\infty};\phi_0)} = 0.
\end{aligned}$$

From (A5), the limit equals to zero if and only if $\phi = \phi_0$. □

Lemma 3 *Assume (A1)-(A5). For any ϕ_1 in a compact subset of Φ and $\phi_1 \ne \phi_0$, there exists a neighborhood $V(\phi_1)$ of ϕ_1 such that*

$$\limsup_{n\to\infty}\sup_{\phi\in V(\phi_1)} R_n(\phi) < 0, \quad a.s. - P_{\phi_0}.$$

Proof. In view of Lemma 2, we will show

$$\limsup_{n\to\infty} \sup_{\phi\in V_r(\phi_1)} \frac{1}{n} \log \frac{L_n(Y_n, ..., Y_1|\bar{\mathbf{Y}}_0^{-\infty}; \phi)}{L_n(Y_n, ..., Y_1|\bar{\mathbf{Y}}_0^{-\infty}; \phi_0)} < 0, \quad a.s. - P_{\phi_0},$$

where $V_r(\phi_1) = \{\phi : \lambda(\phi, \phi_1) \le 1/r\}$, and r is large enough such that $V_r(\phi_1)$ is contained in the compact subset of Φ and Lemma 2 holds for all $\theta \in V_r(\phi_1)$. It is equivalent to show

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n} \log \sup_{\phi\in V_r(\phi_1)} L_n(Y_n, ..., Y_1|\bar{\mathbf{Y}}_0^{-\infty}; \phi) \\ < \lim_{n\to\infty} \frac{1}{n} \log L_n(Y_n, ..., Y_1|\bar{\mathbf{Y}}_0^{-\infty}; \phi_0), \quad a.s. - P_{\phi_0}\end{aligned} \tag{8}$$

provided that the limit in the left hand side of (8) exists.

Define the matrix norm $\| \, . \, \|$ as the sum of all element of the matrix. From equation (4), it follows that

$$\begin{aligned}\min_k \pi(k) q(Y_1|\bar{\mathbf{Y}}_0^{-\infty}; \theta_k(\phi)) \left\| \prod_{t=2}^{n} \mathbf{M}'_t(Y_t; \phi)\mathbf{A}(\phi) \right\| \le L_n(Y_n, ..., Y_1|\bar{\mathbf{Y}}_0^{-\infty}; \phi) \\ \le \max_k \pi(k) q(Y_1|\bar{\mathbf{Y}}_0^{-\infty}; \theta_k(\phi)) \left\| \prod_{t=2}^{n} \mathbf{M}'_t(Y_t; \phi)\mathbf{A}(\phi) \right\|.\end{aligned}$$

From (A1) and (A4), we obtain that

$$\lim_{n\to\infty} \frac{1}{n} \log L_n(\bar{\mathbf{Y}}_n^1|\bar{\mathbf{Y}}_0^{-\infty}; \phi) = \lim_{n\to\infty} \frac{1}{n} \log \left\| \prod_{t=2}^{n} \mathbf{M}'_t(Y_t; \phi)\mathbf{A}(\phi) \right\|, \; a.s. - P_{\phi_0}. \tag{9}$$

Define $S^r_{2,n}(Y_n) = \sup_{\phi\in V_r(\phi_1)} \|\prod_{t=2}^{n} \mathbf{M}'_t(Y_t; \phi)\mathbf{A}(\phi)\|$. Because this matrix norm is multiplicative, it follows that, over $V_r(\phi_1)$,

$$\sup_\phi \left\| \prod_{t=2}^{n+k'} \mathbf{M}'_t(\cdot)\mathbf{A}(\phi) \right\| \le \sup_\phi \left\| \prod_{t=2}^{n} \mathbf{M}'_t(\cdot)\mathbf{A}(\phi) \right\| \cdot \sup_\phi \left\| \prod_{t=n+1}^{n+k'} \mathbf{M}'_t(\cdot)\mathbf{A}(\phi) \right\|,$$

where $\mathbf{M}'_t(\cdot)$ is $\mathbf{M}'_t(Y_t; \phi)$. This is just

$$\log S^r_{2,n+k'}(Y_{n+k'}) \le \log S^r_{2,n}(Y_n) + \log S^r_{n+1,n+k'}(Y_{n+k'})$$

for any positive integers $n(\ge 2)$, k' and r. The process $\{\log S^r_{2,n}(Y_n)\}_{n\ge2}$ is hence subadditive. It is stationary and ergodic following from Assumption

(A2) by, e.g. Stout (1974, Theorem 3.5.8). In addition, $E_{\phi_0} \log S^r_{2,n}(Y_n)$ is finite from (A4). Therefore, the ergodic theorem for the subadditive processes (Kingman, 1973, p.855) or an improved version (e.g. Liggett, 1985) can be applied to declare that $\lim_{n\to\infty} \frac{1}{n} \log S^r_{2,n}(Y_n)$ exists and equals to

$$\gamma_r(\phi_1) = \inf_{n>1} \frac{1}{n} E_{\phi_0} \log S^r_{2,n}(Y_n), \quad a.s. - P_{\phi_0}. \tag{10}$$

At the same time, for the random matrix sequence $\{\mathbf{M}'_t(Y_t;\phi)\mathbf{A}(\phi)\}, t = 2, 3, ...,$ the top Lyanunov exponent can be defined as

$$\begin{aligned} \gamma(\phi) &= \lim_{n\to\infty} \frac{1}{n} E_{\phi_0} \log \left\| \prod_{t=2}^{n} \mathbf{M}'_t(Y_t;\phi)\mathbf{A}(\phi) \right\| \\ &= \inf_{n>1} \frac{1}{n} E_{\phi_0} \log \left\| \prod_{t=2}^{n} \mathbf{M}'_t(Y_t;\phi)\mathbf{A}(\phi) \right\|, \end{aligned} \tag{11}$$

since we have $E_{\phi_0} \log^+ ||\mathbf{M}'_t(Y_t;\phi)\mathbf{A}(\phi)|| < \infty$ by (A4) ($\log^+ x = \max(\log x, 0)$). From Furstenberg and Kesten (1960), we have

$$\gamma(\phi) = \lim_{n\to\infty} \frac{1}{n} \log \left\| \prod_{t=2}^{n} \mathbf{M}'_t(Y_t;\phi)\mathbf{A}(\phi) \right\|, \quad a.s. - P_{\phi_0}. \tag{12}$$

Thus, by equations (9) and (12), Lemma 2 yields $\gamma(\phi) < \gamma(\phi_0)$ for all $\phi \in V_r(\phi_1)$. Particularly, for ϕ_1, there exist $\epsilon > 0$ and $n_\epsilon \in \mathbb{N}$ such that

$$\frac{1}{n_\epsilon} E_{\phi_0} \log \left\| \prod_{t=2}^{n_\epsilon} \mathbf{M}'_t(Y_t;\phi_1)\mathbf{A}(\phi_1) \right\| < \gamma(\phi_0) - \epsilon. \tag{13}$$

Note that by (A4) and the dominated convergence theorem, it follows that

$$\lim_{r\to\infty} \frac{1}{n_\epsilon} E_{\phi_0} \log S^r_{2,n_\epsilon}(Y_{n_\epsilon}) = \frac{1}{n_\epsilon} E_{\phi_0} \log \left\| \prod_{t=2}^{n_\epsilon} \mathbf{M}'_t(Y_t;\phi_1)\mathbf{A}(\phi_1) \right\|. \tag{14}$$

From (10), it follows that $\gamma_r(\phi_1) \le \frac{1}{n_\epsilon} E_{\phi_0} \log S^r_{2,n_\epsilon}(Y_{n_\epsilon})$ for all r. For r large enough, (13) and (14) together with assumption (A3) give us

$$\gamma_r(\phi_1) \le \frac{1}{n_\epsilon} E_{\phi_0} \log \left\| \prod_{t=2}^{n_\epsilon} \mathbf{M}'_t(Y_t;\phi_1)\mathbf{A}(\phi_1) \right\| + \frac{\epsilon}{2} < \gamma(\phi_0) - \frac{\epsilon}{2},$$

which implies (8). □

Lemma 3 implies the strong consistency of the MLE for the GARMS model (1) over any compact subset of Φ containing ϕ_0, denoting it as Φ^*. From (A4), we can at least choose this subset as $V(\phi_0)$ and restricts the results of lemmas to $V(\phi_0)$.

Theorem 1 *For the GARMS model (1), assume (A1)-(A5). Let $\hat{\phi}_n$ be an MLE sequence over Φ^*, satisfying*

$$L_n^*(Y_n, ..., Y_1; \hat{\phi}_n) = \sup_{\phi \in \Phi^*} L_n^*(Y_n, ..., Y_1; \phi), \qquad \text{a.s.} - P_{\phi_0}$$

then $\hat{\phi}_n$ tends to ϕ_0 a.s. as $n \to \infty$.

Proof. Suppose that $\hat{\phi}_n$ didn't tend to ϕ_0 with probability one as $n \to \infty$, i.e., for arbitrarily large integer N, there exist a $\delta > 0$ and at least one $n^*, n^* \geq N$ such that $\lambda(\hat{\phi}_{n^*}, \phi_0) \geq \delta$ with a positive probability. By Lemma 3, it follows that $L_{n^*}^*(Y_{n^*}, ..., Y_1; \hat{\phi}_{n^*})$ is strictly less than $L_{n^*}^*(Y_{n^*}, ..., Y_1; \phi_0)$ with a positive probability. However, by the definition of MLE, with probability one, we have

$$L_{n^*}^*(Y_{n^*}, ..., Y_1; \hat{\phi}_{n^*}) = \sup_{\phi \in \Phi^*} L_{n^*}^*(Y_{n^*}, ..., Y_1; \phi) \geq L_{n^*}^*(Y_{n^*}, ..., Y_1; \phi_0).$$

This contradiction gives our result. □

4 Examples and simulation study

An interesting example of the GARMS is the AR model with Markov switching. It is denoted by MS-AR to distinguish from the ARMS model. Some examples of finite order MS-AR model from Krishnamurthy and Rydén (1998) and Francq and Zakoïan (2001) are re-examined, particularly focusing on the stationarity of the processes and consistency of the MLEs. Such examples are of interest since they showed apparently opposite conclusions between Krishnamurthy and Rydén (1998) and Francq and Zakoïan (2001). In addition, an infinite order MS-AR model is discussed, and the conditions for the consistency, particularly the stationarity condition, are investigated. The simulation study with this model is presented, which confirms the consistency of the MLE and suggests the asymptotic normality.

4.1 Finite order MS-AR models

A finite order MS-AR model is a particular example of GARMS with linear function f, defined as

$$\sum_{i=0}^{p} \mu_i(X_t)Y_{t-i} = c(X_t) + \sigma(X_t)\eta_t, \tag{15}$$

where (X_t) is the underlying Markov chain with d regimes and transition probabilities $(\alpha_{kl}), k, l \in \mathbf{E}$, $\mu_0(k) = 1$ for all values of k, and μ's, c, and σ are regime dependent parameters. η_t is an i.i.d. innovation and independent of (X_t). This model has been studied by Krishnamurthy and Rydén (1998) amongst others.

Krishnamurthy and Rydén (1998) reported a number of simulation studies for (15) and illustrated the consistency of the MLE under different settings. One of their examples (Krishnamurthy and Rydén, 1998, Example 5) has the following setting:

(i) $p = 2$, $d = 2$ in (15) with true parameters $\mu_1(1) = -1$, $\mu_2(1) = -0.8$, $\sigma(1) = 1$, $\mu_1(2) = 2$, $\mu_2(2) = -0.5$, $\sigma(2) = 0.25$, $\alpha_{12} = \alpha_{21} = 0.99$, and $c(1) = c(2) = 0$.

Note that in (i), although the processes $Y_t - Y_{t-1} - 0.8Y_{t-2} = \eta_t$ and $Y_t + 2Y_{t-1} - 0.5Y_{t-2} = 0.25\eta_t$ are non-stationary, the resulting process is strictly stationary and has finite second moment, because of the rapid regime switching, as stated by Krishnamurthy and Rydén (1998). However, besides the rapid regime switching, the different signs of $\mu_1(1)$ and $\mu_1(2)$ play an important role in the stationarity of this process. If we change $\mu_1(1)$ to 1, the resulting process will not be stationary any more (see the condition (19)) in this case. For model (i), the consistency of the MLE can be achieved and verified by our experiments.

By contrast, Francq and Zakoïan (2001) reported a numerical result for model:

(ii) $p = 1$, $d = 2$ in (15) with true parameters $\mu_1(1) = 0$, $\sigma(1) = 1$, $\mu_1(2) = -1.11$, $\sigma(2) = 4$, $\alpha_{12} = 0.9$, $\alpha_{21} = 0.2$, $c(1) = 0$, and $c(2) = 1$.

Although the parameters in (ii) satisfy the stationarity condition (Francq and Zakoïan, 2001, Theorem 2), the estimates they obtained (Francq and Zakoïan, 2001, Table 1) are rather poor. In other experiments, they required the parameter spaces were within the (weakly) stationary region, but produced no better estimates. It questions the result for consistency. However, the sample size that Francq and Zakoïan (2001) used, 50, is surprisingly small, compared

Table 1: The true value, mean and standard deviation of the MLEs of model (ii) in Section 4.1, obtained from 30 replications, each of size 1000.

	$\mu_1(1)$	$\mu_1(2)$	$\sigma(1)$	$\sigma(2)$	α_{12}	α_{21}	$c(1)$	$c(2)$
True value	0	-1.11	1	4	0.9	0.2	0	1
Mean	0	-1.109	1.013	3.98	0.902	0.197	-0.014	1.002
SD	0.004	0.009	0.064	0.092	0.06	0.016	0.087	0.16

with the 5000 used in (i) by Krishnamurthy and Rydén (1998). Moreover, Francq and Zakoïan (2001) randomly selected the start values during the maximization of the likelihood, which may frequently result in local maxima of the likelihood function. This was pointed out by Krishnamurthy and Rydén (1998), who used the true value as the start values in model (i). Therefore, in the first experiment, we re-examine the model (ii), but increase the sample size to 1000 and start the maximization of the likelihood function with the true parameter values. The stationarity conditions are not imposed on the parameters during the maximization. The true values, mean and standard deviation (SD) of the MLE for model (ii) are reported in Table 1. Only 30 replications are used in our study, and the biases and standard deviations are significantly reduced compared with the result of Francq and Zakoïan (2001, Table 1). Note that Francq and Zakoïan (2001) made use of 1000 replications. It is clear that the result for consistency is satisfactory in our study and the apparent 'counter-example' of Francq and Zakoïan (2001) is only due to the inappropriate numerical settings.

4.2 The infinite order MS-AR model

An infinite order MS-AR model can be defined in the mean-square sense by

$$\sum_{i=0}^{\infty} \mu_i(X_t)Y_{t-i} = c(X_t) + \sigma(X_t)\eta_t. \tag{16}$$

It is worth noting that, in a finite order MS-AR model (15), $\{(X_t, Y_t, ..., Y_{t-p+1})\}$ is an Markov process, which plays an important role in the proofs of Krishnamurthy and Rydén (1998) and Douc et al. (2004), while there is no analogue in (16). We confine us to a finite-dimensional parameter space. One way is to assume $\mu_i(k) = 0$ for i larger than some positive integer p and all values of k, which just reduces (16) to the finite order MS-AR model (15). Alternatively, we may also assume that μ_i are functions of some (probably vectorial) parameter ψ. For example, consider the following model

$$Y_t = \sum_{i=1}^{\infty} \psi^i(X_t) Y_{t-i} + \sigma(X_t)\eta_t, \tag{17}$$

where (X_t) is the Markov chain with d regimes. Note that ψ^i is ψ raised to power i, and the absolute values of $\psi(k), k \in \mathbf{E}$, are assumed to be strictly less than 1. A (regime dependent) constant term can also be added to (17).

The conditions associated with Assumptions (A1-A5) will be illustrated under (17) with a frequently used two-regime Markov chain, i.e. $d = 2$. All positive transition probabilities ensure (A1). Assume that the innovation η_t is a standard normal, the continuity assumption (A3) follows. Since assumption (A4) is a moment requirement, it will be satisfied with a standard normal η_t and a stationary solution of (17). For the identifiability assumption (A5), it suffices to label the regimes such that $\sigma(1) \leq \sigma(2)$.

The condition for stationarity of model (17) is not obvious. Francq and Zakoïan (2001, Theorems 1 and 2) obtained the sufficient conditions for strict and second order stationarity, respectively, for an ARMA model with Markov switching (MS-ARMA) defined as

$$\sum_{i=0}^{p} \mu_i(X_t) Y_{t-i} = \sum_{j=0}^{q} \nu_j(X_t)\sigma(X_{t-j})\eta_{t-j}, \tag{18}$$

where $\mu_0(k) = 1$, and $\nu_0(k) = 1, k \in \mathbf{E}$. As in the case of ordinary ARMA, these conditions concern only the parameters in the autoregressive part, as well as the transition probabilities. For the first order MS-AR model, their strict stationarity condition (using notations from (17)) is

$$\sum_{k=1}^{d} \pi(k) \log |\psi(k)| < 0,$$

which is fulfilled with the condition $|\psi(k)| < 1, k \in \mathbf{E}$. However, the strict stationarity condition in the general case requires a strictly negative top Lyapunov exponent of a sequence of designated matrices, which is difficult to verify in practice except by simulation.

For the second order stationarity condition, assume that $d = 2$ and the autoregression in (17) were of p order. Let $\tau(k) = (\psi(k), \cdots, (\psi(k))^{p-1})$, and the $p \times p$ matrix

$$\mathbf{D}(k) = \begin{pmatrix} \tau(k) & (\psi(k))^p \\ \boldsymbol{I}_{p-1} & 0 \end{pmatrix},$$

where $\boldsymbol{I}_{p-1}$ is the identity matrix of size $p-1$. Define then the $2p^2 \times 2p^2$ matrix

$$\tilde{\mathbf{p}} = \begin{pmatrix} \alpha_{11}(\mathbf{D}(1) \otimes \mathbf{D}(1)) & \alpha_{21}(\mathbf{D}(1) \otimes \mathbf{D}(1)) \\ \alpha_{12}(\mathbf{D}(2) \otimes \mathbf{D}(2)) & \alpha_{22}(\mathbf{D}(2) \otimes \mathbf{D}(2)) \end{pmatrix},$$

where $\otimes$ denotes the Kronecker product, and α's are the transition probabilities of the Markov chain. The condition for second order stationarity is that the spectral radius (or maximum eigenvalue) $\rho(\cdot)$ of $\tilde{\mathbf{p}}$ is less than 1, i.e.

$$\rho(\tilde{\mathbf{p}}) < 1. \tag{19}$$

When $p = 1$, (19) reduces to the stationarity condition for a first order MS-AR model. The condition consists of

$$\begin{cases} \alpha_{11}\alpha_{22}\psi^2(1)\psi^2(2) - \alpha_{12}\alpha_{21}\psi^2(1)\psi^2(2) < 1, \\ \alpha_{11}\alpha_{22}\psi^2(1)\psi^2(2) - (\alpha_{11}\psi^2(1) + \alpha_{22}\psi^2(2)) - \alpha_{12}\alpha_{21}\psi^2(1)\psi^2(2) > -1, \\ \alpha_{12}\alpha_{21}\psi^2(1)\psi^2(2) - \alpha_{11}\alpha_{22}\psi^2(1)\psi^2(2) - (\alpha_{11}\psi^2(1) + \alpha_{22}\psi^2(2)) < 1. \end{cases}$$

It turns out that the assumption $|\psi(k)| < 1, k = 1, 2$, suffices for the above three inequalities.

For $p \geq 2$, the size of matrix $\tilde{\mathbf{p}}$ is at least eight and its eigenvalues usually have to be calculated numerically. Recall that an upper bound of the eigenvalues of a matrix is the maximum value of the sums of absolute values in each row and column, and the coefficients in (17) decrease exponentially along the autoregression order. Our numerical experiments show that a process of (17) with $|\psi(k)| < 1/2, k = 1, 2$, will not become explosive no matter what (positive) values the transition probabilities take. It may be worth noting that if X_t has only one regime in (17), the stationarity condition can be shown to be $|\psi| < 1/2$.

Simulation studies using (17) are conducted. In this experiment, sequences that have size 100, 250, 500, and 1000, respectively, are generated from (17). 150 replications for each sample size are generated. Every observation is depending on all previous observations in the same sequence through (17). The true values are used as the start values in the maximization of the likelihood (3), which is necessary for the maximization for such a complicated likelihood with multiple parameters. In real applications, we may maximize the likelihood several times using randomly chosen parameters (with $|\psi(k)| < 1/2$ and $0 < \alpha_{kl} < 1$, $k, l = 1, 2$), and select the parameters that maximize the likelihood to conduct the final estimation.

The true values, and the mean and standard deviation of the MLEs are summarized in Table 2. The regime 1 (i.e. $X_t = 1$) can be used to indicate a more volatile state in practice, with rapid change between each

Table 2: The true value, mean and standard deviation of the MLEs of parameters for model (17), obtained from 150 replications and sequences of size 100, 250, 500, and 1000, respectively.

	$\psi(1)$	$\psi(2)$	$\sigma(1)$	$\sigma(2)$	α_{12}	α_{21}
True value	-0.4	0.3	1	0.5	0.7	0.2
mean: $n = 100$	-0.356	0.299	0.865	0.467	0.715	0.322
$n = 250$	-0.366	0.305	0.946	0.483	0.690	0.270
$n = 500$	-0.394	0.300	0.979	0.491	0.675	0.214
$n = 1000$	-0.394	0.299	0.988	0.497	0.699	0.213
SD: $n = 100$	0.306	0.088	0.264	0.106	0.277	0.247
$n = 250$	0.224	0.068	0.166	0.056	0.227	0.181
$n = 500$	0.172	0.043	0.109	0.038	0.160	0.099
$n = 1000$	0.107	0.026	0.064	0.024	0.099	0.055

time period, stronger persistence, and shorter average duration as implied by $1/(1 - \alpha_{11}) = 10/7$ units of time (see, e.g. Kim and Nelson, 1999); while another one indicates a more normal state. As can be seen from Table 2, the biases of the estimates usually decrease, and the SDs always decrease as the sample size increases, which confirms the consistency of the MLE for the infinite order MS-AR model (17). The simulation study gives also valuable information on the number of observations needed in order to obtain a reasonable estimate. It can be argued that less than 250 observations, corresponding to the daily data of a (business) year, are not enough, two years' data produce estimates that have only marginal biases (at most 7% of the true value in our experiment). The SDs of the estimates are further decreased if four years' data are available. It should be mentioned that the SDs are related to the transition probabilities in the finite sample experiments. The SDs for estimates under regime 1 in our experiments are all larger than the correspondences under regime 2. This can be explained to some extent by the fact that $P(X_t = 1)$ is much less than $P(X_t = 2)$, hence less observations are available in regime 1 than in regime 2.

In addition, the quantile-quantile (QQ) plot of the MLEs is reported in Figure 1. It suggests that the MLEs for this model should be asymptotically normal, by observing that the QQ points usually fall into the zero-one line, the line through the 25% and 75% sample quantiles and the corresponding normal quantiles. It is confirmed by the Kolmogorov-Smirnov normality test. The null hypothesis that the estimate is normal cannot be rejected at at least 95% level. Experiments using other parameters (with $|\psi| < 1/2$), or including

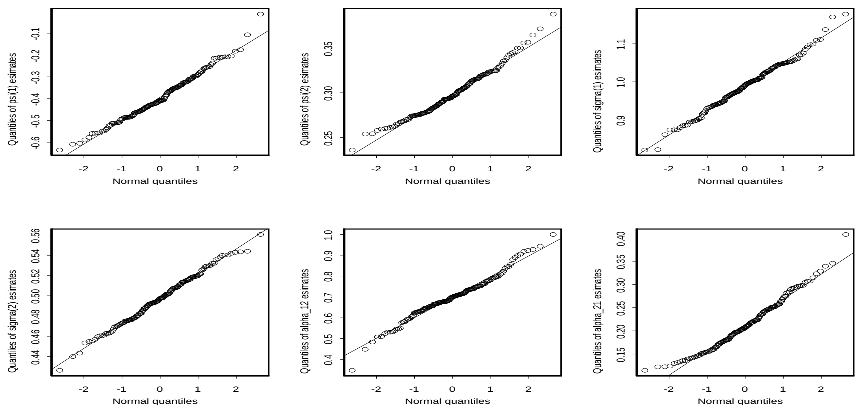

Figure 1: The QQ plot for the MLEs of an infinite order RS-AR model (17), with zero-one line.

intercepts in (17), have been carried out. The results (not reported) also support the consistency and conjecture of asymptotic normality of the MLEs.

5 Conclusions

In this paper, a GARMS model that has the form (1) is considered. The order of autoregression in this model can be infinite. Not only that the popular HMMs are special cases of (1), our model also generalizes the ARMS models studied by Francq and Roussignol (1998), Krishnamerthy and Rydén (1998), and Douc et al. (2004) among others. The MLE of model (1) is studied and proven to be consistent. Our methods are benefitted from Francq and Roussignol (1998). The assumptions are rather common, with the origin to Wald (1949) and followed by Leroux (1992), Francq and Roussignol (1998), Krishnamerthy and Rydén (1998), and Douc et al. (2004). It is perhaps attractive to consider an infinite-dimensional parameter space. However, in an infinite-dimensional space, it is difficult even to define a distance that reserves the compactness of the parameter space and is simultaneously feasible for the proof of consistency. The results for infinite-dimensional parameters are still unknown.

The finite order and infinite order MS-AR models are taken as examples. The conditions under which the general assumptions (A1-A5) are possibly satisfied are discussed for an infinite order MS-AR model. The simulation studies demonstrate the assumptions and the consistency of the MLEs. An apparently 'counter-example' for the consistency of finite order MS-AR model from Francq and Zakoïan (2001) is shown to be a result of inappropriate experimental setups. The numerical experiments using the infinite order MS-AR model also suggest that the MLEs for this model are asymptotically normal.

Douc et al. (2004) proved the asymptotic normality of the MLE when the s is finite in (1). It is the first asymptotic normality result, to the authors' knowledge, for a Markov switching model including an autoregression. Other results, including those of Bickel et al. (1998) and Jensen and Petersen (1999) were all dedicated to the HMMs. However, Douc et al. (2004) made use of the Markov property of process $\{(X_t, \bar{\mathbf{Y}}_t^s)\}$ (s finite), which does not hold any more for the GARMS model (1) with $s = -\infty$. An approach that will be feasible in this case still needs to be investigated.

References

[1] Baum, L.E. and Petrie, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains. *Ann. Math. Statist.* **37**, 1554-1563.

[2] Bickel, P.J. and Ritov, V. (1996). Inference in hidden Markov models I: LAN in the stationary case. *Bernoulli* **2**, 199-228.

[3] Bickel, P.J., Ritov, Y. and Rydén, T. (1998). Asymptotic normality of the maximum likelihood estimator for general hidden Markov models. *Ann. Statist.* **26**, 1614-1635.

[4] Cai, J. (1994). A Markov model of switching-regime ARCH. *J. Busi. Econ. Statist.* **12**, 309-316.

[5] Douc, R., Moulines, É. and Rydén, T. (2004). Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regimes. *Ann. Statist.* **32**, 2254-2304.

[6] Doucet, A., Logothetis, A. and Krishnamurthy, V. (2000). Stochastic sampling algorithms for state estimation of jump Markov linear systems. *IEEE Trans. Automa. Control* **45**, 188-202.

[7] Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of U.K. inflation. *Econometrica* **50**, 987-1008.

[8] Francq, C. and Roussignol, M. (1998). Ergodicity of autoregressive processes with Markov-switching and consistency of the maximum-likelihood estimator. *Statistics* **32**, 151-173.

[9] Francq, C., Roussignol, M. and Zakoïan, J. (2001). Conditional heteroscedasticity driven by hidden Markov chains. *J. Time Ser. Anal.* **22**, 197-220.

[10] Francq, C. and Zakoïan, J. (2001). Stationarity of multivariate Markov-switching ARMA models. *J. Econometrics* **102**, 339-364.

[11] Furstenberg, H. and Kesten, H. (1960). Products of random matrices. *Ann. Math. Statist.* **31**, 457-469.

[12] Hamilton, J.D. (1989). A new approach to the economic analysis of non-stationary time series and the business cycle. *Econometrica* **57**, 357-384.

[13] Jensen, J.L. and Petersen, N.V. (1999). Asymptotic normality of the maximum likelihood estimator in state space models. *Ann. Statist.* **27**, 514-535.

[14] Karlin, S. and Taylor, H.M. (1975). *A first Course in Stochastic Processes.* Academic Press: San Diego.

[15] Kim, C.-J. and Nelson, C.R. (1999). *State-Space Models with Regime-Switching: Classical and Gibbs-Sampling Approaches with Applications.* MIT Press: Cambridge, MA.

[16] Kingman, J.F.C. (1973). Subadditive ergodic theory. *Ann. Probab.* **1**, 883-909.

[17] Krishnamurthy, V. and Rydén, T. (1998). Consistent estimation of linear and non-linear auroregressive models with Markov regime. *J. Time Ser. Anal.* **19**, 291-307.

[18] Leroux, B.G. (1992). Maximum-likelihood estimation for hidden Markov models. *Stoch. Proc. Appl.* **40**, 127-143.

[19] Liggett, T.M. (1985). An improve subadditive ergodic theorem. *Ann. Probab.* **13**, 1279-1285.

[20] MacDonald, I. and Zucchini, W. (1997). *Hidden Markov and Other Models for Discrete-valued Time Series.* Chapman& Hall: London.

[21] McCulloch, R.E. and Tsay, R.S. (1994). Statistical analysis of econometric time series via Markov switching models. *J. Time Ser. Anal.* **15**(5), 523-539.

[22] Petrie, T. (1969). Probabilistic functions of finite state Markov chains. *Ann. Math. Statist.* **40**, 97-115.

[23] Stout, W.F. (1974). *Almost Sure Convergence.* Academic Press: New York.

[24] Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.* **20**, 595-601.

[25] Yao, J. (2001). On square-integrability of an AR process with Markov switching. *Statist. Probab. Lett.* **52**, 265-270.

[26] Yao, J. and Attali, J.-G. (2000). On stability of nonlinear AR processes with Markov switching. *Adv. Appl. Prob.* **32**, 394-407.