Geometric Ergodicity of the Mixture Autoregressive Model

Mary I. Akinyemi

School of Mathematics, University of Manchester

SIAM Manchester Student Chapter Conference 2013

Outline

Introduction

- Mixture autoregressive models provide a flexible framework for modelling time series.
- They capture conditional heterogeneity, multi-modality, skewness, kurtosis and heavy tails using only standard distributions as building blocks.
- Geometric ergodicity is very useful in establishing mixing conditions and central limit results for parameter estimates of a model, it also justifies the use of laws of large numbers and form a basis for exploring the asymptotic theory of the model.

The Mixture Autoregressive model.

Definition and assumptions

Definition

A process $\{y_t\}$ is said to be a mixture autoregressive processes if the conditional distribution function of y_t given past information is given by,

$$F_{t|t-1}(x) = \sum_{k=1}^{g} \pi_k F_k \left(\frac{x - \phi_{k,0} - \sum_{i=1}^{p_k} \phi_{k,i} y_{t-i}}{\sigma_k} \right) \tag{1.1}$$

where

- g is a positive integer representing the number of components in the model;
- $\pi_k > 0$, $k = 1, \ldots, g$, $\sum_{k=1}^{g} \pi_k = 1$ are probabilities and they define a discrete distribution π; $\pi_k > 0$ is also referred to as mixing proportions and can be either be time invariant or functions of observed variables (e.g. lagged observations).
- $\sigma_k > 0$ is a scaling factor for the kth noise component.
- $F_k(\cdot)$ is a (conditional) cumulative distribution function, hence we will denote $f_k(\cdot)$ as the corresponding conditional probability density function for each $k = 1, \ldots, g$
- p_k is the order of the kth autoregressive model, set $p = \max_{1 \leq k \leq g} p_k$
- $\phi_{k,0}$ and $\phi_{k,i}$ $i = 1, \ldots, p_k$ are autoregressive coefficients, also, $\phi_{k,i} = 0$ for $i > p_k$.

RCA representation of the MAR model

Define a vector Z_t such that:

$$Z_{t,k} = \begin{cases} 1 & \text{if } z_t = k \\ 0 & \text{otherwise} \end{cases}$$

Let $\{z_t\}$ be an i.i.d sequence of random variables with distribution π, such that $Pr\{z_t = k\} = \pi_k$, $k = 1, \ldots, g$, y_t can be written as (Boshnakov [2009]),

$$y_t = f_{z_t}(y^{'}) + \sigma_{z_t} \varepsilon_{z_t}(t) \qquad (1.2)$$

where

$$f_{z_t}(y') = \phi_{z_t,0} + \sum_{i=1}^{p} (\phi_{z_t,i} y_{t-i}) \qquad (1.3)$$

Z_t is a simple case of a hidden Markov chain with values in $\{1, \ldots, g\}$ which drives the dynamics of $Y_t = (y_t, \ldots, y_{t-p+1})'$.

- Let $A = a_{ij} \geq 0,\ i = 1, \ldots, g,\ j = 1, \ldots, g-1$ and $\sum_j a_{ij} = 1$ be the corresponding transition probability matrix. Also, let $\theta \in \Theta$ be the combination of all the free parameters of the model, Θ is a compact subset of $\mathbb{R}$.

The conditional densities

- Define the conditional density of y_t given only the past values of y_t as,

$$f_\theta(y \mid y^{'}) = \sum_{k=1}^{g} \frac{\pi_k}{\sigma_k} f_k \left(\frac{y - \phi_{k,0} - \sum_{i=1}^{p_k} \phi_{k,i} y(t-i)}{\sigma_k} \right), \quad (1.4)$$

- and the conditional density of y_t given both the past values of y_t and the chain z_t as,

$$f_\theta(y \mid y^{'}, z_t) = \frac{\pi_{z_t}}{\sigma_{z_t}} f_{z_t} \left(\frac{y - \phi_{z_t,0} - \sum_{i=1}^{p_{z_t}} \phi_{z_t,i} y(t-i)}{\sigma_{z_t}} \right), \quad (1.5)$$

Finally, write

$$Y_t = (y_t, \ldots, y_{t-p+1})' \text{ and } \quad Q_t = (Z_t, Y_t) \tag{1.6}$$

where Q_t is an aperiodic $S \times \mathbb{R}^p$-valued Markov chain on some state space S.

Assumptions A

- The true parameter value θ^* lies in the interior of Θ.
- $\{Z_{t,k} : t \geq 0\}$ with $k = \{1, \ldots, g\}$ is an irreducible, aperiodic Markov chain on a finite space S with probability distribution $\pi_1, \ldots, \pi_g$ and transition probability matrix $A = (a_{ij}) \in S$. So that $Z_{t,k}$ inherits the properties of $\{Z_t\}$.
- The chain (Z_1) is independent of the noise term ε_t and for $\mathscr{F}_{t-1} = \sigma\{Y_r, r \leq t-1\}$,

$$P(z_t = j \mid z_{t-1} = i, \mathscr{F}_{t-1}) = P(z_t = j \mid z_{t-1} = i) \, \forall i, j. \qquad (2.1)$$

- $\{\varepsilon_t\}$ are jointly independent and are independent of past ys.

Assumptions A (cont.)

- $\{\varepsilon_t\}$ has a probability density function that is continuous and positive everywhere.
- $\sigma_{z_t} > 0$ is a scaling for the kth noise component.
- $f_{z_t}(y)$ is non periodic and bounded on compacts sets for all k and $z_t \in S$.

Useful Definitions

- Consider a state space S and a σ-field $\mathscr{F}$. Let (Y_t) be a homogenous Markov chain evolving on S, i.e. for all set $A \in \mathscr{F}$ and all $s, t \in \mathbb{N}$.
- The transition probability $P^t(y, A)$ is defined as,

$$P^t(y, A) := \mathbb{P}(Y_{s+t} \in A \mid Y_r, r < s; Y_s = y). \qquad (0.2)$$

- The *Markov Property* implies that $P^t(y, A)$ does not depend on $Y_r, r < s$, given Y_s. *Time homogeneity* refers to the fact that the transition probability does not depend on s.

A Markov chain is called **ergodic** if it is irreducible, aperiodic and positive Harris recurrent. That is, there exists a probability measure π on $S, \mathscr{F}$, such that,

$$\lim_{t \to +\infty} \|p^t(y, \cdot) - \pi(\cdot)\| = 0, \, y \in S, \tag{0.3}$$

where $\|\cdot\|$ here is the total variation norm (see Meyn and Tweedie [1993]).

The chain (Y_t) is called *geometrically ergodic* if there exists a positive constant $\rho < 1$ such that,

$$\lim_{t \to +\infty} \rho^{-t} \| p^t(y, \cdot) - \pi(\cdot) \| = 0, \forall y \in S \tag{0.4}$$

Mixing coefficients/conditions

(Davidson [1997]) consider a sequence $\{Y_t(\omega)\}_{-\infty}^{\infty}$, let $\mathscr{F}_a^b = \sigma(Y_t, a \leq t \leq b)$, $\mathscr{L}^2(\mathscr{F}_a^b)$ be a set of $\mathscr{F}_a^b-$ measurable random variables with finite and positive definite variance. The sequence is said to be $\alpha-$*mixing* or *strong mixing* if $\lim_{m\to\infty} \alpha_m = 0$, where

$$\alpha_m = \sup_t \alpha(\mathscr{F}_{-\infty}^t, \mathscr{F}_{t+m}^{\infty}). \tag{1.1}$$

and is said to be $\beta-$*mixing* or *absolutely regular* if $\lim_{m\to\infty} \beta_m = 0$, where

$$\beta_m = \sup_t \beta(\mathscr{F}_{-\infty}^t, \mathscr{F}_{t+m}^{\infty}). \tag{1.2}$$

Mixing coefficients/conditions (cont.)

For

$$\alpha(\mathscr{F}^t_{-\infty},\mathscr{F}^{\infty}_{t+m}) = \sup\{|\mathbb{P}(A\cap B)-\mathbb{P}(A)\mathbb{P}(B)| : A\in\mathscr{F}^{\infty}_{t+m}, B\in\mathscr{F}^t_{-\infty}\}$$
$$\beta(\mathscr{F}^t_{-\infty},\mathscr{F}^{\infty}_{t+m}) = \sup\{E(|\mathbb{P}(A\mid\mathscr{F}^t_{-\infty})-\mathbb{P}(A)|) : A\in\mathscr{F}^{\infty}_{t+m}\}$$

Davidson [1997] shows that $\beta-$mixing implies $\alpha-$mixing.

Yu.A.Davydov [1973] and Bradley [2005] show that for an ergodic Markov chain Y_t, of invariant probability measure π,

$$\beta_Y(t) = \int \|P^t(y,.) - \pi\| \pi(dy). \tag{1.3}$$

The rate ρ in Equation (0.4) can be chosen independently of the initial point. If Equation (0.4) holds then it follows that $\beta_Y(t) = O(\rho^t)$. Then (Y_t) is a stationary and geometrically ergodic, hence, $\beta-$mixing and by implication $\alpha-$mixing.

Further Assumptions (Assumptions B)

- For each $z \in S$, there exists $c_i(z), d_i \in \mathbb{R}^p$ and $c_i(z) \geq 0$, $d_i(z) \geq 0$, $i = 1, \ldots, p$ such that for $y = (y_1, \ldots, y_p)$
  - $|f_{z_t}(y)| \leq \sum_{i=1}^{p} c_{i(z)} |y_i| + o(\|y\|)$ as $\|y\| \to \infty$ and
  - $\sigma_{z_t}^2(y) \leq \sum_{i=1}^{p} d_{i(z)} |y_i^2| + o(\|y\|^2)$ as $\|y\| \to \infty$.
- **Drift Condition:** The Foster-Lyapounov drift condition (Tjostheim [1990], Meyn and Tweedie [1993]) There exists a real valued measure function $V \geq 1$ such that for some constant $\varepsilon > 0$, $0 < \rho < 1$, a constant M_1 and a small set $A = \{y \in \mathbb{R} : \|y\| \leq M_1\}$:

$$E[V(Q_t) \mid Q_{t-1} = (q)] \leq \rho V(q) \quad \text{for } y \in A^c \qquad (1.1)$$

$$\sup_{x \in A} E[V(Q_t) \mid Q_{t-1} = q] < \infty \quad \text{for } y \in A \qquad (1.2)$$

We will use the following result by Meyn and Tweedie [1993] to prove the geometric ergodicity of the MAR model.

Lemma (Meyn and Tweedie [1993])

For an aperiodic, $\varphi-$irreducible Markov chain, all petite sets are small sets.

Propositon

For the Markov chain $Q_t = (Z_t, Y_t)$, if for every $Z \in S$, $f_z(\cdot)$ is bounded on all compact sets, then Q_t is $\nu \times \varphi$-irreducible and for every compact set $C \in \mathbb{R}^p$, $S \times C$ is a small set.

To verify the geometric erogodicity of the MAR model, we need to:

- Prove that the process $Q_t = (Z_t, Y_t)$ is $\varphi-$irreducible and aperiodic.
- Show the existence of a test function $V(Q_t)$ satisfying the drift condition (Equation (1.1)) above.

The two steps are summarized in the following theorem.

Theorem

Consider the aperiodic Markov Chain $Q_t = (Z_t, Y_t)$. *For a small set* A *and the aperiodic and* $\varphi-$*irreducible process* $\{Y_t; t \geq 0\}$ *such that* $Y_t = (y_t, \ldots, y_{t-p+1})'$. *Each* y_t *is an MAR process defined by Equation* (1.2). *Suppose that Assumption (1) and Assumption (1) are satisfied and*

$$\sup_z E[\sum_{j=1}^{p} c_i(Z_t)c_j(Z_t) + E(\varepsilon_{Z_t}^2)d_i(Z_t) \mid Z_{t-1} = z] < 1 \qquad (1.3)$$

Theorem

Then

- $\{Y_t; t \geq 0\}$ *is geometrically ergodic with* $V(y) = 1 + \|y\|^2$
- $\{Y_t; t \geq 0\}$ *has a stationary distribution with finite second moments i.e.* $E_{\pi_y}[y_t^2] < \infty$ *and*
- $\{Y_t; t \geq 0\}$ *is* β*-mixing and hence strong mixing at geometric rate.*

where π *is unique invariant distribution of* Y_t *and*
$\pi_y(A) = \pi(S \times A \times \mathbb{R}^{p-1})$, $A \in B(\mathbb{R})$

Proof of Theorem 6.1(i)

To show that the drift condition 1.1 (Equation (1.1)) is satisfied, let $\tau_i(z) = \sum_{j=1}^{p} c_{i(z)} c_{j(z)}$ and choose $\delta > 0$ so that $\sum_{i=1}^{p} \xi_i + \delta = 1$, where

$$\xi_i = \sup_z E[\sum_{j=1}^{p} c_{i(z)} c_{j(z)} + E(\varepsilon_t^2) d_{i(z)} \mid Z_{t-1} = z] \leq (1 - \frac{\delta}{p}) \qquad (1.4)$$

Now define a test function $V : S \times R^p \to R$ by

$$V(z, y) = 1 + \|y^2\| \qquad (1.5)$$

$$E[V(Q_t) \mid Q_{t-1} = q] =\leq E_z[(f_{z_t}(y) + \sigma_{z_t}\varepsilon_t)^2] + \sum_{i=2}^{p} y_{i-1}^2 + 1 \qquad (1.6)$$

$$\leq \sum_{i=1}^{p} E_z[\tau_{i(z)} + E\varepsilon_{z_t}^2 d_{i(z)}]y_i^2 + \sum_{i=2}^{p} y_{i-1}^2 \leq \sum_{i=1}^{p} \xi_i y_1^2 + \sum_{i=2}^{p} y_{i-1}^2 + E_z[L_{z_t}(y)] + 1$$

$$\leq y_1(1 - \frac{\delta}{p}) + \sum_{i=2}^{p} y_{i-1}^2 + E_z[L_{z_t}(y)] + 1 \leq \sum_{i=1}^{p} y_i^2 - \frac{\delta}{p}\sum_{i=1}^{p} y_i^2 + E_z[L_{z_t}(y)] + 1$$

$$= (1 + \sum_{i=1}^{p} y_i^2) - \frac{\delta}{p}(1 + \sum_{i=1}^{p} y_i^2) + E_z[L_{z_t}(y)] + \frac{\delta}{p}$$

$$= V(z,y)\left[1 - \frac{\delta}{p} + \frac{1}{V(z,y)}\left[E_z\left[L_{z_t}(y)\right] + \frac{\delta}{p}\right]\right]$$

where

$$L_{z_t}(y) = (2o(\|y\|)(\sum_{i=1}^{p} c_{i(z)})|y_i|) + (o(\|y\|))^2 + E(\varepsilon_t^2)o(\|y\|^2). \quad (1.7)$$

$\frac{E_z[L_{z_t}(y)]}{V(z,y)} \to 0$ as $\|y\| \to \infty$, also,
$\frac{\delta/p}{V(z,y)} \to 0$ as $\|y\| \to \infty$ so that we have,

$$E[V(Y_t) \mid Y_{t-1} = z, y] \leq V(z,y)[1 - \frac{\delta}{p}] \quad (1.8)$$

Now suppose that $y \in A^c$ and there exists $M_1 > 1$ such that $\|y\| > M_1$ so that $\frac{\delta}{p} < \varepsilon < 1$, ε is a strictly positive constant defined in Equation (1.4).

choose $1 - \frac{\delta}{p} < \rho < 1$ in Equation (1.8), it follows that the first part of Equation (1.1) holds. Furthermore, since $f_{z_t}(y)$ is locally bounded for $y \in A$, the second part of Equation (1.1) holds. Thus,

$$\begin{aligned} E[V(Y_t) \mid Y_{t-1} = (z,y)] &\leq \rho V(z,y) \text{ for } y \in A^c \\ \sup_{y \in A} E[V(Y_t) \mid Y_{t-1} = (z,y)] &< \infty \text{ for } y \in A \end{aligned} \tag{1.9}$$

Therefore, the geometric ergodicity and hence the strict stationarity and $\beta-$mixing property of Y_t and hence, y_t are established.

We prove Theorem 6.1(ii) as follows, by 1 we can write

$$y_t^2 \leq [\sum_{i=1}^{p} c_{i(z)}|y_{i-1}| + o(\|y\|) + (\sum_{i=1}^{p} d_{i(z)}|y_{i-1}^2| + o\|y\|^2)^{\frac{1}{2}} \varepsilon_{z_t}]^2$$

$$= \sum_{i=1}^{p} c_{i(z)}|y_{i-1}|c_{j(z)}|y_{j-1}| + 2\sum_{i=1}^{p} c_{i(z)}y_{i-1}o(\|y\|) + (o\|y\|)^2 \sum_{i=1}^{p} d_{i(z)}|y^2| + o\|y$$

$$= \sum_{i=1}^{p} (\tau_{i(z)} + \varepsilon_{z_t}^2 d_{i(z)} y_{i-1}^2) + 2\sum_{i=1}^{p} c_{i(z)} y_{i-1} o(\|y\|) + (o\|y\|)^2 + o\|y\|^2)\varepsilon_{z_t}^2] + 2f_{z_t}($$

Taking expectation and by the independence of y_{t-1} and ε_{z_t} as well as z_t and ε_{z_t} we have,

$$Ey_t^2 \leq \sum_{i=1}^{p}(\tau_{i(z)} + E\varepsilon_{z_t}^2 d_{i(z)})Ey_{i-1}^2 + L_{z_t}(y) \tag{1.11}$$

$L_{z_t}(y)$ is the same as Equation (1.7) above.

$$EY_t^2 \leq \frac{L_{z_t}(y)}{1 - [\sum_{i=1}^{p}(\tau_{i(z)} + E\varepsilon_{z_t}^2 d_{i(z)})]} \tag{1.12}$$

Now by the proof of Theorem 6.1 i and ii, the Foster criterion F1 and F2 of Tweedie [1988] hold. Hence, by Tweedie [1988, Theorem 2], there exists a finite invariant measure π and Tweedie [1988, Theorem 1(iii)] holds. Hence, the RHS of Equation (1.12) is finite. and $E_\pi(y_t^2) < \infty$ as required.

Summary

Geometric ergodicity is very useful in establishing mixing conditions and central limit results for parameter estimates of a model. It also justifies the use of laws of large numbers and forms part of the basis for exploring the asymptotic theory of the model. A consequence of geometrically ergodicity is $\beta-$mixing.

Since $\beta-$mixing implies $\alpha-$mixing we can say that geometric ergodicity entails both $\alpha-$mixing and $\beta-$mixing. So that the absolute regularity and hence strong mixing of the Markov chain Y_t is a major consequence of geometric ergodicity.

Summary (cont.)

We have established the geometric ergodicity of the MAR model and by implication shown that it satisfies the absolutely regular and strong mixing conditions. In addition, we show that the process $\{y_t\}$ has a stationary distribution with finite second moments.

References

Bhattacharya, R. and Lee, C. (1999). On geometric ergodicity of nonlinear autoregressive models. *Statistics and amp; Probability Letters*, 41(4):439 – 440.

Boshnakov, G. N. (2009). Analytic expressions for predictive distributions in mixture autoregressive models. *Stat. Prob. Let*, 79(15):1704–1709.

Bradley, R. C. (2005). Basic properties of strong mixing conditions. a survey and some open questions. *Probability Surveys*, 2(2):107–144.

Davidson, J. (1997). *Stochastic Limit Theory*. Oxford University Press, New York, 3rd edition. Advanced texts in Econometrics.

References (cont.)

Dedecker, J. (2007). *Weak Dependence: With Examples and Applications*. Springer, New York.

Masanobu Taniguchi, Y. K. (2000). *Asymptotic Theory of Statistical Inference for Time Series*. Springer, New York.

Meyn, S. and Tweedie, R. (1993). *Markov Chains and Stochastic Stability*. Springer, Verlag, London.

Stockis, J.-P., Franke, J., and Kamgaing, J. T. (2010). On geometric ergodicity of charme models. *Journal of Time Series Analysis*, 31(3):141–152.

Tjostheim, D. (1990). Non-linear time series and markov chains. *Advances in Applied Probability*, 22(3):587–611.

References (cont.)

Tweedie, R. L. (1988). Invariant measures for markov chains with no irreducibility assumptions. *Journal of Applied Probability*, 25:pp. 275–285.

Yu.A.Davydov (1973). Mixing conditions for markov chains. *Theory Probability and its Applications*, 18(2):321–338.

QUESTIONS?