On First and Second Order Stationarity of Random Coefficient Models

Georgi N. Boshnakov

First version: 7 April 2009

Research Report No. 3, 2009, Probability and Statistics Group
School of Mathematics, The University of Manchester

On first and second order stationarity of random coefficient models

Georgi N. Boshnakov,[a]

[a]*School of Mathematics, The University of Manchester, Oxford Road, Manchester M13 9PL, UK*

Abstract

We give conditions for first and second order stationarity of mixture autoregressive processes and discuss related, sometimes confusing, results for such models. We obtain a simple condition for positive definiteness of the solution of a generalisation of the Stein's equation with semidefinite right-hand side and apply it to second order stationarity. The said condition may be of independent interest.

Key words: mixture AR models, mixture PAR models, random coefficient models, Stein's equation, extended Stein's equation

1. Introduction

We give a simple and natural sufficient condition for the solution to an extension of the Stein's equation with semidefinite righ-hand side to be positive definite and use it to study second order stationarity in mixture autoregressive (MAR) models and other random coefficient models.

Mixture autoregressive models have been introduced in the influential work Wong and Li (2000), see also Wong (1998). Previous work directly related to this class of models has been done by Le et al. (1996). Extensions of the MAR class to conditionally heteroscedastic, logistic and multivariate processes can be found in Wong and Li (2001a), Wong and Li (2001b), Wong (1998), Fong et al. (2007). A very general model which subsumes most of these models has been studied by Saikkonen (2007). Extention of MAR models to periodically correlated time series has been considered by Shao (2006).

MAR models are closely related to random coefficient autoregressive models. Andel (1976) obtains necessary and sufficient conditions for stationarity and Yule-Walker equations in the univariate case. Conlisk (1974) gives a sufficient condition for stability for some multivariate models. Further developments and a comprehensive exposition of the theory may be found in Nicholls and Quinn (1982).

Wong and Li (2000) give first and second order stationarity conditions for MAR models of orders one and two. Similar conditions have been given by Le et al. (1996) for mixture transition distribution models which are a subset of the MAR models. Strong results for stability and ergodicity are obtained by Saikkonen (2007) for a more general class of models than the ones considered here.

In this paper we consider first and second order stationarity of some of these models. MAR models are non-linear but their representation as random coefficient autoregressive models makes it possible to use matrix theory and linear algebra for their analysis. We use such a representation of the MAR model. This was not mentioned in the original paper Wong and Li (2000) but was used by later authors (see Saikkonen, 2007, and the references therein). This representation allows for using the theory and the methods developed by Nicholls and Quinn (1982).

We use the terms *stationary* and *strictly stationary* for weakly stationary processes and strictly stationary processes, respectively.

We will use also the the following weel known property of the vec operator

$$\operatorname{vec}(ABC) = (C^T \otimes A) \operatorname{vec} B. \tag{1}$$

A vector of ones is denoted by $\mathbf{1}$. $\boldsymbol{P} > 0$ and $\boldsymbol{P} \geq 0$ specify that the symmetric matrix $\boldsymbol{P}$ is positive definite and positive semidefinite, respectively. If $\boldsymbol{M}$ is a matrix, then $\lambda(\boldsymbol{M})$ is the maximum of the moduli of its eigenvalues.

2. An extension of Stein's equation

Consider the following pair of equations

$$\boldsymbol{R} - \boldsymbol{A}\boldsymbol{R}\boldsymbol{A}' = \boldsymbol{Q}, \tag{2}$$

$$\boldsymbol{P} - \boldsymbol{A}\boldsymbol{P}\boldsymbol{A}' - \mathrm{E}(\boldsymbol{U}'\boldsymbol{P}\boldsymbol{U}) = \boldsymbol{Q}, \tag{3}$$

where $\boldsymbol{A}$ and $\boldsymbol{Q}$ are $p \times p$ non-random matrices while $\boldsymbol{U}$ is a random matrix with finite second moments. The first equation is known as Stein's equation. In this paper we refer to the second equation as *extended Stein's equation.*

Conlisk (1974) has shown that if $\lambda(\boldsymbol{A} \otimes \boldsymbol{A} + \mathrm{E}(\boldsymbol{U}' \otimes \boldsymbol{U}')) < 1$, then there exists a pair of positive definite matrices $\boldsymbol{Q}$ and $\boldsymbol{P}$ satisfying (3). The same author later showed (Conlisk, 1976) that the inverse is also true. A more refined result was obtained by Nicholls and Quinn (1982, p. 35) and Quinn and Nicholls (1981). Here we give a generalisation to cover the case of semidefinite $\boldsymbol{Q}$.

Lemma 1. 1. *If* $\boldsymbol{Q} \geq 0$ *and* $\lambda(\boldsymbol{A} \otimes \boldsymbol{A} + \mathrm{E}(\boldsymbol{U}' \otimes \boldsymbol{U}')) < 1$*, then* $\boldsymbol{P} \geq 0$*.*

2. *If* $\boldsymbol{Q} \geq 0$*,* $\lambda(\boldsymbol{A} \otimes \boldsymbol{A} + \mathrm{E}(\boldsymbol{U}' \otimes \boldsymbol{U}')) < 1$*, and the solution,* $\boldsymbol{R}$*, to Equation* (2) *is positive definite, then* $\boldsymbol{P} > 0$*.*

Proof. The first part follows as in Conlisk (1974) or using a sequence of positive definite matrices $\boldsymbol{Q}$. To prove the second part, subtract Equation (2) from (3) to get

$$(\boldsymbol{P} - \boldsymbol{R}) - \boldsymbol{A}(\boldsymbol{P} - \boldsymbol{R})\boldsymbol{A}' = \mathrm{E}(\boldsymbol{U}'\boldsymbol{P}\boldsymbol{U}) \geq 0. \tag{4}$$

The condition $\lambda(\boldsymbol{A} \otimes \boldsymbol{A} + \mathrm{E}(\boldsymbol{U}' \otimes \boldsymbol{U}')) < 1$ implies that the eigenvalues of $\boldsymbol{A}$ have moduli smaller than 1 (Conlisk, 1974). If $\mathrm{E}(\boldsymbol{U}'\boldsymbol{P}\boldsymbol{U})$ is positive definite, then $\boldsymbol{P} - \boldsymbol{R}$ is also p.d. as the solution of the Stein's equation with a stable matrix and positive definite right-hand side. Therefore $\boldsymbol{P} > 0$ in this case.

If $\mathrm{E}(\boldsymbol{U}'\boldsymbol{P}\boldsymbol{U})$ is only positive semidefinite, then we can rewrite Equation (4) as follows:

$$\begin{aligned} 0 &= (\boldsymbol{P} - \boldsymbol{R}) - \boldsymbol{A}(\boldsymbol{P} - \boldsymbol{R})\boldsymbol{A}' - \mathrm{E}(\boldsymbol{U}'\boldsymbol{P}\boldsymbol{U}) \\ &= (\boldsymbol{P} - \boldsymbol{R}) - \boldsymbol{A}(\boldsymbol{P} - \boldsymbol{R})\boldsymbol{A}' - \mathrm{E}(\boldsymbol{U}'(\boldsymbol{P} - \boldsymbol{R})\boldsymbol{U}) - \mathrm{E}(\boldsymbol{U}'\boldsymbol{R}\boldsymbol{U}). \end{aligned}$$

Hence,

$$(\boldsymbol{P} - \boldsymbol{R}) - \boldsymbol{A}(\boldsymbol{P} - \boldsymbol{R})\boldsymbol{A}' - \mathrm{E}(\boldsymbol{U}'(\boldsymbol{P} - \boldsymbol{R})\boldsymbol{U}) = \mathrm{E}(\boldsymbol{U}'\boldsymbol{R}\boldsymbol{U}) \geq 0.$$

From the first part of the lemma it follows that $\boldsymbol{P} - \boldsymbol{R} \geq 0$, i.e. $\boldsymbol{P} \geq \boldsymbol{R} > 0$, since $\boldsymbol{R}$ is positive definite. Hence, $\boldsymbol{P} > 0$. □

3. Models

To avoid confusion we give here definitions for first and second order stationarity, see Section 8 for comments. The definitions below apply to both univariate and multivariate processes. In the latter case the mean is a vector and the covariances are matrices.

Definition 1. *The process* $\{y_t\}$ *is said to be stationary in the mean (or first order stationary) if* $\mathrm{E}\, y_t$ *is constant.*

Definition 2. *The process* $\{y_t\}$ *is said to be stationary (weakly stationary, covariance stationary, second order stationary) if* $\mathrm{E}\, y_t$ *is constant and the covariances* $\mathrm{Cov}(y_t, y_{t-k})$ *depend only on the lag* k*.*

To clarify the ideas we first look at the basic constant coefficient model

$$\boldsymbol{X}_{t+1} = \boldsymbol{c} + \boldsymbol{A}\boldsymbol{X}_t + \boldsymbol{\varepsilon}_{t+1}, \qquad t = 1, 2, \ldots \tag{5}$$

where $\boldsymbol{c} \sim (p \times 1)$ and $\boldsymbol{A} \sim (p \times p)$ are non-random and $\boldsymbol{\varepsilon}_t$, $t = 1, 2, \ldots$ is a sequence of independent random vectors which are also independent from the initial state $\boldsymbol{X}_0$ and have variance matrix $\operatorname{Var} \boldsymbol{\varepsilon}_t = \boldsymbol{Q}$. It is clear from Equation (5) that the existence of $\operatorname{E} \boldsymbol{\varepsilon}_t$ for all t is necessary for the existence of the mean $\operatorname{E} \boldsymbol{Y}_t$ for all t. So, we will assume that this condition holds.

The process defined by Equation (5) is first order stationary if and only if the mean, $\boldsymbol{\mu}_0 = \operatorname{E} \boldsymbol{X}_0$, of the initial value is a solution to the equation $\boldsymbol{\mu}_0 = \boldsymbol{c} + \boldsymbol{A}\boldsymbol{\mu}_0$. For second order stationarity it is necessary also that none of the eigenvalues of $\boldsymbol{A}$ has modulus larger than 1 and $\operatorname{Var} \boldsymbol{X}_0$ is a solution with respect to $\boldsymbol{R}$ of Stein's Equation (2). Usually it is also required that $\operatorname{Var} \boldsymbol{X}_0$ be positive definite. If $\boldsymbol{Q}$ is positive definite this happens if and only if the moduli of all eigenvalues of $\boldsymbol{A}$ are strictly smaller than 1. There may be positive definite solutions when $\boldsymbol{Q}$ is only semidefinite but the description is more delicate (see Lancaster and Tismenetsky, 1985, Theorem 13.2.4).

Random coefficient models can be obtained by allowing for the coefficients of model (5) to be random. Let z_t, $t = 1, 2, \ldots$, be a sequence of random variables such that $\operatorname{E}(\boldsymbol{\varepsilon}_t | z_t) = \boldsymbol{0}$ for all t and the bivariate sequence $(z_t, \boldsymbol{\varepsilon}_t)$, $t = 1, 2, \ldots$, is i.i.d. It is important that $\boldsymbol{\varepsilon}_t$ is not necessarilly independent of z_t. Suppose also that for each t the random vector $\boldsymbol{c}_{z_t}$ is a function of z_t such that $\operatorname{E} \boldsymbol{c}_{z_t} = \boldsymbol{c}$, and $\boldsymbol{A}_{z_t}$ is a random function of z_t such that $\operatorname{E} \boldsymbol{A}_{z_t} = \boldsymbol{A}$. In other words, there are collections of $(p \times 1)$ vectors, $\{\boldsymbol{c}_k\}$, and $(p \times p)$ matrices, $\{\boldsymbol{A}_k\}$, such that $\boldsymbol{c}_{z_t} = \boldsymbol{c}_k$ and $\boldsymbol{A}_{z_t} = \boldsymbol{A}_k$ when $z_t = k$. It is advantageous for some purposes to work with centred variables $\boldsymbol{U}_{z_t} = \boldsymbol{A}_{z_t} - \boldsymbol{A}$ which are such that $\operatorname{E} \boldsymbol{U}_{z_t} = \boldsymbol{0}$.

Consider now the random coefficient model

$$\boldsymbol{Y}_{t+1} = \boldsymbol{c}_{z_{t+1}} + \boldsymbol{A}_{z_{t+1}} \boldsymbol{Y}_t + \boldsymbol{\varepsilon}_{t+1}, \qquad t = 1, 2, \ldots, \tag{6}$$

with initial value $\boldsymbol{Y}_0$ independent of all other random variables.

Let $\boldsymbol{\mu}_t = \operatorname{E} \boldsymbol{Y}_t$ be the mean of $\boldsymbol{Y}_t$. From Equation (6) we get

$$\begin{aligned} \boldsymbol{\mu}_{t+1} &= \operatorname{E} \boldsymbol{c}_{z_{t+1}} + \operatorname{E}\left(\boldsymbol{A}_{z_{t+1}} \boldsymbol{Y}_t\right) + \operatorname{E} \boldsymbol{\varepsilon}_{t+1} \\ &= \boldsymbol{c} + \operatorname{E} \boldsymbol{A}_{z_{t+1}} \operatorname{E} \boldsymbol{Y}_t + \boldsymbol{0} \\ &= \boldsymbol{c} + \boldsymbol{A}\boldsymbol{\mu}_t. \end{aligned}$$

So, the conditions for first order stationarity of this model are the same as for the model with constant coefficients.

Turning to second order stationarity, assume that the mean is constant, i.e. $\operatorname{E} \boldsymbol{Y}_t = \boldsymbol{\mu}$. Subtracting $\boldsymbol{\mu}$ from both sides of (6) we obtain

$$\begin{aligned} \boldsymbol{Y}_{t+1} - \boldsymbol{\mu} &= \boldsymbol{c}_{z_{t+1}} - \boldsymbol{\mu} + \boldsymbol{A}_{z_{t+1}} \boldsymbol{\mu} - \boldsymbol{A}_{z_{t+1}} \boldsymbol{\mu} + \boldsymbol{A}_{z_{t+1}} \boldsymbol{Y}_t + \boldsymbol{\varepsilon}_{t+1} \\ &= (\boldsymbol{c}_{z_{t+1}} - \boldsymbol{\mu} + \boldsymbol{A}_{z_{t+1}} \boldsymbol{\mu}) + \boldsymbol{A}_{z_{t+1}} (\boldsymbol{Y}_t - \boldsymbol{\mu}) + \boldsymbol{\varepsilon}_{t+1} \end{aligned}$$

$$= \boldsymbol{d}_{z_{t+1}} + \boldsymbol{A}_{z_{t+1}}(\boldsymbol{Y}_t - \boldsymbol{\mu}) + \boldsymbol{\varepsilon}_{t+1},$$

where

$$\boldsymbol{d}_{z_{t+1}} = \boldsymbol{c}_{z_{t+1}} - \boldsymbol{\mu} + \boldsymbol{A}_{z_{t+1}}\boldsymbol{\mu}$$

and $\mathrm{E}\, \boldsymbol{d}_{z_{t+1}} = \boldsymbol{0}$. So, the centred process obeys a model similar to that of the non-centred one:

$$\boldsymbol{Y}_{t+1} - \boldsymbol{\mu} = \boldsymbol{d}_{z_{t+1}} + \boldsymbol{A}_{z_{t+1}}(\boldsymbol{Y}_t - \boldsymbol{\mu}) + \boldsymbol{\varepsilon}_{t+1}. \tag{7}$$

The centred process has a zero-mean random intercept which can be aggregated with the noise term if desired.

Let $C_{t,s} \equiv \mathrm{Cov}(\boldsymbol{Y}_t, \boldsymbol{Y}_s) = \mathrm{E}(\boldsymbol{Y}_t - \boldsymbol{\mu})(\boldsymbol{Y}_s - \boldsymbol{\mu})^T$ be the covariance between $\boldsymbol{Y}_t$ and $\boldsymbol{Y}_s$. Let $l > 0$. Multiplying both sides of Equation (7) by $(\boldsymbol{Y}_{t+1-l} - \boldsymbol{\mu})^T$ and taking expectation we get $C_{t+1,t+1-l} = \mathrm{E}\, \boldsymbol{A}_{z_{t+1}} C_{t,t+1-l} = \boldsymbol{A} C_{t,t+1-l}$. Iterating this we obtain

$$C_{t+l,t} = \boldsymbol{A}^l C_{t,t}, \qquad \text{for } l > 0. \tag{8}$$

The same relation holds for the autocovariances of the constant coefficient model (5).

If $\boldsymbol{A}$ has eigenvalues outside the unit circle, then for fixed t some elements of $C_{t+l,t}$ will become arbitrarily large for large l, and larger than the diagonal elements of $C_{t,t}$. Together with the Cauchy-Schwartz inequality this shows that some diagonal elements of $C_{t+l,t+l}$ will also be larger then the corresponding elements of $C_{t,t}$, i.e. $C_{t,t}$ cannot be chosen to be constant. It also shows that $C_{t,t}$ cannot have a finite limit as $t \to \infty$. We therefore conclude that the moduli of all eigenvalues of $\boldsymbol{A}$ should not exceed one if the process is to be second order stationary.

We also need the variance (lag 0 autocovariance) of $\boldsymbol{Y}_t$. Taking variances on both sides of Equation (7) and noting that the expected values of crossproducts are zero we get

$$\mathrm{Var}\left\{\boldsymbol{Y}_{t+1} - \boldsymbol{\mu}\right\} = \mathrm{Var}\left\{\boldsymbol{d}_{z_{t+1}}\right\} + \mathrm{Var}\left\{\boldsymbol{A}_{z_{t+1}}(\boldsymbol{Y}_t - \boldsymbol{\mu})\right\} + \mathrm{Var}\left\{\boldsymbol{\varepsilon}_{t+1}\right\}$$

Denoting $\Delta = \mathrm{Var}\left\{\boldsymbol{d}_{z_{t+1}}\right\}$ we obtain

$$\begin{aligned} C_{t+1,t+1} &= \Delta + \mathrm{E}\left\{\boldsymbol{A}_{z_{t+1}} C_{t,t} \boldsymbol{A}_{z_{t+1}}^T\right\} + \Sigma \\ &= \mathrm{E}\left\{\boldsymbol{A}_{z_{t+1}} C_{t,t} \boldsymbol{A}_{z_{t+1}}^T\right\} + \Delta_1, \end{aligned} \tag{9}$$

where $\Delta_1 = \Delta + \Sigma$. Using the centred variables $\boldsymbol{U}_{z_t}$ introduced earlier we obtain

$$\mathrm{E}\left\{\boldsymbol{A}_{z_{t+1}} C_{t,t} \boldsymbol{A}_{z_{t+1}}^T\right\} = \mathrm{E}\left\{(\boldsymbol{A} + \boldsymbol{U}_{z_{t+1}}) C_{t,t} (\boldsymbol{A} + \boldsymbol{U}_{z_{t+1}}^T)\right\}$$

$$= \boldsymbol{A} C_{t,t} \boldsymbol{A}^T + \mathrm{E}\left\{\boldsymbol{U}_{z_{t+1}} C_{t,t} \boldsymbol{U}_{z_{t+1}}^T\right\}.$$

So, Equation (9) can be written also in the form

$$C_{t+1,t+1} = \boldsymbol{A} C_{t,t} \boldsymbol{A}^T + \mathrm{E}\left\{\boldsymbol{U}_{z_{t+1}} C_{t,t} \boldsymbol{U}_{z_{t+1}}^T\right\} + \Delta_1.$$

For stationarity we need $C_{t+1,t+1} = C_{t,t}$, i.e.

$$C_{t,t} = \boldsymbol{A} C_{t,t} \boldsymbol{A}^T + \mathrm{E}\left\{\boldsymbol{U}_{z_{t+1}} C_{t,t} \boldsymbol{U}_{z_{t+1}}^T\right\} + \Delta_1.$$

Comparing the last equation with (3) we see that this is the extended Stein's equation with $\boldsymbol{Q} = \Delta_1$, $\boldsymbol{P} = C_{t,t}$, and $\boldsymbol{U}' = \boldsymbol{U}_{z_{t+1}}$. It is natural to impose the restriction that Equation (2) has a positive definite solution since it corresponds to a model with non-random coefficients where $\boldsymbol{A}_k = \boldsymbol{A}$ for all k.

Using Lemma 1 we can prove the following theorem. The semidefinite case was not considered by Conlisk. The related result of (Nicholls and Quinn, 1982, Theorem 2.2 on p. 21, first half of p. 24) is more general but does not have the full rank conclusion in the semidefinite case.

We will say that a process is of full rank if its lag 0 covariance matrix is of full rank, i.e. positive definite.

Theorem 1. *Let $\lambda(\boldsymbol{A} \otimes \boldsymbol{A} + \mathrm{E}\left\{\boldsymbol{U}_{z_{t+1}} \otimes \boldsymbol{U}_{z_{t+1}}\right\}) < 1$ and Δ_1 be a positive definite or positive semidefinite matrix such that the solution to the equation*

$$\boldsymbol{R} - \boldsymbol{A}\boldsymbol{R}\boldsymbol{A}' = \Delta_1 \tag{10}$$

is positive definite. The process $\{\boldsymbol{Y}_t\}$, $t = 0, 1, 2, \ldots$ is second order stationary and of full rank if and only if the initial vector $\boldsymbol{Y}_0$ has mean $\boldsymbol{\mu} = (\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{c}$ and its covariance matrix, $C_{0,0}$, is the solution of the equation

$$C_{0,0} = \boldsymbol{A} C_{0,0} \boldsymbol{A}^T + \mathrm{E}\left\{\boldsymbol{U}_{z_{t+1}} C_{0,0} \boldsymbol{U}_{z_{t+1}}^T\right\} + \Delta_1. \tag{11}$$

Informally, the theorem states that to every autoregression model with non-random coefficients there is a range of random coefficient models obtained by replacing its coefficients with random variables which do not vary "too much". The theorem also makes it clear that the existence of full rank stationary solution (equivalently, positive definite solution to Equation (11)) cannot be expressed entirely in terms of the eigenvalues of the matrix $\boldsymbol{A} \otimes \boldsymbol{A} + \mathrm{E}\left\{\boldsymbol{U}_{z_{t+1}} \otimes \boldsymbol{U}_{z_{t+1}}\right\}$. Indeed, if $\operatorname{Var} \boldsymbol{\varepsilon}_t = \boldsymbol{Q}$ is such that the solution to the Stein's equation is only positive semidefinite it may happen that the solution to (11) is semidefinite, as well. However, allowing for enough variation of the random intercept (in particular, giving it a positive definite covariance matrix) we can obtain a random coefficient model with positive definite covariance matrix.

The case of positive definite Δ_1 follows from the results of Nicholls and Quinn (1982) and Conlisk (1974). Strictly speaking the model here is not particular case of theirs since it contains a random intercept and unlike Quinn's model the coefficient matrix and the innovations are not assumed independent of each other. The main point of the theorem presented here however is that it covers the semidefinite case in a neat way. Moreover, the condition that the corresponding constant coeffcient model is of full rank is often easy to verify. It would hardly be exaggeration to claim that all models of interest possess this property. On the other hand, allowing for a semidefinite right-hand side is essential since models like this one often represent lower dimensional processes and thus have singular innovation variance matrices. This is the case with MAR models considered in the following sections.

4. The MAR model

A process $\{y(t)\}$ is said to be a mixture autoregressive process if the conditional distribution function of $y(t+1)$ given the information from the past of the process is a mixture of the following form:

$$\begin{aligned} F_{t+1|t}(x) &\equiv \Pr(y(t+1) \le x|\mathcal{F}_t) \\ &= \sum_{k=1}^{g} \pi_k F_k \left(\frac{x - \phi_{k,0} - \sum_{i=1}^{p_k} \phi_{k,i} y(t+1-i)}{\sigma_k} \right), \end{aligned} \tag{12}$$

where g is a positive integer, the number of components in the model; the probabilities $\pi_k > 0$, $k = 1, \dots, g$, $\sum_{k=1}^{g} \pi_k = 1$, define a discrete distribution, $\boldsymbol{\pi}$; $\sigma_k > 0$ and F_k is a distribution function for each $k = 1, \dots, g$. It is convenient to set $p = \max_{1 \le k \le g} p_k$ and $\phi_{k,i} = 0$ for $i > p_k$. We assume also that $t > p$. For many applications it is sufficient to consider standard normal noise components.

The MAR model is such that at each time t one of g autoregressive-like equations is picked up at random to generate $y(t)$. This observation can be used to give the model in another form. Let $\{z_t\}$ be an iid sequence of random variables with distribution $\boldsymbol{\pi}$, such that $\Pr\{z_t = k\} = \pi_k$, $k = 1, \dots, g$. Now the model (12) can be written as

$$y(t+1) = \phi_{z_{t+1},0} + \sum_{i=1}^{p} \phi_{z_{t+1},i} y(t+1-i) + \sigma_{z_{t+1}} \varepsilon_{z_{t+1}}(t+1), \tag{13}$$

where the distribution function of $\varepsilon_k(t)$ is F_k for each $k = 1, \dots, g$.

When the MAR model is represented by Equation (13), the dependence structure needs to be specified. Let $\mathcal{F}_t$ be the sigma field generated by the process $\{y(t)\}$ up to and including time t. We assume that $\varepsilon_k(t)$ are jointly independent and are also independent of past ys in the sense that for each t the σ-field generated by the set of random variables $\{\varepsilon_k(t+n), n \geq 1,\ 1 \leq k \leq g\}$ is independent of $\mathcal{F}_t$. Further, we assume that the choice of the component at time t (i.e., z_t) does not depend on $\mathcal{F}_{t-1}$ and $\{\varepsilon_k(t), t \geq 1,\ 1 \leq k \leq g\}$.

For $k = 1, \ldots, g$ define $\boldsymbol{A}_k$ by

$$\boldsymbol{A}_k = C[\phi_{k,1}, \ldots, \phi_{k,p}] \equiv \begin{pmatrix} \phi_{k,1} & \phi_{k,2} & \cdots & \phi_{k,p-1} & \phi_{k,p} \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{pmatrix}.$$

Let $\boldsymbol{A}$ be the expected value of $\boldsymbol{A}_{z_{t+1}}$. Then

$$\boldsymbol{A} \equiv \mathrm{E}\left(\boldsymbol{A}_{z_{t+1}}\right) = \sum_{k=1}^{g} \pi_k \boldsymbol{A}_k.$$

For $t \geq 0$, let $\boldsymbol{Y}_t = (y_t, \ldots, y_{t+1-p})^T$ be a vector of p values of the time series $\{y_t\}$. Then the vector process $\{\boldsymbol{Y}_t\}$ is a first order random coefficient autoregressive process:

$$\boldsymbol{Y}_{t+1} = \boldsymbol{c}_{z_{t+1}} + \boldsymbol{A}_{z_{t+1}} \boldsymbol{Y}_t + \boldsymbol{\varepsilon}_{\boldsymbol{t+1}, \boldsymbol{z_{t+1}}}, \tag{14}$$

where $\boldsymbol{\varepsilon}_{\boldsymbol{t+1}, \boldsymbol{z_{t+1}}} = (\sigma_{z_{t+1}} \varepsilon_{z_{t+1}}(t+1), 0, \ldots, 0)^T$,

$$\boldsymbol{c}_{z_{t+1}} = \begin{pmatrix} \phi_{z_{t+1},0} \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \boldsymbol{c} = \mathrm{E}\, \boldsymbol{c}_{z_{t+1}} = \begin{pmatrix} \mathrm{E}\, \phi_{z_{t+1},0} \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} c \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

The model specified by Equation (14) is a random coefficient autoregressive model (see Nicholls and Quinn, 1982) but two of its features need to be noted: the random intercept and the dependence between the coefficient and the noise term. The intercept of course is a trivial difference. The absence of independence between the coefficient and the noise term is a more serious difference but it does not affect the first and second order properties of the model which we discuss in this paper.

5. First order stationarity

For the mean of the process $\{y_t\}$ to be constant we obviously need that the mean of the initial vector $\boldsymbol{Y}_0$ be $\boldsymbol{\mu}_0 = \mu\mathbf{1}$ for some scalar constant μ. This is so without additional conditions if $\boldsymbol{c} = \mathbf{0}$ and $\boldsymbol{\mu}_0 = \mathbf{0}$, since then $\boldsymbol{\mu}_1 = \boldsymbol{c} + \boldsymbol{A}\boldsymbol{\mu}_0 = \mathbf{0}$ and by induction $\boldsymbol{\mu}_t = \mathbf{0}$ for all t. Non-zero constant mean is possible in the case $\boldsymbol{c} = \mathbf{0}$, if $\mathbf{1}$ is an eigenvector of $\boldsymbol{A}$ associated with the eigenvalue 1. If this is the case we can set $\boldsymbol{\mu}_0 = \mu\mathbf{1}$. Then $\boldsymbol{\mu}_1 = \boldsymbol{c} + \boldsymbol{A}\boldsymbol{\mu}_0 = \mathbf{0} + \mu\mathbf{1} = \mu\mathbf{1}$, and by induction the mean is constant. Finally, when $\boldsymbol{c} \neq 0$ note that $\boldsymbol{A}(\mu\mathbf{1}) = \mu\boldsymbol{A}\mathbf{1} = \mu(\sum\phi_i, 1, \dots, 1)^T$. So, $(\boldsymbol{I}-\boldsymbol{A})(\mu\mathbf{1}) = \mu(1 - \sum\phi_i, 0, \dots, 0)^T$. In this case, $\boldsymbol{\mu}_t$ will be equal to $\mu\mathbf{1}$ if and only if $\mu = c/(1 - \sum\phi_i)$. For the denominator in the last expression to be non-zero it is also necessary that 1 is not an eigenvalue of $\boldsymbol{A}$. This analysis exhausts the possibilities for constant mean and we summarise our findings in the following theorem.

Theorem 2. *The process* $\{y_t\}$, $t = 1-p, \dots, 0, 1, 2, \dots$ *is stationary in the mean if and only if* $\mathrm{E}\,\varepsilon_{z_{t+1}}(t+1)$ *exists and one of the following three cases holds:*

1. $\boldsymbol{c} = \mathbf{0}$ *and* $\boldsymbol{\mu}_0 = \mathbf{0}$*,*
2. $\boldsymbol{c} = \mathbf{0}$*,* $\mathbf{1}$ *is an eigenvector of* $\boldsymbol{A}$ *associated with eigenvalue 1, and* $\boldsymbol{\mu}_0 = \mu\mathbf{1}$ *for some constant* μ*,*
3. $\boldsymbol{c} \neq \mathbf{0}$*, 1 is not an eigenvalue of* $\boldsymbol{A}$*, and* $\boldsymbol{\mu}_0 = \mu\mathbf{1}$ *where* $\mu = c/(1 - \sum\phi_i)$*.*

Note the limited importance of the eigenvalues of $\boldsymbol{A}$ for first order stationarity, contrary to claims in some papers, see Section 8 for comments and references.

When the process is started at the inifinite past the requirement that the eigenvalues of $\boldsymbol{A}$ should be inside the unit circle becomes necessary from the outset. However this is not to so much to ensure first order stationarity but to ensure that the process can be defined at all. Indeed, Equation (13) shows that the process is causal, i.e. innovations are independent of past values of the process. Therefore iterating the equation we obtain a series which does not converge in any sense if some of the eigenvalues of $\boldsymbol{A}$ are greater than or equal to 1.

6. Second order stationarity of the MAR model

For second order stationarity it is necessary that $\{y(t)\}$ has constant mean $\mu = \mathrm{E}\,y(t)$. Then $\mathrm{E}\,\boldsymbol{Y}_t = \mu\mathbf{1}$. Equation (7) for the centred process specialises to

$$\boldsymbol{Y}_{t+1} - \mu\mathbf{1} = \boldsymbol{d}_{z_{t+1}} + \boldsymbol{A}_{z_{t+1}}(\boldsymbol{Y}_t - \mu\mathbf{1}) + \boldsymbol{\varepsilon_{t+1,z_{t+1}}},$$

where

$$\boldsymbol{d}_{z_{t+1}} = \boldsymbol{c}_{z_{t+1}} - \mu\mathbf{1} + \boldsymbol{A}_{z_{t+1}}\mu\mathbf{1}$$

and $\mathrm{E}\,\boldsymbol{d}_{z_{t+1}} = \mathbf{0}$. Note that $\boldsymbol{A}_{z_{t+1}}$ is a companion matrix and only the top element of $\boldsymbol{c}_{z_{t+1}}$ is non-zero, see 4. Also, $\boldsymbol{d}_{z_{t+1}}$ has the same pattern as $\boldsymbol{c}_{z_{t+1}}$, i.e. only its first element may be non-zero. So, the centred process obeys a MAR model as well. As before, the centred process has a zero-mean random intercept which can be aggregated with the noise term if desired.

Theorem 1 can be specialised to this case as follows.

Theorem 3. *Let* $\lambda(\boldsymbol{A} \otimes \boldsymbol{A} + \mathrm{E}\left\{\boldsymbol{U}_{z_{t+1}} \otimes \boldsymbol{U}_{z_{t+1}}\right\}) < 1$ *and* $\Delta_1 \neq \mathbf{0}$.

The process $\{y_t\}$, $t = 1-p, \dots, 0, 1, 2, \dots$ *is second order stationary if and only if the initial vector* $(y_0, y_{-1}, \dots, y_{1-p})^T$ *has mean* $\mu\mathbf{1}$, *where* μ *is some scalar, and covariance matrix* $C_{0,0}$ *which is the solution of the equation*

$$C_{0,0} = \boldsymbol{A}C_{0,0}\boldsymbol{A}^T + \mathrm{E}\left\{\boldsymbol{U}_{z_{t+1}}C_{0,0}\boldsymbol{U}_{z_{t+1}}^T\right\} + \Delta_1.$$

Informally, the theorem states that to every autoregression model with non-random coefficients there is a range of random coefficient autoregressive models obtained by replacing its coefficients with random variables which do not vary "too much".

7. Stability of MAR models

8. Remarks

Nicholls and Quinn (1982) discuss comprehensively various notions of stationarity and stability or random coefficient autoregressive models, see also Andel (1976) and Conlisk (1974). It is somewhat surpirising to find confusing theorems about stationarity in the mean and second order stationarity in later papers. Our comments here aim to bring attention to such inaccuracies. We would like to stress that the stationarity conditions discussed below are of marginal importance in the cited papers, especially the influential works Le et al. (1996) and Wong and Li (2000), and our remarks do not aim to diminish their contribution in any way.

In the innovative work Le et al. (1996) which introduced the mixture transition distribution model (a special case of the MAR model) the process is started at $t = 1$ and therefore results about stationarity should say something about the initial values. The authors give necessary and sufficient conditions for first and second order stationarity Le et al. (1996)[Theorems 1 and 2] which clearly should be interpreted as existence results. In fact, the proofs in Appendix A of that paper mention this. However, these proofs use a result from Beneš (1967) to prove that the conditions of their theorems are sufficient. They do not discuss necessity. Moreover Beneš's result gives strict stationarity which leaves the necessity condition even more in need of proof.

Another influential paper Wong and Li (2000) which introduced the mixture autoregression model is more vague on this issue. Its authors do not state the time span but their use of the Beneš's result mentioned above and the spirit of the paper suggest that they are starting the process at a finite moment (not $-\infty$). They do not give any details as to how the Beneš's result is applied. Similar statements appear also in Fong et al. (2007) in an extention of the MAR model to multivariate processes.

The proliferation of this kind of result is continued by Shao (2006) who generalises the MAR model to periodically correlated processes. That author also mentions vaguely that his results follow from Beneš's theorem.

All of the above papers use interchangeably the terms *stationary in the mean* and *first order stationary* so that there is no confusion with terminology in this respect. In respect to second order stationarity, they seem to actually mean stability.

The results mentioned above are formulated as theorems in the respective papers. As such they are confusing and vague at best. At least Le et al. (1996) clarify what they mean in their proof, albeit in an appendix. The other authors simply refer to Beneš's result (almost) out of the blue.

References

Andel, J., 1976. Autoregressive series with random parameters. Math. Operationsforsch. Stat. 7, 735–741.

Beneš, V., 1967. Existence of finite invariant measures for Markov processes. Proc. Am. Math. Soc. 18, 1058–1061.

Conlisk, J., 1974. Stability in a random coefficient model. Int. econom. Review 15, 529–533.

Conlisk, J., 1976. A further note on stability in a random coefficient model. Int. Econ. Rev. 17, 759–764.

Fong, P., Li, W., Yau, C., Wong, C., 2007. On a mixture vector autoregressive model. Can. J. Stat. 35 (1), 135–150.

Lancaster, P., Tismenetsky, M., 1985. The theory of matrices. 2nd ed., with applications. Computer Science and Applied Mathematics. Orlando etc.: Academic Press (Harcourt Brace Jovanovich, Publishers).

Le, N. D., Martin, R., Raftery, A. E., 1996. Modeling flat stretches, bursts, and outliers in time series using mixture transition distribution models. J. Am. Stat. Assoc. 91 (436), 1504–1515.

Nicholls, D. F., Quinn, B. G., 1982. Random coefficient autoregressive models: an introduction. Lecture Notes in Statistics, 11. New York - Heidelberg - Berlin: Springer-Verlag. V, 154 p. DM 26.00; $ 10.90 .

Quinn, B., Nicholls, D., 1981. The stability of random coefficient autoregressive models. Int. Econ. Rev. 22, 741–744.

Saikkonen, P., 2007. Stability of mixtures of vector autoregressions with autoregressive conditional heteroskedasticity. Stat. Sin. 17 (1), 221–239.

Shao, Q., 2006. Mixture periodic autoregressive time series models. Stat. Probab. Lett. 76 (6), 609–618.

Wong, C., Li, W., 2001a. On a logistic mixture autoregressive model. Biometrika 88 (3), 833–846.

Wong, C. S., 1998. Statistical inference for some nonlinear time series models. Ph.D. thesis, University of Hong Kong, Hong Kong.

Wong, C. S., Li, W. K., 2000. On a mixture autoregressive model. J. R. Stat. Soc., Ser. B, Stat. Methodol. 62 (1), 95–115.

Wong, C. S., Li, W. K., 2001b. On a mixture autoregressive conditional heteroscedastic model. J. Am. Stat. Assoc. 96 (455), 982–995.