Statistica Sinica **17**(2007), 221-239

# STABILITY OF MIXTURES OF VECTOR AUTOREGRESSIONS WITH AUTOREGRESSIVE CONDITIONAL HETEROSKEDASTICITY

Pentti Saikkonen

*University of Helsinki*

*Abstract:* This paper gives necessary and sufficient conditions for stationarity and existence of second moments in mixtures of linear vector autoregressive models with autoregressive conditional heteroskedasticity. Sufficient conditions are also provided for a more general model in which the mixture components are permitted to exhibit limited forms of nonlinearity. When specialized to the corresponding non-mixture case these sufficient conditions improve on their previous counterparts obtained for nonlinear autoregressions with nonlinear conditional heteroskedasticity. In this context, a previous conjecture is also disproved. The results of the paper are proved by using the stability theory of Markov chains. Stationarity, existence of second moments of the stationary distribution, and $\beta$-mixing are obtained by establishing an appropriate version of geometric ergodicity.

*Key words and phrases:* Autoregressive conditional heteroskedasticity, $\beta$-mixing, geometric ergodicity, Markov chain, mixture autoregressive model, nonlinear vector autoregressive model, stability.

## 1. Introduction

Most of the traditional time series analysis assumes that the underlying data generation process is stationarity or stable in the sense that the possible nonstationarity is only due to transient effects caused by initial values. For conventional linear time series models, necessary and sufficient conditions for stationarity are well known and straightforward to obtain. However, for nonlinear time series models the situation is much more difficult. For such models, the most convenient approach for studying stationarity is based on the theory of Markov chains. When the considered model has a Markovian structure it suffices to establish a property known as geometric ergodicity. Once this has been done the desired stationarity or stability follows, along with useful mixing results. Chan and Tong (1985), Chan, Petrucelli, Tong and Woolford (1985), Feigin and Tweedie (1985) and Pham (1986) were among the first authors to use this approach for such nonlinear time series models as threshold autoregressive models, random coefficient autoregressive models, and bilinear models.

In more recent work, Le, Martin and Raftery (1996) and Wong and Li (2000) introduced nonlinear time series models that can be viewed as mixtures of conventional linear autoregressive models. Wong and Li (2001) extended these models to allow for autoregressive conditionally heteroskedastic (ARCH) errors. In these papers necessary and sufficient conditions for second order stationarity were also obtained in some simple special cases. A major purpose of this paper is to provide necessary and sufficient conditions for geometric ergodicity and existence of second moments in a multivariate extension of the model considered by Wong and Li (2001).

Our sufficient conditions for geometric ergodicity are based on an approach already used by Feigin and Tweedie (1985) for random coefficient autoregressive models. It turns out that this approach also applies to mixtures of certain nonlinear autoregressive models with a nonlinear ARCH term. The most general model considered in the paper is therefore a multivariate mixture extension of the univariate models studied by Masry and Tjøstheim (1995), Lu (1998) and Chen and Chen (2001). The nonlinearity assumed in the previous work, and hence also in this paper, is rather limited, for a number of commonly used nonlinear models are ruled out. However, it is still reasonable to include mixture extensions of these nonlinear models in the paper because, when specialized to the corresponding non-mixture case, our sufficient condition for geometric ergodicity provides a clear improvement on those of Masry and Tjøstheim (1995), Lu (1998) and Chen and Chen (2001). Furthermore, we are able to disprove Lu's (1998) conjecture on a sufficient condition for the existence of second moments in his model.

The paper is organized as follows. Section 2 presents the considered model and related assumptions. Section 3 contains the main results and Section 4 concludes. Proofs of the theorems are deferred to an appendix.

The following notation is used throughout the paper. The symbol vec denotes the usual column vectorizing operator which stacks the columns of a matrix in a column vector. The half vectorization operator, denoted by vech, stacks only the columns from the principal diagonal of a square matrix downwards in a column vector. For any symmetric $k \times k$ matrix $A$, the symbol $D_k$ is used for the $k^2 \times [k(k+1)/2]$ duplication matrix defined by $\text{vec}(A) = D_k \text{vech}(A)$ whereas $L_k$ signifies the $[k(k+1)/2] \times k^2$ elimination matrix such that $\text{vech}(A) = L_k \text{vec}(A)$. The symbol $\otimes$ is used for Kronecker's product. The largest and smallest eigenvalues of a square matrix $A$ are denoted by $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$, respectively, whereas $\rho(A) = |\lambda_{\max}(A)|$ is the spectral radius of $A$. Furthermore, $\text{tr}(A)$ and $\det(A)$ signify the trace and determinant of the square matrix $A$, respectively. Finally, $\mathbf{1}(\cdot)$ denotes the indicator function and $\|\cdot\|$ is used for the Euclidean norm defined by $\|B\| = [\text{tr}(B'B)]^{1/2}$ when $B$ is a matrix.

## 2. Model and Assumptions

We start with a very general mixture of nonlinear vector autoregressions with a nonlinear ARCH term. Let $\eta_t$ be a sequence of independent, identically distributed random variables with a discrete probability distribution given by

$$P(\eta_t = s) = \pi_s, \quad s = 1, \ldots, m, \tag{1}$$

where each $\pi_s > 0$ and $\pi_1 + \cdots + \pi_m = 1$. The random variable $\eta_t$ determines which one of $m$ autoregressions at time $t$ generates the value of the considered $n$-dimensional stochastic process $z_t$ $(t = 1, 2, \ldots)$. Specifically, suppose $z_t$ is generated by

$$z_t = \sum_{s=1}^{m} \mathbf{1}(\eta_t = s)\left(f_s(z_{t-1}, \ldots, z_{t-p}) + H_s(z_{t-1}, \ldots, z_{t-p})^{\frac{1}{2}} \varepsilon_t^{(s)}\right), \tag{2}$$

where $f_s$: $\mathbb{R}^{np} \to \mathbb{R}^n$ and $H_s$: $\mathbb{R}^{np} \to \mathbb{R}^{n\times n}$ are (possibly) nonlinear functions with $H_s$ positive definite, and $\varepsilon_t^{(s)}$ $(n \times 1)$ is a sequence of independent and identically distributed random vectors with zero mean and identity covariance matrix or, briefly, $\varepsilon_t^{(s)} \sim i.i.d.(0, I_n)$. Moreover, the sequences $\varepsilon_t^{(s)}$, $s = 1, \ldots, m$, and $\eta_t$ are independent and $z_{t-j}$ is independent of $\varepsilon_t^{(s)}$ and $\eta_t$ for all $t, s$ and $j > 0$. Notice that the same number of lags $p$ can be assumed in (2) because this case can always be achieved by an appropriate redefinition of the functions $f_s$ and $H_s$ (cf., Lu (1998)).

For $m = 1$ the indicator function can be dropped from (2) and the model becomes a conventional nonlinear vector autoregression with a nonlinear ARCH term. The stability of the model in this special case was recently studied by Lu and Jiang (2001) without assuming that the errors $\varepsilon_t^{(1)}$ have finite second moments. Assuming the existence of second moments is convenient and mostly not restrictive. The existence of second moments was assumed by Lu (1998) and Chen and Chen (2001, Corollary 4.2) who obtained stability results for univariate models similar to (2) with $n = m = 1$. These results extended the previous results of Masry and Tjøstheim (1995, Lemma 3.1). Our work makes use of ideas similar to those employed by these previous authors.

When $m > 1$, (2) becomes a mixture of nonlinear vector autoregressions with a nonlinear ARCH term. Models of this kind have recently been considered by Le, Martin and Raftery (1996) and, more generally, by Wong and Li (2000, 2001). This previous work has been confined to the scalar case $n = 1$ with the error terms $\varepsilon_t^{(s)}$, $s = 1, \ldots, m$, identically distributed. Moreover, the $f_s$ have been linear and the $H_s$ have been either constant or they have specified a conventional ARCH structure for the regression errors $z_t - f_s(z_{t-1}, \ldots, z_{t-p})$. These special cases presumably have most practical interest although our results also apply to more general mixtures of nonlinear models. However, as mentioned in the introduction, the nonlinearity permitted is rather limited. As in Masry and Tjøstheim (1995), Lu (1998) and Chen and Chen (2001), it is assumed that the functions $f_s$ and $H_s$ are (strictly) dominated by a linear function and a quadratic function, respectively, at regions sufficiently far away from the origin (see conditions (4) and (5) below).

We assume that the error terms $\varepsilon_t^{(s)}$ have (Lebesgue) densities $\phi_\varepsilon^{(s)}$. Then setting $Z_t = \left[z_t' \cdots z_{t-p+1}'\right]'$, it straightforwardly follows from the imposed assumptions that the conditional distribution of $z_t$ given its past only depends on $Z_{t-1}$ and, given $Z_{t-1} = x$, this conditional distribution is characterized by the density

$$\phi_z\left(z_t \,|x\right) = \sum_{s=1}^{m} \pi_s \det\left(H_s\left(x\right)^{-\frac{1}{2}}\right) \phi_\varepsilon^{(s)}\left(H_s\left(x\right)^{-\frac{1}{2}}\left(z_t - f_s\left(x\right)\right)\right). \tag{3}$$

This conditional density corresponds to the definition of the mixture autoregressive models adopted by Le, Martin and Raftery (1996) and Wong and Li (2000, 2001). The definition based on (2) makes it explicit that the models considered by these authors can be viewed as random coefficient autoregressive models and, therefore, their stability can be studied by using methods developed for such models. Well known references are Nicholls and Quinn (1982), Feigin and Tweedie (1985) and Pham (1986).

We now discuss assumptions imposed on the above model. Set $x = [x_1' \cdots x_p']'$ where $x_i \in \mathbb{R}^n$ $(i = 1, \ldots, p)$. Similarly to Masry and Tjøstheim (1995) and Lu (1998) we assume that the functions $f_s$ satisfy

$$f_s\left(x\right) = \sum_{j=1}^{p} B_{sj} x_j + o\left(\|x\|\right) \quad \text{as} \quad \|x\| \to \infty \quad (s = 1, \ldots, m). \tag{4}$$

Here $B_{sj}$ $(j = 1, \ldots, p)$ are $n \times n$ matrices which may all be zero. Thus, for $\|x\|$ large, the functions $f_s$ are dominated by linear functions.

As for the functions $H_s$, we assume that

$$H_s\left(x\right) = \sum_{i=1}^{k} K_{si} x x' K_{si}' + o\left(\|x\|^2\right) \quad \text{as} \quad \|x\| \to \infty \quad (s = 1, \ldots, m), \tag{5}$$

where $K_{si}$ $(i = 1, \ldots, k)$ are $n \times np$ matrices, possibly all zero. This implies that for $\|x\|$ large the functions $H_s$ are dominated by quadratic functions similar to those used in the so called BEKK formulation of the multivariate ARCH model (see Engle and Kroner (1995)). Indeed, in our context, this special case corresponds to choosing

$$H_s(x) = \Phi_{s0} + \sum_{i=1}^{l}\sum_{j=1}^{p} \Phi_{sij} x_j x_j' \Phi_{sij}', \tag{6}$$

where $\Phi_{s0}$ and $\Phi_{sij}$ ($i = 1, \ldots, l, j = 1, \ldots, p$) are $n \times n$ parameter matrices with $\Phi_{s0}$ (typically) positive definite, and the lag length $p$ has been used to make the formulation conformable to that in (2). The latter term on the right hand side of (6) can be written as $\sum_{i=1}^{l}\sum_{j=1}^{p} K_{sij} x x' K_{sij}'$, where $K_{sij} = [0 : \cdots : 0 : \Phi_{sij} : 0 : \cdots : 0]$ and $\Phi_{sij}$ is located in the $j$th position. Thus for the BEKK model (6), condition (5) holds and the term $o(\|x\|^2)$ therein is only due to the constant matrices $\Phi_{0s}$. Note that if the conditional heteroskedasticity appears in the regression errors $z_t - f_s(z_{t-1}, \ldots, z_{t-p})$, a redefinition of the functions $f_s$ and $H_s$ is required to make (2) with the same lag length $p$ appropriate.

The matrices $B_{sj}$ and $K_{si}$ in (4) and (5) are required to satisfy suitable assumptions. To this end, define the companion matrices

$$B_s = \begin{bmatrix} B_{s1} & B_{s2} & \cdots & B_{s(p-1)} & B_{sp} \\ I_n & 0 & \cdots & 0 & 0 \\ 0 & I_n & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I_n & 0 \end{bmatrix} \quad (np \times np) \tag{7}$$

and the matrices

$$K_{si0} = \begin{bmatrix} K_{si} \\ 0 \end{bmatrix} \quad (np \times np). \tag{8}$$

We also need the duplication matrix $D_{np}$ of dimension $n^2p^2 \times (1/2)np(np+1)$ and the elimination matrix $L_{np}$ of dimension $(1/2)np(np+1) \times n^2p^2$.

Now we can introduce our assumptions. We call a function locally bounded if it is bounded on compact subsets of its domain.

**Assumption 1.** (i) For each $s = 1, \ldots, m$, the i.i.d.$(0, I_n)$ random vectors $\varepsilon_t^{(s)}$ have Lebesgue densities $\phi_\varepsilon^{(s)}$ and, for some $s = s_0$, $\phi_\varepsilon^{(s)}$ is bounded away from zero on compact subsets of $\mathbb{R}^n$.
(ii) For each $s = 1, \ldots, m$, the functions $f_s$: $\mathbb{R}^{np} \to \mathbb{R}^n$ and $H_s$: $\mathbb{R}^{np} \to \mathbb{R}^{n\times n}$ are Borel measurable and locally bounded with $H_s$ positive definite. In addition, for the same $s = s_0$ as in (i), $\lambda_{\min}(H_s(x))$ is bounded away from zero on compact subsets of $\mathbb{R}^{np}$.
(iii) The functions $f_s$ and $H_s$ satisfy (4) and (5) for all $s = 1, \ldots, m$ and, furthermore,

$$\rho\Big(\sum_{s=1}^{m} \pi_s L_{np}\left(B_s' \otimes B_s' + K_{s0}'\right) D_{np}\Big) < 1, \tag{9}$$

where $K_{s0} = \sum_{i=1}^{k} (K_{si0} \otimes K_{si0})$ and $\pi_s$ is as in (1).

In the case $m = 1$, Assumptions 1(i) and (ii) or their close variants have previously been used by several authors to study the stability of nonlinear time series models (see e.g., Lu (1998), Cline and Pu (1998), Chen and Chen (2001) and Lu and Jiang (2001)). From a practical point of view they are mild and met in most applications. The eigenvalue condition in Assumption 1(ii) usually holds because an additive positive definite constant part is typically included in models for conditional covariance matrix (cf. the matrix $\Phi_{s0}$ in (6)). Theoretically it is interesting that the additional conditions imposed in the latter parts of Assumptions 1(i) and (ii) are only required to hold for the single choices $\phi_{\varepsilon}^{(s_0)}$ and $H_{s_0}$.

Conditions (4) and (5) in Assumption 1(iii) are similar to those employed by Masry and Tjøstheim (1995), Lu (1998) and Chen and Chen (2001, Corollary 4.2). They restrict the class of permitted nonlinearity and actually rule out cases of interest. For instance, condition (4) does not hold for the general threshold autoregressive model studied by Chan and Tong (1985, Section 3) or its smooth variants introduced by Chan and Tong (1986) and Luukkonen, Saikkonen, and Teräsvirta (1988). However, as will be demonstrated in the next section, the eigenvalue condition (9), and hence Assumption 1(iii) as a whole, is still considerably weaker than its analogs used by the aforementioned authors.

An eigenvalue condition akin to (9) has previously been employed by Nicholls and Quinn (1982, Chap.2) in linear random coefficient autoregressive models. These authors point out (see p.35 of their monograph) that an equivalent alternative condition is obtained by suppressing the matrices $L_{np}$ and $D_{np}$ from (9). This alternative condition was subsequently adopted by Feigin and Tweedie (1985). We prefer the form in (9) because it appears more convenient in some derivations and because, due to a smaller dimension, the involved eigenvalue is easier to compute in practice. Notice that the elimination matrix $L_{np}$ on the left hand side of (9) can be replaced by $D_{np}^{+} = \left(D_{np}' D_{np}\right)^{-1} D_{np}'$, the Moore-Penrose inverse of $D_{np}$ (see the definition of the matrix $K_{s0}$ and result 9.6.5(1)(a) in Lütkepohl (1966)). From this fact it can be deduced that the transpose signs could be suppressed from (9).

The eigenvalue condition (9) can also be formulated by using concepts employed by Pham (1986) in the context of generalized random coefficient autoregressive models. As in that paper, let $(\mathbb{R}^{np})^{\odot 2}$ be the space of $(np)^2$ vectors $(x_{ij}, i, j = 1, \ldots, np)$ which are invariant with respect to any permutation of the subscripts $i$ and $j$. Clearly, $(\mathbb{R}^{np})^{\odot 2}$ can be identified with the space of symmetric matrices of order $np$. If $M$ is a square matrix of order $np$ its symmetric tensor product with itself, $M^{\odot 2}$, is defined on $(\mathbb{R}^{np})^{\odot 2}$ by the operator which, in obvious matrix notation, maps the symmetric matrix $X$ to the symmetric matrix $Y = MXM'$. Well-known properties of the vech operator (see Lütkepohl (1996, Chap.7.3)) imply that this mapping can be identified with $\text{vech}(Y) = L_{np}(M \otimes M)D_{np}\text{vech}(X)$ and, consequently, $L_{np}(M \otimes M)D_{np}$ can be identified with a matrix representation of the operator $M^{\odot 2}$ (cf. the proof of Lemma 3 of Pham (1986)). Thus, it follows that condition (9) can be expressed as

$$\rho\Big(\sum_{s=1}^{m} \pi_s \Big(B_s'^{\odot 2} + \sum_{i=1}^{k} K_{si0}'^{\odot 2}\Big)\Big) < 1, \tag{10}$$

where the transpose signs could again be omitted.

## 3. Results

The following theorem shows that Assumption 1 guarantees the existence of initial values which make the process $z_t$ stationary. The proof and formulation of this theorem is based on the theory of Markov chains. To this end, we cast the model into state space form. Recall that $Z_t = \left[z_t' \cdots z_{t-p+1}'\right]'$ and define

$$\begin{aligned} F_1\left(Z_{t-1}, \eta_t, \varepsilon_t\right) &= \sum_{s=1}^{m} \mathbf{1}\left(\eta_t = s\right)\Big(f_s\left(Z_{t-1}\right) + H_s\left(Z_{t-1}\right)^{\frac{1}{2}} \varepsilon_t^{(s)}\Big), \\ F_i\left(Z_{t-1}, \eta_t, \varepsilon_t\right) &= z_{t-i}, \quad i = 2, \ldots, p, \end{aligned}$$

where $\varepsilon_t = [\varepsilon_t^{(1)\prime} \cdots \varepsilon_t^{(m)\prime}]'$. Setting $F\left(Z_{t-1}, \eta_t, \varepsilon_t\right) = [F_1\left(Z_{t-1}, \eta_t, \varepsilon_t\right)' \cdots F_p(Z_{t-1}, \eta_t, \varepsilon_t)']'$, we have $Z_t = F\left(Z_{t-1}, \eta_t, \varepsilon_t\right)$ and $z_t = \Upsilon' Z_t$, where $\Upsilon' = [I_n : 0 : \cdots : 0]$ $(n \times np)$. Since $\varepsilon_t$ and $\eta_t$ are independent of $Z_{t-1}$ this shows that $Z_t$ is a Markov chain on $\mathbb{R}^{np}$.

In what follows, the concept of $V$-geometric ergodicity of a Markov chain will be employed (see Meyn and Tweedie (1993, p.355)). Here $V$ signifies a real valued measurable function defined on the state space of the considered Markov chain and such that $V(\cdot) \geq 1$. For such a function $V$, the Markov chain $Z_t$ is said to be $V$-geometrically ergodic if there exists a probability measure $\pi$ on the Borel sets of $\mathbb{R}^{np}$ and a constant $\varrho > 1$ such that

$$\sum_{t=1}^{\infty} \varrho^t \sup_{h:|h| \leq V} \left| E\left(h\left(Z_t\right) | Z_0 = x\right) - \int_{\mathbb{R}^{np}} \pi\left(dy\right) h\left(y\right) \right| < \infty \quad \text{for all } x \in \mathbb{R}^{np}. \tag{11}$$

The definition also assumes that the function $V$ is integrable with respect to the probability measure $\pi$. The weakest form of this definition results when $V \equiv 1$. Then the Markov chain $Z_t$ is said to be geometrically ergodic. Geometric ergodicity entails that the $t$-step transition probability measure $P^t\left(x, \cdot\right)$ defined on the Borel sets of $\mathbb{R}^{np}$ by $P^t\left(x, A\right) = P\left(Z_t \in A \,|Z_0 = x\right)$ converges at a geometric rate, and for all $x \in \mathbb{R}^{np}$, to the probability measure $\pi(\cdot)$ with respect to the total variation norm.

As is well known, geometric ergodicity implies stationarity of the process $Z_t$ if the distribution of the initial value $Z_0$ is defined by the probability measure $\pi$ (see Meyn and Tweedie (1993, pp.230-231)). Of course, in this case the process $z_t$ is also stationary. These results are contained in the following theorem.

**Theorem 1.** *Suppose the process $z_t$ $(t = 1, 2, \ldots)$ is generated by* (2). *If Assumption* 1 *holds, the process $Z_t = \left[z_t' \cdots z_{t-p+1}'\right]'$ is a $V$-geometrically ergodic Markov chain with $V(x) = 1 + \|x\|^2$.*

Theorem 1 and our previous discussion imply that, in addition to being strictly stationary, the process $z_t$ is also second order stationary when initialized at the stationary distribution. A further convenient implication of Theorem 1 is that it provides known conditions for second order stationarity in simple special cases. In particular, when $m = 1$ and when (2) defines a linear homoskedastic vector autoregressive model, the eigenvalue condition (9) in Assumption 1(iii) is necessary and sufficient for the existence of a causal second order stationary solution of the stochastic difference equation defining the process $z_t$ (see e.g., Brockwell and Davis (1991, p.418)). This can be seen by observing that in this case the eigenvalue in (9) can be written as $\rho\left(L_{np}\left(B_1 \otimes B_1\right) D_{np}\right) = \rho\left(B_1\right)^2 < 1$ (see results 9.5.4(2) and 9.5.5.(1)(a) in Lütkepohl (1996)). Choosing $m = 1$ and $f_1(x) = 0$ shows that a similar result is obtained in the special case of the BEKK model (6) (cf. Proposition 2.7 and the subsequent discussion in Engle and Kroner (1995)). The previous conditions employed by Masry and Tjøstheim (1995), Lu (1998) and Chen and Chen (2001, Corollary 4.2), in models similar to ours with $m = 1$, are different in this respect. For instance, when applied to a linear homoskedastic vector autoregressive model, they are considerably more restrictive than (9). A further advantage of Theorem 1 over these previous results is that it also applies to mixture models.

Thus, (9) has some advantages over its previous counterparts. This point can even be strengthened by considering the model

$$z_t = \sum_{s=1}^{m} \mathbf{1}\left(\eta_t = s\right) \Big( \sum_{j=1}^{p} B_{sj} z_{t-j} + \Big( \Sigma_s + \sum_{i=1}^{k} K_{si} Z_{t-1} Z_{t-1}' K_{si}' \Big)^{\frac{1}{2}} \varepsilon_t^{(s)} \Big), \tag{12}$$

where $\Sigma_s$ $(s = 1, \ldots, m)$ are positive semidefinite $n \times n$ matrices and $\Sigma_{s_0}$ is positive definite with $s_0$ the same as in Assumption 1. The remaining notation is as before. This model is clearly a special case of our general model and subsumes conventional linear autoregressive models with ARCH errors and their mixture extensions. For this model we can prove the following theorem where (9) assumes that the matrices $B_s$ and $K_{s0}$ are obtained from (12).

**Theorem 2.** *Suppose the process* $z_t$ $(t = 1, 2, \ldots)$ *is generated by* (12) *where* $\varepsilon_t^{(s)}$, $s = 1, \ldots, m$, *satisfy Assumption 1(i). Then the following results hold for the Markov chain* $Z_t = \left[z_t' \cdots z_{t-p+1}'\right]'$*:*

(i) *if* (9) *holds,* $Z_t$ *is* $V$*-geometrically ergodic with* $V(x) = 1 + \|x\|^2$*;*

(ii) *if* $Z_t$ *has a stationary distribution with finite second moments,* (9) *holds.*

Thus, given Assumption 1(i), (9) is necessary and sufficient for the $(1 + \|x\|^2)$-geometric ergodicity of the Markov chain associated with $z_t$ in (12). It is worth noting, however, that in the case of the general model (2), the conditions used in Theorem 1 are far from necessary for geometric ergodicity and existence of second moments of the stationary distribution, as examples on first order threshold autoregressive models in Chan, Petrucelli, Tong and Woolford (1985) show. Further examples on this, confined to geometric ergodicity, can be obtained from Battacharya and Lee (1995).

Eigenvalue conditions similar to ours have previously been employed by Nicholls and Quinn (1982), and Pham (1986) although, due to differences in models, the precise forms of these conditions and obtained results vary from case to case and, as already mentioned, Pham (1986) formulates his eigenvalue condition in an alternative way (see (10)). In Theorem 2.5, Nicholls and Quinn (1982) give necessary and sufficient conditions for the 'stability' of a random coefficient autoregressive model with the term 'stability' referring to nonvanishing impact of initial values on the first and second conditional moments of the process. Pham (1986) studies geometric ergodicity of a generalized random coefficient autoregressive model. His Lemma 3 is similar to part (ii) of Theorem 2 and his Theorem (on p.295) shows that the employed eigenvalue condition (with some other assumptions) is sufficient for geometric ergodicity and existence of second moments (see also Feigin and Tweedie (1985)).

The sufficiency part of Theorem 2 extends Theorem 1 of Hansen and Rahbek (1998), where a similar result is obtained for a multivariate first order ARCH model with Gaussian errors. Theorem 2 also extends the results of Wong and Li (2001), who obtained necessary and sufficient conditions for the second order stationarity of a linear real valued mixture autoregressive model with ARCH errors. Similar results were previously obtained by Le, Martin and Raftery (1996) and Wong and Li (2000) for models without an ARCH term. All these previous results on mixture models were explicitly presented in simple special cases.

Although the result of Theorem 2 is of interest in its own right, it is also useful because it readily shows that previous alternatives to our eigenvalue condition (9) are generally inferior. Consider a univariate special case of our general model without any mixtures so that, in (2), $n = m = 1$ and $x_j$, a typical component of the vector $x$, is scalar. For convenience, take $f_1 = f$ and $H_1 = h$, and rewrite (4) and (5) as

$$f(x) = \sum_{j=1}^{p} b_j x_j + o(\|x\|) \quad \text{as} \quad \|x\| \to \infty, \tag{13}$$

$$h(x) = \sum_{j=1}^{p} \phi_j^2 x_j^2 + o\left(\|x\|^2\right) \quad \text{as} \quad \|x\| \to \infty, \tag{14}$$

respectively. In this special case, (9) is obtained as follows. First, set $m = 1$ and $\pi_1 = 1$ and define the matrix $B_1$ $(p \times p)$ by using (7) with $n = 1$ and $B_{1j} = b_j$ $(j = 1, \ldots, p)$. Then choose $k = p$ in the definition of the matrix $K_{10}$ and define the matrix $K_{1i0}$ $(p \times p)$ in (8) by setting $K_{1i} = [0 \cdots 0 \phi_i 0 \cdots 0]$ $(1 \times p)$ where $\phi_i$ is in the $i$th position. Thus the matrix $K_{1i0}$ $(p^2 \times p^2)$ has zero elements except for $\phi_i^2$ in column $(i-1)p + i$ of the first row $(i = 1, \ldots, p)$. The matrices $L_{np}$ and $D_{np}$ are defined as explained at the end of the introduction or, if desired, they can be suppressed without changing (9) (see the discussion at the end of Section 2).

The preceding special case of our general model is considered in Theorem 1 of Lu (1998), where it is shown that a sufficient condition for geometric ergodicity and existence of second moments of the stationary distribution is

$$\left(\sum_{j=1}^{p} |b_j|\right)^2 + \sum_{j=1}^{p} \phi_j^2 < 1. \tag{15}$$

This condition is different from (9). Since arguments used in the proof of Theorem 1 also apply in the context of Lu's (1998) Theorem 1, it is straightforward to show that (15) is even sufficient for $(1 + \|x\|^2)$-geometric ergodicity. Thus, since both Lu's (1998) result and our Theorem 1 also apply to (12) with $n = m = 1$, it follows from Theorem 2 that our eigenvalue condition is implied by (15). However, in general these conditions are not equivalent, for it is easy to find examples in which (9) holds but (15) fails. For instance, suppose $p = 2$ and $b_1 = 0.8$, $b_2 = -0.7$, $\phi_1^2 = 0.2$, and $\phi_2^2 = 0.15$. Then the left hand side of (15) is 2.6 whereas the left hand side of (9) is 0.965. However, it is worth noting that, especially for large values of $p$, (15) is considerably easier to apply in practice than (9), and the same is true for the multivariate analog of (15) given by Lu and Jiang (2001). In practice, these previous conditions can thus be used as simple first checks to complement our computationally more complicated (9).

In Theorem 3 and Remark 4.1 of Lu (1998) it is (essentially) shown that replacing (9) by (15) yields a necessary and sufficient condition for the $(1+\|x\|^2)$-

geometric ergodicity of the Markov chain associated with the model

$$z_t = b_i z_{t-i} + \Big(\sigma^2 + \sum_{j=1}^{p} \phi_j^2 z_{t-j}\Big)^{\frac{1}{2}} \varepsilon_t, \tag{16}$$

where $1 \le i \le p$, $\sigma^2 > 0$, and $\varepsilon_t \sim i.i.d.(0,1)$. Thus, for this model (15) is equivalent to our eigenvalue condition.

Based on the above mentioned result on model (16), Lu (1998) made the conjecture that, in the general univariate model defined by (2) with $n = m = 1$, (15) could be replaced by the condition $\rho(B)^2 + \sum_{j=1}^{p} \phi_j^2 < 1$, where $B$ is an analog of the companion matrix $B_s$ defined in terms of $b_1, \ldots, b_p$. As with (9), this condition would subsume several known results on geometric ergodicity. Our Theorem 2 can be used to shed light on this conjecture, which has also been discussed by Liebscher (2005). Consider model (12) with $n = m = 1$ and $p = 2$, and choose $b_1 = 1.4$, $b_2 = -0.7$, and $\phi_1^2 = \phi_2^2 = 0.1$, where the notation is as in (13) and (14). Then, the left hand side of (9) is 1.056 which, by Theorem 2, means that the Markov chain associated with the considered model cannot have a stationary distribution with finite second moments. However, if Lu's (1998) conjecture were true the opposite would hold because $\rho(B)^2 + \phi_1^2 + \phi_2^2 = 0.7 + 0.1 + 0.1 = 0.9$. Note that this example only disproves the part of Lu's (1998) conjecture concerning existence of a stationary distribution with finite second moments. It is still possible that Lu's conjectured condition can be sufficient for geometric ergodicity. Indeed, in a recent paper, Ling (2006) obtains necessary and sufficient conditions for geometric ergodicity in a model that is a special case of (12) with $n = m = 1$. He presents an example in which the region in the parameter space guaranteeing geometric ergodicity and finiteness of second moments of the stationary distribution is strictly contained in the region guaranteeing geometric ergodicity.

Our result improves not only on Lu (1998), but also on Corollary 4.2 of Chen and Chen (2001), where (15) is used in essentially the same model as Lu's. A model similar to (2) with $n = m = 1$ is considered in Theorem 2 of Chen and Chen (2000) where, however, a slightly different formulation of (13) is used and the corresponding analog of (14) is formulated for $h^{1/2}$ instead of $h$. This latter difference makes a general comparison of our result with that of Chen and Chen (2000) somewhat difficult, but when $\phi_1 = \cdots = \phi_p = 0$ comparisons are straightforward. Then Theorem 2 of Chen and Chen (2000) applies in our setting and (15) is relevant. For instance, when $p = 2$, $b_1 = 0.8$, and $b_2 = -0.7$, the left hand side of (15) is 2.25 whereas the left hand side of our eigenvalue condition (9) is 0.7. Thus, our Theorem 1 implies $(1+\|x\|^2)$-geometric ergodicity, whereas Theorem 2 of Chen and Chen (2000) is inconclusive. It may be noted that the same is true even for Theorem 1 of Chen and Chen (2000), which only assumes existence of the expectation of $\varepsilon_t$.

In addition to the papers discussed above, results on geometric ergodicity of nonlinear autoregressions with an ARCH term are also obtained in Cline and Pu (1999) and Liebscher (2005). Neither of these papers allows for mixtures and the former assumes an ARCH term which, when specialized to our case, satisfies (5) only with $K_{11} = \cdots = K_{1k} = 0$ (assuming $m = 1$). This is also the case in some of Liebscher's (2005) results. An exception is his Theorem 4 which, in a special case, provides a result rather close to Lu's (1998) conjecture. Because this theorem involves an auxiliary matrix that needs to be chosen, other interesting special cases are more difficult to find. This feature also hampers the application of this result in practice.

In addition to Theorems 1 and 2, useful mixing results can also be obtained. We use the concept of $\beta$-mixing, also known as absolute regularity (for a definition of $\beta$-mixing and its relation to other mixing concepts, see e.g., Doukhan (1994)).

**Theorem 3.** *Suppose the conditions of Theorem* 1 *or* 2 *hold and that the distribution of the initial value* $Z_0$ *has finite second moments. Then the process* $z_t$ *is* $\beta$*-mixing with geometrically decaying mixing numbers.*

Theorem 3 is useful because it makes it possible to apply conventional limit theorems needed in the development of asymptotic estimation and testing procedures. Although strong mixing is often sufficient for this purpose, there are cases where the stronger concept of $\beta$-mixing is needed. Some results in the theory of empirical processes provide examples on this (see Hansen (1996, 2000)).

## 4. Conclusion

This paper provides necessary and sufficient conditions for $(1 + \|x\|^2)$-geometric ergodicity of mixtures of linear vector autoregressive models with a conventional ARCH term. Stationarity, existence of second moments of the stationary distribution, and $\beta$-mixing are consequently obtained. Similar sufficient conditions are also provided for more general models in which the mixture components exhibit limited forms of nonlinearity. When specialized to the corresponding non-mixture case these sufficient conditions improve on their previous counterparts. The results of the paper are useful in the development of asymptotic estimation and testing theory for the considered models.

## Acknowledgements

Part of this research was carried out while the author was visiting the Humboldt University in Berlin. Financial support by the Yrjö Jahnsson Foundation and the Alexander von Humboldt Foundation under a Humboldt research award is gratefully acknowledged. The first version of this paper was completed in June 2003 and entitled "Stability of mixtures of nonlinear vector autoregressions with autoregressive conditional heteroskedasticity". The author thanks two anonymous referees and Zudi Lu for useful comments on earlier versions of this paper.

## Appendix

Before proving Theorem 1 we state the following auxiliary lemma.

**Lemma 4.** *If condition* (9) *holds, the sum*

$$vec\,(U_t) = \sum_{j=0}^{t} \Big( \sum_{s=1}^{m} \pi_s \left( B_s' \otimes B_s' + K_{s0}' \right) \Big)^j vec\,(S) \tag{17}$$

*converges for any symmetric $np \times np$ matrix $S$ and the limit can be expressed as vec$(U)$ where $U$ is a symmetric matrix which is positive definite* (*semidefinite*) *if $S$ is positive definite* (*semidefinite*). *The same conclusions hold with the matrices $B_s'$ and $K_{s0}'$ replaced by $B_s$ and $K_{s0}$, respectively.*

**Proof.** The convergence statement can be proved by using arguments similar to those in the proofs of Theorems 2.4 and 2.5 of Nicholls and Quinn (1982). Next note that (17) implies

$$\begin{aligned}
\mathrm{vec}\,(U_t) &= \sum_{s=1}^{m} \pi_s \left( B_s' \otimes B_s' + K_{s0}' \right) \mathrm{vec}\,(U_{t-1}) + \mathrm{vec}\,(S) \\
&= \sum_{s=1}^{m} \pi_s \mathrm{vec}\left( B_s' U_{t-1} B_s \right) + \sum_{s=1}^{m} \pi_s \sum_{i=1}^{k} \mathrm{vec}\left( K_{si0}' U_{t-1} K_{si0} \right) + \mathrm{vec}\,(S)\,, 
\end{aligned} \tag{18}$$

where $U_0 = 0$ and the latter equation is based on the definition of the matrix $K_{s0}$ and well-known properties of the vec operator (see Lütkepohl (1996, Chap.7.2)). Since $U_1 = S$ is symmetric, it follows by induction that each $U_t$ is symmetric as is the limit of (17). If $S$ is positive definite (semidefinite) it follows similarly that each $U_t$ is positive definite (semidefinite) and, since $\lambda_{\min}(U_t) \geq \lambda_{\min}(S) > 0$ $(\geq 0)$, the limit of (17) is positive definite (semidefinite). The proof of the final statement is similar because in (9) the transpose signs can be suppressed (see the discussion at the end of Section 2).

**Proof of Theorem 1.** The idea of the proof is to apply Theorem 15.0.1 of Meyn and Tweedie (1993). First we demonstrate that the Markov chain $Z_t$ is an irreducible and aperiodic $T$-chain. To this end, conclude from the definition of the model (cf.(3)) that the transition probability kernel of $Z_t$, denoted by $P_Z(x, A)$, satisfies

$$P_Z(x, A) = \sum_{s=1}^{m} \pi_s P_s(x, A) \geq \pi_{s_0} P_{s_0}(x, A),$$

where $P_s(x, A)$ signifies the transition probability kernel of the Markov chain $Z_t^{(s)} = [z_t^{(s)\prime} \cdots z_{t-p+1}^{(s)\prime}]'$ associated with the model

$$z_t^{(s)} = f_s\left(z_{t-1}^{(s)}, \ldots, z_{t-p}^{(s)}\right) + H_s\left(z_{t-1}^{(s)}, \ldots, z_{t-p}^{(s)}\right)^{\frac{1}{2}} \varepsilon_t^{(s)}.$$

Iterating the preceding inequality it can be seen that the corresponding $t$-step transition probabilities satisfy $P_Z^t(x, A) \geq \pi_{s_0}^t P_{s_0}^t(x, A)$ for all $t \geq 1$. Assumption 1 and Theorem 2.2(ii) of Cline and Pu (1998) imply that the Markov chain $Z_t^{(s_0)}$ is an irreducible and aperiodic $T$-chain (see also Example 2.1 of the same reference). From this and the inequality $P_Z^t(x, A) \geq \pi_{s_0}^t P_{s_0}^t(x, A)$, $t \geq 1$, it then readily follows that the same is true for $Z_t$ (for irreducibility and aperiodicity, see Proposition 4.2.1(ii) of Meyn and Tweedie (1993) and Proposition A1.1 of Chan (1990), respectively, and for the T-chain property, see the definition in Meyn and Tweedie (1993, p.127). This also implies that all compact sets of $\mathbb{R}^{np}$ are small (see Theorems 5.5.7 and 6.2.5 of Meyn and Tweedie (1993)). Given these facts, it suffices to establish condition (15.3) of Meyn and Tweedie (1993, p.355).

We proceed by making use of ideas employed in the proof of Theorem 3 of Feigin and Tweedie (1985). First, let $W$ $(np \times np)$ be an auxiliary positive definite matrix, and define the matrix $U$ $(np \times np)$ by the equation

$$\operatorname{vec}(U) = \sum_{j=0}^{\infty} \left( \sum_{s=1}^{m} \pi_s \left(B_s' \otimes B_s' + K_{s0}'\right) \right)^j \operatorname{vec}(W), \tag{19}$$

where $B_s$ and $K_{s0}$ are as in Assumption 1. By Lemma 4 the right hand side is well defined and the matrix $U$ is positive definite. Next, define the real-valued function $q(x) = 1 + x'Ux$. Since the matrix $U$ is positive definite we clearly have $q(x) \geq 1$ for all $x \in \mathbb{R}^{np}$. We also define the compact set $C = \{x \in \mathbb{R}^{np} : x'Ux \leq M_1\}$, where $M_1$ is a constant to be determined later.

We show that there exist constants $0 < \rho < 1$ and $0 < M_2 < \infty$ such that

$$E\left(q\left(Z_t\right) | Z_{t-1} = x\right) < \rho q(x), \quad x \notin C, \tag{20}$$

$$E\left(q\left(Z_t\right) | Z_{t-1} = x\right) < M_2, \qquad x \in C. \tag{21}$$

It is easy to see that then condition (15.3) of Meyn and Tweedie (1993, p.355) holds with the function $V(\cdot)$ therein given by $V(\cdot) = q(\cdot)$. This implies that the Markov chain $Z_t$ is $q$-geometrically ergodic. Thus, since we clearly have $1 + \|x\|^2 \leq Mq(x)$ for some $1 \leq M < \infty$, and since conditions (20) and (21) hold with $q(\cdot)$ replaced by $Mq(\cdot)$, the desired $(1 + \|x\|^2)$-geometric ergodicity follows (see (11)).

To establish conditions (20) and (21), first write $Z_t = F(Z_{t-1}, \eta_t, \varepsilon_t)$ as

$$Z_t = \sum_{s=1}^{m} \mathbf{1}(\eta_t = s)\left(B_s Z_{t-1} + R_s\left(Z_{t-1}, \varepsilon_t^{(s)}\right)\right), \tag{22}$$

where

$$\begin{aligned} R_s\left(Z_{t-1}, \varepsilon_t^{(s)}\right) &= \begin{bmatrix} f_s(Z_{t-1}) - \sum_{j=1}^{p} B_{sj} z_{t-j} \\ 0 \end{bmatrix} + \begin{bmatrix} H_s(Z_{t-1})^{\frac{1}{2}} \varepsilon_t^{(s)} \\ 0 \end{bmatrix} \\ &\equiv R_{s1}(Z_{t-1}) + R_{s2}\left(Z_{t-1}, \varepsilon_t^{(s)}\right). \end{aligned}$$

Using the independence of $Z_{t-1}$ on $\varepsilon_t^{(s)}$ and $\eta_t$, and the independence of $\varepsilon_t^{(s)}$ and $\eta_t$, we find from (22) that

$$E(q(Z_t)|Z_{t-1} = x) = 1 + \sum_{s=1}^{m} \pi_s E\left(\left(B_s x + R_s(x, \varepsilon_t^{(s)})\right)' U\left(B_s x + R_s(x, \varepsilon_t^{(s)})\right)\right).$$

It follows from the definitions that $E(R_{s2}(x, \varepsilon_t^{(s)})) = 0$. Thus, simple calculations show that the preceding equation can be written as

$$\begin{aligned} E(q(Z_t)|Z_{t-1} = x) = 1 + \sum_{s=1}^{m} \pi_s \Big(& x'B_s'UB_s x + 2x'B_s'UR_{s1}(x) + R_{s1}(x)'UR_{s1}(x) \\ & + \operatorname{tr}\left(\overline{H}_s(x) U\right)\Big), \end{aligned} \tag{23}$$

where $\overline{H}_s(x) = \operatorname{diag}[H_s(x) \quad 0]$. Writing

$$R_{s3}(x) = \operatorname{tr}\Big(\Big(\overline{H}_s(x) - \sum_{i=1}^{k} K_{si0} x x' K_{si0}'\Big) U\Big),$$

we can express (23) as

$$E(q(Z_t)|Z_{t-1} = x) = 1 + \sum_{s=1}^{m} \pi_s \Big(x'B_s'UB_s x + \operatorname{tr}\Big(\sum_{i=1}^{k} K_{si0} x x' K_{si0}' U\Big) + R_{s4}(x)\Big),$$

where $R_{s4}(x) = 2x'B_s'UR_{s1}(x) + R_{s1}(x)'UR_{s1}(x) + R_{s3}(x)$. In parentheses on the right hand side we can use the definitions and well-known properties of the vec operator (see Lütkepohl (1996, Chap.7.2)) to obtain $x'B_s'UB_s x = (x' \otimes x')(B_s' \otimes B_s')\operatorname{vec}(U)$ and $\operatorname{tr}(\sum_{i=1}^{k} K_{si0} x x' K_{si0}' U) = (x' \otimes x') K_{s0}' \operatorname{vec}(U)$. Thus, setting $R_4(x) = \sum_{s=1}^{m} \pi_s R_{s4}(x)$ we can conclude that

$$\begin{aligned} E(q(Z_t)|Z_{t-1} = x) &= 1 + \left(x' \otimes x'\right) \sum_{s=1}^{m} \pi_s (B_s' \otimes B_s' + K_{s0}')\operatorname{vec}(U) + R_4(x) \\ &= 1 + \left(x' \otimes x'\right)(\operatorname{vec}(U) - \operatorname{vec}(W)) + R_4(x) \\ &= q(x) - x'Wx + R_4(x). \end{aligned}$$

Here the second equality is based on the definition of the matrix $U$ in (19) and the third one on the definition of the function $q$. Thus, we have shown that

$$E\left(q\left(Z_t\right)|Z_{t-1}=x\right)=q\left(x\right)\left[1-\frac{x'Wx-R_4\left(x\right)}{q\left(x\right)}\right]. \tag{24}$$

Now suppose that $x\notin C$ so that $x'Ux>M_1$. Then, assuming $M_1>1$,

$$q\left(x\right)\leq\frac{x'Ux}{M_1}+x'Ux=\frac{M_1+1}{M_1}x'Ux\leq 2x'Ux,$$

and therefore

$$\begin{aligned}\frac{x'Wx-R_4\left(x\right)}{q\left(x\right)}&\geq\frac{x'x\lambda_{\min}\left(W\right)}{2x'Ux}-\frac{R_4\left(x\right)}{1+x'Ux}\\&\geq\frac{\lambda_{\min}\left(W\right)}{2\lambda_{\max}\left(U\right)}-\frac{\left|R_4\left(x\right)\right|}{\left\|x\right\|^2\lambda_{\min}\left(U\right)}.\end{aligned}$$

From the assumed asymptotic behavior of the functions $f_s$ and $H_s$ and the definition of $R_4\left(x\right)$, it follows that $\left|R_4\left(x\right)\right|=o(\|x\|^2)$ as $\|x\|\to\infty$. Thus, there exists a value of $M_1>1$ such that for $\|x\|>M_1$ the last expression above can be bounded from below by a positive constant $\epsilon<1$. Setting $\rho=1-\epsilon$ we can see from (24) that (20) holds.

As for (21), since the functions $f_s$ and $H_s$ are locally bounded, $R_{s1}\left(x\right)$ and $\overline{H}_s\left(x\right)$ $(s=1,\ldots,m)$ are bounded for $x\in C$, and we can conclude from (23) that (21) also holds. Thus, the proof is complete.

**Proof of Theorem 2.** Part (i) follows directly from Theorem 1, so we only need prove part (ii). The process $Z_t$ is now generated by

$$Z_t=\sum_{s=1}^{m}\mathbf{1}\left(\eta_t=s\right)\left(B_sZ_{t-1}+\begin{bmatrix}H_s\left(Z_{t-1}\right)^{\frac{1}{2}}\\0\end{bmatrix}\varepsilon_t^{(s)}\right),$$

where $H_s\left(Z_{t-1}\right)=\Sigma_s+\sum_{i=1}^{k}K_{si}Z_{t-1}Z_{t-1}'K_{si}'$. Straightforward calculations show that

$$Z_tZ_t'=\sum_{s=1}^{m}\mathbf{1}\left(\eta_t=s\right)\left(B_sZ_{t-1}Z_{t-1}'B_s'+\overline{H}_s\left(Z_{t-1}\right)+R_{s5}\left(Z_{t-1},\varepsilon_t^{(s)}\right)\right), \tag{25}$$

where $\overline{H}_s\left(x\right)=\mathrm{diag}[H_s\left(x\right)\quad 0]$ as before, and

$$\begin{aligned}R_{s5}\left(Z_{t-1},\varepsilon_t^{(s)}\right)&=B_sZ_{t-1}\varepsilon_t^{(s)\prime}\begin{bmatrix}H_s\left(Z_{t-1}\right)^{\frac{1}{2}}\\0\end{bmatrix}'+\begin{bmatrix}H_s\left(Z_{t-1}\right)^{\frac{1}{2}}\\0\end{bmatrix}\varepsilon_t^{(s)}Z_{t-1}'B_s'\\&\quad+\begin{bmatrix}H_s\left(Z_{t-1}\right)^{\frac{1}{2}}\\0\end{bmatrix}\left(\varepsilon_t^{(s)}\varepsilon_t^{(s)\prime}-I_n\right)\begin{bmatrix}H_s\left(Z_{t-1}\right)^{\frac{1}{2}}\\0\end{bmatrix}'.\end{aligned}$$

Let $\overline{\Sigma}_s = \text{diag}[\Sigma_s \;\; 0]$, and define the matrices $K_{si0}$ and $K_{s0}$ from $K_{si}$ in the same way as before. By the definitions, we can then rewrite (25) as

$$Z_t Z_t' = \sum_{s=1}^{m} \mathbf{1}\,(\eta_t = s) \Big( \overline{\Sigma}_s + B_s Z_{t-1} Z_{t-1}' B_s' + \sum_{i=1}^{k} K_{si0} Z_{t-1} Z_{t-1}' K_{si0}' + R_{s5}\big(Z_{t-1}, \varepsilon_t^{(s)}\big) \Big).$$

Now assume that $Z_t$ is second order stationary. From the definition of $R_{s5}(Z_{t-1}, \varepsilon_t^{(s)})$ and the independence assumptions imposed on $z_t$, $\varepsilon_t^{(s)}$ and $\eta_t$, it straightforwardly follows that $E(\mathbf{1}\,(\eta_t = s)\, R_{s5}(Z_{t-1}, \varepsilon_t^{(s)})) = 0$. Thus, taking expectations from both sides of the preceding equation yields

$$\Gamma = \sum_{s=1}^{m} \pi_s \overline{\Sigma}_s + \sum_{s=1}^{m} \pi_s B_s \Gamma B_s' + \sum_{s=1}^{m} \pi_s \sum_{i=1}^{k} K_{si0} \Gamma K_{si0}', \tag{26}$$

where $\Gamma = E\,(Z_t Z_t')$ is independent of $t$.

We show next that the (positive semidefinite) matrix $\Gamma$ is positive definite. Suppose this is not the case and let $a = \left[a_1' \cdots a_p'\right]'$ be a nonzero $np \times 1$ vector such that $a'\Gamma a = 0$ (each component $a_i$ is of dimension $n \times 1$). From (26) it then follows that $\sum_{s=1}^{m} \pi_s a' \overline{\Sigma}_s a = 0$. By the definition of $\overline{\Sigma}_s$ and the assumption that $\Sigma_{s_0}$ is positive definite, this entails $a_1 = 0$ and, if $p = 1$, we have the contradiction $a = 0$. Thus, assume $p > 1$. By the preceding discussion the vector $a$ is of the form $a = \left[0' \;\; a_2' \cdots a_p'\right]'$ and, hence, $a' B_s = \left[a_2' \cdots a_p' 0'\right]' \equiv b_1$ (see (7)) and $a' K_{si0} = 0$ (see (8)) for all $s$. Since we also have $a' \overline{\Sigma}_s = 0$ for all $s$ we find from (26) that $b_1' \Gamma b_1 = 0$. As above this implies that $a_2 = 0$ so that, if $p = 2$, we again get the contradiction $a = 0$. If $p > 2$ we have $a = \left[0' \;\; 0' a_3' \cdots a_p'\right]'$ and, continuing as above, it can be seen that $a_3$ and the possible remaining vectors $a_4, \ldots, a_p$ are all zero. Thus we get the contradiction $a = 0$ and, hence, the matrix $\Gamma$ is positive definite. In the same way it can be seen that any positive semidefinite matrix satisfying (26) must be positive definite.

Because (26) admits a positive definite solution the sum

$$\sum_{j=0}^{t} \Big( \sum_{s=1}^{m} \pi_s L_{np} \left( B_s' \otimes B_s' + K_{s0}' \right) D_{np} \Big)^{j} \text{vech}\Big( \sum_{s=1}^{m} \pi_s \overline{\Sigma}_s \Big) \tag{27}$$

converges as $t \to \infty$ (see Pham (1986, pp.294-296) or Pham (1985, Theorem 4.1)). Denote the limit by $\text{vech}(\Gamma_*)$. Using the fact that the $\overline{\Sigma}_s$ are positive semidefinite, and arguments similar to those in the proof of Lemma 4, it can be seen that the matrix $\Gamma_*$ is positive semidefinite and satisfies (26). By the above discussion, $\Gamma_*$ is thus positive definite. That (9) holds follows from the facts that the sum at (27) converges and that the limit $\Gamma_*$ is positive definite. A justification of this can be obtained from Lemma 3 of Pham (1986) by observing that the formulation therein only seems different from the formulation used here. Indeed, in the proof of Lemma 3, Pham (1986) uses a formulation which is similar to that in (27) except for a slight (and inessential) difference in the definition of the vech operator. In our notation, the proof makes use of the Jordan canonical form of the matrix $\sum_{s=1}^{m} \pi_s L_{np} \left(B_s' \otimes B_s' + K_{s0}'\right) D_{np}$, but the precise form of this matrix is only needed to ensure that the image of the mapping $\mathrm{vech}(S) \to \sum_{s=1}^{m} \pi_s L_{np} \left(B_s' \otimes B_s' + K_{s0}'\right) D_{np}\mathrm{vech}(S)$ with $S$ $(np \times np)$ positive semidefinite and vech defined as in Pham (1986) corresponds to a positive semidefinite matrix. This, however, can be seen in the same way as in the case of the matrix $\Gamma_*$ or the matrix $U_t$ in the proof of Lemma 4, so further details will not be given. This completes the proof.

**Proof of Theorem 3.** The proof is obtained from Proposition 4 of Liebscher (2005), because under the stated assumptions the Markov chain $Z_t$ is $(1 + \|x\|^2)$-geometrically ergodic and hence $Q$-geometrically ergodic in Liebscher's (2005) sense with $Q(x) = 1 + \|x\|^2$.

## References


Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. Springer-Verlag, New York.

Battacharya, R. N. and Lee, C. (1995). Ergodicity of nonlinear first order autoregressive models. *J. Theoret. Probab.* **8**, 207-219.

Chan, K. S. (1990). Deterministic stability, stochastic stability, and ergodicity. Appendix 1 of H. Tong (1990). *Non-linear Time Series: A Dynamical System Approach*, Clarendon Press, Oxford.

Chan, K. S., Petrucelli, J., Tong, H. and Woolford, S. W. (1985). A multiple threshold AR(1) model. *J. Appl. Probab.* **22**, 267-279.

Chan, K. S. and Tong, H. (1985). On the use of the deterministic Lyapunov function for the ergodicity of stochastic difference equations. *Adv. Appl. Probab.* **17**, 666-678.

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

Chen, M. and Chen, G. (2000). Geometric ergodicity of nonlinear autoregressive models with changing conditional variances. *Canad. J. Statist.* **28**, 605-613.

Chen, G. and Chen, M. (2001). On a class of nonlinear AR(p) models with nonlinear ARCH errors. *Austral. N. Z. J. Statist.* **43**, 445-454.

Cline, D. B. H. and Pu, H. H. (1998). Verifying irreducibility and continuity of a nonlinear time series. *Statist. Probab. Lett.* **40**, 139-148.

Cline, D. B. H. and Pu, H. H. (1999). Geometric ergodicity of nonlinear time series. *Statist. Sinica* **9**, 1103-1118.

Doukhan, P. (1994). *Mixing, Properties and Examples.* Springer-Verlag, New York.

Engle, R. F. and Kroner, K. F. (1995). Multivariate simultaneous generalized GARCH. *Econom. Theory* **11**, 122-150.

Feigin, P. D. and Tweedie, R. L. (1985). Random coefficient autoregressive processes: A Markov chain analysis of stationarity and finiteness of moments. *J. Time Ser. Anal.* **6**, 1-14.

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

Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica* **68**, 575-603.

Hansen, E. and Rahbek, A. (1998). Stationarity and asymptotics of multivariate ARCH time series with an application to robustness of cointegration analysis. Preprint No. 12, 1998. Department of Statistics. University of Copenhagen.

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

Liebscher, E. (2005). Towards a unified approach for proving geometric ergodicity and mixing properties of nonlinear autoregressive processes. *J. Time Ser. Anal.* **26**, 669-689.

Ling, S. (2006). A double AR(p) model: structure and estimation. *Statist Sinica*, accepted.

Lu, Z. (1998). On the geometric ergodicity of a non-linear autoregressive model with an autoregressive conditional heteroskedastic term. *Statist. Sinica* **8**, 1205-1217.

Lu, Z. and Jiang, Z. (2001). $L_1$ geometric ergodicity of a multivariate nonlinear AR model with an ARCH term. *Statist. Probab. Lett.* **51**, 121-130.

Lütkepohl, H. (1996). *Handbook of Matrices*. John Wiley and Sons. Chichester.

Luukkonen, R., Saikkonen, P. and Teräsvirta, T. (1988). Testing linearity against smooth transition autoregressive models. *Biometrika* **75**, 491-499.

Masry, E. and Tjøstheim, D. (1995). Nonparametric estimation and identification of nonlinear ARCH time series. *Econom. Theory* **11**, 258-289.

Meyn, S. P. and Tweedie, R. L. (1993). *Markov Chains and Stochastic Stability.* Springer-Verlag, New York.

Nicholls, D. F. and Quinn, B. G. (1982). *Random Coefficient Autoregressive Models: An Introduction*. Springer-Verlag, New York.

Pham, T. D. (1985). Bilinear Markovian representation and bilinear models. *Stochastic Process. Appl.* **20**, 295-306.

Pham, D. T. (1986). The mixing property of bilinear and generalised random coefficient autoregressive models. *Stochastic Process. Appl.* **23**, 291-300.

Wong, C. S. and Li, W. K. (2000). On a mixture autoregressive model. *J. Roy. Statist. Soc. Ser. B* **62**, 95-115.

Wong, C. S. and Li, W. K. (2001). On a mixture autoregressive conditional heteroscedastic model. *J. Amer. Statist. Assoc.* **96**, 982-995.

Department of Mathematics and Statistics, P.O. Box 68 (Gustaf Hällströmin katu 2b), FIN-00014 University of Helsinki, Finland.

E-mail: pentti.saikkonen@Helsinki.Fi

(Received August 2005; accepted June 2006)