Saikkonen, Pentti

Working Paper

Stability results for nonlinear vector autoregressions with an application to a nonlinear error correction model

SFB 373 Discussion Paper, No. 2001,93

Provided in Cooperation with:
Collaborative Research Center 373: Quantification and Simulation of Economic Processes, Humboldt University Berlin

Suggested Citation: Saikkonen, Pentti (2001) : Stability results for nonlinear vector autoregressions with an application to a nonlinear error correction model, SFB 373 Discussion Paper, No. 2001,93, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, Berlin,
https://nbn-resolving.de/urn:nbn:de:kobv:11-10051103

This Version is available at:
https://hdl.handle.net/10419/62736

Stability Results for Nonlinear Vector Autoregressions with an Application to a Nonlinear Error Correction Model

by

Pentti Saikkonen *

Department of Statistics, P.O. Box 54 (Unioninkatu 37)

FIN-00014 University of Helsinki, Finland

E-mail: Pentti.Saikkonen@Helsinki.Fi

October, 2001

*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.

Abstract

This paper improves previous sufficient conditions for stationarity obtained in the context of a general nonlinear vector autoregressive model with nonlinear autoregressive conditional heteroskedasticity. The results are proved by using the stability theory developed for Markov chains. Stationarity, existence of second moments of the stationary distribution, and useful mixing results are obtained by establishing appropriate versions of geometric ergodicity. The results are applied to a nonlinear error correction model to obtain an analog of Granger's representation theorem.

Keywords: Geometric ergodicity, Markov chain, Mixing, Nonlinear error correction model, Nonlinear vector autoregressive process, Stability

1 Introduction

Most of the traditional time series analysis assumes that the underlying data generation process is stationarity or asymptotically stationary in the sense that the non-stationarity 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, as far as nonlinear time series models are concerned, the situation is much more difficult. For such models, the most convenient approach for studying stationarity is apparently 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 asymptotic stationarity or, as we shall say, 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 nonlinear time series models. A good reference of the whole approach and its applications till the early nineties can be found in the comprehensive book of Meyn and Tweedie (1993).

Most of the stability results referred to above were obtained for specific nonlinear time series models such as threshold autoregressive models, random coefficient autoregressive models, and bilinear models. In subsequent work, Masry and Tjøstheim (1995), Lu (1998), and Lu and Jiang (2001) obtained similar results for general nonlinear autoregressive processes with nonlinear autoregressive conditional heteroskedasticity (ARCH). A drawback of these results is, however, that the given sufficient conditions for stability can be very far from necessary.

In this paper we follow Masry and Tjøstheim (1995), Lu (1998), and Lu and Jiang (2001) and consider a general nonlinear vector autoregressive (VAR) process with a nonlinear ARCH component. Of our two major contributions the first one is that the stability results obtained by these previous authors will be considerably improved.

Indeed, our sufficient condition for stability is even better than a condition Lu (1998) only conjectured to hold in a corresponding univariate context. In the special case of a linear VAR model with a conventional ARCH component necessary and sufficient conditions for stability will be provided. These results can actually be obtained in a fairly simple manner by adopting a method of proof already employed by Feigin and Tweedie (1985).

The second major contribution of this paper is that we show how stability results obtained for nonlinear VAR models can be applied to nonlinear error correction models similar to those discussed by Granger and Swanson (1996), Granger and Haldrup (1997), and Granger (2001). The main result obtained in this context can be seen as an extension of Granger's representation theorem given by Johansen (1995, p. 49) for linear VAR error correction models. In particular, just like in the previous linear case we provide conditions which ensure that the considered nonlinear VAR process is nonstationary and integrated of order one (or I(1)) but some of its linear combinations are stationary. As recently indicated by de Jong (2001), such results are of great importance in the development of asymptotic estimation and testing theory for nonlinear error correction models.

Results, which at this stage can be readily obtained for nonlinear error correction models, are unfortunately not of as broad applicability as one would hope. Several interesting types of nonlinearity are ruled out including most, though not all, of the threshold autoregressive error correction models and their smooth versions discussed by Balke and Fomby (1997) and van Dijk, Teräsvirta, and Frances (2001). On the other hand, our error correction model allows for a nonlinear ARCH component and the form of nonlinearity is also otherwise slightly more general than permitted in the related recent work of Corradi, Swanson, and White (2000). The error correction model considered by these authors is somewhat different from our model and, as we shall point out, more difficult to handle. A point, potentially related to this difficulty,

is that some of the proofs of Corradi, Swanson, and White (2000) appear to be in flaw and some of their results are not true as stated.

The plan of the paper is as follows. The considered nonlinear VAR model and related stability results are presented in Section 2. In Section 3 these results are applied to a nonlinear VAR error correction model. Section 4 presents conclusions of the paper and a mathematical appendix contains proofs of the theorems.

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 $(k \times k)$ matrix A, the symbol D_k is used for the $\left(k^2 \times \frac{1}{2}k\left(k+1\right)\right)$ duplication matrix defined by $\text{vec}(A) = D_k \text{vech}(A)$ whereas L_k will signify the $\left(\frac{1}{2}k\left(k+1\right) \times k^2\right)$ elimination matrix such that $\text{vech}(A) = L_k \text{vec}(A)$. The largest and smallest eigenvalues of the square matrix A are denoted by $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$, respectively, whereas $\rho(A) = |\lambda_{\max}(A)|$ is the spectral radius of A. Moreover, $\text{rk}(A)$, $\text{tr}(A)$, and $\det(A)$ signify the rank, trace, and determinant of A, respectively. The symbols L and Δ stand for the lag and differencing operators, respectively, so that for a sequence y_t we have $Ly_t = y_{t-1}$ and $\Delta y_t = y_t - y_{t-1}$. If A is a $(k \times l)$ matrix of full column rank $(k > l)$ its orthogonal complement is denoted by $A_\perp$ so that $A_\perp$ is a $(k \times (k-l))$ matrix of full column rank and such that $A'A_\perp = 0$. Finally, $\|\cdot\|$ is used for the Euclidean norm and a sum is defined to be zero if the lower bound of the summation index exceeds the upper bound.

2 Nonlinear Vector Autoregression

Consider the n-dimensional stochastic process z_t $(t = 1, 2, ...)$ generated by the nonlinear VAR process

$$z_t = f\left(z_{t-1}, ..., z_{t-p}\right) + u_t \tag{1}$$

where f: $\mathbb{R}^{np} \to \mathbb{R}^n$ is a (generally) nonlinear function and u_t is a zero mean error term. We allow for the possibility that the errors are conditionally heteroskedastic and assume that

$$u_t = H\left(z_{t-1}, ..., z_{t-p}\right)^{1/2} \varepsilon_t \tag{2}$$

where $\varepsilon_t \sim i.i.d.\left(0, I_n\right)$ is independent of z_s $(s < t)$ and H: $\mathbb{R}^{np} \to \mathbb{R}^{n \times n}$ is a positive definite matrix valued function. Notice that it means no loss of generality to use the same lag length p in equations (1) and (2) because this case can always be achieved by a redefinition of the involved functions (cf. Lu (1998)).

The stability of the process defined by equations (1) and (2) was recently studied by Lu and Jiang (2001) without assuming that the errors ε_t have finite second moments. Assuming the existence of second moments is convenient because from a practical point of view it is mostly not restrictive and needed to develop conventional asymptotic estimation and testing theory. The existence of second moments was assumed by Lu (1998) who obtained stability results for univariate models similar to (1) and (2) with $n = 1$. Lu's (1998) results extended those previously obtained by Masry and Tjøstheim (1995, Lemma 3.1). Our work makes use of ideas similar to those employed by these previous authors.

We shall now discuss assumptions imposed on the above model. Denote $x = \left[x_1' \ ... \ x_p'\right]'$ where $x_i \in \mathbb{R}^n$ $(i = 1, ..., p)$. Following Masry and Tjøstheim (1995) and Lu (1998) we assume that the function f satisfies

$$f\left(x\right) = \sum_{j=1}^{p} B_j x_j + o\left(\|x\|\right) \quad \text{as} \quad \|x\| \to \infty. \tag{3}$$

Here B_j $(j = 1, ..., p)$ are $(n \times n)$ matrices which may all be zeros. Thus, it is assumed that for large $\|x\|$ the behavior of the function f is dominated by a linear function.

As for the function H, we assume that

$$H\left(x\right) = \sum_{i=1}^{k} \sum_{j=1}^{l} K_{ij} x x' K_{ij}' + o\left(\|x\|^2\right) \quad \text{as} \ \|x\| \to \infty, \tag{4}$$

where K_{ij} $(i = 1, ..., k,\ j = 1, .., l)$ are $(n \times np)$ matrices possibly all zero. This implies that for large $\|x\|$ the nonlinearity in the conditional covariance matrix of z_t is dominated by a quadratic function similar to that used in the so called BEKK formulation of the multivariate ARCH model (see Engle and Kroner (1995)). Indeed, if $f(x) = 0$ we have $z_t = u_t$ and a special case of the BEKK model can be written as

$$E(u_t u_t' \mid u_s,\ s < t) = \Phi_0 + \sum_{i=1}^{k} \sum_{j=1}^{p} \Phi_{ij} u_{t-j} u_{t-j}' \Phi_{ij}'. \tag{5}$$

Here Φ_0 and Φ_{ij} $(i = 1, ..., k,\ j = 1, .., p)$ are $(n \times n)$ parameter matrices with Φ_0 (typically) positive definite and the lag length p has been used to make the formulation conformable to that in (2). Thus, in this case equation (4) holds with $l = p$ and $K_{ij} = [0 : \cdots : 0 : \Phi_{ij} : 0 : \cdots : 0]$ where Φ_{ij} is in the jth position. Furthermore, the term $o\left(\|x\|^2\right)$ in (4) is only due to the constant matrix Φ_0.

Note that, although we presented the BEKK model by assuming $f(x) = 0$, our results also apply to a model defined by equations (1) and (5) or, more generally, to models obtained from (1) and (2) by replacing z_t in the latter by u_t. Of course, a redefinition of the original model is then required to obtain the same lag length p to (1) and (2). It should be noted, however, that our formulation does not allow for generalized ARCH (GARCH) type conditional heteroskedasticity.

To be able to obtain stability results for our general model, suitable assumptions have to be imposed on the matrices B_j and K_{ij} in equations (3) and (4). To this end, define the companion matrix

$$B = \begin{bmatrix} B_1 & B_2 & \cdots & B_{p-1} & B_p \\ I_n & 0 & \cdots & 0 & 0 \\ 0 & I_n & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I_n & 0 \end{bmatrix} \quad (np \times np)$$

and the matrix

$$K_{ij0} = \begin{bmatrix} K_{ij} \\ 0 \end{bmatrix} \quad (np \times np).$$

We also need the duplication matrix $D_{n^2p^2}$ of dimension $\left(n^4p^4 \times \frac{1}{2}n^2p^2\left(n^2p^2+1\right)\right)$ and the elimination matrix $L_{n^2p^2}$ of dimension $\left(\frac{1}{2}n^2p^2\left(n^2p^2+1\right) \times n^4p^4\right)$.

Now we can introduce our assumptions which also include conditions required from the error term ε_t.

Assumption 1. (i) The distribution of the $i.i.d\left(0, I_n\right)$ random vectors ε_t is absolutely continuous with respect to the Lebesgue measure and has a density function which is positive almost everywhere in $\mathbb{R}^n$.
(ii) The functions f and H are Borel measurable and bounded on bounded subsets of $\mathbb{R}^{np}$. In addition, the matrix valued function H is such that $H\left(x\right)$ is positive definite for any $x \in \mathbb{R}^{np}$ and one of the following conditions holds: (a) $\inf_{x\in\mathcal{K}} \lambda_{\min}\left(H\left(x\right)\right) \geq \lambda_{\mathcal{K}} > 0$ for any compact set $\mathcal{K} \subset \mathbb{R}^{np}$ or (b) H is continuous and $\lambda_{\min}\left(H\left(x\right)\right) > 0$ for any $x \in \mathbb{R}^{np}$.
(iii) The functions f and H satisfy (3) and (4) and, furthermore,

$$\rho\left(L_{n^2p^2}\left(B' \otimes B' + K_0'\right) D_{n^2p^2}\right) < 1 \tag{6}$$

where $K_0 = \sum_{i=1}^{k}\sum_{j=1}^{l}\left(K_{ij0} \otimes K_{ij0}\right)$.

□

Parts (i) and (ii) of Assumption 1 have been adopted from Lu and Jiang (2001). From a practical point of view they are very mild and met in most applications. Of the two alternative conditions in Assumption 1(ii) the former may be more useful in practice because it permits the function H to be discontinuous which is needed in threshold type models. Moreover, the eigenvalue assumption usually holds in this case because an additive positive definite constant part is typically included in conditional covariance matrix models (cf. the matrix Φ_0 in the BEKK model (5)). Assumption

1(iii) restricts the class of permitted nonlinear functions and actually rules out cases of interest, as will be exemplified in the next section. However, relaxing this assumption seems difficult unless the type of nonlinearity is specified in more detail (cf. the discussion in Remark 3.1 of Masry and Tjøstheim (1995)). We shall return to this point later and demonstrate that Assumption 1(iii) is nonetheless considerably weaker than its previous counterparts.

An eigenvalue condition of the same character as (6) has previously been used by Nicholls and Quinn (1982, Chapter 2) in random coefficient autoregressive models. These authors also considered the alternative condition $\rho\left(B' \otimes B' + K_0'\right) < 1$ subsequently adopted by Feigin and Tweedie (1985). For our purposes condition (6) appears more convenient, however. Notice that the elimination matrix $L_{n^2p^2}$ on the left hand side of (6) can be replaced by $D^{+}_{n^2p^2} = \left(D'_{n^2p^2} D_{n^2p^2}\right)^{-1} D'_{n^2p^2}$, the Moore-Penrose inverse of $D_{n^2p^2}$. This follows from the definition of the matrix K_0 and well-known properties of elimination and duplication matrices (see result 9.6.5(1)(a) in Lütkepohl (1996)).

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 the state space form by defining $X_t = \left[z_t' \ \cdots \ z_{t-p+1}'\right]'$ and, furthermore,

$$\begin{aligned} F_1\left(X_{t-1}, \varepsilon_t\right) &= f\left(X_{t-1}\right) + H\left(X_{t-1}\right)^{1/2} \varepsilon_t \\ F_i\left(X_{t-1}, \varepsilon_t\right) &= z_{t-i}, \quad i = 2, ..., p \end{aligned}$$

and

$$F\left(X_{t-1}, \varepsilon_t\right) = \left[F_1\left(X_{t-1}, \varepsilon_t\right)' \ \cdots \ F_p\left(X_{t-1}, \varepsilon_t\right)'\right]'.$$

With these definitions we have

$$z_t = J' X_t \quad \text{and} \quad X_t = F\left(X_{t-1}, \varepsilon_t\right)$$

where $J' = [I_n : 0 : \cdots : 0]$ $(n \times np)$. Since ε_t is independent of X_{t-1} this shows that X_t is a Markov chain whose state space is $\mathbb{R}^{np}$.

In what follows, the concept of V-geometric ergodicity of a Markov chain will be employed (see Meyn and Tweedie (1993, p. 356)). Here V signifies a real valued 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 X_t is said to be V-geometrically ergodic if there exists a probability measure π on the Borel sets of $\mathbb{R}^{np}$ and a constant $\varrho > 1$ such that

$$\sum_{t=1}^{\infty} \varrho^t \sup_{h:|h|\leq V} \left| E\left(h\left(X_t\right) \mid X_0 = x\right) - \int_{\mathbb{R}^{np}} \pi\left(dy\right) h\left(y\right) \right| < \infty \quad \text{for all } x \in \mathbb{R}^{np}. \tag{7}$$

The definition also assumes that the function V is integrable with respect to the probability measure π. The weakest form of this definition results when $V \equiv 1$. Then the Markov chain X_t is said to be geometrically ergodic. Geometric ergodicity entails that the t-step transition probability measure $P^t(x, \cdot)$ defined on the Borel sets of $\mathbb{R}^{np}$ by $P^t(x, A) = P\left(X_t \in A \mid X_0 = x\right)$ 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.

It is straightforward to show that geometric ergodicity implies stationarity of the process X_t if the distribution of the initial value X_0 is defined by the probability measure π (see Meyn and Tweedie (1993), p. 230-231). Of course, in this case the process z_t is also stationary. These results are contained the following theorem.

Theorem 1 *Suppose that the process z_t $(t \geq 1)$ is generated by equations (1) and (2). Then, if Assumption 1 holds the process $X_t = \left[z_t' \ \cdots \ z_{t-p+1}'\right]'$ is $\|x\|^2$-geometrically ergodic.*

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 from 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 equations (1) and (2) define a linear homoskedastic VAR process the eigenvalue condition (6) 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 special case $K_0 = 0$ and $\rho\left(L_{n^2p^2}\left(B \otimes B\right) D_{n^2p^2}\right) = \rho\left(B\right)^2$ (see results 2.4(11)(e), 9.5.4(2) and 9.5.5.(1)(a) in Lütkepohl (1996)). A similar result is obtained in the special case of the BEKK model (5) (cf. Proposition 2.7 and the subsequent discussion in Engle and Kroner (1995)). The previous conditions given for similar general models by Masry and Tjøstheim (1995), Lu (1998), and Lu and Jiang (2001) are different in this respect. For instance, when applied to a linear homoskedastic VAR model, they are considerably more restrictive than our eigenvalue condition (6).[1] A further advantage of Theorem 1 over these previous results is that it shows the existence of second moments of the stationary distribution although at the cost of assuming that the errors ε_t have finite second moments.

Thus, our eigenvalue condition (6) has some advantages over its previous counterparts. This point can be made even clearer by considering the model

$$z_t = \sum_{j=1}^{p} B_j z_{t-j} + \left(\Sigma + \sum_{i=1}^{k}\sum_{j=1}^{l} K_{ij} X_{t-1} X_{t-1}' K_{ij}'\right)^{1/2} \varepsilon_t \tag{8}$$

where Σ is a positive definite $(n \times n)$ matrix and the other notation is as before. This model is clearly a special case of our general model and subsumes the linear VAR model with ARCH errors. For this model the following result can be proved.

Theorem 2 *Suppose that the process z_t $(t \geq 1)$ is generated by equation (8) where ε_t is independent of z_s $(s < t)$ and satisfies Assumption 1(i). Then, the process $X_t =$*

[1] Note that the condition given in Lemma 3.1 of Masry and Tjøstheim (1995) is generally vacuous, as pointed out by Lu (1998). In the notation of this paper it is nonvacuous only when $K_{ij} = 0$ for all $i = 1, ..., k$, $j = 1, .., l$.

$\left[z_t' \ \cdots \ z_{t-p+1}'\right]'$ *is* $\|x\|^2$*-geometrically ergodic if and only if condition (6) holds with the matrices* B *and* K_0 *defined from model (8).*

Thus, given Assumption 1(i), the eigenvalue condition (6) is necessary and sufficient for the $\|x\|^2$-geometric ergodicity of the process (8). A similar result has previously been obtained in Hansen and Rahbek (1998, Theorem 1) in the special case of a multivariate first order ARCH model with $\varepsilon_t \sim N(0, I_n)$.

Although the result of Theorem 2 is of interest in its own right it is also useful because it readily shows that previous alternatives of our eigenvalue condition (6) are generally inferior. We exemplify this by considering a univariate special case of our general model. Thus, suppose that in equations (1) and (2) $n = 1$ so that x_j, a typical component of the vector x, is a scalar. For convenience, rewrite the conditions in equations (3) and (4) as

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

and

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

respectively. This special case is considered in Theorem 1 of Lu (1998) where it is shown that a sufficient condition for geometric ergodicity is

$$\left(\sum_{j=1}^{p} |b_j|\right)^2 + \sum_{j=1}^{p} \phi_j^2 < 1. \tag{9}$$

This condition is different from our eigenvalue condition (6). Since arguments used in the proof of our Theorem 1 also apply in the context of Lu's (1998) Theorem 1 it is straightforward to show that condition (9) is even sufficient for $\|x\|^2$-geometric ergodicity. Thus, since both Lu's (1998) result and our Theorem 1 also apply to model (8) with $n = 1$ it follows from Theorem 2 that our eigenvalue condition is implied by (9). In Theorem 3 and Remark 4.1 of Lu (1998) it is (essentially) shown

that replacing the eigenvalue condition (6) by (9) yields a necessary and sufficient condition for the $\|x\|^2$-geometric ergodicity of the process

$$z_t = b_i z_{t-i} + \left(\sigma^2 + \sum_{j=1}^{p} \phi_j^2 z_{t-j} \right)^{1/2} \varepsilon_t \tag{10}$$

where $1 \leq i \leq p$ and $\sigma^2 > 0$. Thus, for this model condition (9) is equivalent to our eigenvalue condition. However, in general this is not the case. It is easy to find examples in which our eigenvalue condition (6) holds although the first sum on the left hand side of (9) is larger than one. Then condition (9) fails and its failure can be very clear.

Based on the above mentioned result of model (10) Lu (1998) made the conjecture that in the general univariate model defined by (1) and (2) with $n = 1$ condition (9) could be replaced by the condition $\rho(B)^2 + \sum_{j=1}^{p} \phi_j^2 < 1$ which, like our eigenvalue condition (6), would subsume several known results of geometric ergodicity. However, even if this conjecture were proved the obtained result would be inferior to that obtained in Theorem 1. This can be seen in the same way as in the case of condition (9) because it is possible to find examples in which the eigenvalue condition (6) holds but the condition $\rho(B)^2 + \sum_{j=1}^{p} \phi_j^2 < 1$ fails. In fact, Theorem 2 and the assumption that, for large values of $\|X_{t-1}\|$, the considered general process is supposed to be dominated by the process (8) suggest that it may be difficult to weaken the eigenvalue condition (6) without making specific assumptions about the functions f and H in (1) and (2). Yet, it is worth emphasizing that stationarity may obtain even if the eigenvalue condition (6) does not hold. An example of this will be given in the next section.

We close this section by showing that Theorem 1 implies useful mixing results for the process z_t. Specifically, the following theorem can be readily obtained.

Theorem 3 *Under the conditions of Theorem 1 the following results hold.*
(i) If the distribution of the initial values has finite second moments the process z_t is

strong mixing with geometrically decaying mixing numbers.
(ii) If the initial values have the stationary distribution the process z_t is β-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 β-mixing, also known as absolute regularity, is needed (for definitions, see e.g. Doukham (1994)). For instance, the application of some results in the theory of empirical processes requires β-mixing (see Hansen (1996, 2000) for recent applications of such results). It would be possible to relax the assumption imposed on initial values in part (ii) of Theorem 3 but something stronger than assumed in part (i) is needed (see Doukham (1994, p. 89 and 92)).

3 Nonlinear Error Correction Model

In this section, the results obtained in the previous section will be applied to a nonlinear error correction model. It appears convenient to first discuss the conventional linear error correction model and change the notation so that y_t $(t = 1, 2, ...)$ denotes the n-dimensional stochastic process of interest. Then, the usual linear VAR error correction model can be written as

$$\Delta y_t = \Pi y_{t-1} + \sum_{j=1}^{p} \Gamma_j \Delta y_{t-j} + u_t. \tag{11}$$

Here Π and Γ_j are $(n \times n)$ matrices of unknown parameters and u_t is an error term which for the moment is assumed to be a zero mean i.i.d. sequence with a finite positive definite covariance matrix. Define the polynomial matrix

$$A(z) = (1 - z) I_n - \Pi z - \sum_{j=1}^{p} \Gamma_j (1 - z) z^j \tag{12}$$

and assume that the following conditions hold.

Assumption 2. (i) If $\det A(z) = 0$ then $|z| > 1$ or $z = 1$.
(ii) $\text{rk}(\Pi) = r < n$
(iii) The number of unit roots in condition (i) equals $n - r > 0$.

□

Assumption 1(ii) implies that we can write

$$\Pi = \alpha\beta' \tag{13}$$

where α and β are $(n \times r)$ matrices of full column rank. A well-known consequence of Assumption 1 is that y_t is a nonstationary I(1) process but, with an appropriate specification of initial values, $\beta' y_t$ and Δy_t are stationary. These results can be found in Theorem 4.2 and Corollary 4.3 of Johansen (1995) which also show that Assumption 1(iii) can be replaced by the requirement that the matrix $\alpha'_\perp \left(I_n - \sum_{j=1}^{p} \Gamma_j\right) \beta_\perp$ is nonsingular.

The stationary linear combinations $\beta' y_t$ can be interpreted as long run equilibrium relations between the components of the nonstationary process y_t. Using equation (13) they can be incorporated into the error correction model (11) to obtain

$$\Delta y_t = \alpha\beta' y_{t-1} + \sum_{j=1}^{p} \Gamma_j \Delta y_{t-j} + u_t. \tag{14}$$

In the nonlinear error correction model we shall mainly consider the linear combinations $\beta' y_{t-1}$ are augmented by a nonlinear term. Specifically, we consider the model

$$\Delta y_t = \alpha\left(\beta' y_{t-1} + g\left(\beta' y_{t-1}\right)\right) + \sum_{j=1}^{p} \Gamma_j \Delta y_{t-j} + u_t, \tag{15}$$

where g: $\mathbb{R}^r \to \mathbb{R}^r$ is a general nonlinear function to be described in more detail later. Here we just note that $g\left(\beta' y_{t-1}\right)$ may only depend on some components of the vector $\beta' y_{t-1}$ and it may contain an additive constant term. In particular, one may have

$$g\left(\beta' y_{t-1}\right) = \mu + \overline{g}\left(\beta' y_{t-1}\right)$$

where some components of $\overline{g}(\beta' y_{t-1})$ can be identically equal to zero.

General nonlinear error correction models of the form (15) have previously been discussed by Granger and Swanson (1996), Granger and Haldrup (1997) and Granger (2001). Of course, one could absorb the linear combinations $\beta' y_{t-1}$ into the nonlinear function $g(\beta' y_{t-1})$ but we have not done this because the achieved additional generality appears rather limited. The reason is that, as far as the proofs of the subsequent stability results are concerned, it is 'almost necessary' that in (15) the error correction term $\beta' y_{t-1} + g(\beta' y_{t-1})$ contains the linear part $\beta' y_{t-1}$. A more precise discussion of this point will be given later. Here it may be noted that there are cases of practical interest where this additive linear part is present anyway. For instance, consider the bivariate model

$$\begin{bmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} (y_{1t-1} - b y_{2t-1} - \mu - \delta G(y_{1t-1} - b y_{2t-1} - \mu)) + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} \Delta y_{1t-1} \\ \Delta y_{2t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} \tag{16}$$

where G is the cumulative distribution function of a continuous random variable. If $\delta = 0$ we have a conventional linear error correction model in which the stationary long run relation between y_{1t} and y_{2t} varies around a fixed level determined by the parameter μ. However, when $\delta \neq 0$ a smooth level shift in the long run relation occurs. Specifically, if the value of the parameter δ is positive the level of the long run relation is small when $y_{1t-1} - b y_{2t-1} - \mu$ is small but when $y_{1t-1} - b y_{2t-1} - \mu$ increases sufficiently a smooth shift to a higher level occurs. Thus, the specification of the long run relation in model (16) is based on an idea similar to that used in smooth transition regression models discussed by Granger and Teräsvirta (1993) amongst others. Note that more than one level shift can be readily allowed for by redefining the function G or its argument in a suitable way. For examples, see van Dijk, Teräsvirta, and Frances (2001).

As a limiting case of model (16) one obtains a threshold type model in which G is a step function implying an abrupt change in the constant term. This is a special case of the threshold error correction model of Balke and Fomby (1997). In a univariate case stationary threshold autoregressive models based on this idea were recently considered by Lanne and Saikkonen (2001) who demonstrated that neglecting such a level shift can easily lead to the incorrect conclusion that the considered series is generated by an I(1) process. This indicates that in the case of model (16) an application of the corresponding standard linear model may similarly lead to the conclusion that the processes y_{1t} and y_{2t} are not cointegrated and no long run relation between them exists. However, if $\delta \neq 0$ this is not a correct interpretation because a nonlinear stationary long run relation can still exist, as will be seen shortly. Similar conclusions can be drawn from the simulation results of Balke and Fomby (1997) based on other types of threshold autoregressive error correction models.

As the preceding discussion has already made clear, an essential feature of our nonlinear error correction model is that the nonlinear function $g(\beta' y_{t-1})$ is absorbed into the error correction term. As an alternative model, not necessarily having this feature, one could consider

$$\Delta y_t = \alpha\beta' y_{t-1} + g_0(\beta' y_{t-1}) + \sum_{j=1}^{p} \Gamma_j \Delta y_{t-j} + u_t, \tag{17}$$

where g_0: $\mathbb{R}^r \to \mathbb{R}^n$ is some nonlinear function. In the special case where the lagged differences are absent this model was recently considered by Corradi, Swanson, and White (2000). Although our main interest is in model (15) we shall also briefly discuss this alternative specification. The reason for concentrating on model (15) is that we have not been able to obtain similar results for model (17) whose time series properties turn out to be less straightforward to handle than those of model (15).

To be able to apply Theorem 1 to the nonlinear error correction models discussed above a simple transformation of the polynomial matrix $A(z)$ in (12) is needed. In

order to develop this transformation, let c $(n \times r)$ be a matrix of full column rank and such that $c'\beta = I_r$. One possible choice for the matrix c is $\beta(\beta'\beta)^{-1}$ but any other will also do. It is easy to see that the matrix $Q = [\beta : c_\perp]'$ is nonsingular and its inverse can be written as $Q^{-1} = [c : \beta_\perp]$ where $c'_\perp \beta_\perp = I_{n-r}$ is assumed, as the notation indicates. We also define the related $(n \times n)$ polynomial matrix

$$P(z) = \begin{bmatrix} \beta' \\ (1-z)\, c'_\perp \end{bmatrix}$$

and introduce the notation

$$\Gamma(z) = I_n - \sum_{j=1}^{p} \Gamma_j z^j$$

where Γ_j $(j = 1, ..., p)$ are as before. Now we can state the following lemma which gives a representation for the polynomial matrix $A(z)$.

Lemma 4 *Assumption 2 holds if and only if the polynomial matrix $A(z)$ can be expressed as*

$$A(z) = Q^{-1} B_*(z)\, P(z)$$

where $B_*(z)$ is a polynomial matrix defined by

$$B_*(z) = Q\left[\Gamma(z)\, c\,(1-z) - \alpha z : \Gamma(z)\, \beta_\perp\right]$$

and such that $B_*(0) = I_n$ and $\det B_*(z) \neq 0$, $|z| \leq 1$.

Using the notation introduced above we now define the polynomial matrix

$$B(z) = I_n - \sum_{j=1}^{p} B_j z^j = Q^{-1} B_*(z)\, Q$$

and transform the process y_t as

$$z_t = Q^{-1} P(L)\, y_t = Q^{-1} \begin{bmatrix} \beta' y_t \\ \Delta c'_\perp y_t \end{bmatrix}. \tag{18}$$

By the definitions and Lemma 4 we then have $\beta' z_t = \beta' y_t$ and $A(L) y_t = B(L) z_t$, and it follows that model (15) can be transformed to

$$z_t = \sum_{j=1}^{p} B_j z_{t-j} + \alpha g(\beta' z_{t-1}) + u_t. \tag{19}$$

Thus, we have transformed the original error correction model to a special case of the nonlinear VAR model considered in the previous section. Two points are worth noting in this nonlinear VAR model. First, its variables are nonsingular linear transformations of the vectors $\beta' y_t$ and $\Delta c_\perp' y_t$ which in the linear case are known to be stationary when Assumption 1 holds. Second, the lag polynomial matrix $B(L)$ related to the nonlinear VAR model (19) has its roots outside the unit circle. This means that we can apply Theorem 1 to show the stationarity of the nonlinear VAR process (19) and thereby the stationarity of the processes $\beta' y_t$ and $c_\perp' \Delta y_t$ in our nonlinear error correction model (15).[2] This clearly means that the same approach could be applied even if we extended the nonlinearity of the model by including the lagged values $\beta' y_{t-2}, ..., \beta' y_{t-p}$ in the argument of the function g in (15). Since $c_\perp' z_t = \Delta c_\perp' y_t$ we could similarly include the variables $\Delta c_\perp' y_{t-1}, ..., \Delta c_\perp' y_{t-p}$ or, more generally, make the function g depend on the lagged differences $\Delta y_{t-1}, ..., \Delta y_{t-p}$. However, for simplicity and for the lack of a reasonable motivation we have preferred not to make these extensions explicit.

It is also possible to allow for conditional heteroskedasticity in the nonlinear error correction model (15). Since the related stability results are proved by applying Theorem 1 to the transformed model (19) a general form of conditional heteroskedasticity that can be easily handled is

$$u_t = H(z_{t-1}, ..., z_{t-p})^{1/2} \varepsilon_t. \tag{20}$$

[2]Of course it is also possible to use Lemma 4 to prove these results in the linear case where $g(\beta' y_{t-1})$ reduces to a constant vector. In this way it is possible to obtain an alternative to Johansen's (1995, p. 49-52) proof of Granger's representation theorem.

Here the error term ε_t and the function H are as in the previous section. Although this formulation is convenient from the viewpoint of formulating and proving the desired stability results it may be somewhat disturbing that the conditional heteroskedasticity is assumed to depend on lagged values of the transformed variable z_t. Alternative formulations would be possible but, since they seem to complicate the exposition and formulation of the needed assumption, we shall not try to discuss them here. Another reason for this is that the above specification still covers the important special case where a (possibly nonlinear) ARCH model is specified for the error term u_t. This follows from the fact that the same error term u_t appears both in the original model (15) and in the transformed model (19). Therefore, after an appropriate reformulation, lagged values of u_t can be used in (20) (cf. the corresponding discussion in the previous section). Below we shall see that there is also another case where the occurrence of lagged values of the transformed variable z_t in (20) is not restrictive.

In addition to Assumption 2 the following conditions are needed to prove the stability of the nonlinear error correction model (15) combined with (20).

Assumption 3. (i) The function g in (15) is Borel measurable, bounded on bounded subsets of $\mathbb{R}^r$, and such that $\|g(x)\| = o(\|x\|)$ as $\|x\| \to \infty$ $(x \in \mathbb{R}^r)$.
(ii) The function H in (20) satisfies condition (4) and the conditions in Assumption 1(ii).
(iii) The eigenvalue condition (6) in Assumption 1(iii) holds with the matrix K_0 as before and the matrix B defined in terms of the matrices B_j $(j = 1, ..., p)$ in equation (19).

□

This assumption makes it easy to apply Theorems 1 and 3 but the condition $\|g(x)\| = o(\|x\|)$ as $\|x\| \to \infty$ is somewhat restrictive. For instance, in the bivariate model (16) one might wish to allow for the possibility that, instead of the

constant term μ, the values of the adjustment parameters α_1 and α_2 depend on the value of the linear combination $y_{1t-1} - by_{2t-1} - \mu$. An example of a model with such a feature is obtained by replacing the term $\delta G\left(y_{1t-1} - by_{2t-1} - \mu\right)$ in (16) by $\delta G\left(y_{1t-1} - by_{2t-1} - \mu\right)\left(y_{1t-1} - by_{2t-1} - \mu\right)$. However, even this simple specification is ruled out by the condition imposed on the function g in Assumption 3(i). Being able to relax this condition would clearly be of great interest. However, that would require extensions of Theorem 1 and, as discussed in the previous section, such extensions seem difficult to obtain without making more specific assumptions of the form of the considered nonlinearity.

The following theorem contains the main result of this section.

Theorem 5 *Suppose that the process y_t is generated by equations (15) and (20) with ε_t in satisfying Assumption 1(i). Then, if Assumptions 2 and 3 are also satisfied, the following results hold.*
(i) The conclusions of Theorems 1 and 3 hold with the process z_t defined as $z_t = [y_t'\beta \;\; \Delta y_t' c_\perp]'$.
(ii) Suppose that the initial values $y_{-p+1}, ..., y_0$ are such that the process z_t defined in (i) is stationary. Then, the process y_t is nonstationary and such that $T^{-1/2}y_{[Ts]}$ $(0 \leq s \leq 1)$ converges weakly to a Brownian motion with covariance matrix of rank $n - r$.

Thus, Theorem 5 shows that results entirely similar to those obtained in the standard linear VAR error correction model also hold for our nonlinear extension which may even involve conditional heteroskedasticity. Notice that it is not uncommon to define an I(1) process by requiring that a functional central limit theorem similar to that in Theorem 5 holds (see e.g. Corradi, Swanson, and White (2000)). Thus, the latter result of Theorem 5 has the interpretation that y_t is an I(1) process. The

assumption of stationary initial values is probably not necessary but used to simplify the proof.

Earlier we noted that it is 'almost necessary' that the error correction term $\beta' y_{t-1} + g(\beta' y_{t-1})$ in model (15) contains the linear part $\beta' y_{t-1}$. From the proof of Theorem 5 it can be seen that we could replace the sum $\beta' y_{t-1} + g(\beta' y_{t-1})$ by a function $g_*(\beta' y_{t-1})$, say. However, since the proof of Theorem 5 relies on Lemma 4, we would need the condition $g_*(x) = \beta' x + o(\|x\|)$ as $\|x\| \to \infty$. Moreover, since the $(n \times r)$ matrix β should be of full column rank the additional generality achieved by this formulation appears rather limited.

When the conditional heteroskedasticity in Theorem 5 is of the most general form assumed the parameters of the polynomial matrix $B(z)$ have to satisfy conditions not required in the standard linear error correction model (14). These additional conditions can be dispensed with if the form of conditional heteroskedasticity is restricted. The following corollary makes this precise.

Corollary 6 *Suppose that the assumptions of Theorem 5 hold with Assumption 3(ii) strengthened by requiring that $K_{ij} = 0$, $i = 1, ..., k$, $j = 1, .., l$, in equation (4). Then the results of Theorem 5 hold without Assumption 3(iii).*

This corollary applies when there is no conditional heteroskedasticity but also when the conditional heteroskedasticity can be modeled by a bounded function. The latter case is relevant, for instance, when the conditional heteroskedasticity is similar to that in the threshold models of Balke and Fomby (1997). When the assumptions made of the function H in Corollary 6 hold the lagged values of z_t in equation (20) can be replaced with any linear transformations without affecting the stated result. More generally, if we again denote $X_{t-1} = \left[z'_{t-1} \cdots z'_{t-p}\right]'$ and set $X^*_{t-1} = CX_{t-1}$ with C an nonrandom matrix of appropriate dimension the result of Corollary 6 holds even if the function $H(z_{t-1}, ..., z_{t-p}) = H(X_{t-1})$ is replaced by $H\left(X^*_{t-1}\right)$.

We close this section by considering the alternative model specification (17). Arguments used to transform model (15) to (19) show that model (17) can be transformed to

$$z_t = \sum_{j=1}^{p} B_j z_{t-j} + g_0\left(\beta' z_{t-1}\right) + u_t \tag{21}$$

where the notation is as before. This model combined with (20) is a special case of the nonlinear VAR model of the previous section. This implies that, after modifying Assumption 3(i) to concern the function g_0, the first result of Theorem 5 can again be concluded from Theorem 1. However, we have not been able to prove the second result of Theorem 5 in the case of model (17). In fact, it seems that such a result cannot be proved without making more specific assumptions of the form of nonlinearity.

To see the difficulty alluded to above, let $\mathbf{1}(\cdot)$ signify the indicator function and consider the univariate first order threshold autoregressive model

$$y_t = y_{t-1} + \mu_1 \mathbf{1}\left(y_{t-1} \leq 0\right) + \mu_2 \mathbf{1}\left(y_{t-1} > 0\right) + \varepsilon_t \tag{22}$$

where the error term ε_t satisfies Assumption 1(i) and the intercept terms μ_1 and μ_2 are such that $\mu_1 > 0$ and $\mu_2 < 0$. The value of the threshold parameter is assumed to be zero just for convenience and without any loss of generality. This model is a univariate special case of (17) with the function g_0 bounded and with the counterpart of the polynomial matrix $A(z)$ containing a unit root. However, it is well-known that, in spite of the unit root, equation (22) has a stationary solution (see Chan, Petrucelli, Tong, and Woolford (1985) for a proof and Balke and Fomby (1997) for a discussion). The same is true even if the number of regimes exceeds two and the intercept terms related to the highest and lowest regimes are negative and positive, respectively. Note also that a simple extension of this example to the vector case is obtained by taking independent copies of processes of the type (22).

The above example shows that for nonlinear processes unit roots do not necessarily mean nonstationarity and it also demonstrates why it is difficult to prove the second

result of Theorem 5 without specifying the function g_0 in more detail. Furthermore, this example implies that the result given in Proposition 2.4 of Corradi, Swanson, and White (2000) is not true as stated. In this proposition the authors assume that the function g_0 is bounded and consider model (17) with $u_t = \varepsilon_t$ and without the term $\alpha\beta' y_{t-1}$ and the lagged differences on the right hand side.[3] They argue that the process explodes with a positive probability or, more specifically, that $\Pr(\|y_t\| \to \infty) > 0$ as $t \to \infty$. However, the process defined by equation (22) provides a counter example because an explosive behavior of this kind is not possible for an (asymptotically) stationary process.

It may be noted that the difficulty with Proposition 2.4 of Corradi, Swanson, and White (2000) does not occur in the alternative specification the authors discuss on page 44 (line 3) of their paper. The reason is that then the nonlinear function disappears from the model because the model behaves exactly in the same way as our model (15). Using this alternative specification would have been reasonable in the stationarity test considered in Section 3 of Corradi, Swanson, and White (2000) because otherwise the consistency result given in Theorem 3.1 of that paper is not true. This is an immediate consequence of the existence of stationary nonlinear processes such as (22) which have a unit root in the autoregressive polynomial.

Since the proofs of Propositions 2.3(i) and 2.4 of Corradi, Swanson, and White (2000) are based on the same argument the above discussion also casts doubts on the validity of the former proposition and the related test procedure considered in Section 4 of that paper. However, in this case we have not found a counter example which

[3]Corradi, Swanson, and White (2000) actually assume that the function g_0 satisfies some smoothness properties which rule out discontinuous functions such as the indicator functions in model (22). However, the proof they give for their Propositon 2.4 makes no use of these smoothness properties so that, if the proof were valid for the considered smooth functions, it would also be valid for the indicator functions used in (22).

would exactly match the considered model.[4]

The reason why proving the result of Theorem 5(ii) in the case of model (17) is more difficult than in the case of model (15) may be seen by considering the familiar linear special cases where the functions g and g_0 are constants. As is well known, the deterministic parts obtained for the levels y_t in these two models are very different. In model (15) the deterministic part only consists of a constant term but in model (17) a linear deterministic trend term is generally obtained. In the proof of Theorem 5(ii) the nonlinear term $g(\beta' y_{t-1})$ is treated in the same way as the intercept term in the standard linear model so that, if this approach is attempted in the case of model (17), a rather different situation is obtained.

4 Conclusion

In this paper previous stability results obtained for general nonlinear VAR models have been improved and it has also been shown how these results can be applied to a general nonlinear VAR error correction model to obtain an analog of Granger's representation theorem. However, more research is still needed in this matter because for several nonlinear models of interest the question of stability remains a totally or at least partially open problem. As already indicated, in the future it may be best to concentrate on models in which the form of nonlinearity is at least somewhat more specific than assumed in this paper.

[4] The proof of Proposition 2.3(i) of Corradi, Swanson, and White (2000) seems to contain a flaw. In the notation of that paper, the problem is that in the last step the authors use the result $\widetilde{g}_{0,T}/T^{1/2} \Rightarrow G$ where $\Rightarrow$ signifies weak convergence and G is said to be either a nondegenerate or a degenerate random variable. Then, using a previously defined nondegenerate normal random vector B_1, they claim that the desired result follows from the equality $P(\omega{:}B_1(\omega) = -G(\omega)) = 0$. When G is degenerate this equality is of course valid. However, the problem is that it is not valid if $B_1 = -G$ (a.s.) and the authors give no explanation why this possibility is ruled out.

More research is particularly desirable in the case of nonlinear error correction models because for them we mainly showed how the problem can be approached. As discussed in the context of the nonlinear the error correction model (17), the real difficulty may not be proving that the differences and some linear combinations of the considered process are stationary but proving that the process itself is I(1).

Appendix

Before proving Theorem 1 we shall prove the following auxiliary lemma.

Lemma 7 *(i) Condition (6) holds if and only if, for any symmetric* $(n^2p^2 \times n^2p^2)$ *matrix* S,

$$(B' \otimes B' + K_0')^t \, vec\,(S) \to 0, \quad as\ t \to \infty. \tag{23}$$

(ii) The convergence in (23) is equivalent to the convergence of the series

$$vec\,(U_t) = \sum_{j=0}^{t-1} (B' \otimes B' + K_0')^j \, vec\,(S)\,. \tag{24}$$

When this series converges its limit can be expressed as $vec(U)$ *where* U *is a symmetric matrix which is positive definite if* S *is positive definite.*

Proof. Part (i) can be proved by using the same arguments as in the proof of Theorem 2.5 of Nicholls and Quinn (1982). That the convergence of (24) is implied by (23) is seen in the same way as in the proof of Theorem 2.4 of Nicholls and Quinn (1982) whereas the converse is obvious. Finally, since S is symmetric it follows by induction that each U_t is symmetric. If S is positive definite it similarly follows by induction that each U_t is positive definite and, moreover, $\lambda_{\min}(U_t) \geq \varepsilon$ for some $\varepsilon > 0$. Thus, the limit of (23) is symmetric (positive definite) when S is symmetric (positive definite). ■

Proof of Theorem 1. The idea of the proof is to apply Theorem 15.0.1 of Meyn and Tweedie (1993). From Lemmas 1 and 2 of Lu and Jiang (2001) we first find that the Markov chain X_t is irreducible and aperiodic and that all bounded sets of $\mathbb{R}^{np}$ with positive Lebesgue measure are small sets. Given this, it suffices to show that condition (15.3) of Meyn and Tweedie (1993, p. 355) holds.

We proceed by making use of ideas employed in the proof of Theorem 3 of Feigin and Tweedie (1985). First, let W $(n^2p^2 \times n^2p^2)$ be an auxiliary positive definite matrix and define the matrix U $(n^2p^2 \times n^2p^2)$ by the equation

$$\operatorname{vec}(U) = \sum_{j=0}^{\infty} (B' \otimes B' + K_0')^j \operatorname{vec}(W) \tag{25}$$

where K_0 is as defined in the formulation of the theorem. By Lemma 7 the right hand side is well defined and the matrix U is positive definite. Next, define the real valued function q by

$$q(x) = 1 + x'Ux.$$

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 bounded set

$$C = \{x \in \mathbb{R}^{np} : x'Ux \leq M_1\}$$

where is M_1 an appropriate constant to be determined later.

Using the above definitions, we shall show that there exist constants $0 < \rho < 1$ and $0 < M_2 < \infty$ such that

$$E(q(X_t) \mid X_{t-1} = x) < \rho q(x), \quad x \notin C \tag{26}$$

and

$$E(q(X_t) \mid X_{t-1} = x) < M_2, \quad x \in C. \tag{27}$$

It is easy to see that then condition (15.3) of Meyn and Tweedie (1993, p. 355) holds with $V(\cdot) = q(\cdot)$ and, consequently, the Markov chain X_t is q-geometrically

ergodic. Thus, since we clearly have $\|x\|^2 \leq Mq(x)$ for some $0 < M < \infty$, the desired $\|x\|^2$-geometric ergodicity follows from conditions (26) and (27).

To establish conditions (26) and (27), we first write equation $X_t = F(X_{t-1}, \varepsilon_t)$ as

$$X_t = BX_{t-1} + R(X_{t-1}, \varepsilon_t) \tag{28}$$

where the matrix B is as in Assumption 1 and

$$\begin{aligned} R(X_{t-1}, \varepsilon_t) &= \begin{bmatrix} f(X_{t-1}) - \sum_{j=1}^{p} B_j z_{t-j} \\ 0 \end{bmatrix} + \begin{bmatrix} H(X_{t-1})^{1/2} \varepsilon_t \\ 0 \end{bmatrix} \\ &\stackrel{def}{=} R_1(X_{t-1}) + R_2(X_{t-1}, \varepsilon_t). \end{aligned}$$

Using the definitions we can then write

$$\begin{aligned} q(X_t) &= 1 + (BX_{t-1} + R(X_{t-1}, \varepsilon_t))' U (BX_{t-1} + R(X_{t-1}, \varepsilon_t)) \\ &= 1 + X_{t-1}' B'UBX_{t-1} + 2X_{t-1}' B'UR_1(X_{t-1}) + 2X_{t-1}' B'UR_2(X_{t-1}, \varepsilon_t) \\ &\quad + R_1(X_{t-1})' UR_1(X_{t-1}) + 2R_1(X_{t-1})' UR_2(X_{t-1}, \varepsilon_t) \\ &\quad + R_2(X_{t-1}, \varepsilon_t)' UR_2(X_{t-1}, \varepsilon_t). \end{aligned}$$

Thus, since ε_t and X_{t-1} are independent this equality and the definitions readily show that

$$\begin{aligned} E(q(X_t) \mid X_{t-1} = x) &= 1 + x'B'UBx + 2x'B'UR_1(x) \\ &\quad + R_1(x)' UR_1(x) + \mathrm{tr}\overline{H}(x) U \end{aligned} \tag{29}$$

where $\overline{H}(x) = \mathrm{diag}[H(x) \quad 0]$. Denoting

$$R_3(x) = \mathrm{tr}\left(\overline{H}(x) - \sum_{i=1}^{k} \sum_{j=1}^{l} K_{ij0} xx' K_{ij0}' \right) U$$

we can write equation (29) as

$$E(q(X_t) \mid X_{t-1} = x) = 1 + x'B'UBx + \mathrm{tr} \sum_{i=1}^{k} \sum_{j=1}^{l} K_{ij0} xx' K_{ij0}' U + Z(x)$$

where $Z(x) = 2x'B'UR_1(x) + R_1(x)'UR_1(x) + R_3(x)$. For the second and third terms on the right hand side we can use well-known properties of the vec operator (see Lütkepohl (1996, Chapter 7.2)) and find that

$$x'B'UBx = (x' \otimes x')(B' \otimes B')\,\mathrm{vec}(U)$$

and

$$\mathrm{tr}\sum_{i=1}^{k}\sum_{j=1}^{l} K_{ij0}xx'K_{ij0}'U = (x' \otimes x')\,K_0'\mathrm{vec}(U).$$

Thus, we can conclude that equation (29) can be written as

$$\begin{aligned} E(q(X_t) \mid X_{t-1} = x) &= 1 + (x' \otimes x')(B' \otimes B' + K_0')\mathrm{vec}(U) + Z(x) \\ &= 1 + (x' \otimes x')(\mathrm{vec}(U) - \mathrm{vec}(W)) + Z(x) \\ &= q(x) - x'Wx + Z(x). \end{aligned}$$

Here the second equality is based on the definition of the matrix U in (25) and the third one on the definition of the function q. Thus, we have shown that

$$E(q(X_t) \mid X_{t-1} = x) = q(x)\left[1 - \frac{x'Wx - Z(x)}{q(x)}\right]. \tag{30}$$

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

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

and therefore

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

From the assumed asymptotic behavior of the functions f and H and the definition of $Z(x)$ it follows that $|Z(x)| = o(\|x\|^2)$ as $\|x\|^2 \to \infty$. 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 $\varepsilon < 1$. Setting $\rho = 1 - \varepsilon$ we can then see from (30) that condition (26) holds.

As for condition (27), since the functions f and H are bounded on bounded subsets of $\mathbb{R}^{np}$, $R_1(x)$ and $\overline{H}(x)$ are bounded for $x \in C$ and we can conclude from equality (29) that condition (27) also holds. Thus, the proof is complete. ■

Proof of Theorem 2. If condition (6) holds the stated result follows from Theorem 1, so we only need to prove the converse.

The process X_t is now generated by

$$X_t = BX_{t-1} + \begin{bmatrix} H(X_{t-1})^{1/2} \\ 0 \end{bmatrix} \varepsilon_t$$

where

$$H(X_{t-1}) = \Sigma + \sum_{i=1}^{k} \sum_{j=1}^{l} K_{ij} X_{t-1} X_{t-1}' K_{ij}'.$$

Straightforward calculations show that

$$X_t X_t' = BX_{t-1} X_{t-1}' B' + \overline{H}(X_{t-1}) + R_4(X_{t-1}, \varepsilon_t) \tag{31}$$

where $\overline{H}(x) = \text{diag}[H(x) \quad 0]$ as before, and

$$\begin{aligned} R_4(X_{t-1}, \varepsilon_t) &= BX_{t-1}\varepsilon_t' \begin{bmatrix} H(X_{t-1})^{1/2} \\ 0 \end{bmatrix}' + \begin{bmatrix} H(X_{t-1})^{1/2} \\ 0 \end{bmatrix} \varepsilon_t X_{t-1}' B' \\ &\quad + \begin{bmatrix} H(X_{t-1})^{1/2} \\ 0 \end{bmatrix} (\varepsilon_t \varepsilon_t' - I_n) \begin{bmatrix} H(X_{t-1})^{1/2} \\ 0 \end{bmatrix}'. \end{aligned}$$

Denote $\overline{\Sigma} = \text{diag}[\Sigma \quad 0]$ and use the definitions to rewrite equation (31) as

$$X_t X_t' = \overline{\Sigma} + BX_{t-1} X_{t-1}' B' + \sum_{i=1}^{k} \sum_{j=1}^{l} K_{ij0} X_{t-1} X_{t-1}' K_{ij0}' + R_4(X_{t-1}, \varepsilon_t).$$

Using well-known properties of the vec operator (see Lütkepohl (1996, Chapter 7.2)) to both sides of this equation yields

$$\text{vec}\left(X_t X_t'\right) = \text{vec}\left(\overline{\Sigma}\right) + \left(B' \otimes B' + K_0'\right) \text{vec}\left(X_{t-1} X_{t-1}'\right) + \text{vec}\left(R_4\left(X_{t-1}, \varepsilon_t\right)\right).$$

In the preceding equations K_{ij0} and K_0 are defined from K_{ij} in the same way as before. Since we clearly have $E\left(R_4\left(X_{t-1}, \varepsilon_t\right) \mid X_{t-1}\right) = 0$ we can conclude from this that, for $t \geq 1$,

$$E\left(\text{vec}\left(X_t X_t'\right) \mid X_0 = x\right) = \text{vec}\left(\overline{\Sigma}\right) + \left(B' \otimes B' + K_0'\right) E\left(\text{vec}\left(X_{t-1} X_{t-1}'\right) \mid X_0 = x\right).$$

Thus, by repetitive substitution,

$$E\left(\text{vec}\left(X_t X_t'\right) \mid X_0 = x\right) = \left(B' \otimes B' + K_0'\right)^t \text{vec}\left(xx'\right) + \sum_{j=0}^{t-1} \left(B' \otimes B' + K_0'\right)^j \text{vec}\left(\overline{\Sigma}\right).$$

Now suppose that $t \to \infty$ and conclude from the definition of $\|x\|^2$-geometric ergodicity that the expression on the right hand side of this equation converges to a limit independent of the initial value x. This implies that $(B' \otimes B' + K_0')^t \text{vec}(xx') \to 0$ as $t \to \infty$ for every $x \in \mathbb{R}^{np}$. Using the spectral decomposition of symmetric matrices in the same way as in the proof of Theorem 2.4 of Nicholls and Quinn (1982) we then find that, for any symmetric $(n^2p^2 \times n^2p^2)$ matrix S, $(B' \otimes B' + K_0')^t \text{vec}(S) \to 0$ as $t \to \infty$. By Lemma 7 this implies condition (6). ■

Proof of Theorem 3. To prove (i), first recall that in the proof of Theorem 1 we showed that the Markov chain X_t is irreducible and aperiodic and that condition (15.3) of Meyn and Tweedie (1993, p. 355) holds with an appropriate small set and the function V given by $V(x) = q(x) = 1 + x'Ux$. Since it is clear that $M^{-1}\|x\|^2 \leq q(x) \leq M\|x\|^2$ for some $M \geq 1$ we can therefore conclude from Theorem 16.0.1 of Meyn and Tweedie (1993) that the Markov chain X_t is $\|x\|^2$-uniformly ergodic (for a definition of this concept, see page 382 of the same reference). Given the $\|x\|^2$-uniform ergodicity, the stated strong mixing property of X_t follows from Theorem 16.1.5 Meyn and Tweedie (1993, p. 388) and the discussion given after its proof.

That geometric ergodicity implies β-mixing in the case of stationary initial values has been pointed out by Pham (1986) (see also Doukham (1994, p. 4 and 89).■

Proof of Lemma 4. Suppose that Assumption 2 holds and use the representation of the matrix Π in (13) to express the polynomial matrix $A(z)$ defined in equation (12) as

$$A(z) = (\Gamma(z)(1-z) - \alpha\beta' z) Q^{-1} Q.$$

Next note that

$$\alpha\beta' Q^{-1} = \alpha\beta' [c : \beta_\perp] = \alpha [I_r : 0]$$

and hence

$$A(z) = [\Gamma(z) c(1-z) - \alpha z : \Gamma(z) \beta_\perp] \begin{bmatrix} \beta' \\ (1-z) c_\perp' \end{bmatrix} = Q^{-1} B_*(z) P(z).$$

Here the latter equality follows from the definitions which also show that $B_*(0) = I_n$. Thus, since

$$\det A(z) = \det Q^{-1} \det B_*(z) \det P(z)$$

and since the number of unit roots in $\det P(z)$ is clearly $n - r$ it follows from Assumption 1(iii) that $\det B_*(z) \neq 0$ for $|z| \leq 1$.

To prove the converse, suppose the polynomial matrix $A(z)$ defined in equation (12) has the representation given in the lemma. Then, Assumptions 2(i) and (iii) follow from analyzing the determinant of $A(z)$ in the same way as in the proof of the first part whereas the validity of Assumption 2(ii) is seen by calculating $A(1) = -\Pi = Q^{-1} B_*(1) P(1) = -\alpha\beta'$. ■

Proof of Theorem 5. We shall prove the first assertion by applying Theorems 1 and 3 to the transformed process defined by equations (19) and (20). An inspection of the assumptions reveals that we only need to show that Assumption 3(i) implies that the function $g(\beta' x)$, $x \in \mathbb{R}^{np}$, satisfies $\|g(\beta' x)\| = o(\|x\|)$ as $\|x\| \to \infty$. To see this, notice that we can write $x = \beta a + \beta_\perp b$ for some $a \in \mathbb{R}^r$ and $b \in \mathbb{R}^{n-r}$. Without loss

of generality we can also assume that $\beta'\beta = I_r$ and $\beta_\perp'\beta_\perp = I_{n-r}$. Thus, if $\|x\| \to \infty$ at least one of the relations $\|a\| \to \infty$ or $\|b\| \to \infty$ must hold. If $\|a\| \to \infty$ holds the equality $g(\beta'x) = g(a)$ implies the desired conclusion. On the other hand, if $\|b\| \to \infty$ and $\|a\|$ remains bounded the assumption that the function g is bounded on bounded subsets of $\mathbb{R}^r$ gives $g(\beta'x) = g(a) = O(1)$ as $\|x\| \to \infty$. Thus, again the desired conclusion follows, and we have established the first assertion of the theorem.

As for the second assertion, first conclude from part (i) that the stationary process z_t has finite second moments. Thus, since Assumption 3.1(i) implies that $\|g(\beta'x)\| \leq M\|x\|$, $0 < M < \infty$, for $\|x\|$ large enough, it follows that the process $g(\beta'z_{t-1})$ is also stationary with finite second moments. Further, since the function H is assumed to satisfy condition (4) one can similarly show that the process u_t is a stationary martingale difference sequence with finite second moments. Finally, Theorem 3(i) implies that u_t is strong mixing and hence also ergodic (see e.g. Doukham (1994, p. 21)). These facts will be used in the subsequent proof.

By repetitive substitution one obtains from equation (19)

$$z_t = k_t + B(L)^{-1}\alpha g(\beta'z_{t-1}) + B(L)^{-1}u_t, \quad 1, 2, ...$$

where k_t is due to the initial values $z_{-p+1}, ..., z_0$ and $g(\beta'z_t) = 0$ and $u_t = 0$ for $t \leq 0$. Next use the Beveridge-Nelson decomposition $B(L)^{-1} = B(1)^{-1} + \Delta\widetilde{F}(L)$ to write this equation as

$$z_t = k_t + B(1)^{-1}\alpha g(\beta'z_{t-1}) + \Delta v_t + e_t, \quad 1, 2, ... \tag{32}$$

where, for simplicity, $v_t = \widetilde{F}(L)\alpha g(\beta'z_{t-1})$ and $e_t = B(L)^{-1}u_t$. By the definitions,

$$\begin{aligned} B(1)^{-1}\alpha &= ([-\alpha : \Gamma(1)\beta_\perp]Q)^{-1}\alpha \\ &= Q^{-1}[-\alpha : \Gamma(1)\beta_\perp]^{-1}\alpha \\ &= [c : \beta_\perp]\begin{bmatrix} -I_r \\ 0 \end{bmatrix} \\ &= -c. \end{aligned}$$

Thus, since $c_{\perp}'z_t = \Delta c_{\perp}'y_t$ we can solve equation (32) for $c_{\perp}'y_t$ to get

$$c_{\perp}'y_t = c_{\perp}'y_0 + c_{\perp}'\sum_{j=1}^{t} k_j + c_{\perp}'v_t - c_{\perp}'v_0 + c_{\perp}'\sum_{j=1}^{t} e_j.$$

From this and the stated initial value assumption it is easy to see that

$$T^{-1/2}c_{\perp}'y_{[Ts]} = T^{-1/2}c_{\perp}'\sum_{j=1}^{[Ts]} e_j + T^{-1/2}c_{\perp}'v_{[Ts]} + T^{-1/2}c_{\perp}'v_0 + o_p(1) \tag{33}$$

where the term $o_p(1)$ is uniform in $0 \leq s \leq 1$.

Now consider the sequence v_t and conclude from the definition that

$$v_t = \sum_{j=1}^{t-2} \widetilde{d}_j g(\beta' z_{t-1-j})$$

where $\widetilde{d}_j$ is the jth coefficient in the power series representation of $\widetilde{F}(L)\alpha$. Since Lemma 4 implies that the roots of $\det B(z)$ lie outside the unit circle it follows that $\left\|\widetilde{d}_j\right\|$ converges to zero at a geometric rate as $j \to \infty$. Thus, since the process $g(\beta' z_{t-1})$ is stationary with finite second moments we have $T^{-1/2}c_{\perp}'v_{[Ts]} = o_p(1)$ uniformly in $0 \leq s \leq 1$. The second and third terms on the right hand side of (33) can therefore be absorbed into the last term and, using the Beveridge-Nelson decomposition $B(L)^{-1} = B(1)^{-1} + \Delta\widetilde{F}(L)$ in the definition of e_t, we can write

$$T^{-1/2}c_{\perp}'y_{[Ts]} = T^{-1/2}c_{\perp}'B(1)^{-1}\sum_{j=1}^{[Ts]} u_j + T^{-1/2}c_{\perp}'\widetilde{u}_{[Ts]} - T^{-1/2}c_{\perp}'\widetilde{u}_0 + o_p(1) \tag{34}$$

where $\widetilde{u}_t = \widetilde{F}(L)u_t$. Note that, apart from the impact of initial values, $\widetilde{u}_t$ is a zero mean stationary process with finite second moments. This implies that the third term on the right hand side of (34) can be replaced by $o_p(1)$. The same is true for the second term if $T^{-1/2}\max_{1\leq j\leq T}\|\widetilde{u}_{tj}\| = o_p(1)$. If the impact of initial values is ignored and $\widetilde{u}_t$ is treated as a stationary process this follows from the fact that the Lindeberg condition holds for the array $T^{-1/2}\widetilde{u}_t$, $t, ..., T$ (see Hall and Heyde (1980, p. 53)).

Since this argument can clearly be extended to allow for the present specification of initial values we have

$$T^{-1/2}c_{\perp}'y_{[Ts]} = T^{-1/2}c_{\perp}'B(1)^{-1}\sum_{j=1}^{[Ts]}u_j + o_p(1).$$

Since u_t is a stationary and ergodic martingale difference sequence with finite second moments a standard functional central limit theorem shows that the right hand side converges weakly to a Brownian motion (cf. Theorem 23.1 of Billingsley (1968)). To see the assertion concerning the covariance matrix, use the definitions to conclude that

$$\begin{aligned} c_{\perp}'B(1)^{-1} &= c_{\perp}'Q^{-1}\left[-\alpha : \Gamma(1)\beta_{\perp}\right]^{-1} \\ &= \left[0 : I_{n-r}\right]\left[-\alpha : \Gamma(1)\beta_{\perp}\right]^{-1} \\ &= \left(\alpha_{\perp}'\Gamma(1)\beta_{\perp}\right)^{-1}\alpha_{\perp}' \end{aligned}$$

where the last equality follows from straightforward matrix calculus. This shows that the $((n-r)\times n)$ matrix $c_{\perp}'B(1)^{-1}$ is of rank $n-r$ and we only need to show that the covariance matrix of the process u_t is positive definite. To demonstrate this, recall the notation $X_{t-1} = \left[z_{t-1}'\cdots z_{t-p}'\right]'$ and let a be any $(n\times 1)$ vector of unit length. Then use the law of iterated expectations to conclude that

$$a'Eu_tu_t'a = Ea'H(X_{t-1})a \geq E\lambda_{\min}(H(X_{t-1})) \geq E\lambda_{\min}(H(X_{t-1}))\mathbf{1}(X_{t-1}\in\mathcal{K})$$

where $\mathcal{K}\subset R^{np}$ is a compact set. By Assumption 3(ii) the set $\mathcal{K}$ can be chosen so that the last expectation is positive. Thus, the proof is complete. ■

Proof of Corollary 6. The stated result follows because when $K_{ij} = 0$, $i = 1,...,k$, $j = 1,..,l$, we have $\rho\left(L_{n^2p^2}\left(B'\otimes B' + K_0'\right)D_{n^2p^2}\right) = \rho(B)^2$, as discussed after Theorem 1. Assumption 2, Lemma 4, and well-known properties of companion matrices imply that $\rho(B)^2 < 1$ (see e.g. Johansen (1995, p. 16)). Thus, the eigenvalue condition (6) holds and Assumption 3(iii) is automatically satisfied.■

References

Balke, N.S. and T.B. Fomby (1997). Threshold cointegration. International Economic Review 38, 627-645.

Billingsley, P. (1968). Convergence of Probability Measures. John Wiley & Sons. New York.

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

Chan, K.S., J. Petrucelli, H. Tong, and S.W. Woolford (1985). A multiple threshold AR(1) model. Journal of Applied Probability 22, 267-279.

Corradi, V., N. Swanson, and H. White (2000). Testing for stationarity-ergodicity and for comovements between nonlinear discrete time Markov processes. Journal of Econometrics 96, 39-73.

De Jong, R. (2001). Nonlinear minimization estimators in the presence of cointegrating relations. Journal of Econometrics (forthcoming).

Doukham, P. (1994). Mixing. Springer-Verlag. New York.

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

Granger, C.W.J. (2001). An overview of nonlinear macroeconometric empirical models. Macroeconomic Dynamics (forthcoming).

Granger, C.W.J. and N. Haldrup (1997). Separation in cointegrated systems, long memory components and common trends. Oxford Bulletin of Economics and Statistics 59, 449-463.

Granger, C.W.J. and N. Swanson (1996). Future developments in the study of cointegrated variables. Oxford Bulletin of Economics and Statistics 58, 537-553.

Granger, C.W.J. and T. Teräsvirta (1993). Modeling Nonlinear Economic Relationships. Oxford University Press. Oxford.

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

Hall, P. and C.C. Heyde (1980). Martingale Limit Theory and Its Application. Academic Press. New York.

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 A. Rahbek (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.

Johansen, S. (1995). Likelihood Based Inference on Cointegration in the Vector Autoregressive Model. Oxford University Press. Oxford.

Lanne, M. and P. Saikkonen (2001). Threshold autoregressions for strongly autocorrelated time series. Journal of Business and Economic Statistics (forthcoming).

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

Lu, Z. and Z. Jiang (2001). L_1 geometric ergodicity of a multivariate nonlinear AR model with an ARCH term. Statistics and Probability Letters 51, 121-130.

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

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

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

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

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

van Dijk, D., T. Teräsvirta and P.H. Frances (2001). Smooth transition autoregressive models - A survey of recent developments. SSE/EFI Working Paper Series in Economics and Finance, No. 380. Stockholm School of Economics