Econometric Theory, **24**, 2008, 294–318. Printed in the United States of America.
DOI: 10.1017/S0266466608080122

STABILITY OF REGIME SWITCHING ERROR CORRECTION MODELS UNDER LINEAR COINTEGRATION

PENTTI SAIKKONEN
University of Helsinki

The paper obtains conditions that ensure stationarity of linear long-run equilibrium relations and differenced observations in vector autoregressive error correction models with nonlinear short-run dynamics. The considered models include various threshold error correction models and their smooth transition counterparts. These models assume that the form of the short-run dynamics depends on values of observable transition functions that determine the regime in which the considered process evolves. In related models studied in the paper the transition functions are unobservable. These models are obtained by making the transition functions of threshold error correction models dependent on an unobservable random term. Previous stationarity conditions obtained for these kinds of regime switching error correction models are extended by using recent developments on nonlinear autoregressive models based on the theory of Markov chains and the concept of joint spectral radius of a set of square matrices. In addition to stationarity, existence of second-order moments and beta mixing is also established. The results of the paper enhance the understanding of the considered nonlinear error correction models and pave the way for the development of their asymptotic estimation and testing theory.

1. INTRODUCTION

After its invention in the 1980s, the theory of cointegration developed rapidly, and the link between cointegration and linear vector autoregressive (VAR) error correction models was found to be particularly useful. More recently, nonlinear extensions of these models have also been considered. We refer to Bec and Rahbek (2004) and Kapetanios, Shin, and Snell (2006) for up-to-date discussions on the relevant literature of nonlinear cointegration and nonlinear error correction models.

Since the late 1990s, interest in nonlinear error correction models has been steadily growing, but the theory of these models is still rather undeveloped.

Financial support from the Research Unit of Economic Structures and Growth (RUESG) in the University of Helsinki and the Yrjö Jahnsson Foundation is gratefully acknowledged. The author thanks Anders Rahbek for stimulating discussions on the topic of this paper and Helmut Lütkepohl and an anonymous referee for useful comments. Address correspondence to Pentti Saikkonen, 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.

 © 2008 Cambridge University Press 0266-4666/08 $15.00

https://doi.org/10.1017/S0266466608080122 Published online by Cambridge University Press

The theory of linear error correction models was greatly advanced by the early development of a basic structural result often referred to as Granger's representation theorem (see Johansen, 1995, p. 49). This theorem provides conditions under which a linear VAR process is nonstationary and integrated of order one (or I(1)) but some of its linear combinations are stationary and can be interpreted as long-run equilibrium relations. In addition to its theoretical importance this result is also of practical significance because it links the linear error correction model to the concept of cointegration. For the theory and understanding of nonlinear error correction models analogues of this result are of utmost importance. Such results have recently been obtained by Corradi, Swanson, and White (2000), Escribano and Mira (2002), Bec and Rahbek (2004), Escribano (2004), and Saikkonen (2005). The purpose of this paper is to provide extensions of this previous work.

Threshold error correction models introduced by Balke and Fomby (1997) have probably been the most popular nonlinear error correction models considered so far, and recently their smooth transition counterparts have also attracted attention (see Kapetanios et al., 2006, and references therein). A smooth transition model assumes that the abrupt regime switches implied by a threshold model are gradual. In both models the regime switches are determined by observable transition functions. An unobservable transition function results when the threshold specification is modified by making its transition function dependent on an unobservable random term. Modifications of threshold error correction models based on this idea have been studied by Bec and Rahbek (2004) and Saikkonen (2005), who also provide references of earlier time series models based on the same idea. In this paper, fairly general versions of all these nonlinear error correction models, characterized by regime switches, will be considered, and conditions that ensure stationarity of the long-run equilibrium relations and differenced observations will be provided. Partial extensions of Granger's representation theorem are thereby obtained. The extensions are only partial because the precise I(1) nature of the considered process remains to be worked out.

Our work is closely related to that of Bec and Rahbek (2004) and Saikkonen (2005). Following these authors we use the theory of Markov chains and the concept of geometric ergodicity. However, we can allow for more general switching mechanisms or more general forms of nonlinearity than previously. This is achieved by exploiting the recent advances of Liebscher (2005) to establish geometric ergodicity and, hence, stationarity in nonlinear autoregressive models. These advances are based on a novel idea of using the concept of joint spectral radius of a set of square matrices to extend previous results on geometric ergodicity in nonlinear autoregressive models.

The paper proceeds as follows. The nonlinear error correction models studied in the paper are first introduced in Section 2, after which the main results are presented in Section 3. Section 4 concludes. Proofs are contained in the Appendix.

https://doi.org/10.1017/S0266466608080122 Published online by Cambridge University Press

A few words on the notation. The symbol L is used for the usual lag operator, whereas $\Delta = 1 - L$ is the corresponding differencing operator. The trace and determinant of the (square) matrix A are denoted by $\mathrm{tr}(A)$ and $\det(A)$, respectively. If A is positive semidefinite its unique positive semidefinite square root is denoted by $A^{1/2}$. The rank of the $k \times l$ matrix B is signified by $\mathrm{rk}(B)$, and the orthogonal complement of B is a $k \times (k - l)$ matrix $B_\perp$ of full column rank and such that $B'B_\perp = 0$ $(k > l)$. An identity matrix of order n is denoted by I_n. Finally, $\|\cdot\|$ is used for the euclidean norm, and $\mathbf{1}(\cdot)$ signifies the indicator function.

2. REGIME SWITCHING ERROR CORRECTION MODELS

2.1. Illustrative Examples

For illustrative purposes this section provides an informal discussion of simple special cases of the general nonlinear error correction model to be considered in the paper. In these examples only a single long-run equilibrium relation is assumed, and the nonlinearity manifests itself as a regime dependent adjustment to disequilibrium errors.

Let y_t $(t = 1, 2, \ldots)$ be an n-dimensional time series that can be modeled by the threshold error correction model

$$\Delta y_t = \sum_{s=1}^{3} \mathbf{1}(\beta' y_{t-1} \in R_s)\alpha_s \beta' y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + \Omega^{1/2}\varepsilon_t, \tag{1}$$

where α_s $(n \times 1)$, β $(n \times 1)$, Γ_j $(n \times n)$, and Ω $(n \times n)$ are parameters with β nonzero and Ω positive definite and $R_1 = (-\infty, c_1]$, $R_2 = (c_1, c_2]$, and $R_3 = (c_2, \infty)$. Moreover, $\varepsilon_t \sim i.i.d.(0, I_n)$, that is, a sequence of independent and identically distributed (i.i.d.) n-dimensional random vectors with zero mean and identity covariance matrix.

Thus, the model assumes three regimes with (possibly) different adjustment parameters. When the disequilibrium error $\beta' y_{t-1}$ is in regime R_s a conventional linear error correction model with adjustment parameter α_s applies. An abrupt change in the adjustment parameter results when the disequilibrium error enters another regime. For simplicity and in line with some recent applications the number of regimes has been restricted to three. Bec and Rahbek (2004) consider the case where $\alpha_1 = \alpha_3$ is nonzero and provide conditions under which an analogue of Granger's representation theorem holds, implying stationarity of the linear combination $\beta' y_t$ and the differences Δy_t. The same conclusion can also be inferred from a general result in Saikkonen (2005).

When $\alpha_1 \neq \alpha_3$ stationarity of $\beta' y_t$ and Δy_t cannot be obtained from the results referred to earlier. The purpose of this paper is to provide extensions in this direction. In addition to threshold error correction models we also consider their smooth transition analogues. A smooth transition analogue of the threshold error

https://doi.org/10.1017/S0266466608080122 Published online by Cambridge University Press

correction model (1) is obtained by replacing the indicator function by a suitable smooth function. For instance, consider the logistic function

$$L_s(x) = (1 + \exp\{-\gamma(x - c_s)\})^{-1} \qquad (x \in \mathbb{R}),$$

where $\gamma > 0$, $s = 1, 2$, and $c_1 < c_2$. Instead of the threshold error correction model (1) one may wish to consider the model

$$\Delta y_t = \sum_{s=1}^{3} h_s(\beta' y_{t-1})\alpha_s \beta' y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + \Omega^{1/2}\varepsilon_t, \tag{2}$$

where

$$h_s(\beta' y_{t-1}) = \begin{cases} 1 - L_1(\beta' y_{t-1}), & s = 1 \\ L_1(\beta' y_{t-1}) - L_2(\beta' y_{t-1}), & s = 2 \\ L_2(\beta' y_{t-1}), & s = 3 \end{cases}$$

and the notation is otherwise as in (1). Because of the smoothness of the functions h_s the regime switches implied by this model take place gradually over a longer period of time as opposed to the abrupt regime switches implied by the threshold specification. The time it takes for the regime switches to realize depends on the value of the parameter γ. When the value of γ increases the regime switches get faster, and the threshold error correction model (1) is obtained in the limit as $\gamma \to \infty$. If $\alpha_1 = \alpha_3$ is nonzero the stationarity of $\beta' y_t$ and Δy_t can be concluded from Saikkonen (2005), but when $\alpha_1 \neq \alpha_3$ no similar result is available.

Smooth transition models provide one way to modify threshold models. Another modification is obtained by adding randomness to the regime switches of threshold models. Thus, instead of the threshold error correction model (1) one may consider

$$\Delta y_t = \sum_{s=1}^{3} \mathbf{1}(\beta' y_{t-1} + \eta_t \in R_s)\alpha_s \beta' y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + \Omega^{1/2}\varepsilon_t, \tag{3}$$

where η_t is an unobservable i.i.d. sequence totally independent of the sequence ε_t and the remaining notation is as in (1). Thus, the only difference between this model and the threshold error correction model (1) is the occurrence of the random error η_t in the indicator function. When $\beta' y_{t-1} + \eta_t \in R_s$ a linear error correction model with adjustment parameter α_s applies. However, given $\beta' y_{t-1}$ this only happens with some probability between zero and one. This probability depends on the value of the disequilibrium error $\beta' y_{t-1}$ and the probability distribution of the random variable η_t. For instance, when the value of $\beta' y_{t-1}$ is close to the value of the threshold parameter c_1 both of the events $\{\beta' y_{t-1} + \eta_t \in R_1\}$ and $\{\beta' y_{t-1} + \eta_t \in R_2\}$ can have a nonnegligible probability to realize. Thus, the regime switching mechanism is more flexible than in the corre-

https://doi.org/10.1017/S0266466608080122 Published online by Cambridge University Press

sponding threshold specification where $\eta_t = 0$ is assumed and one of the aforementioned probabilities is always zero.

Error correction models of the type (3) have recently been considered by Bec and Rahbek (2004) and Saikkonen (2005). As shown by these authors the conditional distribution of Δy_t given the past of the process is a mixture distribution, and consequently one may refer to the model as a mixture error correction model. On the other hand, model (3) implies that the conditional expectation of Δy_t given the past of the process is similar to that in the smooth transition error correction model (2), and therefore this model may also be seen as a smoothed version of the threshold error correction model. If $\alpha_1 = \alpha_3$ is nonzero the stationarity of $\beta' y_t$ and Δy_t can again be obtained from Bec and Rahbek (2004) and Saikkonen (2005), but these results do not apply when $\alpha_1 \neq \alpha_3$, a case to be covered in this paper.

In addition to the case $\alpha_1 \neq \alpha_3$ neither the results of Bec and Rahbek (2004) nor those of Saikkonen (2005) apply when $\alpha_1 = \alpha_2 \neq \alpha_3$ or $\alpha_1 \neq \alpha_2 = \alpha_3$, that is, when we have a model with two regimes and a single threshold. The extensions obtained in this paper can also be used for such models.

Although various regime switching error correction models have arguably been the most commonly used nonlinear error correction models considered so far, other specifications may also be of interest. For instance, Escribano (2004) suggests modeling the adjustment to the disequilibrium error $\beta' y_{t-1}$ by a rational function with certain restrictions imposed on the degrees and coefficients of the numerator and denominator. Escribano (2004) studies the stability of this kind of nonlinear error correction model in a single equation context. The results of this paper cannot be directly used for models of this type.

2.2. General Regime Switching Error Correction Model

We shall now introduce a general nonlinear error correction model that encompasses the models discussed in the previous section. As before, y_t $(t = 1, 2, \ldots)$ is an n-dimensional stochastic process and $\varepsilon_t \sim i.i.d.(0, I_n)$ an n-dimensional error term. The scalar valued i.i.d. error term η_t used in model (3) is generalized to a vector of dimension l but is otherwise as before. The model is defined by

$$\Delta y_t = \sum_{s=1}^{m} h_s(S' Y_{t-1}, \eta_t) \left(\mu_s + \alpha_s \beta' y_{t-1} + \sum_{j=1}^{p-1} \Gamma_{sj} \Delta y_{t-j} \right) + \left(\sum_{s=1}^{m} h_s(S' Y_{t-1}, \eta_t) \Omega_s \right)^{1/2} \varepsilon_t, \tag{4}$$

where α_s $(n \times r)$, β $(n \times r)$, Γ_{sj} $(n \times n)$, μ_s $(n \times 1)$, and Ω_s $(n \times n)$ are parameters with $\mathrm{rk}(\beta) = r$ and Ω_s positive definite. Moreover, $Y_{t-1} = [(\beta' y_{t-1})' \ \Delta y'_{t-1} \ \ldots \ \Delta y'_{t-p+1}]'$ and S $((n(p-1) + r) \times k)$ is a (known) non-

https://doi.org/10.1017/S0266466608080122 Published online by Cambridge University Press

random matrix of full column rank used to determine linear combinations of the vector Y_{t-1} on which the transition functions h_s: $\mathbb{R}^{k+l} \to [0,1]$ depend. The matrix S could be imbedded in the functions h_s, but its explicit use is convenient. Typically S' is a selection matrix, but this is not necessary. Properties required for the functions h_s will be described shortly.

If $m = 1$, $h_1 \equiv 1$, and the parameter matrix α_1 is of full column rank model (4) reduces to a standard linear error correction model. If $m > 1$ extensions of the models discussed in the previous section are obtained with suitable choices of the functions h_s. For instance, a general threshold error correction model is obtained by choosing

$$h_s(S'Y_{t-1}, \eta_t) = \mathbf{1}(\varphi_d(S'Y_{t-1}) \in R_s), \tag{5}$$

where φ_d: $\mathbb{R}^k \to \mathbb{R}^d$ is a (Borel) measurable function and the sets $R_s \subseteq \mathbb{R}^d$ form a partition of $\mathbb{R}^d$. The transition function used in (1) is obtained by setting $r = k = d = 1$ and $m = 3$ and by choosing $S' = [1 \quad 0]$, $\varphi_1(S'Y_{t-1}) = \beta' y_{t-1}$, and $R_1 = (-\infty, c_1]$, $R_2 = (c_1, c_2]$, and $R_3 = (c_2, \infty)$. Thus, S' is a row vector that selects the first component from the vector Y_{t-1}, and φ_1 is an identity function. Even with these choices the resulting special case of model (4) is more general than (1) because, in addition to the adjustment parameters α_s, also other short-run parameters and the error covariance matrix can be regime dependent and there is also a regime dependent intercept term. For $r > 1$ and $d = 1$ conceivable choices are $S' = [I_r \quad 0]$ and $\varphi_1(S'Y_{t-1}) = \|\beta' y_{t-1}\|$ resulting in a transition function considered by Bec and Rahbek (2004). Like these authors and Saikkonen (2005) we also allow for the possibility that, in addition to the disequilibrium errors $\beta' y_{t-1}$, the regime switches can be determined by the differences $\Delta y_{t-1}, \ldots, \Delta y_{t-p+1}$. Note, however, that the disequilibrium errors $\beta' y_{t-1}$ are supposed to be linear and the same in each regime. This feature is common to all the models considered in the paper.

A smooth transition analogue of the threshold error correction model discussed previously can be obtained in the same way as in the example of the previous section. For instance, suppose that in (5) $d = 1$ and define

$$h_s(S'Y_{t-1}, \eta_t) = \begin{cases} 1 - L_1(\varphi_1(S'Y_{t-1})), & s = 1 \\ L_{s-1}(\varphi_1(S'Y_{t-1})) - L_s(\varphi_1(S'Y_{t-1})), & s = 2, \ldots, m-1 \\ L_{m-1}(\varphi_1(S'Y_{t-1})), & s = m, \end{cases} \tag{6}$$

where we now have $m - 1$ logistic functions and threshold parameters $c_1 < \cdots < c_{m-1}$. Conceivable choices are again $S' = [I_r \quad 0]$ and $\varphi_1(S'Y_{t-1}) = \|\beta' y_{t-1}\|$.

Note that the equality $h_1(Y_{t-1}, \eta_t) + \cdots + h_m(Y_{t-1}, \eta_t) = 1$ holds in both of the two cases (5) and (6). This property is convenient for our developments and will therefore be assumed even generally. Among other things it implies

https://doi.org/10.1017/S0266466608080122 Published online by Cambridge University Press

that the commonly used assumption of conditionally homoskedastic errors in smooth transition error correction models can simply be obtained by assuming that the covariance matrices Ω_s, $s = 1, \ldots, m$, are equal.

A general extension of the mixture error correction model (3) is obtained by modifying equation (5) to the form

$$h_s(S'Y_{t-1}, \eta_t) = \mathbf{1}(\varphi_d(S'Y_{t-1}, \eta_t) \in R_s), \tag{7}$$

where the domain of the (Borel) measurable function φ_d is now $\mathbb{R}^{k+l}$ and the notation is otherwise as in (5). Similarly to the examples discussed in the previous section, the only difference between this choice and its counterpart (5) used in the threshold specification is the occurrence of the random error η_t, which now appears as an argument of the function φ_d. Assuming $d = 1$ one may have $S' = [I_r \;\; 0]$ and $\varphi_1(S'Y_{t-1}, \eta_t) = \|\beta' y_{t-1}\| + \eta_t$, which gives a mixture version of the threshold specification discussed after equation (5). Even generally, the interpretation of the model defined by equations (4) and (7) and its relation to the corresponding threshold specification is basically the same as in the examples of the previous section. In what follows, the specification (7) will always be assumed when the functions h_s are supposed to be dependent on the error term η_t.

It will be convenient to generalize model (4) slightly. To this end, let I_1 be a subset of $\{1, \ldots, m\}$ and $\mathsf{I}_2 = \{1, \ldots, m\} \backslash \mathsf{I}_1$, the complement of I_1 in $\{1, \ldots, m\}$. Choose an $s_1 \in \mathsf{I}_1$ and define

$$\bar{\alpha}_s = \begin{cases} \alpha_s, & \text{if } s \in \mathsf{I}_1 \\ \alpha_{s_1}, & \text{if } s \in \mathsf{I}_2 \end{cases} \quad \text{and} \quad \bar{\Gamma}_{sj} = \begin{cases} \Gamma_{sj}, & \text{if } s \in \mathsf{I}_1 \\ \Gamma_{s_1 j}, & \text{if } s \in \mathsf{I}_2. \end{cases}$$

Model (4) can now be written as

$$\Delta y_t = \sum_{s=1}^{m} h_s(S'Y_{t-1}, \eta_t) \left(\bar{\alpha}_s \beta' y_{t-1} + \sum_{j=1}^{p-1} \bar{\Gamma}_{sj} \Delta y_{t-j} \right) + g(Y_{t-1}, \eta_t) + \left(\sum_{s=1}^{m} h_s(S'Y_{t-1}, \eta_t) \Omega_s \right)^{1/2} \varepsilon_t, \tag{8}$$

where $g(Y_{t-1}, \eta_t)$ is defined in such a way that the right-hand sides of (4) and (8) match. Specifically, setting $\Xi_s = [\alpha_s : \Gamma_{s1} : \cdots : \Gamma_{s,p-1}]$ we have

$$g(Y_{t-1}, \eta_t) = \sum_{s \in \mathsf{I}_2} h_s(S'Y_{t-1}, \eta_t)(\Xi_s - \Xi_{s_1}) Y_{t-1} + \sum_{s=1}^{m} h_s(S'Y_{t-1}, \eta_t) \mu_s. \tag{9}$$

Of course, model (8) is just a reparameterized form of the original model (4), which can be obtained from (8) by assuming that the set I_2 is empty or that $\bar{\alpha}_s = \alpha_s$, $\bar{\Gamma}_{sj} = \Gamma_{sj}$ ($s = 1, \ldots, m$, $j = 1, \ldots, p-1$), and $g(Y_{t-1}, \eta_t) = \sum_{s=1}^{m} h_s(S'Y_{t-1}, \eta_t) \mu_s$. From now on we shall consider model (8) and recall that the i.i.d. error terms ε_t ($n \times 1$) and η_t ($l \times 1$) are assumed to be totally

https://doi.org/10.1017/S0266466608080122 Published online by Cambridge University Press

independent of each other and, moreover, such that both ε_t and η_t are independent of y_{t-j}, $j > 0$. Occasionally it may be useful to be explicit about the fact that the functions h_s and g are supposed to be independent of the random error η_t. With a slight abuse of notation we then denote $h_s(\cdot, \eta_t) = h_s(\cdot)$ and $g(\cdot, \eta_t) = g(\cdot)$.

To see the motivation of the preceding extension, consider the threshold error correction model (1) and choose $\mathsf{I}_1 = \{1,3\}$. In this case $\Gamma_{sj} = \Gamma_j$ and $\mu_s = 0$ for all $s = 1,2,3$, whereas $g(Y_{t-1}) = \mathbf{1}(c_1 < \beta' y_{t-1} \leq c_2)\alpha_2 \beta' y_{t-1}$. Thus, the function $g(\cdot)$ is bounded. This is beneficial because it means that the function $g(\cdot)$ can be ignored in the critical steps of our proofs. This in turn implies that the value of the parameter vector α_2 is totally irrelevant and nothing needs to be assumed about this parameter. This widens the scope of our results. In particular, although our results require that the parameter vectors α_1 and α_3 are nonzero the parameter vector α_2 can be totally unrestricted. The process y_t may thus exhibit unit root type nonstationarity or even explosive behavior in the middle regime, and yet the disequilibrium error $\beta' y_t$ and the differences Δy_t are stationary. With appropriate assumptions this can be generalized so that only the parameter matrices α_s, $s \in \mathsf{I}_1$, are required to be of full column rank but the parameter matrices α_s, $s \in \mathsf{I}_2$, can be totally unrestricted. The models studied by Bec and Rahbek (2004) and Saikkonen (2005) already allowed for this feature but only in the special case where a single element could be included in the set I_1. In the case of the mixture model (3), where the functions h_s depend on the error term η_t, the relevant boundedness is that of the expectation $E\|g(\cdot, \eta_t)\|$. As in Saikkonen (2005), we can actually be somewhat more general than assumed here because slightly less than boundedness will suffice.[1]

In the example discussed earlier an important point is that the function $g(Y_{t-1}, \eta_t)$ depends on $S'Y_{t-1}$ and not on the whole vector Y_{t-1}. To achieve the same in the general case we have to assume that

$$(\Xi_s - \Xi_{s_1})S_\perp = 0 \quad \text{for all } s \in \mathsf{I}_2. \tag{10}$$

To see the point of this assumption, notice that we can always write $(\Xi_s - \Xi_{s_1})Y_{t-1} = (\Xi_s - \Xi_{s_1})P_S Y_{t-1} + (\Xi_s - \Xi_{s_1})P_{S_\perp} Y_{t-1}$ where the orthogonal projection matrices P_S and $P_{S_\perp}$ project vectors on the column spaces of the matrices S and $S_\perp$, respectively. When condition (10) holds we have $(\Xi_s - \Xi_{s_1})Y_{t-1} = (\Xi_s - \Xi_{s_1})S(S'S)^{-1}S'Y_{t-1}$ for all $s \in \mathsf{I}_2$. Thus, $g(Y_{t-1}, \eta_t)$ depends on the vector Y_{t-1} only through $S'Y_{t-1}$, and equation (9) becomes

$$\begin{aligned} g(Y_{t-1}, \eta_t) &= \sum_{s \in \mathsf{I}_2} h_s(S'Y_{t-1}, \eta_t)(\Xi_s - \Xi_{s_1})S(S'S)^{-1}S'Y_{t-1} \\ &\quad + \sum_{s=1}^{m} h_s(S'Y_{t-1}, \eta_t)\mu_s \\ &\overset{def}{=} g^\dagger(S'Y_{t-1}, \eta_t). \end{aligned} \tag{11}$$

https://doi.org/10.1017/S0266466608080122 Published online by Cambridge University Press

We shall now introduce assumptions needed in the discussion that follows. The first assumption concerns the error term ε_t.

Assumption 1. The distribution of the *i.i.d.*$(0, I_n)$ random vectors ε_t has a (Lebesgue) density that is bounded away from zero on compact subsets of $\mathbb{R}^n$.

From a practical point of view this assumption is innocuous. Regarding the functions h_s, it appears convenient to first assume that they are independent of the random error η_t.

Assumption 2. The (Borel) measurable functions h_s: $\mathbb{R}^k \to [0,1]$ do not depend on the error term η_t and satisfy the following conditions.

(a) $h_1(x) + \cdots + h_m(x) = 1$ for all $x \in \mathbb{R}^k$.
(b) Either the set I_2 is empty or, for all $s \in \mathsf{I}_2$, $h_s(x)x = o(\|x\|)$ as $\|x\| \to \infty$.

Assumption 2(a) is trivially satisfied by threshold error correction models and also by typical smooth transition error correction models, possibly after reparameterization. For instance, the logistic specification (6) automatically satisfies this assumption. It is worth noting, however, that because the functions h_s are assumed to take values on the unit interval many conceivable choices, including rational functions, are ruled out. If the set I_2 is empty Assumption 2(b) becomes redundant. If not and condition (10) holds it implies that the function $g^\dagger$ is dominated by a linear function when $\|x\|$ gets large or that $g^\dagger(x) = o(\|x\|)$ as $\|x\| \to \infty$. This condition is the slight weakening of the boundedness of the function g referred to before. In the threshold error correction model (1) and its extension discussed after equation (9) Assumption 2(b) clearly holds and guarantees the boundedness of the function $g^\dagger$. The same is true for the corresponding smooth transition analogues based on logistic functions.

As already mentioned, the functions h_s are supposed to be as in (7) when they also depend on the random error η_t. In this case we need to consider the probabilities

$$\pi_s(S'x) = P(\mathbf{1}(\varphi_d(S'x, \eta_t) \in R_s)), \qquad s = 1, \ldots, m. \tag{12}$$

Given $Y_{t-1} = x$, $\pi_s(S'x)$ is the conditional probability that the model defined by the regime R_s applies. Because the sets R_s form a partition of $\mathbb{R}^d$ we have $\pi_1(S'x) + \cdots + \pi_m(S'x) = 1$. If the function φ_d is independent of the random error η_t the function $\pi_s(S'x)$ is zero-one valued so that the previous threshold error correction model is obtained as a special case.

Instead of Assumption 2 we now have the following assumptions

Assumption 2′. The functions h_s are as defined in (7), and the functions π_s: $\mathbb{R}^k \to [0,1]$ satisfy the following conditions.

(a) For some $1 \le s_0 \le m$, $\pi_{s_0}(x)$ is bounded away from zero on compact subsets of $\mathbb{R}^k$.
(b) Either the set I_2 is empty or, for all $s \in \mathsf{I}_2$, $\pi_s(x)x = o(\|x\|)$ as $\|x\| \to \infty$.

https://doi.org/10.1017/S0266466608080122 Published online by Cambridge University Press

Assumption 2′(a) is technical and makes it possible to apply available results in the proofs. It appears reasonably mild and expected to hold if the random vector η_t has a density with properties similar to those required for the density of ε_t in Assumption 1. Assumption 2′(b) is entirely similar to Assumption 2(b). When the set I_2 is nonempty and condition (10) holds it implies that $E\|g^{\dagger}(x,\eta_t)\| = o(\|x\|)$ as $\|x\| \to \infty$ but not necessarily boundedness of this function.

Next define the polynomial matrices

$$A_s(z) = (1-z)I_n - \alpha_s\beta' z - \sum_{j=1}^{p-1} \Gamma_{sj}(1-z)z^j, \qquad s = 1,\ldots,m,$$

and make the following assumption.

Assumption 3. For all $s \in \mathsf{I}_1$, the polynomial matrices $A_s(z)$ satisfy the following three conditions.

(a) If $\det A_s(z) = 0$ then $|z| > 1$ or $z = 1$.
(b) The $n \times r$ matrices α_s and β are of full column rank.
(c) The number of unit roots in condition (a) equals $n - r$.

Thus, we assume that the polynomial matrices $A_s(z)$, $s \in \mathsf{I}_1$, satisfy conditions required for the polynomial matrix of a conventional linear VAR error correction model. In that special case, these conditions guarantee the validity of Granger's representation theorem (see Johansen, 1995, Thm. 4.2 and Cor. 4.3). The fact that similar conditions are only assumed for $s \in \mathsf{I}_1$ instead of all $s \in \{1,\ldots,m\}$ is related to the previously mentioned fact that the parameter matrices α_s, $s \in \mathsf{I}_2$, are not required to be of full column rank.

Because we shall work with model (8) we also define the polynomial matrices

$$\bar{A}_s(z) = (1-z)I_n - \bar{\alpha}_s\beta' z - \sum_{j=1}^{p-1} \bar{\Gamma}_{sj}(1-z)z^j, \qquad s = 1,\ldots,m.$$

By the definitions of $\bar{\alpha}_s$ and $\bar{\Gamma}_{sj}$, the polynomial matrices $\bar{A}_s(z)$ satisfy the three conditions of Assumption 3 for all $s = 1,\ldots,m$. In addition to condition (10) and Assumption 3 a further condition has to be imposed on the parameters of the model. This condition will be introduced in the next section. It actually implies Assumption 3, but it is convenient to state both.

3. RESULTS

Following Saikkonen (2005) simple transformations of the polynomial matrices $\bar{A}_s(z)$ are employed. To this end, let c $(n \times r)$ be a matrix of rank r and such that $c'\beta = I_r$. One can choose $c = \beta(\beta'\beta)^{-1}$, but any other choice will also do. Then the matrix $Q = [\beta : c_\perp]'$ is nonsingular, and its inverse is $Q^{-1} = [c : \beta_\perp]$ where $c_\perp'\beta_\perp = I_{n-r}$. We also define the $n \times n$ polynomial matrices

https://doi.org/10.1017/S0266466608080122 Published online by Cambridge University Press

$P(z) = [\beta : (1 - z)c_\perp]'$ and $\bar{\Gamma}_s(z) = I_n - \sum_{j=1}^{p-1} \bar{\Gamma}_{sj} z^j$. By Assumption 3 and Lemma 1 of Saikkonen (2005) we can write

$$\bar{A}_s(z) = Q^{-1}\bar{B}_s^*(z)P(z), \qquad s = 1, \ldots, m, \tag{13}$$

where $\bar{B}_s^*(z) = Q[\bar{\Gamma}_s(z)c(1 - z) - \bar{\alpha}_s z : \bar{\Gamma}_s(z)\beta_\perp]$ is a polynomial matrix such that $\bar{B}_s^*(0) = I_n$ and $\det \bar{B}_s^*(z) \neq 0$, $|z| \leq 1$. This result is now used to transform the model.

Recall that $h_s(\cdot, \eta_t)$ stands for either $h_s(\cdot)$ or $\mathbf{1}(\varphi_d(\cdot, \eta_t) \in R_s)$. Using condition (10) and the definition of the function $g^\dagger$ (see (11)) we can write model (8) as

$$\sum_{s=1}^{m} h_s(S'Y_{t-1}, \eta_t)\bar{A}_s(L)y_t = g^\dagger(S'Y_{t-1}, \eta_t) + \left(\sum_{s=1}^{m} h_s(S'Y_{t-1}, \eta_t)\Omega_s\right)^{1/2} \varepsilon_t, \tag{14}$$

where use has been made of the fact that, by assumption, $\sum_{s=1}^{m} h_s(S'Y_{t-1}, \eta_t) = 1$. Transform the process y_t as

$$z_t = Q^{-1}P(L)y_t = c\beta' y_t + \beta_\perp c_\perp' \Delta y_t \tag{15}$$

and observe that the vectors $Y_t = [(\beta' y_t)' \ \Delta y_t' \ \ldots \ \Delta y_{t-p+2}']'$ ($(n(p - 1) + r) \times 1$) and $Z_t = [z_t' \ \ldots \ z_{t-p+1}']'$ ($np \times 1$) are related by

$$Y_t = JZ_t, \tag{16}$$

where the matrix J is of full row rank and given by

$$J = \begin{bmatrix} \beta' & 0 & 0 & \ldots & 0 \\ I_n & -c\beta' & 0 & \ddots & \vdots \\ 0 & I_n & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & I_n & -c\beta' \end{bmatrix} \quad ((n(p-1) + r) \times np)$$

(for details, see Saikkonen, 2005).

Now define the polynomial matrices $\bar{B}_s(z) = I_n - \sum_{j=1}^{p} \bar{B}_{sj} z^j = Q^{-1}\bar{B}_s^*(z)Q$ ($s = 1, \ldots, m$). By equation (13), $\bar{A}_s(z) = \bar{B}_s(z)Q^{-1}P(z)$, which combined with equation (15) shows that $\bar{A}_s(L)y_t = \bar{B}_s(L)z_t$. From this and equations (14) and (16) we therefore find that[2]

$$z_t = \sum_{s=1}^{m} h_s(S'JZ_{t-1}, \eta_t) \sum_{j=1}^{p} \bar{B}_{sj} z_{t-j} + g^\dagger(S'JZ_{t-1}, \eta_t) + \left(\sum_{s=1}^{m} h_s(S'JZ_{t-1}, \eta_t)\Omega_s\right)^{1/2} \varepsilon_t. \tag{17}$$

https://doi.org/10.1017/S0266466608080122 Published online by Cambridge University Press

As in Saikkonen (2005), we have thus shown that our nonlinear error correction model can be transformed to a nonlinear VAR model. We shall first obtain results for this nonlinear VAR model and use them to infer similar results for our nonlinear error correction model.

To be able to obtain our results, suitable assumptions have to be imposed on the matrices $\bar{B}_{sj}$ in (17). To this end, define the companion matrices

$$\bar{B}_s = \begin{bmatrix} \bar{B}_{s1} & \bar{B}_{s2} & \dots & \bar{B}_{s(p-1)} & \bar{B}_{sp} \\ I_n & 0 & \dots & 0 & 0 \\ 0 & I_n & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I_n & 0 \end{bmatrix}, \qquad s = 1, \dots, m.$$

We need the concept of joint spectral radius of a set of square matrices. If $\mathcal{A}$ is such a set of square matrices its joint spectral radius is defined by

$$\rho(\mathcal{A}) = \limsup_{j \to \infty} \left(\sup_{A \in \mathcal{A}^j} \|A\| \right)^{1/j},$$

where $\mathcal{A}^j = \{A_1 A_2 \dots A_j : A_i \in \mathcal{A},\ i = 1, \dots, j\}$ and the euclidean norm $\|\cdot\|$ could be replaced by any matrix norm (the value of $\rho(\mathcal{A})$ does not depend on the choice of this norm). If the set $\mathcal{A}$ only contains a single matrix A then the joint spectral radius of $\mathcal{A}$ coincides with the ordinary spectral radius or the largest absolute eigenvalue of A. In our applications the set $\mathcal{A}$ is always bounded, in which case the joint spectral radius can also be defined as

$$\rho(\mathcal{A}) = \limsup_{j \to \infty} \left(\sup_{A \in \mathcal{A}^j} \rho(A) \right)^{1/j}, \tag{18}$$

where $\rho(A)$ signifies the spectral radius of the matrix A. This and several other useful results about the joint spectral radius can be found in a recent paper by Liebscher (2005), which also contains further references.

We need to consider the joint spectral radius of the (finite) set of matrices $\mathcal{A}_1 = \{\bar{B}_s : s \in \mathsf{I}_1\}$ and assume that

$$\rho(\mathcal{A}_1) < 1. \tag{19}$$

Note that from the definition of the joint spectral radius or from (18) one readily finds that this condition could equivalently be given by using the coefficient matrices of $\bar{B}_s^*(z)$ instead of $\bar{B}_s(z)$ ($s \in \mathsf{I}_1$). Condition (19) implies that $\rho(\bar{B}_s) < 1$ for every $s \in \mathsf{I}_1$, which in turn is equivalent to Assumption 3. This follows from Lemma 1 of Saikkonen (2005) and the fact that the matrices $\bar{B}_s$, $s = 1, \dots, m$, are determined by the coefficient matrices of the polynomial matrices $A_s(z)$, $s \in \mathsf{I}_1$. This latter fact also makes clear that condition (19) could equivalently be stated as $\rho(\{\bar{B}_s : s = 1, \dots, m\}) < 1$. It is worth noting, how-

https://doi.org/10.1017/S0266466608080122 Published online by Cambridge University Press

ever, that even if $\rho(\bar{B}_s) < 1$ holds for every $s \in \mathsf{I}_1$ it need not be the case that $\rho(\mathcal{A}_1) < 1$ (cf. the example in Liebscher, 2005, Sect. 4). Thus, in general condition (19) is a genuine strengthening of Assumption 3. An exception is the special case where the set I_1 contains a single element s_1 because then condition (19) is equivalent to $\rho(\bar{B}_{s_1}) < 1$. This special case is relevant for the models studied by Bec and Rahbek (2004) and Saikkonen (2005) and also for conventional linear models. Thus, condition (19) appears a natural extension of these previous conditions.

We shall show that the assumptions made in the previous section combined with condition (19) guarantee the existence of initial values that make the process z_t stationary. By equation (15) we then get the stationarity of the linear combinations $\beta' y_t$ and the differences Δy_t. The proof and formulation of this result are based on the theory of Markov chains. Therefore, we first note that the process $Z_t = [z_t' \ \dots \ z_{t-p+1}']'$ can be viewed as a Markov chain. To see this, conclude from equation (17) that

$$Z_t = \sum_{s=1}^{m} h_s(S'JZ_{t-1}, \eta_t)\bar{B}_s Z_{t-1} + \Upsilon g^{\dagger}(S'JZ_{t-1}, \eta_t) + \Upsilon H(S'JZ_{t-1}, \eta_t)^{1/2}\varepsilon_t, \tag{20}$$

where $\Upsilon = [I_n : 0 : \dots : 0]'$ $(np \times n)$ and, for simplicity,

$$H(S'JZ_{t-1}, \eta_t) = \sum_{s=1}^{m} h_s(S'JZ_{t-1}, \eta_t)\Omega_s.$$

Because ε_t and η_t are independent of Z_{t-1} by assumption this shows that Z_t is a Markov chain on $\mathbb{R}^{np}$.

We shall use the concept of V-geometric ergodicity of a Markov chain (see Meyn and Tweedie, 1993, p. 355). Here V is a real valued (Borel) 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 $\boldsymbol{\pi}$ on the Borel sets of $\mathbb{R}^{np}$ and constants $\varrho < 1$ and $M_x < \infty$ (depending on x) such that

$$\sup_{v:|v|\leq V} \left| E(v(Z_t)|Z_0 = x) - \int_{\mathbb{R}^{np}} \boldsymbol{\pi}(dy)v(y) \right| \leq \varrho^t M_x \quad \text{for all } x \in \mathbb{R}^{np} \quad \text{and} \quad \text{all } t \geq 1. \tag{21}$$

The definition also assumes that the function V is integrable with respect to the probability measure $\boldsymbol{\pi}$. 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 $P^t(x, \cdot)$ defined on the Borel sets of $\mathbb{R}^{np}$ by $P^t(x, A) = P(Z_t \in A | Z_0 = x)$ converges at a geometric rate and for all $x \in \mathbb{R}^{np}$ to the probability measure $\boldsymbol{\pi}(\cdot)$ with respect to the total variation norm.

https://doi.org/10.1017/S0266466608080122 Published online by Cambridge University Press

Geometric ergodicity implies stationarity of the process Z_t if the distribution of the initial value Z_0 is defined by the probability measure $\boldsymbol{\pi}$ (see Meyn and Tweedie, 1993, pp. 230–231). Of course, in this case the process z_t is also stationary. Another convenient feature of V-geometric ergodicity is that it implies existence of moments. For instance, if the Markov chain Z_t is initialized from the stationary distribution, existence of the expectation of $v(Z_t)$ for all v such that $|v(\cdot)| \leq V(\cdot)$ follows immediately from the definition of V-geometric ergodicity. Of course, it is a direct consequence of (21) that, for such a v, the conditional expectation $E(v(Z_t)|Z_0 = z)$ converges at a geometric rate to the corresponding expectation taken with respect to the stationary distribution. Finally, in the case of stationary initial values geometric ergodicity is equivalent to β-mixing with geometrically decaying mixing numbers, and, with appropriate conditions on initial values, it implies the previously mentioned β-mixing property even when the initial values are nonstationary (see Liebscher, 2005). As is well known, β-mixing implies strong mixing and that conventional limit theorems apply (for definitions of these mixing concepts, see, e.g., Doukhan, 1994).

Now we can state the main result of this paper. Its proof is based on a slight modification of an idea due to Liebscher (2005). This modification makes it possible to obtain $(1 + \|x\|^2)$-geometric ergodicity and, hence, existence of second moments instead of mere geometric ergodicity.

THEOREM 1. *Let $y_t (t = 1, 2, \ldots)$ be a stochastic process generated by (8) and z_t as defined in (15). Suppose that Assumptions 1 and 2, or alternatively 2′, are satisfied. Then, if conditions (10) and (19) also hold the process $Z_t = [z_t' \ \ldots \ z_{t-p+1}']'$ is a V-geometrically ergodic Markov chain with $V(x) = 1 + \|x\|^2$. Thus, there exists a choice of initial values $z_{-p+1}, \ldots, z_0$ such that the process z_t is both strictly stationary and second-order stationary and, moreover, β-mixing with geometrically decaying mixing numbers. Hence, there exists a choice of initial values $y_{-p}, \ldots, y_0$ such that the same is true for the process $[(\beta' y_t)' \ \ (c_\perp' \Delta y_t)']'$.*

Theorem 1 shows that in our nonlinear context the linear combinations $\beta' y_t$ and the differences Δy_t are stationary just as in the standard linear case. Special cases are obtained with different choices of the matrix S. The most interesting special case probably assumes that the functions h_s only depend on the disequilibrium errors $\beta' y_{t-1}$. Then $S' = [I_r \ \ 0]$, and, to meet condition (10), we have to assume that $\Gamma_{sj} = \Gamma_{s_1 j}$ $(j = 1, \ldots, p-1)$ holds for some $s_1 \in \mathsf{I}_1$ and all $s \in \mathsf{I}_2$. Of course, these restrictions also need to be taken into account in condition (19). Note that equality of Γ_{sj} $(j = 1, \ldots, p-1)$ is not required for all $s = 1, \ldots, m$ although this may be usual in practice and necessarily holds when the set I_1 only contains a single element.

The preceding discussion particularly shows how to obtain sufficient conditions for the stationarity of $\beta' y_t$ and Δy_t in models (1)–(3) discussed in

https://doi.org/10.1017/S0266466608080122 Published online by Cambridge University Press

Section 2 and also their extensions obtained by allowing for regime dependent intercept terms and error covariance matrices. In these cases the index set I_1 is defined by $\mathsf{I}_1 = \{1,3\}$. Omitting the middle regime gives similar sufficient conditions for the stationarity of $\beta' y_t$ and Δy_t in the corresponding two-regime models with only one threshold discussed at the end of Section 2. Neither of these cases can be handled with the previous results of Bec and Rahbek (2004) and Saikkonen (2005).

Theorem 1 does not provide a full extension of Granger's representation theorem. For that we would need to show that the levels y_t are I(1) or, more precisely, that there are exactly $n - r$ linearly independent linear combinations of y_t that form an I(1) process, possibly around a deterministic trend. Such results are obtained by Bec and Rahbek (2004) and Saikkonen (2005) but only in special cases where the index set I_1 contains a single element. As equations (8) and (11) show, the nonlinearity is then only due to the function $g^{\dagger}(\cdot, \eta_t)$, which is dominated by a linear function in the sense that $E\|g^{\dagger}(x, \eta_t)\| = o(\|x\|)$ as $\|x\| \to \infty$ (see the discussions after Assumptions 2(b) and 2′(b)). The nonlinearity is therefore "small," and methods similar to those used in the linear case can be applied to show the existence of exactly $n - r$ (linearly independent) I(1) components. Obtaining similar results in the present case is more difficult because considerably more complicated nonlinearity is allowed.

The definition of an I(1) process used by Bec and Rahbek (2004) and Saikkonen (2005) is that, possibly after appropriate centering, the random vector $T^{-1/2} y_{[Ts]}$ obeys a functional central limit theorem, that is, converges weakly to a Brownian motion. The rank of the limiting Brownian motion equals the rank of the long-run covariance matrix of the differenced process Δy_t and therefore determines the number of searched I(1) components. However, in our context Δy_t can be a very complicated nonlinear process, and therefore the rank of its long-run covariance matrix is not easy to find. The functional central limit theorem itself does not pose problems because it is implied by Theorem 1. Because $\beta' y_t$ is stationary the number of I(1) components is obviously at most $n - r$, but whether it is $n - r$ is not obvious. This is due to the well-known fact that in nonlinear processes unit roots need not mean nonstationarity (see, e.g., Balke and Fomby, 1997, p. 630, and the references therein).[3]

We shall leave the problem of finding the number of I(1) components for future research. In practice one may try to find it by simulation. For given values of the parameters of the model and distributions of the error terms ε_t and η_t (or only ε_t) one can simulate a long realization of the considered process and estimate the long-run covariance matrix of the differences Δy_t. Using this estimate one can then study whether the rank of the long-run covariance matrix of Δy_t is $n - r$ or some smaller value. Statistical tests may also be used here.

As already pointed out, Theorem 1 applies to models not covered in previous work. It is also known that, in some cases, the assumed condition (19) restricts the parameters of the model less than some alternative sufficient conditions for geometric ergodicity (see Liebscher, 2005, Sect. 7). Unfortunately,

https://doi.org/10.1017/S0266466608080122 Published online by Cambridge University Press

however, condition (19) is generally hard to check in practice, and its implications on the parameters of the model are difficult to illustrate. As a very simple illustration, consider the case $p = 1$. Using the previously mentioned fact that condition (19) can equivalently be expressed in terms of the coefficient matrices of $\bar{B}_s^*(z)$ one obtains the condition $\rho(\{I_r + \beta'\bar{\alpha}_s,\ s \in \mathsf{I}_1\}) < 1$. For $r > 1$ even this condition is unillustrative, but in the special case $r = 1$ it reduces to $-2 < \beta'\bar{\alpha}_s < 0$, $s \in I_1$. To simplify further, suppose that only the first components of the vectors $\bar{\alpha}_s$, denoted by $\bar{\alpha}_{s1}$ ($s \in \mathsf{I}_1$), are nonzero and normalize the first (nonzero) component of β to unity. Then the disequilibrium error only affects the first component of the process, and condition (19) becomes $-2 < \bar{\alpha}_{s1} < 0$, $s \in I_1$. In the special cases of Section 2 we have $m = 3$ and $I_1 = \{1, 3\}$. Then only the first equation of the model is error correcting with negative coefficients α_{11} and α_{31} (larger than -2) in the lowest and highest regimes, respectively, and with the forms of adjustment depending on the slightly different adjustment mechanisms of the three models (1)–(3). Whether adjustment takes place in some way in the middle regime depends on the value of the parameter α_{21}, which is not involved in condition (19) and can take any value.

In more general models an algorithm and MATLAB program provided by Gripenberg (1997) can, in principle, be used to find an arbitrarily small interval containing the joint spectral radius $\rho(\mathcal{A}_1)$. However, the needed computations can be infeasible unless the dimension of the matrices $\bar{B}_s$ ($s \in I_1$) is reasonably small. This is reflected by the known fact that the problem of checking condition (19) is NP-hard, that is, cannot be solved in a number of steps that is a polynomial function of the dimension (see Blondel and Nesterov, 2005, Sect. 6). Upper and lower bounds of the joint spectral radius are therefore of interest. As already noticed, $\max_{i \in \mathsf{I}_1} \rho(\bar{B}_i)$ provides a simple lower bound for $\rho(\mathcal{A}_1)$ that can be used as a first check of condition (19).[4] Other lower bounds and also upper bounds are discussed subsequently in the practically relevant case of two matrices.

For convenience and without loss of generality, consider the matrices $\bar{B}_1$ and $\bar{B}_2$. We denote by D_q the duplication matrix of dimension $q^2 \times \frac{1}{2}q(q+1)$ and by L_q the elimination matrix of dimension $\frac{1}{2}q(q+1) \times q^2$ (for definitions, see Lütkepohl, 1996, Chs. 9.5 and 9.6). From Theorem 5 of Blondel and Nesterov (2005) one obtains the inequalities

$$\left(\frac{1}{2}\right)^{1/2} \rho^{1/2}(\bar{B}_1^{[1]} + \bar{B}_2^{[1]}) \le \rho(\bar{B}_1, \bar{B}_2) \le \rho^{1/2}(\bar{B}_1^{[1]} + \bar{B}_2^{[1]}),$$

where $\bar{B}_i^{[1]} = L_{np}(\bar{B}_i \otimes \bar{B}_i)D_{np}$ with $\otimes$ the Kronecker product. Thus, we have an upper bound and lower bound for the joint spectral radius $\rho(\bar{B}_1, \bar{B}_2)$. The computation of these bounds only requires computing the spectral radius of a single matrix, $\bar{B}_1^{[1]} + \bar{B}_2^{[1]}$. However, for typical values of n and p the order of the matrix $\bar{B}_1^{[1]} + \bar{B}_2^{[1]}$ is much larger than that of the original matrices $\bar{B}_1$ and $\bar{B}_2$. Specifically, the matrices $\bar{B}_1$ and $\bar{B}_2$ are of order $np \times np$ whereas

https://doi.org/10.1017/S0266466608080122 Published online by Cambridge University Press

$\bar{B}_1^{[1]} + \bar{B}_2^{[1]}$ is of order $q_1 \times q_1$ where $q_1 = \frac{1}{2}np(np+1)$. This dimensionality problem becomes even more critical if better bounds are needed. To see this, set $\bar{B}_i^{[2]} = L_{q_1}(\bar{B}_i^{[1]} \otimes \bar{B}_i^{[1]})D_{q_1}$ and note that the preceding bounds can be improved to

$$\left(\frac{1}{2}\right)^{1/4} \rho^{1/4}(\bar{B}_1^{[2]} + \bar{B}_2^{[2]}) \leq \rho(\bar{B}_1, \bar{B}_2) \leq \rho^{1/4}(\bar{B}_1^{[2]} + \bar{B}_2^{[2]})$$

(see Blondel and Nesterov, 2005, discussion following Thm. 5). Further improvements are possible by continuing the preceding idea iteratively so that the jth step results in bounds of the form

$$\left(\frac{1}{2}\right)^{1/2^j} \rho^{1/2^j}(\bar{B}_1^{[j]} + \bar{B}_2^{[j]}) \leq \rho(\bar{B}_1, \bar{B}_2) \leq \rho^{1/2^j}(\bar{B}_1^{[j]} + \bar{B}_2^{[j]}).$$

Clearly, the joint spectral radius $\rho(\bar{B}_1, \bar{B}_2)$ can be approximated by arbitrary accuracy by increasing the number of steps. However, the order of the matrix whose spectral radius needs to be computed at each step increases quickly, and even for moderate values of n and p the required computations become infeasible after the first few steps. For instance, if $np = 10$ (e.g., $n = 5$ and $p = 2$) the orders of the matrices $\bar{B}_i^{[1]}$ and $\bar{B}_i^{[2]}$ are already as large as 55×55 and $1{,}540 \times 1{,}540$, respectively.

4. CONCLUSION

In this paper we have extended previous results on the stationarity of long-run equilibrium relations and differenced observations in nonlinear VAR error correction models characterized by regime switches. Our extensions are based on a recent idea of using the joint spectral radius of a (bounded) set of square matrices to establish geometric ergodicity in nonlinear autoregressive models. This approach yields results that, in some cases, are known to be less restrictive than counterparts obtained with previous alternatives, but it involves the practical difficulty that the computation of the joint spectral radius is generally hard. One may hope, however, that the active research on joint spectral radius will be able to alleviate these computational difficulties in the future (cf. Blondel and Nesterov, 2005, Sect. 8). Future research is also needed to determine theoretically the exact number of I(1) components in the considered models, which were not found in this paper. Finally, it would be interesting to explore extensions to models in which the long-run equilibrium relations are allowed to be regime dependent.

NOTES

1. Whether unit root or explosive behavior in the middle regime is reasonable or not depends on the considered application. The former was found relevant in Bec and Rahbek (2004), but the latter may be less appealing. Indeed, a referee pointed out that if the adjustment to disequilibrium

https://doi.org/10.1017/S0266466608080122 Published online by Cambridge University Press

errors is explosive in the middle regime the economically interesting equilibrium is almost never observed, even under stationarity.

2. When $n = r$, $z_t = y_t$, and, in the definition of the matrix J, $\beta = c = I_n$. When $r = 0$, $z_t = \Delta y_t$, and, in the definition of the matrix J, the first r rows are deleted and $c\beta' = I_n$ is assumed.

3. Specifically, Balke and Fomby (1997, p. 630) note that first-order scalar threshold autoregressive processes with autoregressive coefficient equal to unity in each regime are stationary (or geometrically ergodic) if the signs of the intercept terms in the highest and lowest regimes are chosen suitably. The application of Liebscher's condition based on the joint spectral radius to this special case requires that the autoregressive coefficients in the highest and lowest regimes are smaller than one in absolute value (see Liebscher, 2005, Thm. 5). This shows that, in general, Liebscher's condition is not necessary for stationarity (or geometric ergodicity).

4. One can also consider general convex combinations of the matrices $\bar{B}_s$, $s \in \mathrm{I}_1$, because it is known that $\rho(\sum_{s\in \mathrm{I}_1} a_s \bar{B}_s) \leq \rho(\mathcal{A}_1)$ for all a_s such that $\sum_{s\in \mathrm{I}_1} a_s = 1$ (see Blondel and Nesterov, 2005, Lem. 1).

REFERENCES

Balke, N.S. & T.B. Fomby (1997) Threshold cointegration. *International Economic Review* 38, 627–645.

Bec, F. & A. Rahbek (2004) Vector equilibrium correction models with non-linear discontinuous adjustments. *Econometrics Journal* 7, 628–651.

Blondel, V.D. & Y. Nesterov (2005) Computationally efficient approximations of the joint spectral radius. *SIAM Journal of Matrix Analysis* 27, 256–272.

Cline, D.B.H. & H.H. Pu (1998) Verifying irreducibility and continuity of a nonlinear time series. *Statistics and Probability Letters* 40, 139–148.

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

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

Escribano, A. (2004) Nonlinear error correction: The case of money demand in the U.K. (1878–2000). *Macroeconomic Dynamics* 8, 76–116.

Escribano, A. & S. Mira (2002) Nonlinear error correction models. *Journal of Time Series Analysis* 23, 509–522.

Gripenberg, G. (1997) Computing the joint spectral radius. *Linear Algebra and Its Applications* 234, 43–60.

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

Kapetanios, G., Y. Shin, & A. Snell (2006) Testing for cointegration in nonlinear smooth transition error correction models. *Econometric Theory* 22, 279–303.

Liebscher, E. (2005) Towards a unified approach for proving geometric ergodicity and mixing properties of nonlinear autoregressive processes. *Journal of Time Series Analysis* 26, 669–689.

Lütkepohl, H. (1996) *Handbook of Matrices*. Wiley.

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

Saikkonen, P. (2005) Stability results for nonlinear error correction models. *Journal of Econometrics* 127, 69–81.

Shih, M.-H., J.-W. Wu, & C.-T. Pang (1997) Asymptotic stability and generalized Gelfand spectral radius formula. *Linear Algebra and Its Applications* 252, 61–70.

APPENDIX

Proof of Theorem 1. The idea is to apply Theorem 15.0.1 of Meyn and Tweedie (1993), which first requires establishing irreducibility and aperiodicity of the Markov

https://doi.org/10.1017/S0266466608080122 Published online by Cambridge University Press

chain Z_t. In addition to this, identification of appropriate small sets is needed. We shall assume that the set I_2 is nonempty. If this is not the case the needed modifications are straightforward and will not be discussed. Mostly it suffices to interpret summation over $s \in \mathsf{I}_2$ as zero.

First note that a factorization similar to that obtained for the polynomial matrices $\bar{A}_s(z)$ in (13) can also be obtained for the polynomial matrices $A_s(z)$. In other words, we have $A_s(z) = Q^{-1}B_s^*(z)P(z)$, $s = 1, \ldots, m$, where, with obvious notation, $B_s^*(z) = Q[\Gamma_s(z)c(1-z) - \alpha_s z : \Gamma_s(z)\beta_\perp]$ is a polynomial matrix such that $B_s^*(0) = I_n$. However, unlike in (13) the condition $\det B_s^*(z) \neq 0$, $|z| \leq 1$, is not guaranteed to hold for all $s = 1, \ldots, m$ but only for $s \in \mathsf{I}_1$. These facts can be inferred from the assumption $\mathrm{rk}(\beta) = r$ and the proof of Lemma 1 of Saikkonen (2005) by observing that this proof only uses Assumptions 3(a) and (c) to obtain the condition on $\det B_s^*(z)$.

Using the preceding factorization instead of (13) it is straightforward to proceed as in Section 3 and show that the original model can be transformed to

$$z_t = \sum_{s=1}^{m} h_s(S'JZ_{t-1}, \eta_t)\left(\sum_{j=1}^{p} B_{sj}\, z_{t-j} + \mu_s\right) + H(S'JZ_{t-1}, \eta_t)^{1/2}\varepsilon_t, \tag{A.1}$$

where the coefficient matrices B_{js} are obtained from $B_s(z) = I_n - \sum_{j=1}^{p} B_{sj}\, z^j = Q^{-1}B_s^*(z)Q$.

Now consider the case of Assumption 2 so that in (A.1) $h_s(S'Y_{t-1}, \eta_t) = h_s(S'Y_{t-1})$ and $H(S'JZ_{t-1}) = \sum_{s=1}^{m} h_s(S'JZ_{t-1})\Omega_s$ are independent of η_t. Recall that the matrices Ω_s $(s = 1, \ldots, m)$ are assumed to be positive definite so that, by Assumption 2(a), the smallest eigenvalue of the matrix $H(S'Jx)$ is bounded away from zero on compact subsets of $\mathbb{R}^{np}$. Using this fact and Assumption 1 we find from (A.1) and Theorem 2.2(ii) of Cline and Pu (1998) that the Markov chain Z_t is an irreducible and aperiodic T-chain and consequently all compact sets of $\mathbb{R}^{np}$ are small (see Meyn and Tweedie, 1993, Thms. 5.5.7 and 6.2.5). To see that the same is true when $h_s(S'Y_{t-1}, \eta_t) = \mathbf{1}(\varphi_d(S'Y_{t-1}, \eta_t) \in R_s)$, conclude from (A.1) that the conditional distribution of z_t given $Z_{t-1} = x$ is characterized by the density

$$\phi_z(z_t|x) = \sum_{s=1}^{m} \pi_s(S'Jx)\det(\Omega_s^{-1/2})\phi_\varepsilon(\Omega_s^{-1/2}(z_t - f_s(x))),$$

where $f_s(x)$ equals $f_s(Z_{t-1}) = \sum_{j=1}^{p} B_{sj}\, z_{t-j} + \mu_s$ evaluated at $Z_{t-1} = x$ and ϕ_ε signifies the density of ε_t. Clearly,

$$\phi_z(z_t|x) \geq \pi_{s_0}(S'Jx)\det(\Omega_{s_0}^{-1/2})\phi_\varepsilon(\Omega_{s_0}^{-1/2}(z_t - f_{s_0}(x))).$$

Assumptions 1 and 2′(a) imply that the right-hand side is bounded away from zero on compact subsets of $\mathbb{R}^n \times \mathbb{R}^{np}$. Thus, the desired conclusion again follows from Cline and Pu (1998) (see Theorem 2.2(ii) or 2.3 and also Example 2.3 in that paper).

The next step is to verify an appropriate version of condition (15.3) of Meyn and Tweedie (1993, p. 355). This is first done by assuming that $h_s(S'Y_{t-1}, \eta_t) = \mathbf{1}(\varphi_d(S'Y_{t-1}, \eta_t) \in R_s)$. Our proof is based on ideas similar to those in the proof of Theorem 3 of Liebscher (2005). First, because $\rho(\mathcal{A}_1) < 1$ we can find a real number $\rho \in (\rho(\mathcal{A}_1), 1)$ and a positive integer N such that

https://doi.org/10.1017/S0266466608080122 Published online by Cambridge University Press

$$\sup_{A\in\mathcal{A}_1^N}\|A\| \le \rho^N \tag{A.2}$$

(cf. Liebscher, 2005, (33)). Next define the function q: $\mathbb{R}^{np} \to [1,\infty)$ by

$$q(x) = 1 + \|x\|^2 + \sum_{j=1}^{N-1} \rho^{-2j} \sup_{A\in\mathcal{A}_1^j}\|Ax\|^2.$$

We need to consider the conditional expectation of $q(Z_t)$ given $Z_{t-1} = x$. To this end, denote

$$G(Z_{t-1},\eta_t) = \sum_{s=1}^{m} \mathbf{1}(\varphi_d(S'JZ_{t-1},\eta_t)\in R_s)\bar{B}_s Z_{t-1} + \Upsilon g^{\dagger}(S'JZ_{t-1},\eta_t) \tag{A.3}$$

and conclude from (20) that

$$\begin{aligned} E\left(\sup_{A\in\mathcal{A}_1^j}\|AZ_t\|^2 \,|\, Z_{t-1}=x\right) &\le E\left(\sup_{A\in\mathcal{A}_1^j}\|AG(x,\eta_t)\|^2\right) \\ &\quad + E\left(\sup_{A\in\mathcal{A}_1^j}\|A\Upsilon H(S'Jx,\eta_t)^{1/2}\varepsilon_t\|^2\right) \\ &\quad + 2E\left(\sup_{A\in\mathcal{A}_1^j}|G(x,\eta_t)'A'A\Upsilon H(S'Jx,\eta_t)^{1/2}\varepsilon_t|\right). \end{aligned} \tag{A.4}$$

In what follows, c, c_1, c_2, ... will denote positive real numbers. Clearly,

$$\sup\{\|A\| : A \in \cup_{j=1}^{N}\mathcal{A}_1^j\} \le c.$$

Thus, because $H(S'Jx,\eta_t)$ is bounded and ε_t and η_t are independent the second term on the right-hand side of (A.4) satisfies

$$E\left(\sup_{A\in\mathcal{A}_1^j}\|A\Upsilon H(S'Jx,\eta_t)^{1/2}\varepsilon_t\|^2\right) \le \sup_{A\in\mathcal{A}_1^j}\|A\|^2 E\|\Upsilon H(S'Jx,\eta_t)^{1/2}\|^2\|\varepsilon_t\|^2 \le c_1. \tag{A.5}$$

Next consider the third term on the right-hand side of (A.4) and first notice that

$$\sup_{A\in\mathcal{A}_1^j}\|AG(x,\eta_t)\| \le c\|G(x,\eta_t)\| \le c_2\|x\| + c_3,$$

where the latter inequality is an immediate consequence of the definitions (see (11) and (A.3)). Thus,

$$\begin{aligned} &2E\left(\sup_{A\in\mathcal{A}_1^j}|G(x,\eta_t)'A'A\Upsilon H(S'Jx,\eta_t)^{1/2}\varepsilon_t|\right) \\ &\quad\le 2E\left(\sup_{A\in\mathcal{A}_1^j}\|AG(x,\eta_t)\|\,\|A\Upsilon H(S'Jx,\eta_t)^{1/2}\varepsilon_t\|\right) \\ &\quad\le 2(c_2\|x\| + c_3)E\left(\sup_{A\in\mathcal{A}_1^j}\|A\Upsilon H(S'Jx,\eta_t)^{1/2}\varepsilon_t\|\right). \end{aligned}$$

https://doi.org/10.1017/S0266466608080122 Published online by Cambridge University Press

As in (A.5) the last expectation can be bounded by a constant. This fact combined with (A.4) and (A.5) shows that

$$E\left(\sup_{A\in\mathcal{A}_1^j}\|AZ_t\|^2\,|\,Z_{t-1}=x\right)\le E\left(\sup_{A\in\mathcal{A}_1^j}\|AG(x,\eta_t)\|^2\right)+c_4\|x\|+c_5. \tag{A.6}$$

Now consider the first term on the right-hand side of (A.6) and conclude from (A.3) that

$$\begin{aligned}
&E\left(\sup_{A\in\mathcal{A}_1^j}\|AG(x,\eta_t)\|^2\right)\\
&\quad\le E\left(\sup_{A\in\mathcal{A}_1^j}\left\|A\sum_{s=1}^m\mathbf{1}(\varphi_d(S'Jx,\eta_t)\in R_s)\bar{B}_s x\right\|^2\right)+E\left(\sup_{A\in\mathcal{A}_1^j}\|A\Upsilon g^\dagger(S'Jx,\eta_t)\|^2\right)\\
&\qquad+2E\left(\sup_{A\in\mathcal{A}_1^j}\left|\sum_{s=1}^m\mathbf{1}(\varphi_d(S'Jx,\eta_t)\in R_s)x'\bar{B}_s'A'A\Upsilon g^\dagger(S'Jx,\eta_t)\right|\right).
\end{aligned}\tag{A.7}$$

For the second term on the right-hand side we have

$$\begin{aligned}
E\left(\sup_{A\in\mathcal{A}_1^j}\|A\Upsilon g^\dagger(S'Jx,\eta_t)\|^2\right)&\le\sup_{A\in\mathcal{A}_1^j}\|A\|^2\|\Upsilon\|^2E\|g^\dagger(S'Jx,\eta_t)\|^2\\
&\le c_6E\|g^\dagger(S'Jx,\eta_t)\|^2.
\end{aligned}$$

From this and the definition of $g^\dagger(\cdot,\eta_t)$ (see (11)) it follows that

$$\begin{aligned}
&E\left(\sup_{A\in\mathcal{A}_1^j}\|A\Upsilon g^\dagger(S'Jx,\eta_t)\|^2\right)\\
&\quad\le 2c_6E\left\|\sum_{s\in\mathrm{I}_2}\mathbf{1}(\varphi_d(S'Jx,\eta_t)\in R_s)(\Xi_s-\Xi_{s_1})S(S'S)^{-1}S'Jx\right\|^2\\
&\qquad+2c_6E\left\|\sum_{s=1}^m\mathbf{1}(\varphi_d(S'Jx,\eta_t)\in R_s)\mu_s\right\|^2\\
&\quad\le 2c_6\sum_{s\in\mathrm{I}_2}\pi_s(Jx)x'J'S(S'S)^{-1}S'(\Xi_s-\Xi_{s_1})'(\Xi_s-\Xi_{s_1})S(S'S)^{-1}S'Jx+c_7\\
&\quad\le c_8\|x\|\sum_{s\in\mathrm{I}_2}\pi_s(S'Jx)\|S'Jx\|+c_7.
\end{aligned}$$

The third term on the right-hand side of (A.7) is dominated by

$$\begin{aligned}
&2E\left(\sup_{A\in\mathcal{A}_1^j}\sum_{s=1}^m|\mathbf{1}(\varphi_d(S'Jx,\eta_t)\in R_s)x'\bar{B}_s'A'A\Upsilon g^\dagger(S'Jx,\eta_t)|\right)\\
&\quad\le 2E\left(\sup_{A\in\mathcal{A}_1^j}\|A\|^2\|\Upsilon\|\,\|x\|\|g^\dagger(S'Jx,\eta_t)\|\sum_{s=1}^m\|\bar{B}_s\|\mathbf{1}(\varphi_d(S'Jx,\eta_t)\in R_s)\right)\\
&\quad\le c_9\|x\|E\|g^\dagger(S'Jx,\eta_t)\|.
\end{aligned}$$

https://doi.org/10.1017/S0266466608080122 Published online by Cambridge University Press

By the definition of $g^{\dagger}(\cdot,\eta_t)$ we thus obtain

$$
\begin{aligned}
&2E\left(\sup_{A\in\mathcal{A}_1^j}\left|\sum_{s=1}^{m}\mathbf{1}(\varphi_d(S'Jx,\eta_t)\in R_s)x'\bar{B}_s'A'AYg^{\dagger}(S'Jx,\eta_t)\right|\right)\\
&\quad\le c_9\|x\|E\sum_{s\in\mathrm{I}_2}\mathbf{1}(\varphi_d(S'Jx,\eta_t)\in R_s)\|(\Xi_s-\Xi_{s_1})S(S'S)^{-1}\|\|S'Jx\|\\
&\qquad+c_9\|x\|E\sum_{s=1}^{m}\mathbf{1}(\varphi_d(S'Jx,\eta_t)\in R_s)\|\mu_s\|\\
&\quad\le c_{10}\|x\|\sum_{s\in\mathrm{I}_2}\pi_s(S'Jx)\|S'Jx\|+c_{11}\|x\|.
\end{aligned}
$$

Combining the preceding derivations we can bound the right-hand side of (A.7) and hence the right-hand side of (A.6) and obtain

$$
E\left(\sup_{A\in\mathcal{A}_1^j}\|AZ_t\|^2|Z_{t-1}=x\right)\le E\left(\sup_{A\in\mathcal{A}_1^j}\left\|A\sum_{s=1}^{m}\mathbf{1}(\varphi_d(S'Jx,\eta_t)\in R_s)\bar{B}_sx\right\|^2\right)+R_1(x),
$$

where

$$
R_1(x)=c_5+c_7+(c_4+c_{11})\|x\|+(c_8+c_{10})\|x\|\sum_{s\in\mathrm{I}_2}\pi_s(S'Jx)\|S'Jx\|.
$$

By the definition of $\bar{B}_s$, $\sum_{s=1}^{m}\mathbf{1}(\varphi_d(S'Jx,\eta_t)\in R_s)\bar{B}_s\in\mathcal{A}_1$ (a.s.), and therefore

$$
E\left(\sup_{A\in\mathcal{A}_1^j}\left\|A\sum_{s=1}^{m}\mathbf{1}(\varphi_d(S'Jx,\eta_t)\in R_s)\bar{B}_sx\right\|^2\right)\le\sup_{A\in\mathcal{A}_1^{j+1}}\|Ax\|^2.
$$

Hence,

$$
E\left(\sup_{A\in\mathcal{A}_1^j}\|AZ_t\|^2|Z_{t-1}=x\right)\le\sup_{A\in\mathcal{A}_1^{j+1}}\|Ax\|^2+R_1(x),\tag{A.8}
$$

and this holds for all $x\in\mathbb{R}^{np}$. It is easy to check that it also holds with $j=0$ if on the left-hand side we omit the supremum and set $A=I_{np}$. Thus, we can conclude that, for all $x\in\mathbb{R}^{np}$,

$$
\begin{aligned}
E(q(Z_t)|Z_{t-1}=x)\le 1+\sum_{j=1}^{N-1}\rho^{-2(j-1)}\sup_{A\in\mathcal{A}_1^j}\|Ax\|^2\\
+\rho^{-2(N-1)}\sup_{A\in\mathcal{A}_1^N}\|A\|^2\|x\|^2+c_{12}R_1(x).
\end{aligned}
$$

In view of (A.2) the third term on the right-hand side is bounded by $\rho^2\|x\|^2$, and it follows that, for all $x\in\mathbb{R}^{np}$,

$$
\begin{aligned}
E(q(Z_t)|Z_{t-1}=x)&\le\rho^2(q(x)-1)+1+c_{12}R_1(x)\\
&=\rho^2q(x)+R_2(x),
\end{aligned}\tag{A.9}
$$

where $R_2(x)=1-\rho^2+c_{12}R_1(x)$.

https://doi.org/10.1017/S0266466608080122 Published online by Cambridge University Press

By the definition of $q(x)$, we have $q(x) \geq 1 + \|x\|^2$, and therefore for all $x \in \mathbb{R}^{np}$,

$$E(q(Z_t)|Z_{t-1} = x) \leq \rho^2 q(x)\left(1 + \frac{R_2(x)}{\rho^2(1 + \|x\|^2)}\right).$$

From the definition of $R_1(x)$ and Assumption 2′(b) it follows that $R_1(x) = o(\|x\|^2)$ as $\|x\| \to \infty$. Hence, the same is true for $R_2(x)$, and we can find a real number M such that

$$E(q(Z_t)|Z_{t-1} = x) \leq \rho q(x) \quad \text{for } \|x\| > M.$$

On the other hand, the definition of q implies that

$$q(x) \leq 1 + \|x\|^2 + \sum_{j=1}^{N-1} \rho^{-2j} \sup_{A \in \mathcal{A}_1^j} \|A\|^2 \|x\|^2$$

$$\leq 1 + M^2 + c^2 M^2 \sum_{j=1}^{N-1} \rho^{-2j} \quad \text{for } \|x\| \leq M.$$

Thus, because $R_2(x)$ is bounded for $\|x\| \leq M$ the same is true for $E(q(Z_t)|Z_{t-1} = x)$ (see (A.9)). Hence, we can find a constant b such that

$$E(q(Z_t)|Z_{t-1} = x) \leq \rho q(x) + b\mathbf{1}(\|x\| \leq M), \qquad x \in \mathbb{R}^{np}. \tag{A.10}$$

Because compact sets are small this shows that condition (15.3) of Meyn and Tweedie (1993, p. 355) holds with the function q. Hence, Theorem 15.0.1 of the same reference implies that the Markov chain Z_t is q-geometrically ergodic and, because $q(x) \geq 1 + \|x\|^2$, also $(1 + \|x\|^2)$-geometrically ergodic, as was to be shown.

Now consider the case $h_s(S'Y_{t-1}, \eta_t) = h_s(S'Y_{t-1})$ and define the bounded set of matrices $\mathcal{A}_2 = \{\sum_{s=1}^m h_s(S'x)\bar{B}_s : x \in \mathbb{R}^{n(p-1)+r}\}$. By the definition of the matrices $\bar{B}_s$, any $A \in \mathcal{A}_2$ is a linear combination of matrices belonging to $\mathcal{A}_1$. Recall that, by assumption, $\rho(\mathcal{A}_1) < 1$. This is equivalent to the existence of a matrix norm $\|\cdot\|^*$ and a real number $\bar{\rho} \in (0,1)$ such that $\|\bar{B}_s\|^* \leq \bar{\rho}$ for all $s = 1, \ldots, m$ (see Shih, Wu, and Pang, 1997, Thm. 1; Liebscher, 2005, Thm. 1). Thus, for $A \in \mathcal{A}_2$, $\|A\|^* = \|\sum_{s=1}^m h_s(S'x)\bar{B}_s\|^* \leq \sum_{s=1}^m h_s(S'x)\|\bar{B}_s\|^* \leq \bar{\rho} \sum_{s=1}^m h_s(S'x) \leq \bar{\rho}$ and, by the same argument as before, this is equivalent to $\rho(\mathcal{A}_2) < 1$.

The preceding discussion implies that (A.2) can also be assumed with the set $\mathcal{A}_1$ replaced by $\mathcal{A}_2$. This means that the proof can be continued by following the steps in the case $h_s(S'Y_{t-1}, \eta_t) = \mathbf{1}(\varphi_d(S'Y_{t-1}, \eta_t) \in R_s)$. The function q is defined in the same way as before, and in place of $G(Z_{t-1}, \eta_t)$ we have

$$G(Z_{t-1}) = \sum_{s=1}^m h_s(S'Y_{t-1})\bar{B}_s Z_{t-1} + \Upsilon g^\dagger(S'JZ_{t-1}),$$

where

$$g^\dagger(S'JZ_{t-1}) = \sum_{s \in \mathrm{I}_2} h_s(S'JZ_{t-1})(\Xi_s - \Xi_{s_1})S(S'S)^{-1}S'JZ_{t-1} + \sum_{s=1}^m h_s(S'JZ_{t-1})\mu_s.$$

https://doi.org/10.1017/S0266466608080122 Published online by Cambridge University Press

Replacing $G(x, \eta_t)$ in (A.4) by $G(x)$ and proceeding in the same way as in the previous case it is straightforward to check that in place of (A.6) we now have

$$E\left(\sup_{A\in\mathcal{A}_2^j}\|AZ_t\|^2\,|\,Z_{t-1}=x\right)\le \sup_{A\in\mathcal{A}_2^j}\|AG(x)\|^2+c_4\|x\|+c_5, \tag{A.11}$$

where

$$\sup_{A\in\mathcal{A}_2^j}\|AG(x)\|^2\le \sup_{A\in\mathcal{A}_2^j}\left\|A\sum_{s=1}^{m}h_s(S'Jx)\bar{B}_s x\right\|^2+\sup_{A\in\mathcal{A}_2^j}\|A\Upsilon g^{\dagger}(S'Jx)\|^2$$
$$+2\sup_{A\in\mathcal{A}_2^j}\left|\sum_{s=1}^{m}h_s(S'Jx)x'\bar{B}_s'A'A\Upsilon g^{\dagger}(S'Jx)\right|. \tag{A.12}$$

By the definition of $g^{\dagger}(\cdot)$,

$$\|g^{\dagger}(S'Jx)\|\le c_{13}\sum_{s\in \mathrm{I}_2}h_s(S'Jx)\|S'Jx\|+c_{14}.$$

Hence, for the second term on the right-hand side of (A.12) we have

$$\sup_{A\in\mathcal{A}_2^j}\|A\Upsilon g^{\dagger}(S'Jx)\|^2\le \sup_{A\in\mathcal{A}_2^j}\|A\|^2\|\Upsilon\|^2\|g^{\dagger}(S'Jx)\|^2$$
$$\le c_{15}\left(\sum_{s\in \mathrm{I}_2}h_s(S'Jx)\|S'Jx\|\right)^2+c_{16}$$
$$\le c_{17}\|x\|\sum_{s\in \mathrm{I}_2}h_s(S'Jx)\|S'Jx\|+c_{16}.$$

In the last inequality we have used the Cauchy–Schwartz inequality and the assumption $0\le h_s(S'Jx)\le 1$ to get a convenient upper bound.

The third term on the right-hand side of (A.12) can be treated in the same way as the third term on the right-hand side of (A.7). We get

$$2\sup_{A\in\mathcal{A}_2^j}\left|\sum_{s=1}^{m}h_s(S'Jx)x'\bar{B}_s'A'A\Upsilon g^{\dagger}(S'Jx)\right|\le 2\sup_{A\in\mathcal{A}_2^j}\sum_{s=1}^{m}|h_s(S'Jx)x'\bar{B}_s'A'A\Upsilon g^{\dagger}(S'Jx)|$$
$$\le c_{18}\|x\|\,\|g^{\dagger}(S'Jx)\|$$
$$\le c_{19}\|x\|\sum_{s\in \mathrm{I}_2}h_s(S'Jx)\|S'Jx\|+c_{20}\|x\|.$$

Using the preceding inequalities we can bound the right-hand side of (A.12) and hence the right-hand side of (A.11). Thus, we find that, for all $x\in\mathbb{R}^{np}$,

$$E\left(\sup_{A\in\mathcal{A}_2^j}\|AZ_t\|^2\,|\,Z_{t-1}=x\right)\le \sup_{A\in\mathcal{A}_2^j}\left\|A\sum_{s=1}^{m}h_s(S'Jx)\bar{B}_s x\right\|^2+R_3(x),$$

https://doi.org/10.1017/S0266466608080122 Published online by Cambridge University Press

where

$$R_3(x) = c_5 + c_{16} + (c_4 + c_{20})\|x\| + (c_{17} + c_{19})\|x\| \sum_{s \in \mathrm{I}_2} h_s(S'Jx)\|S'Jx\|.$$

Because $\sum_{s=1}^{m} h_s(S'Jx)\bar{B}_s \in \mathcal{A}_2$ this implies that

$$E\left(\sup_{A \in \mathcal{A}_2^j} \|AZ_t\|^2 \,|\, Z_{t-1} = x \right) \leq \sup_{A \in \mathcal{A}_2^{j+1}} \|Ax\|^2 + R_3(x). \tag{A.13}$$

Obviously, (A.13) differs from (A.8) only in that $\mathcal{A}_1$ is replaced by $\mathcal{A}_2$ and that constant terms in $R_3(x)$ are different from their counterparts in $R_1(x)$. Thus, repeating the arguments following (A.8) we can show that an analogue of (A.10) holds and further that the Markov chain Z_t is $(1 + \|x\|^2)$-geometrically ergodic.

As discussed in Section 3, the assertions on stationarity and β-mixing of z_t follow from the $(1 + \|x\|^2)$-geometric ergodicity of Z_t. Finally, the results obtained for the Markov chain Z_t readily yield similar results for the process $[(\beta' y_t)' \quad (c_\perp' \Delta y_t)']'$, which is a linear transformation of z_t (see (15)). This completes the proof.

https://doi.org/10.1017/S0266466608080122 Published online by Cambridge University Press