

NONLINEAR ERROR CORRECTION MODELS



BY ALVARO ESCRIBANO AND SANTIAGO MIRA

Universidad Carlos III de Madrid and Universidad de Las Palmas de Gran Canaria

Final Version received November 2000

Abstract. The relationship between cointegration and error correction (EC) models is well characterized in a linear context, but the extension to the nonlinear context is still a challenge. Few extensions of the linear framework have been done in the context of nonlinear error correction (NEC) or asymmetric and time varying error correction models. In this paper, we propose a theoretical framework based on the concept of near epoch dependence (NED) that allows us to formally address these issues. In particular, we partially extend the Granger Representation Theorem to the nonlinear case.

Keywords. Cointegration; nonlinear error correction; near epoch dependence.

JEL. C15, C22, C32

1. INTRODUCTION

Granger (1981) introduced the concept of cointegration but it was not until Engle and Granger (1987) and Johansen (1988, 1991) that this concept achieved immense popularity among econometricians and applied economists. The great impact those papers had on the profession was due to the fact that they showed how we should work statistically with economic variables that are non-stationary, so as to avoid the problem of spurious regressions (Granger and Newbold, 1974; Phillips, 1986). Furthermore, most of the estimation and inference procedures changed dramatically from the classical statistical frameworks when dealing with variables that have unit roots and are cointegrated. By now, it is clear how to deal with integrated and cointegrated data in a linear context (Watson, 1994), but almost no research has been dedicated to the simultaneous consideration of non-stationarity and nonlinearity, even though many economist agree that those are dominant and likely properties of large amounts of economic data. How can it be possible that so little research has been dedicated to this topic? The answer is clear; it is difficult to work with nonlinear time series models within a stationary and ergodic framework and, therefore, even more difficult within a nonstationary context.

An introduction to the state of the art in econometrics relating nonlinearity and nonstationarity within a time series context can be found in Granger and Teräsvirta (1993) and Granger (1995). Those authors discussed the concepts of long-range dependence in mean and extended memory which generalize the linear

concept of integration, $I(1)$, to a nonlinear framework. The main disadvantage of those definitions is that they have no Laws of Large Numbers (LLN), nor Functional Central Limit Theorems (FCLT) associated with them and, therefore, it is hard to obtain estimation and inference results. On the other hand, there are interesting empirical macroeconomic applications where nonlinearity has been found in a non stationary context and, therefore, there is a need to justify those results econometrically. This paper starts filling this major gap with the analysis of nonlinear error correction models.

As an empirical application of nonlinear error correction (NEC) models. We have the case of the UK money demand from 1878 to 1970. Hendry and Ericsson (1991) used the NEC model suggested by Escribano (1986) in the specification of their money demand as an alternative to the linear money demands suggested by Friedman and Schwartz (1982), Hendry and Ericsson (1991) and Longbottom and Holly (1985). The variables in Hendry and Ericsson (1991) are: m , log money stock (millions); i , log real net national product; p , deflator of i ; rs , log of short term interest rate; rl , log of long term interest rate; and RS , short term interest rate. L is the lag operator such that $L^k x_t = x_{t-k}$. Let \hat{u}_t be the residuals from the cointegrating relationship estimated by OLS, then the two step approach of Engle and Granger (1987) is given by

$$\begin{aligned} \hat{u}_t = & (m \quad p \quad y)_t + 0.309 + 7RS_t \\ (1 - L)(m \quad p)_t = & 0.45(1 - L)(m \quad p)_{t-1} - (1 - L)^2(m \quad p)_{t-2} - 0.60(1 - L)p_t \\ & + 0.39(1 - L)p_{t-1} - 0.021(1 - L)rs_t - 0.062(1 - L^2)rl_t \\ & 2.55(\hat{u}_{t-1} - 0.2)\hat{u}_{t-1}^2 + 0.005 + 3.7(D1 + D3) + \epsilon_t \end{aligned}$$

where D_1 and D_3 are dummy variables for the two world wars. The term \hat{u}_{t-1} enters nonlinearly, and the nonlinear adjustment is a cubic polynomial. Other empirical examples of NECs models or nonlinear cointegration are given by Granger and Lee (1989), Balke and Fomby (1992), Burgess (1992), Kunst (1992), Granger and Swanson (1995), Escribano and Granger (1998) and Escribano and Pfann (1998).

The structure of this paper is as follows: in Section 2, we propose an alternative concept of integration, $I(0)$ and $I(1)$, which could also be extended to nonlinear cointegration. Section 3 presents some auxiliary results. In Section 4, we propose a representation theorem which relates the concept of linear cointegration to the nonlinear error correction introduced by Escribano (1986, 1987). Section 5 suggest some extensions. Section 6 presents the main conclusions.

2. DEFINITIONS

Following Lo (1991), Kwiatowski *et al.* (1992) and Stock (1994), a general concept of $I(0)$ for a sequence $\{m_t\}$ is given by the ‘high level’ condition that m_t verifies a FCLT, i.e. that

$$T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} m_t \xrightarrow{d} B(r)$$

where $B(r)$ is a Brownian motion. In a nonlinear dynamic model, this FCLT holds for functions of the exogeneous variables and underlying disturbances that have a sufficiently fading memory. The concept of mixing is appropriate to modelize the fading memory without restricting the heterogeneity of the process, and our definitions will be based on that concept, which is formalized as follows.

DEFINITION 1. (Strong mixing) *Let $\{v_t\}$ be a sequence of random variables. Let $\mathcal{F}_s^t \equiv \sigma(v_s, \dots, v_t)$ be the generated sigma algebra. Define the α mixing coefficients*

$$\alpha_m \equiv \sup_t \sup_{\{G \in \mathcal{F}_t^\infty, F \in \mathcal{F}_{t-m}^\infty\}} |\mathbf{P}(G \cap F) - \mathbf{P}(G)\mathbf{P}(F)|$$

The process $\{v_t\}$ is said to be strong mixing (also α mixing) if $\alpha_m \rightarrow 0$ as $m \rightarrow \infty$. The coefficient α_m measure the amount of dependence between events involving variables separated by at least m time periods. If $\alpha_m = O(m^{-\lambda})$ for all $\lambda < a$, then α_m is said to be of size a .

However, the mixing property is, for some purposes, a too restrictive one, since a function of a mixing sequence that depends on an infinite number of lags may not be mixing. An alternative concept is needed that allows the application of limit theorems. Different approaches to modelize these dynamics have been developed: Bierens (1981) employs the concept of stochastically stable w.r.t. an α mixing sequence; Gallant and White (1988) or Wooldridge and White (1988) employ the concept of near epoch dependence (NED) w.r.t. an α mixing sequence. Both concepts require the assumption that the exogenous variables and the disturbances are α mixing so as to provide useful results. The definition of I(0) that we are going to use is based on the concept of NED.

DEFINITION 2. (NED) *Let $\{w_t\}$ be a sequence of random variables with $E\{w_t^2\} < \infty$ for all t . It is said that $\{w_t\}$ is NED on the underlying sequence $\{v_t\}$ of size a if $\phi(n)$ is of size a , where $\phi(n)$ given by*

$$\sup_t \|w_t - E_{t-n}^{t+n}(w_t)\|_2 \equiv \phi(n)$$

where $E_{t-n}^{t+n}(w_t) = E(w_t | v_{t-n}, \dots, v_{t+n})$ and $\|\cdot\|_2$ is the L_2 norm of a random variable, defined as $E^{1/2}|\cdot|^2$.

We assume that the future values of v_t do not improve the conditional expectation of w_t , in the sense of Sims (1972), such that the forward values v_{t+r} ($r = 1, \dots, n$) are useless, but harmless. When $\phi(n)$ goes to zero at an appropriate rate, then w_t depends essentially on the recent epoch of v_t . If w_t depends on a finite number of lags of v_t then it is NED of any size. More general definitions of NED can be used – see, for instance, Davidson (1994) – but we use

the one given in Gallant and White (1988). One useful feature of NED sequences is that, under some conditions, functions of NED sequences are NED, which greatly simplifies working with NED sequences. As it was explained above, the existence of a FCLT is the central feature to characterize an $I(0)$ sequence. A simplified version of a FCIT for NED variables is as follows (Wooldridge and White, 1988; Davidson, 1994).

THEOREM 1. (FCLT for NED) *Consider the assumptions:*

- (i) $\{w_t\}$ is a mean zero sequence of random variables, uniformly L_r bounded and NED of size $\frac{1}{2}$ on an α mixing process of size $r/(r-2)$; and
- (ii) $T^{-1}E(\sum_{t=1}^T w_t)^2 \rightarrow \sigma^2$ where $0 < \sigma^2 < \infty$

Then $W_T(r) \rightarrow \sigma^2 B(r)$ where $W_T(r) = T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} w_t$ and $B(r)$ is the standard Brownian motion.

The above considerations motivate the following definition.

DEFINITION 3. *A sequence $\{w_t\}$ is $I(0)$ if it is NED on an underlying α mixing sequence $\{v_t\}$ but the sequence $\{x_t\}$ given by $x_t = \sum_{s=1}^t w_s$ is not NED on $\{v_t\}$. In this case, we will say that x_t is $I(1)$*

Notice that if x_t is $I(1)$ then Δx_t is $I(0)$. This definition excludes $I(-1)$ series as $I(0)$, like $z_t = e_t - e_{t-1}$ for α mixing sequences e_t , since in this case $\sum z_t$ is α mixing. Notice the conditions of Theorem 1 ensure a FCLT for an $I(0)$ series. The following definition of cointegration is based on the concepts presented.

DEFINITION 4. *Two $I(1)$ sequences $\{y_t\}$ and $\{x_t\}$ are (linearly) cointegrated with cointegrating vector $\beta^* = [1, \beta_{12}^*]'$, if $y_t - \beta_{12}^* x_t$ is NED on a particular α mixing sequence but $y_t - \delta_{12} x_t$ is not NED for $\delta_{12} \neq \beta_{12}^*$*

In this definition, we have assumed a normalization of the cointegrating vector β^* as $[1, \beta_{12}^*]$. Notice that this definition allows us to extend the notion of cointegration to a nonlinear context by defining the nonlinear function $g(y_t, w_t, \delta)$ as NED if and only if $\delta = \beta^*$. This approach avoids the difficulties faced by Escribano (1987) or Granger and Hallman (1991) when characterizing the time series properties of nonlinear transformations of series that are $I(0)$ or $I(1)$. The above definitions are the basis of a formulation of NEC mechanisms.

DEFINITION 5. *A NEC model of the $(n \times 1)$ and $I(1)$ vector X_t is a balanced relation between an autoregressive linear model (VAR) for the differences ΔX_t , and a nonlinear term for the lag of the levels, say $F(X_{t-1})$, plus an error term. say v_t .*

The models that we want to generalize are the VAR-EC models. The general model that we study is a NEC in the form

$$\Delta \mathbf{X}_t = \Psi_1 \Delta \mathbf{X}_{t-1} + F(\mathbf{X}_{t-1}) + \mathbf{v}_t \quad (1)$$

Notice that the linear part of the model depends on the differenced variable $\Delta \mathbf{X}_t$, whereas only the nonlinear part depends on the levels \mathbf{X}_t . In this sense, the model generalizes the VAR EC models by allowing a nonlinear error correction but keeping the linear terms in differences. Therefore, the generalization keep the linear modelling for the general specification of the model, but allows a nonlinear specification for the particular role of the correction. Recall that \mathbf{v}_t is α mixing, not NED. There is only one lag but this is not restrictive (a redefinition of \mathbf{X}_t is enough to consider more lags). The following model provides an example of generating mechanisms for NEC models. Consider the series defined as

$$\begin{aligned} \Delta x_t &= \psi_{11} \Delta x_{t-1} + v_{1t} \\ z_t = z_{t-1} + \lambda_1 \Delta x_{t-1} - \theta J_2(z_{t-1}) + (v_{1t} \quad \theta v_{2t}) \\ \Delta y_t &= \psi_{21} \Delta x_{t-1} + J_2(z_{t-1}) + v_{2t} \end{aligned} \quad (2)$$

with $|\psi_{11}| < 1$, $\lambda_1 = \psi_{11} - \theta \psi_{21}$, and $|1 - \theta \frac{\partial J_2}{\partial z}| < 1$. This mechanism provides a NEC as (1) where the cointegrating relation is $z_t = x_t - \theta y_t$ (in this case $\boldsymbol{\beta} = [1 \quad -\theta]'$), $\mathbf{X}_t = [x_t \ y_t]'$, $\mathbf{v}_t = [v_{1t} \ v_{2t}]$, and

$$\Psi_1 = \begin{pmatrix} \psi_{11} & 0 \\ \psi_{21} & 0 \end{pmatrix} \quad F(\mathbf{X}_{t-1}) = \begin{pmatrix} 0 \\ J_2(z_{t-1}) \end{pmatrix}$$

3. AUXILIARY RESULTS

Before characterizing the representation theorem, it would be useful to introduce some results that will be instrumental in the proof. For any vector norm $\|\mathbf{X}\|$ we can define a matrix norm $\|\mathbf{A}\|$, which is a subordinate matrix norm, such that for any vector \mathbf{X} it is true that

$$\|\mathbf{A}\mathbf{X}\| \leq \|\mathbf{A}\| \|\mathbf{X}\|$$

The following theorem finds a suitable matrix norm which will be useful for our purposes.

THEOREM 2. *For any given matrix \mathbf{A} and any number $\epsilon > 0$, there exists at least one subordinate matrix norm $\|\cdot\|_S$ such that*

$$\|\mathbf{A}\|_S \leq SR(\mathbf{A}) + \epsilon$$

where $SR(\mathbf{A})$ is the spectral radius of \mathbf{A} , i.e. the largest eigenvalue of the matrix \mathbf{A} .

PROOF. See the Appendix.

The above norm approximates the spectral radius as closely as we want from above, and this will be the appropriate norm to work with. Now we extend the definition for random variables.

DEFINITION 6. Let \mathbf{Y}_t be a random vector. We define its S_r norm as

$$\|\mathbf{Y}_t\|_{S_r} \equiv (E(\|\mathbf{Y}_t\|_S^r))^{1/r} \equiv E^{1/r}(\|\mathbf{Y}_t\|_S^r)$$

Note that this is usually called the L_r norm when random variables appear instead of random vectors and $\|\cdot\|_S$ is changed by the absolute value.

LEMMA 1. If \mathbf{W} is a random vector, the function defined by $\|\mathbf{W}\|_{S_r}$ is a norm.

PROOF. See the Appendix.

Consider the following nonlinear dynamic model

$$\mathbf{Z}_t = H(\mathbf{Z}_{t-1}; \gamma) + \mathbf{u}_t \quad (3)$$

where \mathbf{Z}_t and \mathbf{u}_t are $r \times 1$, and $H(\cdot; \gamma) : \mathbb{R}^r \rightarrow \mathbb{R}^r$ is a differentiable function of \mathbf{Z} on an open set of \mathbb{R}^r . This nonlinear autoregressive model will play an important role for the study of our basic model (1). In Theorem 3, we prove that there are enough conditions to guarantee that \mathbf{Z}_t is NED. Assumption 1 describe the conditions.

ASSUMPTION 1.

- (a) The sequence $\{\mathbf{u}_t\}$ is α mixing of size $r/(r-2)$ for $r > 2$.
- (b) (Boundedness Condition) We have

$$\text{SR}(\nabla_{\mathbf{Z}} H(\mathbf{Z}; \gamma)) \leq 1 - \delta$$

for all \mathbf{Z} , where $H(\cdot; \gamma)$ is continuously differentiable in each argument in an open set of \mathbb{R}^r , and $\nabla_{\mathbf{Z}} H(\mathbf{Z}; \gamma)$ is the matrix of first partial derivatives w.r.t. \mathbf{Z} .

- (c) For some finite constant Δ_u , $E\|\mathbf{u}_t\|_S^2 \leq \Delta_u$.

Assumption 1(b) says that the spectral radius of the matrix of first partial derivatives is smaller than 1. This boundedness condition imposed on the nonlinear function plays an important role. Notice that taking $\epsilon < \delta$, we obtain $\|\Delta_{\mathbf{Z}} H(\mathbf{Z}; \gamma)\| \leq 1 - \delta + \epsilon < 1$. This is a generalization of the concept of a nonlinear contraction. Theorem 3 ensures that the boundedness condition is sufficient to obtain a NED sequence. The proof extends the ideas of Gallant and White (1988).

THEOREM 3. Under Assumption 1, the sequence $\{\mathbf{Z}_t\}$ given in (3) is NED for $\|\cdot\|_S$, on the underlying α mixing sequence $\{\mathbf{u}_t\}$ of any size.

PROOF. See the Appendix.

We still need a technical Lemma that will be used later on. It essentially ensures that a nonlinear arbitrary function of a NED sequence is still a NED sequence. See Gallant and White (1988) or Davidson (1994) for a proof.

LEMMA 2. Let \mathbf{Z}_t be a vector sequence where each component is NED on $\{v_t\}$ of size a . Assume $J(\mathbf{Z}_t)$ is bounded in L_2 norm of size a , and the generalized Lipschitz condition

$$|J(\mathbf{Z}) - J(\mathbf{Y})| \leq B(\mathbf{Z}, \mathbf{Y})d(\mathbf{Z}, \mathbf{Y}) \text{ a.s.}$$

holds for a non negative measurable $B(\mathbf{Z}, \mathbf{Y})$ and a metric $d(\cdot, \cdot)$, such that for $1 \leq q \leq 2$ $\|B(\mathbf{Z}_t, E_{t-m}^{t+m}\mathbf{Z}_t)\|_{q/(q-1)} < \infty$

$$\|d(\mathbf{Z}_t, E_{t-m}^{t+m}\mathbf{Z}_t)\|_q < \infty$$

and for $r > 2$

$$\|B(\mathbf{Z}_t, E_{t-m}^{t+m}\mathbf{Z}_t)d(\mathbf{Z}_t, E_{t-m}^{t+m}\mathbf{Z}_t)\|_r < \infty$$

Then $\{J(\mathbf{Z}_t)\}$ is L_2 NED on $\{v_t\}$ of size $a(r-2)/(r-1)$.

In Section 4, we provide sufficient conditions to ensure that model (1) is correctly specified in a sense detailed below. This can be understood as a partial generalization of Granger's Representation Theorem presented in Engle and Granger (1987) and Johansen (1991).

4. A REPRESENTATION THEOREM

Now we have the tools to give a representation theorem for a nonlinear error correction with linear cointegration, in the sense that we provide sufficient conditions to ensure a balanced specification of NEC models.

THEOREM 4. (Representation Theorem) Consider the nonlinear error correction model for the $(n \times 1)$ vector \mathbf{X}_t , given by (1). Assume that

- (a) v_t is α mixing of size $s/(s-2)$ for $s > 2$
- (b) $\Sigma_r v_t$ is not NED on an α mixing sequence
- (c) $E\|v_t\|_S^2 \leq \Delta_v$
- (d) $F(\mathbf{X}_{t-1}) = J(\mathbf{Z}_{t-1})$, where $\mathbf{Z}_t \equiv \beta' \mathbf{X}_t$, for some vector $(r \times 1)\beta$, and a continuously differentiable function $J(\cdot)$, which satisfies the generalized Lipschitz conditions of Lemma 2,
- (e) $SR(\Psi_1) < 1$, where $SR(\mathbf{M})$ is the spectral radius of the matrix \mathbf{M} , and
- (f) for some fixed $\delta \in (0, 1)$

$$SR \begin{pmatrix} \Psi_1 & \nabla_{\mathbf{Z}} J(\mathbf{Z}) \\ \beta' \Psi_1 & \mathbf{I}_r + \beta' \nabla_{\mathbf{Z}} J(\mathbf{Z}) \end{pmatrix} \leq 1 - \delta$$

The above conditions ensure that

- (i) ΔX_t and Z_t are simultaneously NED on the α mixing sequence (v_t, u_t) , where $u_t = \beta' v_t$; and
(ii) X_t is $I(1)$.

PROOF. (i) Define the $(n \times 1)$ vector $W_t = \Delta X_t$ and the $(r \times 1)$ vector $Z_t = \beta' X_t$. If we multiply (1) by β' and write both systems we obtain

$$\begin{aligned} W_t &= \Psi_1 W_{t-1} + J(Z_{t-1}) + v(t) \\ Z_t &= Z_{t-1} + \beta' \Psi_1 W_{t-1} + \beta' J(Z_{t-1}) + u_t \end{aligned}$$

where $u_t = \beta' v_t$. This system that can be written

$$Y_t = G(Y_{t-1}) + e_t$$

where $Y = [W' \ Z']'$,

$$G(Y) = \begin{pmatrix} \Psi_1 W_{t-1} + J(Z_{t-1}) \\ Z_{t-1} + \beta' \Psi_1 W_{t-1} + \beta' J(Z_{t-1}) \end{pmatrix}$$

and $e_t[v_t' \ u_t']$. Then we have a Markovian system with α mixing errors. The matrix of partial derivatives with respect to W_t and Z_t given by $\nabla_Y G(Y)$ is

$$\nabla_Y G(Y) = \begin{pmatrix} \nabla_W G_1 & \nabla_Z G_1 \\ \nabla_W G_2 & \nabla_Z G_2 \end{pmatrix} = \begin{pmatrix} \Psi_1 & \nabla_Z J \\ \beta' \Psi_1 & I + \nabla_Z \beta' J \end{pmatrix}$$

where $G_1(\cdot)$ and $G_2(\cdot)$ are defined according to the system above. Notice that ∇_Z and β' commute. We are in the assumptions of Theorem 3, which ensures that $SR(\nabla_Y G(Y)) \leq 1 - \delta$ is a sufficient condition for Y_t to be NED on e_t , since the moment conditions and differentiability hold. This proves that W and Z are NED on an α mixing sequence.

(ii) The Vector X_t can be written

$$X_t = (1 \ L)^{-1} (1 \ \Psi_1 L)^{-1} (J(Z_{t-1}) + v_t)$$

Consider the sequence Q_t , given by

$$Q_t = (1 \ \Psi_1 L)^{-1} (J(Z_{t-1}) + v_t).$$

The result of Lemma 2 ensures that $J(Z_t)$ is NED. Now $SR(\Psi_1) < 1$, implies that the infinite summation

$$\sum_{n=0}^{\infty} \Psi_1^n J(Z_{t-(n+1)}) + v_{t-n}$$

is a NED sequence. However the sequence $(1 \ L)^{-1} Q_t$ is not NED, because $(1 \ L)^{-1} v_t$ is not NED. This completes the proof.

The above condition for the example give in (2) becomes

$$SR \begin{pmatrix} \psi_{11} & 0 & 0 \\ \psi_{21} & 0 & J'_2 \\ \lambda_1 & 0 & 1 \ \theta J'_2 \end{pmatrix} < 1 - \delta$$

and since the characteristic polynomial associated with this is $(\psi_{11} - \lambda)(1 - \theta J_2' - \lambda)$, then the condition becomes $|\psi_{11}| < 1$ and $|1 - \theta J_2'| < 1 - \delta_2$. Of course, cross conditions may be required for more general NEC systems.

5. EXTENSIONS

The results of the former section can be extended to more general specifications but, perhaps, at the expense of a less clear exposition. Consider, for instance, the case when the error correction function depends on say two lags X_{t-1} and X_{t-2} (time varying error correction models). Theorem 4 could be extended to include this case. Consider the NEC model

$$\Delta X_t = \Psi_1 \Delta X_{t-1} + F(X_{t-1}, X_{t-2}) + v_t \quad (4)$$

An example of these types of models is the smooth transition regression (STR) function given in Granger and Teräsvirta (1993), where the transition depends on some equilibrium errors of the long run relationship specified by the cointegrating relation. For example, if we have $X_t = [y_t \quad x_t]'$, then the first equation of (4) may be written as

$$\begin{aligned} \Delta y_t = & \psi_{11} \Delta y_{t-1} + \psi_{12} \Delta x_{t-1} \\ & + (\gamma_{11} \Delta y_{t-1} + \gamma_{12} \Delta x_{t-1})(1 + \exp(-\gamma_{13}(y_{t-1} - \beta_{12}^* x_{t-1}))) + v_{1t} \end{aligned}$$

In this case, the dynamics of Δy_t have an autoregressive representation with exogenous variables, whose parameters change depending on the long run relationship.

6. CONCLUSIONS

There is large evidence of empirical applications in economics and finance where nonlinearities are found in error correction contexts. However, there are no formal studies that justify the empirical use of error correction models within a nonlinear framework. To start filling this gap, we extend certain results of linear integrated and cointegrated variables to a nonlinear framework, by introducing a concept of integration based on near epoch dependence requirements. Within this framework, we are able to generalize certain properties of Granger's representation theorem to the nonlinear case. We found that if the variables are I(1) with a nonlinear error correction system then they are linearly cointegrated under certain conditions on the nonlinear adjustment. In particular, we give sufficient conditions for the NEC to be well specified and balanced.

APPENDIX

MATRIX NORMS Recall that

$$\|A\|_\infty = \max_i \sum_j |a_{ij}| \quad \text{and} \quad \|Y\|_\infty = \max_i |Y_i|$$

Given a matrix A of size $(n \times n)$ let

$$A = MJM^{-1}$$

its Jordan decomposition such that J is a diagonal matrix with boxes in its diagonal. The boxes are of the form

$$J = \begin{pmatrix} J_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & J_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & J_s \end{pmatrix} \quad \text{where} \quad J_i = \begin{pmatrix} \lambda_i & 1 & 0 & 0 \\ 0 & \lambda_i & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & \lambda_i \end{pmatrix}$$

Let us define the matrix D_δ as

$$D_\delta = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \delta & 0 \\ 0 & 0 & \delta^{n-1} \end{pmatrix}$$

If we take the matrix norm $\|\cdot\|_S$ as

$$\|A\|_S \equiv \|(\mathbf{MD}_\delta)^{-1}A(\mathbf{MD}_\delta)\|_\infty$$

then it is clear that $\|A\|_S \leq \text{SR}(A) + \delta$, because $(\mathbf{MD}_\delta)^{-1}A(\mathbf{MD}_\delta)$ is equal to the matrix J where the boxes J_i are substituted by boxes J_i^* of the form

$$J_i^* = \begin{pmatrix} \lambda_i & \delta & 0 & 0 \\ 0 & \lambda_i & \delta & 0 \\ 0 & 0 & 0 & \delta \\ 0 & 0 & 0 & \lambda_i \end{pmatrix}$$

In this case, the vector norm is

$$\|Y\|_S = \|(\mathbf{MD}_\delta)Y\|_\infty$$

and then the matrix norm is a subordinate norm. A very similar definition is given in Ciarlet (1989).

PROOF OF LEMMA 1

(i) Triangular inequality: By the Minkowsky inequality, we have

$$\begin{aligned} E(\|W + Z\|_S^r) &= E(\|W + Z\|_S \|W + Z\|_S^{r-1}) \\ &\leq E(\|W\|_S + \|Z\|_S) (E(\|W + Z\|_S^{r-1})) \\ &= E(\|W\|_S \|W + Z\|_S^{r-1}) + E(\|Z\|_S \|W + Z\|_S^{r-1}) \end{aligned}$$

The Holder inequality states that

$$E(|XY|) \leq E^{1/p}|X|^p E^{1/q}|Y|^q \quad \text{for } \frac{1}{p} + \frac{1}{q} = 1$$

Taking $p = r, q = (r-1)/r, |X| = \|W\|_S$ and $|Y| = \|W + Z\|_S^{r-1}$ (and analogously for the second term in the summation) we have

$$\begin{aligned}
E(\|\mathbf{W} + \mathbf{Z}\|_S^r) &\leq E^{1/r} \|\mathbf{W}\|_S^r E^{(r-1)/r} \left(\|\mathbf{W} + \mathbf{Z}\|_S^{r-1} \right)^{r/(r-1)} \\
&\quad + E^{1/r} \|\mathbf{Z}\|_S^r E^{(r-1)/r} \left(\|\mathbf{W} + \mathbf{Z}\|_S^{r-1} \right)^{r/(r-1)} \\
&\quad \left(E^{1/r} \|\mathbf{W}\|_S^r + E^{1/r} \|\mathbf{Z}\|_S^r \right) E^{(r-1)/r} \left(\|\mathbf{W} + \mathbf{Z}\|_S^{r-1} \right)^{r/(r-1)} \\
&\quad \left(E^{1/r} \|\mathbf{W}\|_S^r + E^{1/r} \|\mathbf{Z}\|_S^r \right) E^{1-1/r} (\|\mathbf{W} + \mathbf{Z}\|_S^r)
\end{aligned}$$

therefore

$$1 \leq \left(E^{1/r} \|\mathbf{W}\|_S^r + E^{1/r} \|\mathbf{Z}\|_S^r \right) E^{-1/r} (\|\mathbf{W} + \mathbf{Z}\|_S^r)$$

or

$$E^{1/r} (\|\mathbf{W} + \mathbf{Z}\|_S^r) \leq \left(E^{1/r} \|\mathbf{W}\|_S^r + E^{1/r} \|\mathbf{Z}\|_S^r \right)$$

as we want

(ii) Scalar multiplication:

$$\|\alpha \mathbf{X}\|_S^r = E^{1/r} |\alpha| \|\mathbf{X}\|_S^r = E^{1/r} |\alpha|^r \|\mathbf{X}\|_S^r = |\alpha| E^{1/r} \|\mathbf{X}\|_S^r = |\alpha| \|\mathbf{X}\|_S^r$$

PROOF OF THEOREM 3 Define

$$\bar{\mathbf{Z}}_t = \begin{cases} H(\mathbf{Z}_{t-1}) & \text{for } t > 0 \\ \mathbf{0} & \text{for } t \leq 0 \end{cases}$$

and

$$\mathbf{Z}_{t,s}^m = \begin{cases} H(\mathbf{Z}_{t-1,s+1}) + \mathbf{u}_t & \text{for } s+1 \leq m \\ \mathbf{Z}_t & \text{for } s+1 > m \end{cases}$$

then it is clear that $\mathbf{Z}_{t,0}^m$ is $\sigma(\mathbf{u}_t, \dots, \mathbf{u}_{t-m+1})$ measurable. The difference between \mathbf{Z}_t and its predictor $\bar{\mathbf{Z}}_t$ is bounded for $t > 0$ because

$$\begin{aligned}
\|\mathbf{Z}_t - \bar{\mathbf{Z}}_t\|_S &= \|H(\mathbf{Z}_{t-1}) + \mathbf{u}_t - H(\bar{\mathbf{Z}}_{t-1})\|_S \\
&\leq \|\mathbf{u}_t\|_S + \|H(\mathbf{Z}_{t-1}) - H(\bar{\mathbf{Z}}_{t-1})\|_S
\end{aligned}$$

and by the Mean Value Theorem

$$\begin{aligned}
\mathbf{H}(\mathbf{Z}_t) - \mathbf{H}(\bar{\mathbf{Z}}_t) &= \begin{pmatrix} H_1(\mathbf{Z}_t) - H_1(\bar{\mathbf{Z}}_t) \\ \vdots \\ H_r(\mathbf{Z}_t) - H_r(\bar{\mathbf{Z}}_t) \end{pmatrix} \\
&= \begin{pmatrix} \frac{\partial H_1}{\partial z_1}(\check{\mathbf{Z}}_t)(z_{1t} - \bar{z}_{1t}) + \dots + \frac{\partial H_1}{\partial z_r}(\check{\mathbf{Z}}_t)(z_{rt} - \bar{z}_{rt}) \\ \vdots \\ \frac{\partial H_r}{\partial z_1}(\check{\mathbf{Z}}_t)(z_{1t} - \bar{z}_{1t}) + \dots + \frac{\partial H_r}{\partial z_r}(\check{\mathbf{Z}}_t)(z_{rt} - \bar{z}_{rt}) \end{pmatrix} \\
&= \begin{pmatrix} \frac{\partial H_1}{\partial z_1}(\check{\mathbf{Z}}_t) & \dots & \frac{\partial H_1}{\partial z_r}(\check{\mathbf{Z}}_t) \\ \vdots & \ddots & \vdots \\ \frac{\partial H_r}{\partial z_1}(\check{\mathbf{Z}}_t) & \dots & \frac{\partial H_r}{\partial z_r}(\check{\mathbf{Z}}_t) \end{pmatrix} \begin{pmatrix} z_{1t} - \bar{z}_{1t} \\ \vdots \\ z_{rt} - \bar{z}_{rt} \end{pmatrix} \\
&= \nabla_z H(\check{\mathbf{Z}}_t)(\mathbf{Z}_t - \bar{\mathbf{Z}}_t)
\end{aligned}$$

Now, since $\|\cdot\|_S$ is a subordinate matrix norm

$$\begin{aligned}\|\mathbf{Z}_t - \bar{\mathbf{Z}}_t\|_S &\leq \|\mathbf{u}_t\|_S + \|\nabla_Z H(\bar{\mathbf{Z}}_t)\|_S \|\mathbf{Z}_{t-1} - \bar{\mathbf{Z}}_{t-1}\|_S \\ &\leq \delta_{u,t} + \delta_Z \|\mathbf{Z}_{t-1} - \bar{\mathbf{Z}}_{t-1}\|_S\end{aligned}$$

for some $\delta_{u,t}$ and since $\mathbf{Z}_0 = \bar{\mathbf{Z}}_0 = \mathbf{0}$. Then, by iteration

$$\begin{aligned}\|\mathbf{Z}_t - \bar{\mathbf{Z}}_t\|_S &\leq \sum_{j=0}^{t-1} \delta_{u,t-j} \delta_Z^j \\ \|\mathbf{Z}_t - \bar{\mathbf{Z}}_t\|_S^2 &\leq \sum_{j=0}^{t-1} \delta_{u,t-j}^2 \delta_Z^{2j} + \sum_{j=0}^{t-1} \sum_{i/j}^{t-1} \delta_{u,t-i} \delta_Z^{i+j} \\ E\|\mathbf{Z}_t - \bar{\mathbf{Z}}_t\|_S^2 &\leq \Delta_Z z\end{aligned}$$

for some bound $\Delta_Z^{(2)}$. Now, likewise we have

$$\|\mathbf{Z}_t - \mathbf{Z}_{t,0}^m\|_S = \|H(\mathbf{Z}_{t-1}) - H(\mathbf{Z}_{t-1,1}^m)\|_S$$

and again by the Mean Value Theorem

$$\|H(\mathbf{Z}_{t-1}) - H(\mathbf{Z}_{t-1,1}^m)\|_S \leq \|\nabla_Z H(\mathbf{Z})\|_S \|\mathbf{Z}_{t-1} - \mathbf{Z}_{t-1,1}^m\|_S$$

However, since $\|\nabla_Z H(\mathbf{Z})\|_S \leq \delta_Z$, we have

$$\|\mathbf{Z}_t - \mathbf{Z}_{t,0}^m\|_S \leq \delta_Z \|\mathbf{Z}_{t-1} - \mathbf{Z}_{t-1,1}^m\|_S$$

and by iteration

$$\|\mathbf{Z}_t - \mathbf{Z}_{t,0}^m\|_S \leq \delta_Z^m \|\mathbf{Z}_t - \bar{\mathbf{Z}}_{t,m}\|_S$$

and taking expectations

$$E\|\mathbf{Z}_t - \mathbf{Z}_{t,0}^m\|_S^2 \leq \delta_Z^{2m} E\|\mathbf{Z}_t - \bar{\mathbf{Z}}_{t,m}\|_S^2$$

and since $0 < \delta_Z < 1$ we obtain

$$\lim_{m \rightarrow \infty} E\|\mathbf{Z}_t - \mathbf{Z}_{t,0}^m\|_S = 0$$

Now, given $E_{t,m}(\mathbf{Z}_t) \equiv E(\mathbf{Z}_t | u_t, \dots, u_{t-m+1})$, we can obtain a bound for $\|\mathbf{Z}_t - E_{t,m}(\mathbf{Z}_t)\|_{S^2}$. Since $\mathbf{Z}_{t,0}^m$ is $\sigma(u_t, \dots, u_{t-m+1})$ measurable then

$$\begin{aligned}\|\mathbf{Z}_t - E_{t,m}(\mathbf{Z}_t)\|_{S^2} &\leq \delta_K \|\mathbf{Z}_t - \mathbf{Z}_{t,0}^m\|_{S^2} \\ &\quad \delta_K E^{1/2} \|\mathbf{Z}_t - \mathbf{Z}_{t,0}^m\|_S^2\end{aligned}$$

and given that $E\|\mathbf{Z}_t - \mathbf{Z}_{t,0}^m\|_S^2 \rightarrow 0$ at exponential rate then $\{\mathbf{Z}_t\}$ is NED on the underlying sequence $\{\mathbf{u}_t\}$ of any size.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the useful comments of A. Ronald Gallant, Manuel Moreno and an anonymous referee. The first author acknowledges financial support from spanish DGICYT PB95 0298 and TMR ERB 4061 PC97 0994.

REFERENCES

- BALKE, N. S. and FOMBY, T. B. (1992) Threshold cointegration. *Working Paper, Economics Department, Southern Methodist University*.
- BIERENS, H. J. (1981) *Robust Methods and Asymptotic Theory in Nonlinear Econometrics*. Lecture Notes in Economics and Mathematical Systems, vol. 192, Berlin: Springer Verlag.
- BURGESS, S. M. (1992) Nonlinear dynamics in a structural model of employment. *Journal of Applied Econometrics* 7, S101 18.
- CLORLET, P. G. (1989) *Introduction to Numerical Linear Algebra and Optimization*. Cambridge UK: Cambridge University Press.
- DAVIDSON, J. (1994) *Stochastic Limit Theory*. New York: Oxford University Press.
- ENGLER, R. F. and GRANGER, C. W. J. (1987) Co-integration and error correction: Representation, estimation, and testing. *Econometrica* 55(2), 251 76.
- ESCRIBANO, A. (1986) *Nonlinear error-correction: The case of money demand in the UK (1878 1970)*. PhD dissertation, University of California, San Diego.
- (1987) Error-correction systems: Nonlinear adjustment to linear long-run relationships, *CORE Discussion Paper 8730, C.O.R.E.*
- and GRANGER, C. W. J. (1998) Investigating the relationship between gold and silver prices. *Journal of Forecasting* 17, 81 107.
- and PFANN, G. A. (1998) Nonlinear error correction, asymmetric adjustment and cointegration. *Economic Modelling* 15, 197 216.
- GALLANT, A. R. and WHITE, H. (1988) *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. New York: Basil Blackwell.
- FRIEDMAN, M. and SCHWARTZ, A. (1982) *Monetary Trends in the United States and the United Kingdom: Their Relation to Income, Prices, and Interest Rates, 1867 1975*. Chicago: University of Chicago Press.
- GRANGER, C. W. J. (1981) Some properties of time series data and their use in employment econometric model specification. *Journal of Econometrics* 16, 121 30.
- (1995) Modelling nonlinear relationships between extended-memory variables. *Econometrica* 63(2), 265 79.
- and HALLMAN, J. (1991) Long memory series with attractors. *Oxford Bulletin of Economics and Statistics* 53, 11 26.
- and LEE, T. H. (1989) Investigation of production, sales and inventory relationships using multicointegration and non-symmetric error correction models. *Journal of Applied Econometrics* 4, S145 59.
- and NEWBOLD, P. (1974) Spurious regressions in econometrics. *Journal of Econometrics* 2, 111 20.
- and SWANSON, N. (1995) Further developments in the study of cointegrated variables. *Mimeo*, University of California.
- and TERASVIRTA, T. (1993) *Modelling Nonlinear Economic Relationships*. New York: Oxford University Press.
- HENDRY, D. F. and ERICSSON, N. R. (1991) An econometric analysis of the UK money demand. In *Monetary Trends in the United States and the United Kingdom* by Milton Friedman and Anna J. Schwartz. *The American Economic Review* 81, 8 38.
- JOHANSEN, S. (1988) Statistical analysis of cointegration vectors. *Journal of Economics Dynamic and Control* 12, 231 54.
- (1991) Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59, 1551 80.
- KUNST, R. M. (1992) Threshold Cointegration in Interest Rates. *Working Paper, Institute for Advanced Studies, Vienna*.
- KWIATKOWSKI, D., PHILLIPS, P. C. B., SCHMIDT, P. and SHIN, Y. (1992) Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics* 54, 159 78.
- LO, A. W. (1991) Long-term memory in stock market prices. *Econometrica* 59, 1279 313.
- LONGBOTTOM, A. and HOLLY, S. (1985) Econometric methodology and monetarism: Professor Friedman and Professor Hendry on the demand for money. *Discussion Paper No. 131, London Business School*.
- PHILLIPS, P. C. B. (1986) Understanding spurious regressions in econometrics. *Journal of Econometrics* 33, 311 40.
- SIMS, C. A. (1972) Money, income, and causality. *The American Economic Review* 62, 540 52.

- STOCK, J. H. (1994) Deciding between I(1) and I(0). *Journal of Econometrics* 63, 105-31.
- WATSON, M. (1994) *Large sample estimation and hypothesis testing*. Vector autoregression and cointegration. In *Handbook of Econometrics*, vol. IV (eds R. F. Engle and D. L. McFadden). Amsterdam: Elsevier Science, pp. 2844-915.
- WOOLDRIDGE, J. M. and WHITE, H. (1988) Some invariance principles and central limit theorems for dependent heterogeneous processes. *Econometric Theory* 4, 210-30.