Working Paper 96-54
Statistics and Econometrics Series 23
July 1996

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (341) 624-9849

NONLINEAR COINTEGRATION AND NONLINEAR ERROR CORRECTION.

Alvaro Escribano and Santiago Mira*

Abstract

The relationships between stochastic trending variables given by the concepts of cointegration and error correction (EC) are well characterized in a linear context, but the extension to a nonlinear context is still a challenge. Few extensions of the linear framework were developed in the context of linear cointegration but nonlinear error correction (NEC) models, and even in this context, there are still many open questions. The theoretical framework is not well developed at this moment and only particular cases have been discussed empirically. In this paper we propose a statistical framework that allow us to address those issues. First, we generalize the notion of integration to the nonlinear case. As a result a generalization of cointegration is feasible, and also a formal definition of NEC models. Within this framework we analyze the nonlinear least squares (NLS) estimation of nonlinear cointegration relations and the extension of the two-step estimation procedures od Engle and Granger (1987) for NEC models. Finally, we discuss a generalization of Granger Representation Theorem to the nonlinear case and discuss the properties of the one-step (NLS) procedure to estimate NEC models.

Keywords:
Nonlinear cointegration; nonlinear error correction; mixing; near epoch dependence; long memory; granger representation theorem.

*Escribano and Mira, Department of Statistics and Econometrics, Universidad Carlos III de Madrid; We gratefully acknowledge the comments of C.W.J. Granger, M.H. Pesaran, and J. Romo.

1 Introduction

Granger (1981) introduced the concept of cointegration but it was not until Engle and Granger (1987) and Johansen (1988) that this concept got an inmense popularity among econometricians and applied economists. The great impact those papers had in the profession was due to the fact that they showed, how to empirically work with economic variables that have unit roots to avoid the problem of spurious regressions. Furthermore, most of the modelling, estimation and inference procedures change dramatically from the classical statistical frameworks when dealing with variables that have unit roots and are cointegrated, see Phillips (1991). That forced a large part of the profession to work within this framework.

It is clear how to deal with integrated and cointegrated data within a linear context, but almost no research has been dedicated to the simultaneous consideration of nonstationarity, I(1), and nonlinearity, even though many macroeconomist agree with the fact that those are realistic and dominant properties of economic data. How can it be possible that almost no research have been dedicated to this topic ? The answer is clear, it is difficult to work with nonlinear time series models in a stationary and ergodic framework and even more difficult in a nonstationary context. Nevertheless there are already empirical examples of nonlinear error correction models with linear cointegration and with nonlinear cointegration. See Hendry and Ericsson (1991) and Granger and Swanson (1995) for some examples.

An introduction to the state of the art in econometrics relating nonlinearity and nonstationarity can be found in a recent paper by Granger (1995). There he discusses the concepts of long-range dependence and extended memory which generalize the linear concept of integration, I(1), to a nonlinear framework. The main disadvantages of those definitions are that there are no Laws of Large Numbers, nor Central Limit Theorems associated to them and therefore there are no easy ways to obtain estimation and inference results. This paper starts filling this mayor gap.

The structure of the paper is the following. In section 2, we propose a definition of nonlinear integration, NI(1), which also allows us to define the concept of nonlinear cointegration. Section 3 deals with the estimation of cointegrating relationships, and presents some Monte Carlo results. Section 4 studies the problem of the two-step estimation procedure in the context of nonlinear error correction models and presents some Monte Carlo results. Section 5 analyzes an extension via the near epoch dependence (NED) concept. Finally, in section 6 we present the main conclusions.

2 Cointegration and Error Corretion: The Non Linear Case

As we have discussed previously if we do not assume that the series follow ARMA models, then the classical definitions of stochastic trends and extended memory are not appropiate. Granger and Terasvista (1993) and Granger (1995) propose a natural generalization of the concepts to the nonlinear case as follows.

Let us take $F_h(x) = P(x_{t+h} \leq x|I_t)$ which provides the conditional distribution of x_{t+h} given the information set $I_t = \{x_{t-j} : j \geq 0\}$. It will be said that the series is "short memory in distribution" (SMD) if

$$\lim_h F_h(x) = \overline{F}(x)$$

i.e. the conditional distribution does not depends on I_t. Therefore,

$$|P(x_{t+h} \in C_1|x_{t-j} \in C_2) - P(x_{t+h} \in C_1)| \to 0 \quad \text{as} \quad m \to \infty$$

for all subsets $C_1, C_2 \in I_t$ such that $P(x_{t-j} \in C_2) \neq 0$. We will consider that the concept of mixing encapsulates the concept of SMD. Since ϕ-mixing implies α-mixing we will consider the concept of α-mixing.

Definition 2.0 (α-Mixing) Let $\{v_t\}$ a sequence of random variables. Let $\mathcal{F}_s^t \equiv \sigma(v_s, ..., v_t)$ and define the α-mixing coefficients as

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

It will be said that the sequence $\{v_t\}$ is α-mixing (or strong mixing) if and only if $\alpha_m \to 0$ as $m \to \infty$. The coefficient α_m measures the dependence between events that depend on v_t's separeted by at least m time periods. The α-mixing property allow simultaneously temporal dependence and heterogeneity in the process. If $\alpha_m = O(m^\lambda)$ for all $\lambda < -\varphi_0$, then it will be said that α_m is of size $-\varphi_0$. Since the concept of α-mixing is based on the σ-algebras generated by the sequence of variables, then the concept is invariant under Borel measurable transformations of a finite number of those variables. See, for instance, White (1984).

2.1 Non Linear Cointegration

Under general conditions there exists a LLN, as the following theorem states.

Theorem 2.1 (McLeish) Let $\{v_t\}$ a scalar α-mixing sequence with α_m of size $r/(r-1)$, $r > 1$, and with finite means $E(v_t) \equiv \mu_t$. If for some δ, $0 < \delta \leq r$, we have

$$\sum_{t=1}^{\infty} \left(E|v_t - \mu_t|^{r+\delta} / t^{r+\delta} \right)^{1/r} < \infty$$

then $T^{-1} \sum_{t=1}^{T} (v_t - \mu_t) \xrightarrow{a.s.} 0$. □

Proof: See Theorem 3.47 of White (1984).

The condition of Theorem 2.1 is essentially a condition of existence of moments of order $(r+\delta)$. See White (1984). Also under general conditions there exists a FCLT which gives the convergence of partial sums of the α-mixing sequences, as establishes the following theorem.

Theorem 2.2 (Herrndorf) Let $\{v_s\}$ be a sequence of random variables and define $S_T = \sum_{t=1}^{T} v_t$, and $V_T(r) = \sum_{t=1}^{[Tr]} v_t$, where $[Tr]$ is the greater integrer smaller than Tr. Then under assumptions

(i) $E(v_t) = 0$, for all t;

(ii) $\sup_t E(|v_t|^{\beta}) < \infty$, for some $\beta > 2$;

(iii) $\sigma^2 = \lim_{T\to\infty} E(T^{-1}(S_T)^2)$, verifies that $0 < \sigma^2 < \infty$; and

(iv) $\{v_t\}$ is α-mixing with α-mixing coefficients α_m satisfaying

$$\sum_{t=1}^{\infty} \alpha_m^{1-2/\beta} < \infty;$$

we have that $T^{-1/2} V_T(\cdot) \xrightarrow{d} \sigma W(\cdot)$, as $T \to \infty$, where $W(\cdot)$ is the SBM in [0,1]. □

Proof: See Herrndorf (1984).

Condition (ii) controls the existence of moments. Condition (iv) controls the temporal dependence of the process. Since β is the same in (ii) and (iv) there exists a trade off between both, see Phillips (1987). Condition (iii) avoids cases such as the following. Let v_t a Gaussian random walk such that Δv_t ($\Delta v_t \equiv (1-L)v_t \equiv v_t - v_{t-1}$) is a non-invertible MA(1). In that case Δv_t and v_t are α-mixing sequences, but v_t does not satisfy (iii). The following definition of strong nonlinear I(1) (SNI(1)) takes this case into account.

Definition 2.3 (SNI(0) y SNI(1)) A sequence $\{v_t\}$ is strongly nonlinear I(0), SNI(0), if it is α-mixing but the sequence $\{y_t\}$ given by $y_t = \sum_{s=1}^{t} v_t$, is not α-mixing. We will say that y_t is SNI(1).

Note that if y_t is SNI(1) then Δy_t is SNI(0). An important property of the above definition is that the α-mixing condition can be tested. There exists some papers that deal with this problem. Some of the more important are Lo (1991), Kwiatowski, Phillips, Schmidt and Shin (1992) (KPSS), and Stock (1994).

In what follows we will consider only sequences without deterministic components, i.e., $x_t = \tilde{x}_t - \mu_t$, where μ_t is the mean of $\tilde{x}_t$, such that $E(x_t) = 0$. Note that the above definition of SNI(0) the size of the sequence is not specified. It will be understood that a vector $X_t = [x_{1t}, ..., x_{nt}]'$ $(n \times 1)$ is SNI(1) (SNI(0)) if each component x_{it} is SNI(1) (SNI(0)).

Definition 2.4 (Non-Linear Cointegration) Let $\{y_t\}$ and $\{x_t\}$ two SNI(1) sequences. We will say that y_t and x_t are strongly nonlinear cointegrated (SNCI) with cointegration function $g(\cdot, \cdot, \gamma_1^*)$, if $g(y_t, x_t, \gamma_1^*)$ is α-mixing and $g(y_t, x_t, \gamma_1)$ is not α-mixing for $\gamma_1 \neq \gamma_1^*$.

Some comments are appropiate. First, note that we define $g(y_t, x_t, \gamma_1)$ as "not α-mixing" for $\gamma_1 \neq \gamma_1^*$, but we do not specify if $g(y_t, x_t, \gamma_1)$ is SNI(1). That definition would be inaccurate in the linear case because in that case $g(y_t, x_t, \gamma_1)$ could be I(-1). In this case, however, if $g(y_t, x_t, \gamma_1)$ is not α-mixing, then the dependence has to be stronger, and not weaker. Second, note that the restriction imposed by the α-mixing condition on the sequence $\{g_t\} = \{g(y_t, z_t, \gamma_1^*)\}$ implies the existence of restrictions on the mean of $\{g_t\}$, but also on every other moment of the sequence. Third, note that the cointegration function is not unique since any measurable function of an α-mixing sequence is α-mixing. Therefore we will consider the functions $f : \Re^2 \rightarrow \Re$ divided into equivalence classes such that two functions f_1 y f_2 are in the same class if there exists a function $g : \Re \rightarrow \Re$ such that $f_1 = g \circ f_2$. The study will be restricted to one function of each class. Fourth, note that with this definition new linear cointegration relations appear that were not allowed within the classical cointegration definition, because the dynamics of the variables are not necessarily represented as ARMA models. Finally, we suppose that the cointegration functions are measurable functions with respect to the appropiate σ-field.

Some extra conditions are implicitly impossed on the cointegration relation in order to avoid non-sense cointegration. The following examples specify the relations that are not considered as cointegration relations. (1) $g(y_t, x_t, \gamma_1) = h(y_t, \gamma_1)$, i.e., in fact it is a function of only one variable; (2) g is such that for any two variables y_t, x_t of some family of SNI(1) variables, $g(y_t, x_t, \gamma_1^*)$ it is always α-mixing, i.e. g gives always cointegration.

The second example tries to avoid "too restrictive" functions. Granger and Hallman (1991) give the following case. If x_t is a Gaussian random walk, then $\sin(x_t)$ has properties of "short memory". Functions such as $g(y_t, x_t, \gamma_1) = \cos(y_t + \gamma_1 x_t)$, or $g(y_t, x_t, \gamma_1) = \sin(\gamma_1(y_t x_t))$, are therefore "too restrictive" if they always produce cointegration. Consider the following example. Let x_t and y_t be scalar variables such that $x_t = \sum_{s=1}^{t} \varepsilon_s$ and

$y_t = \sum_{s=1}^{t} \eta_s$, where ε_s and η_s are α-mixing variables which verify a LLN, and converge in probability to non null values e_x and e_y respectively. If we take the ratio

$$f_t = (x_t/y_t) = (\sum_{s=1}^{t} \varepsilon_s)/(\sum_{s=1}^{t} \eta_s) \quad [2.1]$$

then f_t converges to e_x/e_y. The sequence f_t converges in probabilidad to some constant then, under certain conditions, it is α-mixing. Notice that even if the limit of the sequence is a constant it does not imply that the sequence is α-mixing as the following example illustrates. Let $\{r_t\}$ be a sequence given by $r_1 \sim \mathcal{U}(-1, 1)$ and $r_t \sim \mathcal{U}(-r_{t-1}, 0)$ if r_{t-1} is positive and $r_t \sim \mathcal{U}(0, -r_{t-1})$ if r_{t-1} is negative. The sequence systematically changes the sign. Take the outcomes $H = \{r_2 > 0\}$ and $G = \{r_{2(t+m)} < 0\}$ then $P(H) = \frac{1}{2} = P(G)$ and $P(H \cap G) = 0$. Therefore for every t

$$\sup_{\{H \in \mathcal{F}_{-\infty}^{t}, G \in \mathcal{F}_{t+m}^{\infty}\}} |P(G \cap H) - P(G)P(H)| = \frac{1}{4}$$

and then, although the sequence $\{r_t\}$ converges in probability to 0 it is not α-mixing. Note that hardly a ratio as [2.1] presents a behaviour as systematic as that in r_t, specially if ε_t and η_t are "good enough".

It is of interest to consider the "stability" of the definition SNI(0) for instantaneous transformations. This is due to the fact that the α-mixing property is preserved for such transfomations. The following Lemma formalizes the result.

Lemma 2.6 Let us suppose four SNI(1) series given by $\{y_t\}$, $\{\tilde{y}_t\}$, $\{x_t\}$, and $\{\tilde{x}_t\}$, which are related $\tilde{y}_t = f_y(y_t)$, and $\tilde{x}_t = f_x(x_t)$ for invertible transformations $f_y(\cdot)$ and $f_x(\cdot)$. If there exists a cointegrating function $g_R(\cdot, \cdot)$ for the x_t and y_t series then exists a cointegrating function $g_T(\cdot, \cdot)$ for the $f_x(x_t)$ and $f_y(y_t)$ series. Conversely, if there exists a cointegrating function $g_T(\cdot, \cdot)$ for the transformed series $\tilde{y}_t$ and $\tilde{x}_t$, then there exists a cointegrating function $g_R(\cdot, \cdot)$ for the series y_t and x_t. □

Proof: See Appendix A.

The invertibility condition of f_x and f_y is not necessary if we impose other restrictions. For instance if we know that $x_t > 0$ then we may consider that $x_t^2 = \tilde{x}_t$ is invertible. Finally, we present some possible generalizations of the definitions given above.

An extension of the idea of nonlinear integration can include the notion of the nonlinear trend. For example we can say that the x_t series has a Non-linear Trend (NT) if $x_t = F_x(\tau_t)$ for some τ_t series which is SNI(1) and $F_x(\cdot)$ is in some subset of the set of functions $F : \Re \rightarrow \Re$ (which we will not specify). Therefore, we will say that two NT series x_t and y_t have a non-linear co-trend (NCT) if there exists a funtion $C_{xy}(\cdot, \cdot, \gamma)$ such that $C_{xy}(x_t, y_t, \gamma)$ is α-mixing for $\gamma = \gamma^*$ and it is not for $\gamma \neq \gamma^*$. Consider the following example. Let w_t be an SNI(1) series and let us take

$$\begin{aligned} y_t &= \exp(-\gamma_1^* w_t + u_t) \\ x_t &= w_t + v_t \end{aligned}$$

where ε_t is an α-mixing sequence. Then $F(x_t, y_t) = y_t \exp(\gamma_1^* x_t)$ is a NCT relation. Different appoximations to these issues can be found in Escribano (1986 and 1987) and Granger (1988).

2.2 Non Linear Error Correction Mechanism

A non linear error correction (NEC) mechanism for the $(n \times 1)$ X_t vector is an autoregressive lineal model for the differences ΔX_t plus a nonlinear term for the lag of the levels X_{t-1}. If we take the case $n = 2$ and $X_t = [x_t, y_t]'$, the NEC with only one lag is $\Delta X_t = \Psi^* \Delta X_{t-1} + F(X_{t-1}, \Gamma^*) + \varepsilon_t$, whose first equation can be written in the form

$$\begin{aligned} \Delta x_t &= \psi_{11}^* \Delta x_{t-1} + \psi_{12}^* \Delta y_{t-1} + f(x_{t-1}, y_{t-1}, \gamma^*) + \varepsilon_{1t} \\ &= \psi_{11}^* \Delta x_{t-1} + \psi_{12}^* \Delta y_{t-1} + f(g(x_{t-1}, y_{t-1}, \gamma_1^*), \gamma_2^*) + \varepsilon_{1t} \quad [2.2] \end{aligned}$$

where Δy_t and Δx_t are α-mixing, and the parameter γ^* may be split into $\gamma^* = [\gamma_1^{*'}, \gamma_2^{*'}]'$. The subvector γ_1^* is the cointegration vector and the subvector γ_2^* is the vector of parameters of the error coretion mechanism.

Note the distinction made in [2.2] between the cointegration function $g(y_t, x_t, \gamma_1^*)$ and the error correction function $f(\cdot, \gamma_2^*)$. The function $g(\cdot, \cdot, \gamma_1^*) = 0$ gives the lung run equilibrium relationship and the deviations from this equilibrium $g(y_{t-1}, x_{t-1}, \gamma_1^*)$ are the errors corrected by the model.

A nonlinear error correction mechanism with only one lag is given by

$$\Delta X_t = \Psi_1 \Delta X_{t-1} + H(X_{t-1}) + \varepsilon_t$$

where $H(X_{t-1}) = H(X_{t-1}, \Gamma)$ for some vector of parameters Γ. The following definition allow us to give a necessary condition on the NEC formulation.

Definition 2.7 Given a funtion $F : \Re^p \to \Re^q$ such that $F(X) = Y$ for vectors $X = [x_1, ..., x_p]$ and $Y = [y_1, ..., y_q]$, we will say that F is partially invertible if there exists at least one $i \in \{1, ..., p\}$ and one $g_i : \Re^q \to \Re$ such that $x_i = g_i(Y)$.

The function $H(\cdot)$ is not necessarily a transformation of a finite number of other cointegrating relations, i.e. not necessarily $H(X_{t-1}) = J(P(X_{t-1}))$ for other cointegration function $P(\cdot)$. See Mira (1996) for a longer discussion. As a consequence we do not have a generalization of the Granger Representation Theorem given in Engle and Granger (1987) (in the sense that the existence of cointegration implies an error correction representation where the error correction is a function of the base of the space of cointegration relations) nor the converse formulation given in Johansen (1991). Nevertheless we can give a necessary condition for the NEC representation which will be extended in the last section to a partial generalization of the Granger Representation Theorem to the nonlinear case.

Proposition 2.8 Let us suppose a model of nonlinear time series for the sequence of random vectors $(n \times 1)$ $\{X_t\}$ given by

$$X_t = F(X_{t-1}, X_{t-2}) + \varepsilon_t,$$

where we have taken only two lags for simplicity. We have the following assumptions

(1) ΔX_t and ε_t are SNI(0);

(2) the function $F(X_{t-1}, X_{t-2})$ is non linear only in the first lag, i.e.

$$F(X_{t-1}, X_{t-2}) = G(X_{t-1}) + \Phi_2 X_{t-2}; \quad \text{and}$$

(3) the function $H(X_{t-1})$ given by $H(X_{t-1}) = -(I - \Phi_2)X_{t-1} + G(X_{t-1})$ is not partially invertible.

Then:

(i) under assumptions (1) and (2) we have the following representation

$$\Delta X_t = \Psi_1 \Delta X_{t-1} + H(X_{t-1}) + \varepsilon_t \qquad [2.3]$$

where $\Psi_1 = -\Phi_2$ and $H(X_t) : \Re^n \to \Re^n$ is given by $H(X_{t-1}) = -(I - \Phi_2)X_{t-1} + G(X_{t-1})$; and

(ii) the representation given in [2.3] is a NEC if and only if assumption (3) holds.

□

Proof: See Appendix A.

Some remarks deserve to be mentioned. First, note that condition (2) is intuitively clear, because we do not expect that any nonlinear function of the lags can be transformed into an error correction model, even if there exists a cointegrating function. Second, note that the condition of not partially invertible discardes the case of an SNI(0) variable which enters into the cointegration relation. Third, note that in the linear case the proof of the representation theorem relies in the fact that $A(1)$ is of rank r (the cointegration rank) and then it is not invertible; if that not were the case X_{t-1} can be inverted and we obtain X_t as an ARMA model, which would be a contradiction. See Mira (1996) for a detailled discussion. Fourth, note that the cointegration function depends on the AR representation for X_t as can logically be expected. As a consequence not any cointegration function can appear in the error correction representation, only those related with the AR representation for the levels of X_t. Finally, note that we cannot fully characterize the function $H(\cdot)$ to obtain a Representation Theorem. This question will be solved in Section 5.

If the error correction function depends on say two lags X_{t-1} and X_{t-2}, an extension of Proposition 2.8 can be given. Let us write

$$\begin{aligned} X_t &= G(X_{t-1}, X_{t-2}) + \Phi_2 X_{t-2} + \varepsilon_t \\ \Delta X_t &= G(X_{t-1}, X_{t-2}) - X_{t-1} + \Phi_2 X_{t-2} + \varepsilon_t \\ &= (-\Phi_2)(X_{t-1} - X_{t-2}) - (I - \Phi_2)X_{t-1} + G(X_{t-1}, X_{t-2}) + \varepsilon_t \\ &= \Psi_1 \Delta X_{t-1} + H(X_{t-1}, X_{t-2}) + \varepsilon_t \quad [2.5] \end{aligned}$$

where $\Psi_1 = -\Phi_2$ and $H(X_{t-1}, X_{t-2}) = -(I - \Phi_2)X_{t-1} + G(X_{t-1}, X_{t-2})$. In this case the condition of not partially invertible has to be impossed on the function $H : \Re^{2n} \to \Re^n$. An example of this type of models is the Smooth Transition Regression function (STR) given in Granger y Terasvirta (1993), where the transition depends on some equilibrium errors of the long range relationship specified by the cointegration relation. If we have $X_t = [y_t, z_t]'$, then the first equation of [2.5] may be written as

$$\begin{aligned} \Delta y_t &= \beta_{11}\Delta y_{t-1} + \beta_{12}\Delta z_{t-1} \\ &\quad +(\delta_{11}\Delta y_{t-1} + \delta_{12}\Delta z_{t-1})(1 + \exp(-\gamma_1(y_{t-1} - \gamma_2 z_{t-1})) + \varepsilon_{1t} \end{aligned}$$

In this case the dynamics of Δy_t is given as an autoregressive model with exogenous variables, whose parameters change depending on some equilibrium errors of the long range relationship.

2.3 Linear Cointegration and Non Linear Error Correction

It is of special interest the case where the error correction is nonlinear but the cointegration is linear. The model is

$$\Delta X_t = \Psi_1 \Delta X_{t-1} + H(X_{t-1}) + \varepsilon_t$$

where $H(X_{t-1}) = J(KX_{t-1})$ for some matrix K $(n \times 1)$ and some function $J : \Re^n \to \Re^n$. In this case K may be of full rank if $J(\cdot)$ is not partially invertible. Conversely, if $J(\cdot)$ is invertible then K cannot be of full rank and then is a linear combination of the space of cointegrating relations. Therefore, we partially recover Granger Representation Theorem because we have

$$KX_t = J^{-1}(\Delta X_t - \Psi_1 \Delta X_{t-1} - \varepsilon_t).$$

and K is a linear combination of the base of cointegrating relations. The following example clarify the issue. Let $X_t = [y_t, x_t, r_t]'$, and suppose that K is given by linear combinations of two cointegrating relations, i.e.

$$K = \begin{pmatrix} K_1' \\ K_2' \\ K_3' \end{pmatrix} = \gamma\alpha' = \begin{pmatrix} \gamma_1' \\ \gamma_2' \\ \gamma_3' \end{pmatrix} \begin{pmatrix} \alpha_1' \\ \alpha_2' \end{pmatrix} = \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \\ \gamma_{31} & \gamma_{32} \end{pmatrix} \begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \end{pmatrix}$$

The rank of K is 2, and then $J(\cdot)$ may be invertible. Also it is clear that a function of $(K_1'X_{t-1}, K_2'X_{t-1}, K_3'X_{t-1})$ can be written as a function of $[\alpha_1'X_{t-1}, \alpha_2'X_{t-1}]' = [z_{1,t-1}, z_{2,t-1}]'$. The error correction mechanism with only one lag is given by

$$\begin{aligned} \Delta y_t &= \beta_1 \Delta y_{t-1} + \beta_2 \Delta x_{t-1} + \beta_3 \Delta r_{t-1} + J_1(z_{1,t-1}, z_{2,t-1}) + \varepsilon_{1t} \\ \Delta x_t &= \delta_1 \Delta y_{t-1} + \delta_2 \Delta x_{t-1} + \delta_3 \Delta r_{t-1} + J_2(z_{1,t-1}, z_{2,t-1}) + \varepsilon_{2t} \\ \Delta r_t &= \rho_1 \Delta y_{t-1} + \rho_2 \Delta x_{t-1} + \rho_3 \Delta r_{t-1} + J_3(z_{1,t-1}, z_{2,t-1}) + \varepsilon_{3t} \end{aligned}$$

and the error correction is a function of the base of cointegrating relations.

3 Non-Linear Cointegration

In this section we study the problem of estimation of the cointegrating parameters when the cointegration relation is nonlinear.

3.1 Some Tools

We will introduce some tools from functional analysis. Let $(\mathcal{F}_1, \|\cdot\|_1)$ and $(\mathcal{F}_2, \|\cdot\|_2)$ be normed spaces, and let $\Psi : \mathcal{F}_1 \to \mathcal{F}_2$ be a functional. We will say that Ψ is differentiable at the point $F \in \mathcal{F}_1$ with respect to a collection of subsets $\mathcal{S}$ of $\mathcal{F}_1$ if there exists a linear continuous map $D\Psi(F;\cdot) : \mathcal{F}_1 \to \mathcal{F}_2$ (which we will call the differential of Ψ at F) such that for G in some neighbourhood of zero,

$$\Psi(F+G) = \Psi(F) + D\Psi(F;G) + R_\Psi(F;G)$$

where the remainder R_Ψ satisfies

$$\lim_{t\to 0}\frac{R_\Psi(F;tG)}{t}=0$$

uniformly in $G \in S$ for every $S \in \mathcal{S}$. Special choices for $\mathcal{S}$ give the most interesting differentials. If $\mathcal{S}$ is the family of all singletons of $\mathcal{F}_1$ then $D\Psi(F;G)$ is the Gâteaux differential. If $\mathcal{S}$ is the family of all compact subsets of $\mathcal{F}_1$ then $D\Psi(F;\cdot)$ is the Hadamard differential. If $\mathcal{S}$ is the family of all bounded subsets of $\mathcal{F}_1$ then $D\Psi(F;\cdot)$ is the Fréchet differential. Clearly Fréchet differentiability implies Hadamard differentiability, which in turn implies Gâteaux differentiability. In relation with the former definition we have the following theorem, which is a functional version of the well known delta-method theorem.

Theorem 3.1 Suppose $\Psi : \mathcal{F}_1 \to \mathcal{F}_2$ is Hadamard differentiable at $F \in \mathcal{F}_1$ with differential $D\Psi(F;\cdot)$ and that $\{X_T\}_{T=1}^{\infty}$ is a sequence of random elements in $\mathcal{F}_1$ that satisfies:

(i) $T^{-1/2}X_T \xrightarrow{d} X$ in $\mathcal{F}_1$ as $T \to \infty$; and

(ii) the sequence $\{T^{-1/2}X_T\}_{T=1}^{\infty}$ is tight in $\mathcal{F}_1$;

then $T^{-1/2}\Psi(X_T) \xrightarrow{d} D\Psi(0;X)$ in $\mathcal{F}_2$ as $T \to \infty$. □

Proof: The proof is essentially the same as in Heesterman and Gill (1992) with only few changes.

In our case the spaces $(\mathcal{F}_1, \|\cdot\|_1)$ and $(\mathcal{F}_2, \|\cdot\|_2)$ are $(D[0,1]^2, \|\cdot\|_B^2)$, for $D[0,1]$ the space of right continuous with left limits functions (cadlag functions), and $\|\cdot\|_B$ the norm defined by the Skorohod distance modified as in Billingsley (1984), Section 14, and $D[0,1]^2$ and $\|\cdot\|_B^2$ are the double products. Each element X_T, is a function $X_T(\cdot) : [0,1] \to \Re$, with $X_T(r)$ equal to the partial sumation $\sum_{t=1}^{[Tr]}\xi_t$, with $\{\xi\}_{s=1}^{\infty}$ an α-mixing sequence. Therefore, the operator $\Psi(\cdot)$ is given by $\Psi(X_T) = \Psi(\sum_{t=1}^{[Tr]}\xi_t)$. The element X is $W(\cdot)$, the Standard Brownian Motion.

From Theorem 2.2 we have that $T^{-1/2}X_T(\cdot) \xrightarrow{d} \sigma W(\cdot)$, and $\{T^{-1/2}X_T(\cdot)\}$ is tight in $D[0,1]$ (see for instance Herrndorf (1984)), then we have also that

$$T^{-1/2}\Psi(X_T(\cdot)) \xrightarrow{d} D\Psi(0;\sigma W(\cdot))$$

where $D\Psi(0;\sigma W(\cdot))$ is the Haddamar differential of $\Psi(\cdot)$ at zero in the direction of $\sigma W(\cdot)$.

For instance, if the functional $\Psi(F)$ is $\exp(F)$ to find its Fréchet differential it is enough to find its Gâteaux differential and prove that it is a continuous function. See Kolmogorov and Fomin (1978). On the other hand the Gâteaux differential $D_G\Psi(F;G)$ can be written as $\lim_{t\to 0}\|\frac{\Psi(F+tG)-\Psi(F)}{t}\|_2$. In our case $\Psi(F+tG)-\Psi(F) = \exp(F+tG)-\exp(F) = \exp(F)(\exp(tG)-1)$ and then we have

$$\begin{aligned}\lim_{t\to 0}\frac{\exp(F(r))(\exp(tG(r))-1)}{t} &= \exp(F(r))\lim_{t\to 0}\frac{(\exp(tG(r))-1)}{t}\\ &= \exp(F(r))G(r)\end{aligned}$$

Since this convergence is pointwise it also holds for the Skorohod topology, and the Gâteaux derivative is $D_G\Psi(F(r);G(r)) = \exp(F(r))G(r)$, which is lineal in G and continuous in F and then it is the Fréchet derivative. In general for functionals $\Psi(F)$ which are analogous to functions $\psi(f)$ the Fréchet differential $D_F\Psi$ is analoguos to the usual diferential $D\psi$ of the function ψ.

3.2 Estimation of the Cointegration Relationship

The cointegration function states that $g_t^* = g_t(\gamma^*) = g(y_t, z_t, \gamma^*)$ is α-mixing and that $g_t = g_t(\gamma) = g(y_t, z_t, \gamma)$ is not α-mixing for $\gamma \neq \gamma^*$. Note that, as in the linear case, under some conditions on g_t^* we have

$$(\frac{1}{T}\sum_{t=1}^{T}(g_t^*)^2 - \frac{1}{T}\sum_{t=1}^{T}E(g_t^*)^2) \xrightarrow{p} 0 \quad .$$

Therefore to ensure that a nonlinear least squares estimate provides a consistent estimation of γ^* we have to ensure that $\frac{1}{T}\sum_{t=1}^{T}(g_t)^2 \to \infty$ for $g_t \neq g_t^*$. Recall that $y_t = \sum_{s=1}^{t}\eta_s$, and $z_t = \sum_{s=1}^{t}\varepsilon_s$, then the following assumption states a relation between the function $g(\sum_{s=1}^{t}\eta_s, \sum_{s=1}^{t}\varepsilon_s, \gamma)$ and some function $\Phi(\sum_{s=1}^{t}\phi_s, \sum_{s=1}^{t}\delta_s)$ of some α-mixing sequences $\{\phi_s\}$ and $\{\delta_s\}$. Clearly, in general, these sequences will be some elemental transformation of the sequences $y_t = \sum_{s=1}^{t}\eta_s$, and $z_t = \sum_{s=1}^{t}\varepsilon_s$.

Assumption 3.2 (a) There exist a transformation $\Phi(\cdot)$, which is Haddamar differentiable such that the strongly nonlinear cointegration relation $g(y_t, z_t, \gamma)$ can be writen as $\Phi(\sum_{s=1}^{t}\phi_s, \sum_{s=1}^{t}\delta_s)$ for some strong mixing sequences $\{\phi_s\}$ and $\{\delta_s\}$; and
(b) $T^{-1}\sum_{t=1}^{T}E(g_t^*)^2 \xrightarrow{p} \mu$ as $T\to\infty$.

Lemma 3.3 Under Assumption 3.2, the NLS estimator γ^T which minimizes $\sum_{t=1}^{T}g(y_t, z_t, \gamma)^2$ provides a consistent estimator of the parameter γ^*. □

Proof: See Appendix B.

In the context of linear cointegrating relationships we know that if the $(n \times 1)$ variable X_t is SNI(1) and the linear combination $\gamma' X_t$ is SNI(0), then the OLS estimator $\hat{\gamma}$ of γ is obtained by

$$\hat{\gamma} \in \arg\min_{\gamma\in\Gamma} \sum_{t=1}^{T} \gamma' X_t$$

where the restriction Γ is a normalization of the cointegrating vector, such that the linear space generated by the restricted vector $\gamma \in \Gamma$ has to be the same as the space generated by the true γ. The restriction given by $\Gamma = \{\gamma : \gamma = [1, \tilde{\gamma}_2, ..., \tilde{\gamma}_n]'\}$ verifies the required condition and allows us to obtain the estimation by OLS. In the case of r linear cointegrating relationships many possible restrictions are allowed.

In the nonlinear case if the cointegration is given by $f(X_t, \gamma)$ where $f(\cdot, \gamma) : \Re^n \rightarrow \Re$ then the estimation is given by

$$\hat{\gamma} \in \arg\min_{\gamma} \sum_{t=1}^{T} (f(X_t, \gamma))^2.$$

In this case if $f(X_t, \gamma)$ is α-mixing then also $h(f(X_t, \gamma))$ is α-mixing for Borel measurable $h(\cdot)$ functions. New problems arise related to, but different from, those obtained in the linear case. First, the function $h(f(X_t, \gamma))^2$ may be a function with a maximum around γ^* and then when we find $\min_\gamma \sum_{t=1}^{T} h(f(X_t, \gamma))^2$ the objective function may be flat around the true value γ^* and then the algorithm provides an estimated value quite different from the true value. With an infinite sample the problem vanishes but not with finite samples. With finite samples the normalization proposed is the minimization of $h(f(X_t, \gamma))^2$ for some $h(\cdot)$ which may depend on $f(\cdot)$. Mira (1996) provides an example. Second, as in the linear case the function $h(\cdot)$ may depend on a set of parameters γ_2 such that $h(\cdot, \gamma_2^*) = 0$ and then we have an identification problem. For instance, in the linear case the problem is $\min_{\alpha,\beta} \sum_{t=1}^{T} (\alpha(y_t + \beta x_t))^2$, whose minimum is at $\alpha = 0$.

3.3 Asymptotic Distribution of the Estimator

For the nonlinear case the estimator γ^T of the parameter γ is given by the NLS algorithm. In this case the objective function is

$$\min_{\gamma} \frac{1}{2} \sum_{t=1}^{T} g_t(\gamma)^2 \equiv \min_{\gamma} \frac{1}{2} G(\gamma)' G(\gamma) \equiv \min_{\gamma} Q(\gamma)$$

where the vector $G(\gamma)$ is given by $G(\gamma) = [g_1(\gamma), ..., g_T(\gamma)]'$. The following assumption will help us to deal with the nonlinear function $G(\gamma)$.

Assumption 3.4 The functions $\frac{dG}{d\gamma}$ are Lipschitz.

If we assume that $\frac{dQ}{d\gamma}(\gamma^*) = 0$, for the true value γ^*, then applying a first order Taylor expansion around γ^0 we get,

$$0 = \frac{dQ}{d\gamma}(\gamma^*) \approx \frac{dQ}{d\gamma}(\gamma^0) + (\gamma^* - \gamma^0)' \frac{d^2Q}{d\gamma d\gamma'}(\gamma^0) \quad [3.1]$$

therefore

$$(\gamma^*)' \approx (\gamma^0)' - \left(\frac{dQ}{d\gamma}(\gamma^0)\right)\left(\frac{d^2Q}{d\gamma d\gamma'}(\gamma^0)\right)^{-1}$$

and the iteration of the Newton-Raphson algorithm is given by

$$(\gamma^{j+1})' = (\gamma^j)' - \left(\frac{dQ}{d\gamma}(\gamma^j)\right)\left(\frac{d^2Q}{d\gamma d\gamma'}(\gamma^j)\right)^{-1} \quad .$$

If we approximate $(\frac{d^2Q}{d\gamma d\gamma'}(\gamma^0))$ by $(\frac{dG}{d\gamma}(\gamma^0)'\frac{dG}{d\gamma}(\gamma^0))$ and impose that this matrix is invertible in γ^* and γ^T for all T, then we obtain the relation

$$\begin{aligned}(\gamma^0 - \gamma^*)' &\approx \left(\frac{dQ}{d\gamma}(\gamma^0)\right)\left(\frac{dQ}{d\gamma}(\gamma^0)'\frac{dQ}{d\gamma}(\gamma^0)\right)^{-1} \\ &\approx \left(G(\gamma^0)'\frac{dG}{d\gamma}(\gamma^0)\right)\left(\frac{dG}{d\gamma}(\gamma^0)'\frac{dG}{d\gamma}(\gamma^0)\right)^{-1}\end{aligned}$$

Since γ^T is consistent for γ^*, the approximation in [3.1] becomes a equality in the limit, and the asymptotic distribution is given by

$$\begin{aligned}\lim_{T\to\infty} T(\gamma^T - \gamma^*)' &= \lim_{T\to\infty}\left(T^{-1}G(\gamma^T)'\frac{dG}{d\gamma}(\gamma^T)\right)\left(T^{-2}\frac{dG}{d\gamma}(\gamma^T)'\frac{dG}{d\gamma}(\gamma^T)\right)^{-1} \\ &= \lim_{T\to\infty}\left(T^{-1}G(\gamma^*)'\frac{dG}{d\gamma}(\gamma^*)\right)\left(T^{-2}\frac{dG}{d\gamma}(\gamma^*)'\frac{dG}{d\gamma}(\gamma^*)\right)^{-1} \quad [3.2] \\ &= \lim_{T\to\infty}(T^{-1}V'X)(T^{-2}X'X)^{-1}\end{aligned}$$

for $V = G(\gamma^*)$ and $X = \frac{dG}{d\gamma}(\gamma^*)$. The second equality is ensured by Assumption 3.4 and the continuous mapping theorem. The following theorem introduces the convergence to the Standard Brownian Motion in the vectorial case.

Theorem 3.5 (Phillips and Durlauf) Let $\{x_s\}$ be a sequence of $(k \times 1)$ vectors and let $X_T(r) = \sum_{t=1}^{[Tr]} x_t$, and define $S_T = \sum_{t=1}^{T} x_t = X_T(1)$, then if

(i) $E(x_t) = 0$ for all t;

(ii) $E(T^{-1}S_TS_T') \to \Sigma$, a positive definite matrix, as $T \to \infty$ and $E(T^{-1}(S_{K+T}-S_K)(S_{K+T}-S_K)') \to \Sigma$ as $\min\{K,T\} \to \infty$.

(iii) $\{x_{it}^2\}$ is uniformly integrable for all $i = 1, ..., k$;

(iv) $\sup_t E(|x_{it}|^\beta) < \infty$ for some $2 \le \beta < \infty$ and all $i = 1, ..., k$;

(v) $\beta > 2$ and α_m is of size $-\beta/(\beta-2)$;

then for $W(r)$ the k-dimensional Standard Brownian Motion, and for the decomposition

$$\begin{aligned}
\Sigma_0 &= \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} E(x_t x_t') \\
\Sigma_1 &= \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{t-1} E(x_j x_t') \\
\Sigma &= \lim_{T\to\infty} E(T^{-1}S_TS_T') = \Sigma_0 + \Sigma_1 + \Sigma_1'
\end{aligned}$$

we have the following results as $T \to \infty$,

(a) $T^{-1/2}X_T(r) \xrightarrow{d} \Sigma^{-1/2}W(r) \equiv B(r)$;

(b) $T^{-2}\sum_{t=1}^T S_t(r)S_t(r)' \xrightarrow{d} \int_0^1 B(r)B(r)'dr$;

(c) $T^{-1}\sum_{t=1}^T S_{t-1}x_t' \xrightarrow{d} \int_0^1 B(r)dB(r)' + \Sigma_1$;

(d) $T^{-3/2}\sum_{t=1}^T S_t \xrightarrow{d} \int_0^1 B(r)dr$.

for $W(\cdot)$ the SBM k-dimensional. □

Proof: See Lemma 3.1 in Phillips and Durlauf (1986).

Note that in this case $X_T \in D[0,1]^k$, the product metric space of all cadlag real valued functions on $[0,1]$. In this case the definition of α-mixing has to be extended appropiately to the n-dimensional space. The results (a)-(d) hold for the scalar case under assumptions of Theorem 2.2. Write the matrix X as

$$X = \begin{pmatrix} d_1^1 & d_1^2 & \cdots & d_1^k \\ \vdots & \vdots & \ddots & \vdots \\ d_T^1 & d_T^2 & \cdots & d_T^k \end{pmatrix} \equiv \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_T \end{bmatrix}$$

where $d_t^j = \frac{\partial g_t}{\partial \gamma_j}$, then, the $(k \times k)$ matrix $(X'X)$ can be writen as $\sum_{t=1}^T \mathbf{x}_t'\mathbf{x}_t$, for a $(1 \times k)$ vector $\mathbf{x}_t$. Analogously the $(1 \times k)$ vector $(V'X)$ can be written as $\sum_{t=1}^T g_t^*\mathbf{x}_t$. Let us suppose the following assumption

Assumption 3.6 The derivative $\mathbf{x}_t$ can be written as $\mathbf{x}_t = \sum_{s=1}^{t} \mathbf{l}_s$ for $\{\mathbf{l}_s\}$ a α-mixing vector sequence, with $\mathbf{l}_s = [l_{1s}, ..., l_{ks}]'$.

Therefore for each $j = 1, ..., k$ we have $\frac{\partial g_t}{\partial \gamma_j} = \sum_{s=1}^{t} \lambda_{js}$. Consider the example of the cointegrating function $g_t = (y_t - \gamma_1)(z_t - \gamma_2)$ in this case $\frac{\partial g_t}{\partial \gamma_1} = -z_t + \gamma_2$ and $\frac{\partial g_t}{\partial \gamma_2} = -y_t + \gamma_1$. Now we have the following theorem.

Theorem 3.7 Under Assumptions 3.4 and 3.6 and if the vector $[g_{t-1}^*, \mathbf{l}_t']'$ verify the assumptions of Theorem 3.5, then the asymptotic distribution of the estimator γ^T is given by

$$\begin{aligned}\lim_{T\to\infty} T(\gamma^T - \gamma^*)' &= \lim_{T\to\infty} (T^{-1}V'X)(T^{-2}X'X)^{-1} \\ &= \lim_{T\to\infty} (T^{-1}\sum_{t=1}^{T} g_t^* \mathbf{x}_t)(\sum_{t=1}^{T} \mathbf{x}_t'\mathbf{x}_t)^{-1} \\ &\xrightarrow{d} \Big(\int_0^1 \mathbf{B}_2(r)dB_1(r) + \Sigma_{12}\Big)\Big(\int_0^1 \mathbf{B}_2(r)\mathbf{B}_2(r)'dr\Big)^{-1}\end{aligned}$$

□

Proof: See Appendix B.

Note that the former theorem ensures the superconsistency of the estimator.

3.4 Bias in the Estimation of the Cointegrating Parameters

Let us consider two α-mixing series $\{\eta_s\}$ and $\{\varepsilon_s\}$ and two series $\{y_t\}$ and $\{x_t\}$ given by $y_t = \sum_{s=1}^{t} \eta_s$ and $x_t = \sum_{s=1}^{t} \varepsilon_s$, such that there exists a function $g(\cdot,\cdot,\gamma^*)$ such that $g(y_t, x_t, \gamma^*)$ is α-mixing. As a by-product, when the function $g(\cdot,\cdot,\gamma^*)$ is linear this approach allows linear cointegration relations that were not allowed in the classical cointegration approach. Whithin the usual framework, a cointegration relation given by $x_t + \alpha y_t$ implies that both x_t and y_t follow ARMA models. With the approach proposed here, those variables may follow any linear or nonlinear model. This section studies the biases that appear in the estimation of linear and nonlinear relationships.

3.4.1 Model 1

This case studies the bias that appear when cointegration is linear and the series are nonlinear transformations of i.i.d. $\mathcal{N}(0,1)$ series. Let us define $\eta_s = v_s + \phi^* v_{s-1} + (m_s - m_{s-1})$ and $\varepsilon_s = v_{s-1} + (n_s - n_{s-1})$, where m_s, n_s and a_s are i.i.d. $\mathcal{N}(0,1)$ and v_s is defined below. In this case the cointegration parameter is $\gamma^* = \phi^* + 1$ since

$$\begin{aligned} y_t - \gamma^* x_t &= \sum_{s=1}^{t}(v_s + \phi^* v_{s-1}) - \gamma^* \sum_{s=1}^{t} v_{s-1} \\ &= \sum_{s=1}^{t}(v_s + (\phi^* - \gamma^*)v_{s-1}) \\ &= \sum_{s=1}^{t}(v_s - v_{s-1}) = v_t + v_0. \end{aligned}$$

where $v_s = \text{sign}(-a_s)(\beta_2 - \exp(\text{sign}(-a_s)\beta_1 a_s)$, for model 1.1, and v_s is $a_s^3/(a_s^2+1)$ for model 1.2, The values are $\phi_1^* = 1$, $\beta_1 = 0.8$ and $\beta_2 = 0.5$.

To analyze the behaviour of the estimators we generate N=1000 samples of sizes T=100, T=200 y T=1000, (with 100 extra data discarded) and we estimate the values γ^T. The following table presents the bias (estimated as the mean $\overline{\gamma}^T = \frac{1}{N}\sum_{i=1}^{N}(\gamma_i^T - \gamma^*)$) and the standard deviation (given by $\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\gamma_i^T - \overline{\gamma}^T)^2}$).

Comparing model 1.1 and model 1.2 we see that the nonlinearity affects the OLS estimation. In model 1.1 when T is smaller or equal to 500 the bias is a large part of the value of the parameter. For T=1000 the bias if about 10% of the value of the parameter. However, the bias in model 1.2 is smaller for size T=100 and even smaller for larger sizes.

Table 1	T=100	T=500	T=1000
Model 1.1	1.1339	0.4936	0.2865
	(0.3599)	(0.2607)	(0.1870)
Model 1.2	0.3768	0.0932	0.0500
	(0.2400)	(0.0708)	(0.0388)

3.4.2 Model 2

In this case we study the bias that appears when the cointegration relation is linear but the series are nonlinear transformations of ARMA series. Consider $\{v_t\}$ and $\{a_t\}$ as series i.i.d. $\mathcal{N}(0,1)$ and define $w_t = \delta w_{t-1} + v_t$, $\rho_t = \log(1 + (0.1)w_t)$ and $\phi_t = a_t - a_{t-1}$. Now define ε_t and η_t as $\eta_t = \pi\rho_t$ and $\varepsilon_t = \rho_t + \lambda\phi_t$. Then y_t and x_t are generated as the acumulation of η_t and ε_t respectively. If we take $y_t - \gamma^* x_t$ then

$$\begin{aligned} y_t - \gamma x_t &= \sum_{s=1}^{t}\left(\eta_s - \gamma\varepsilon_s\right) \\ &= \sum_{s=1}^{t}\left(\pi\rho_s - \gamma\rho_s - \gamma\lambda\phi_s\right) \end{aligned}$$

which is α-mixing for $\gamma = \pi$. The values of $\pi = 0.8$, and $\delta = 0.5$ are maintained in both models. Model 2.1 will have $\lambda = 0.4$ and Model 2.2 will have $\lambda = 0.16$. The following table presents the bias of γ^T in the same way as we did in table 1.

With this simulation we see that when λ is large (i.e., ϕ_t has more importance in the errors) the bias is larger. In both cases for size T=500 or greater the biases are smaller than 10% of the value of the parameter.

Table 2	T=100	T=500	T=1000
Model 2.1	0.2133	0.0521	0.0245
	(0.1205)	(0.0404)	(0.0228)
Model 2.2	0.0506	0.0091	0.0041
	(0.0440)	(0.0086)	(0.0044)

3.4.3 Model 3

Consider a fully nonlinear model. We take series n_s and a_s generated by an ARMA(1,1) given by $n_s = 0.6n_{s-1} + 0.8e_{t-1} + e_t$ with e_t i.i.d. $\mathcal{N}(0,1)$ and we define $z_t = \sum_{s=1}^{t} a_s$ and $w_t = \gamma_2(v_t - \gamma_1 z_t)$, with v_s i.i.d. $\mathcal{N}(0,1)$. Now define

$$\begin{aligned} x_t &= \exp(\gamma_2\gamma_1(z_t + 100)/100) + \lambda_1 \\ y_t &= \exp((w_t + n_t + 100)/100) + \lambda_2. \end{aligned}$$

Then the relation g_t given by $g_t = (x_t - \lambda_1)(y_t - \lambda_2)$ is a nonlinear cointegration relation since $g_t = \exp((\gamma_2\gamma_1 + 1) + (\gamma_2 v_t + n_t)/100)$. Note that if the values λ_1 and λ_2 were known in advance the relation $\log(x_t - \lambda_1) + \log(y_t - \lambda_2)$ could have been estimated, but in general they are not known. The parameters λ_1 and λ_2 are estimated by γ_1 and γ_2 given by

$$\min_{\gamma_1,\gamma_2,\gamma_3} \sum_{s=1}^{t}((x_t - \gamma_1)(y_t - \gamma_2) - \gamma_3).$$

In this case the comment about the normalization in the nonlinear case is applied. Instead of estimating $(x - \gamma_1)(y - \gamma_2)$ it is better to estimate the modification previously proposed.

The procedure used to minimize is the function $ms(\cdot)$ of *S-plus*. In this procedure the initial values given for iterations has been: the mean of x for λ_1, the mean of y for λ_2, and the mean of $x \times y$ for λ_3. The values of the parameters are $\gamma_1 = 0.8$, $\gamma_2 = 0.7$, $\lambda_1 = \lambda_2 = 1$. The following table presents the results in the same way that tables 1 and 2. It can be seen in the table that for $T = 100$ the bias are quite large and they decrease slowly. For $T = 1000$ the bias is still around a 25% of the value of the parameter.

Table 3	T=100	T=500	T=1000
λ_1	1.4758	0.5345	0.2207
	(0.5500)	(0.4644)	(0.2689)
λ_2	2.3163	0.7929	0.3272
	(0.8506)	(0.6713)	(0.3969)

4 Two-Step Estimation Procedure for NEC

In the former section we show that the NLS estimator of the nonlinear cointegration relationship is superconsistent under certain conditions. In this section, we study the question of generalizing the two-step estimation procedure of Engle and Granger (1987) to the nonlinear case. The nonlinear error correction model that we want to estimate is a single equation model with a nonlinear error correction which depends on a single cointegrating relationship, given by

$$\Delta y_t = \beta^* \Delta y_{t-1} + \delta^* \Delta x_{t-1} + f(g(y_{t-1}, x_{t-1}, \gamma_1^*), \gamma_2^*) + v_t$$

which can be written as

$$r_t = \beta^* r_{t-1} + \delta^* w_{t-1} + f(z_{t-1}^*, \gamma_2^*) + v_t \qquad [4.1]$$

for $z_{t-1}^* \equiv g(y_{t-1}, x_{t-1}, \gamma_1^*)$, $\Delta y_t \equiv r_t$, and $\Delta x_t \equiv w_t$. If we stack all the observations in vector form we get

$$\overline{R} - KB^* - F^*(\gamma_2^*) = V \qquad [4.2]$$

$$G^*(\theta^*) = V \qquad [4.3]$$

where $\overline{R} = [r_1, ..., r_T]'$, $R = [r_0, ..., r_{T-1}]'$, $W = [w_0, ..., w_{T-1}]'$, $K = [R, W]$, $B^* = [\beta^*, \delta^*]'$, $F^*(\gamma_2^*) = [f(z_0^*, \gamma_2^*), ..., f(z_{T-1}^*, \gamma_2^*)]'$, and $\theta^* = [\beta^*, \delta^*, \gamma_2^*]'$.

We define also $z_{t-1}^T = g(y_{t-1}, x_{t-1}, \gamma_1^T)$ for γ_1^T the NLS estimation of the cointegration parameter γ_1^*, and $F^T(\gamma_2^*) = [f(z_0^T, \gamma_2^*), ..., f(z_{T-1}^T, \gamma_2^*)]'$.

The two-step estimation procedure proposed by Engle and Granger (1987) consists in estimating the cointegration parameter in a first step, say γ_1^T, generate the residuals, and then use those residuals in a second step for estimating the remaining parameters of the nonlinear error correction model [5.1] but substituting z_{t-1}^* by z_{t-1}^T. For instance in a linear case we would substitute $z_t^T = y_t - \gamma_1^T x_t$ for $z_t^* = y_t - \gamma_1^* x_t$. In order to obtain a similar result for the nonlinear case we consider the following assumption.

Assumption 4.2 Define the function

$$G^T(\theta^*) = \left(F^*(\gamma_2^*) - F^T(\gamma_2^*)\right) + V$$

and assume that the following conditions hold,

$$\begin{aligned}
\lim_{T\to\infty}(T^{-1}G_\theta^T(\theta^*)'G_\theta^T(\theta^*)) = \lim_{T\to\infty}(T^{-1}G_\theta^*(\theta^*)'G_\theta^*(\theta^*)) &= O_p(1) \quad [4.4] \\
\lim_{T\to\infty}(T^{-1/2}V'G_\theta^T(\theta^*)) = \lim_{T\to\infty}(T^{-1/2}V'G_\theta^*(\theta^*)) &= O_p(1), \quad \text{and} \quad [4.5] \\
\lim_{T\to\infty}(T^{-1/2}(F^*(\gamma_2^*) - F^T(\gamma_2^*))'G_\theta^T(\theta^*)) &= o_p(1). \quad [4.6]
\end{aligned}$$

These assumptions have clear implications in the linear case, see Mira (1996). With the above assumption we can prove the following theorem.

Theorem 4.2 Let us suppose that model [4.1] can be estimated consistently by NLS. Under Assumption 4.1, the estimation of model [4.1] with the cointegration parameter estimated by NLS γ_1^T, instead of the true parameter γ_1^*, provides the same asymptotic distribution for the NLS estimations θ^T of the rest of parameters θ^*, than those obtained with the true value γ_1^*.

Proof: See Appendix C.

4.1 Bias in the Estimation of NEC Models with Linear Cointegration

In this section we present an example of a non linear error correction model with linear cointegration, and analyze the bias that appears in the two step estimation.

The data generating process is the following. Let $\{a_t\}$ and $\{v_t\}$ be two independent α-mixing sequences and define

$$\begin{aligned}
x_t &= x_{t-1} + a_t \quad [4.7] \\
z_t^* &= z_{t-1}^* + \delta_1^* a_t + f(z_{t-1}^*, \gamma_2^*) + v_t \quad [4.8] \\
y_t &= \gamma_1^* x_t + z_t^* \quad [4.9]
\end{aligned}$$

where the parametric function $f(z_{t-1}^*, \gamma_2^*)$ is the function that we want as nonlinear error correction. If z_t defined in [4.8] is α-mixing then we have that x_t is SNI(1), y_t is SNI(1) and they are cointegrated with linear cointegration function $y_t - \gamma_1^* x_t$. Now taking the difference operator in [4.9] we obtain

$$\Delta y_t \;=\; (\gamma_1^* + \delta_1^*)\Delta x_t + f(z_{t-1}^*, \gamma_2^*) + v_t \quad [4.10]$$

which is a nonlinear error correction mechanism (NEC) with linear cointegration given by $z_t^* = y_t - \gamma_1^* x_t$. We will impose the common factor restriction to simplify the model, such that $\delta_1^* = 0$ and obtain the model

$$\Delta y_t \;=\; \gamma_1^* \Delta x_t + f(z_{t-1}^*, \gamma_2^*) + v_t \quad [4.11]$$

The errors in the linear error correction mechanisms are given by $z_{t-1}^* = y_{t-1} - \gamma_1^* x_{t-1}$, and the estimated OLS residuals are given by $z_{t-1}^T = y_{t-1} - \gamma_1^T x_{t-1}$, where γ_1^T is the value β estimated in the regression $y_t = \alpha + \beta x_t + \varepsilon_t$, since $y_t = \gamma_1^* x_t + (z_t^0 + \mu)$, where $z_t = z_t^0 + \mu$ and z_t^0 has zero mean.

If we take the derivative in [4.7] with respect to z_{t-1}^* we obtain

$$\frac{d}{dz_{t-1}^*} z_t^* = 1 + \frac{d}{dz_{t-1}^*} f(z_{t-1}^*, \gamma_2^*) \quad [4.12]$$

such that impossing the boundness condition $-1 < \frac{d}{dz_{t-1}^*} z_t^* < 1$ we have a sufficient condition that ensure that the series z_t^* is near epoch dependence (NED). See Mira and Escribano (1995) for a discussion of this condition. Several models verify this condition, here a brief analysis of one of them is exposed. See Mira (1996) for a detailled analysis of those models.

Consider the following parametric nonlinear function

$$f(s, \beta_1, \beta_2, \gamma_2) = -\gamma_2 \ \arctan(\beta_1 s + \beta_2)$$

for $\gamma_2 > 0$. The derivative is

$$1 \;-\; \gamma_2 \frac{\beta_1}{1 + (\beta_1 s + \beta_2)^2}$$

and the derivative is in the region of interest for the appropiate values of the parameters. For instance for values of $(\beta_1, \beta_2, \gamma_2)$ equal to $(1, 0, 1)$ the derivative is $1 - \frac{1}{1+s^2}$ which clearly is always between 0 and 1. The set of values that we are going to consider is $(\beta_1, \beta_2, \gamma_2)$ equal to $(2, 0, 0.7)$.

Now we present the estimation results of model [1.5] for the nonlinear case. The estimation procedure is the function *ms*$(\cdot)$ of *S-plus*. The sizes of the samples are 100, 500 and 1000 where 100 previous observations have been disregarded. The value of γ_1 has been set to 0.7 and it initial value to 1. The set of initial values for $(\beta_1, \beta_2, \gamma_2)$ are $(1, 1, 1)$.

Table 4	γ_1	γ_2	β_1	β_2
T=100	-0.004934	0.1782	-133.386	6.77
	(0.10478)	(0.70808)	(2890.57)	(318.079)
T=500	0.00199	0.0150	-0.10739	-0.00534
	(0.04448)	(0.1144)	(0.585)	(0.159)
T=1000	-0.00184	0.00631	-0.04648	0.00221
	(0.0307)	0.0786	0.38279	(0.10068)

From table 4 we conclude that a sample size of 100 is too small to get a satisfactory (small bias) estimation. The biases are greater for the parameters of the nonlinear terms than for the linear ones.

5 The NED Extension

The definition of NI(0) introduced in section 2 is based in the concept of α-mixing. This concept imposses restrictions on the whole set of outcomes of the σ-algebras, which may be a too strong assumption. There are several ways of relaxing this concept whithout loosing the useful structure that it contains; see for instance Bierens (1983), Gallant and White (1988) and Potscher y Prucha (1991) for a detailed discussion. One of the more interesting alternatives is the concept of near epoch dependence (NED).

Definition 5.0 (NED) Let $\{z_t : \Omega \to \Re\}$ be a sequence $(\mathcal{F},\mathcal{B})$-medible with $E(z_t^2) < \infty$ for all t. Then it will be said that $\{z_t\}$ is near epoch dependent (NED) on the underlying sequence v_t iff $\{\phi_m\}$ is of size $-a$, for ϕ_m given by

$$\phi_m \equiv \sup_t \|z_t - E_{t-m}^{t+m}(z_t)\|_{L2}$$

and where $E_{t-m}^{t+m}(z_t) = E(z_t|v_{t-m},\cdots,v_{t+m})$ and $\|\cdot\|_{L2}$ is the norm L_2 of a random variable, defined as $E^{1/2}|\cdot|^2$.

We will assume that the future values of v_t will not improve the conditional expectation of z_t, in the sense of Sims (1972), such that the forward values v_{t+r} $(r = 1, ..., m)$ are useless, but harmless. From the definition we can say that ϕ_m is the worst mean square forecast error when z_t is predicted by $E_{t-m}^{t+m}(z_t)$. When ϕ_m goes to zero at an appropiate rate, then z_t depends essentially on the recient epoch of v_t. If z_t depends on a finite number of lags of v_t then it is NED of any size.

The property of NED is maintained under sums and products (see Gallant and White (1988)) and verifies a LLN and a CLT (see Wooldridge and White (1988)). Under the concept of NED we can rewrite almost exactly the same results given in the previous sections writing NED where it was written α-mixing (with appropiate assumptions). This motivates the following definition.

Definition 5.1 A sequence $\{\varepsilon_t\}$ is weakly nonlinear I(0) (WNI(0)) if it is NED on an underlying α-mixing sequence $\{v_t\}$ but the sequence $\{x_t\}$ given by $x_t = \sum_{s=1}^{t} \varepsilon_t$ is not NED. We will say that x_t is WNI(1).

Definition 5.2 Two sequences $\{y_t\}$ and $\{x_t\}$ which are WNI(1) are weakly nonlinear cointegrated (WNCI) with cointegration function $g(\cdot,\cdot,\gamma)$, if $g(y_t, x_t, \gamma^*)$ is NED on some α-mixing sequence but the sequence $g(y_t, x_t, \gamma)$, it is not NED for $\gamma \neq \gamma^*$.

Notice that if x_t is WNI(1) then Δx_t is WNI(0). With these alternative definitions we can give a representation theorem, providing sufficient conditions for a model to be a NEC. Furthermore, we can also give sufficient conditions to ensure that the one-step (NLS) estimation of single equation NEC is consistent.

Let us suppose the following model

$$Z_t = \Phi_1 W_{t-1} + F(Z_{t-1}, \gamma) + U_t \quad [5.1]$$

where Z_t and U_t are $(r \times 1)$, W_t is $(n \times 1)$ Φ_1 is $(r \times n)$, and $F_\gamma : \Re^r \to \Re^r$ as a function of Z. The assumption and theorem that follows will be useful later.

Assumption 5.3

(a) The sequence $\{U_t\}$ is α-mixing of size $-v/(v-2)$ for $v > 2$, and the sequence $\{W_t\}$ given in [5.1] is NED on an underlying α-mixing sequence $\{A_t\}$, of size $-v/(v-2)$ for $v > 2$, in the sense that for ψ_m given as

$$\psi_m \equiv \sup_t E\|W_t - E(W_t|A_t, ..., A_{t-m})\|_S^2$$

it holds that $\psi_m \to 0$ as $m \to \infty$, where the norm $\|\cdot\|_S$ is introduced in Mira and Escribano (1995). See Appendix D.

(b) For the norm $\|\cdot\|_S$ we have

$$\|\nabla_Z F(Z, \gamma)\|_S \equiv \delta_Z < 1.$$

(c) The following moment conditions hold for $i = 2$

(i) $E\|W_t\|_S^i \leq \Delta_W^{(i)}$,

(ii) $E\|U_t\|_S^i \leq \Delta_U^{(i)}$,

(iii) $E\|U_t\|_S^i \|W_t\|_S^i \leq \Delta_{WU}^{(i)}$.

(d) $F(\cdot, \gamma)$ is continuously differentiable in each argument.

Assumption 5.3 (b) says that the spectral radious of the matrix of first partial derivatives is smaller than 1.

Theorem 5.4 Under Assumption 5.3 the sequence $\{Z_t\}$ given in [5.1] is NED on the underlying sequence $\{(U_t, A_t)\}$ of any size. □

Proof: See Appendix D.

The core of the proof is that if Z_t is NED on W_t and W_t is NED on A_t then Z_t is NED on A_t. Now we have the tools to give a representation theorem for a nonlinear error correction with linear cointegration, in the sense that we give sufficient conditions that ensure a balanced specification of the NEC.

Theorem 5.5 (Representation Theorem) Consider a nonlinear time series model for the sequence of $(n \times 1)$ vectors $\{X_t\}$, given by

$$X_t = F(X_{t-1}, X_{t-2}) + \varepsilon_t \quad [5.2]$$

where for simplicity only two lags are supposed. Let us suppose the following assumptions

(1) ε_t and ΔX_t are WNI(0);

(2) the function $F(X_{t-1}, X_{t-2})$ is nonlinear only in the first lag, i.e.

$$F(X_{t-1}, X_{t-2}) = G(X_{t-1}) + \Phi_2 X_{t-2};$$

(3) the function $H(X_{t-1})$ given by $H(X_{t-1}) = -(I - \Phi_2)X_{t-1} + G(X_{t-1})$ is not partially invertible; and

(4) $H(X_{t-1}) = J(\alpha' X_{t-1})$.

Then

(i) under Assumption (2) we have the following representation

$$\Delta X_t = \Psi_1 \Delta X_{t-1} + H(X_{t-1}) + \varepsilon_t \quad [5.3]$$

where $\Psi_1 = -\Phi_2$ and $H(X_t) : \Re^n \rightarrow \Re^n$ is given by $H(X_{t-1}) = -(I - \Phi_2)X_{t-1} + G(X_{t-1})$;

(ii) Assumption (3) is a necessary condition to ensure that [5.3] is a NEC;

(iii) Under Assumption (4), if we multiply [5.3] by α' we obtain

$$Z_t = \Phi_1 W_{t-1} + F(Z_{t-1}) + U_t \quad [5.4]$$

where $Z_t = \alpha' X_t$, $W_t = \Delta X_t$, $\Phi_1 = \alpha' \Psi_1$, and $F(Z_{t-1}) = \alpha' J(\alpha' X_{t-1}) + \alpha' X_{t-1}$;

(iv) under Assumptions (1)-(4) plus Assumption 5.3 for model [5.4] we have that Z_t is NED.

□

Proof: See Appendix D.

Note that (1) implies on [5.3] that $(1 - \Psi_1 L)$ cannot have a unit root. The result (iv) of the former theorem ensures that under Assumptions (1) to (4) plus 5.3 we have that [5.4] is a correctly specified NED. Consider the example. If model [2.7] with linear cointegration and nonlinear error correction which coincides with [5.3], then the expression [5.4] is given by

$$\begin{aligned} z_{1t} &= \phi_{11} w_{1,t-1} + \phi_{12} w_{2,t-1} + \phi_{13} w_{3,t-1} + z_{1,t-1} \\ &\quad + \alpha_{11} J_1(z_{1,t-1}, z_{2,t-1}) + \alpha_{12} J_2(z_{1,t-1}, z_{2,t-1}) + \alpha_{13} J_3(z_{1,t-1}, z_{2,t-1}) + u_{1t} \\ z_{2t} &= \phi_{21} w_{1,t-1} + \phi_{22} w_{2,t-1} + \phi_{23} w_{3,t-1} + z_{2,t-1} \\ &\quad + \alpha_{21} J_1(z_{1,t-1}, z_{2,t-1}) + \alpha_{22} J_2(z_{1,t-1}, z_{2,t-1}) + \alpha_{23} J_3(z_{1,t-1}, z_{2,t-1}) + u_{2t}. \end{aligned}$$

The condition given by Assumption 5.3 (b) says that $\text{RSpec}(\nabla_Z F(Z)) < 1$ where the function $\text{RSpec}(M)$ is the spectral radious of the matrix M. In this example we have

$$\nabla_Z F(Z) = \begin{pmatrix} 1 + \alpha_{11} \frac{\partial J_1}{\partial z_1} + \alpha_{12} \frac{\partial J_2}{\partial z_1} + \alpha_{13} \frac{\partial J_3}{\partial z_1} & \alpha_{11} \frac{\partial J_1}{\partial z_2} + \alpha_{12} \frac{\partial J_2}{\partial z_2} + \alpha_{13} \frac{\partial J_3}{\partial z_2} \\ \alpha_{21} \frac{\partial J_1}{\partial z_1} + \alpha_{22} \frac{\partial J_2}{\partial z_1} + \alpha_{23} \frac{\partial J_3}{\partial z_1} & 1 + \alpha_{21} \frac{\partial J_1}{\partial z_2} + \alpha_{22} \frac{\partial J_2}{\partial z_2} + \alpha_{23} \frac{\partial J_3}{\partial z_2} \end{pmatrix}.$$

For instance if we have only one equation and only one cointegration relation then $J_2 = J_3 = 0$ y $z_{2t} = 0$ (since α' is (1×2)) and the matrix $\nabla_Z F(Z)$ is

$$\nabla_Z F(Z) = \left(1 + \alpha_{11} \frac{\partial J_1}{\partial z_1} \right).$$

Therefore condition $\text{RSpec}(\nabla_Z F(Z)) < 1$ reduces to $|1 + \alpha_{11} \frac{\partial J_1}{\partial z_1}| < 1$. See Mira (1996) for some comments about the case of nonlinear cointegration and nonlinear error correction.

Theorem 5.5 can be as well stated replacing Assumtion (1) by Assumption (1') given by

(1') ε_t is WNI(0);

and in this case we obtain Theorem 5.6. This theorem provides sufficient conditions to jointly ensure that ΔX_t and $\alpha' X_t$ are NED, based again in Mira and Escribano (1995).

Theorem 5.6 Let us suppose (1') plus (2) to (4) of Theorem 5.5, then Assumptions CT, CN, and LR from Mira and Escribano (1985) applied to model

$$\xi_t = \Xi_1 \xi_{t-1} + \Xi_2 \xi_{t-2} + F(\xi_{t-1}) + \eta_t \quad [5.5]$$

where $\xi_t' = [Z_t', \Delta X_{2t}]$, ensure jointly that ΔX_t and $\alpha' X_t$ are NED.

Proof: For the specification of the variables and parameters as well as a sketch of the proof see Appendix D.

Lastly, once model [5.3] is ensured to be a correctly specified NEC, it is of interest to give sufficient conditions that ensure its one step consistent estimation, in the sense of Stock (1994). The following theorem is about this issue.

Theorem 5.7 Suppose the Assumptions of Theorem 5.5 are satisfied for [5.3] and [5.4]. Now, Assumptions of Mira and Escribano (1995) on each equation of [5.3] allow its consistent estimation.

Proof: See Appendix D.

6 Conclusions

We have shown how, by working with the concept of α-mixing, we can estimate several types of interesting nonlinear time series models in a nonstationary framework. By doing that, we extended the concept of I(1) to strongly nonlinear I(1), SNI(1), and of cointegration to strongly nonlinear cointegration. Using results from functional analysis, we give sufficient conditions to obtain a super-consistent estimator of a nonlinear cointegration relationship estimated by nonlinear least squares (NLS). This framework allowed us to extend the two-step estimator of Engle and Granger(1987) to nonlinear error correction models (NEC). In these class of models the cointegrating relationship can be linear or nonlinear. There are available some statistics that can be used to test the hypothesis of α-mixing. A weaker concept of nonlinear I(1) is introduced based on the concept of near epoch dependence (NED). With this concept of weakly nonlinear I(1), WNI(1), we can give a representation theorem for NEC models with linear cointegration and we can justify a one-step (NLS) estimation of NEC models. Finally, the small sample biases are studied by running Monte Carlo simulations. It is found that for samples of size 100, the biases in the estimation of the parameters of the model can be large, but that those biases are substantially reduced when the sample size increases to 500 observations or higher.

A Appendix to Section 2

A.1 Proof of Lemma 2.6

For the first part define $w_t = f_x(x_t)$ y $r_t = f_y(y_t)$. Now, define $g_T(w_t, r_t) = g_R(f_x^{-1}(w_t), f_y^{-1}(r_t))$. Clearly $g_R(f_x^{-1}(w_t), f_y^{-1}(r_t)) = g_R(x_t, y_t)$ and then it is α-mixing.

The second part is more straightforward. Define $g_R(x_t, y_t) = g_T(f_x(x_t), f_y(y_t))$ and the result follows.

Q.E.D.

A.2 Proof of Proposition 2.8

Let us write

$$\begin{aligned} X_t &= F(X_{t-1}, X_{t-2}) + \varepsilon_t \\ &= G(X_{t-1}) + \Phi_2 X_{t-2} + \varepsilon_t \\ \Delta X_t &= G(X_{t-1}) - X_{t-1} + \Phi_2 X_{t-2} + \varepsilon_t \\ &= (-\Phi_2)(X_{t-1} - X_{t-2}) - (I - \Phi_2)X_{t-1} + G(X_{t-1}) + \varepsilon_t \\ &= \Psi_1 \Delta X_{t-1} + H(X_{t-1}) + \varepsilon_t \end{aligned}$$

where $\Psi_1 = -\Phi_2$, and $H(X_{t-1}) = -(I - \Phi_2)X_{t-1} + G(X_{t-1})$. Now, since ε_t and ΔX_t are SNI(0) then $H(X_{t-1})$ is also SNI(0), eventhough X_t is not. If that not were the case then $H(\cdot)$ would be invertible and then X_t would be a function of α-mixing variables and therefore it were not SNI(1). Then given (1) and (2), (3) is a necessary condition.

Q.E.D.

B Appendix to Section 3

B.1 Proof of Lemma 3.3

We will prove that $T^{-1}\sum_{t=1}^{T} g(y_{t-1}, z_{t-1}, \gamma)^2 \to \infty$. To do that we will use Theorem 3.1. We will write g_{t-1}^2 instead of the expression $g(y_{t-1}, z_{t-1}, \gamma)^2$. Then from the assumptions we can write

$$\begin{aligned} T^{-2}\sum_{t=1}^{T} g_{t-1}^2 &= T^{-1}\sum_{t=1}^{T}(\Phi(\sum_{s=1}^{t-1}\phi_s, \sum_{s=1}^{t-1}\delta_s))^2 T^{-1} \\ &= T^{-1}\int_0^1 M_T(r)dr \end{aligned}$$

where $M_T(r)$ is given by

$$M_T(r) = \begin{cases} 0 & \text{for } 0 \le r < \frac{1}{T} \\ (\Phi(\phi_1, \delta_1))^2 & \text{for } \frac{1}{T} \le r < \frac{2}{T} \\ \vdots & \vdots \\ (\Phi(\sum_{s=1}^{T-1}\phi_s, \sum_{s=1}^{T-1}\delta_s))^2 & \text{for } \frac{T-1}{T} \le r < 1 \\ (\Phi(\sum_{s=1}^{T}\phi_s, \sum_{s=1}^{T}\delta_s))^2 & \text{for } r = 1 \end{cases}.$$

Now we have the following convergences

$$\begin{aligned} T^{-1/2}\sum_{s=1}^{[Tr]}\phi_s &\xrightarrow{d} \sigma_1 W_1(r) \\ T^{-1/2}\sum_{s=1}^{[Tr]}\delta_s &\xrightarrow{d} \sigma_2 W_2(r) \\ T^{-1/2}\Phi(\sum_{s=1}^{[Tr]}\phi_s, \sum_{s=1}^{[Tr]}\delta_s) &\xrightarrow{d} D\Phi(0; \sigma_1 W_1(r), \sigma_2 W_2(r)) \\ T^{-1}M_T(r) \equiv T^{-1}(\Phi(\sum_{s=1}^{[Tr]}\phi_s, \sum_{s=1}^{[Tr]}\delta_s))^2 &\xrightarrow{d} (D\Phi(0; \sigma_1 W_1(r), \sigma_2 W_2(r)))^2 \equiv \widetilde{W}(r)^2 \\ \int_0^1 T^{-1}M_T(r)dr &\xrightarrow{d} \int_0^1 \widetilde{W}(r)^2 dr \end{aligned}$$

Since $T^{-2}\sum_{t=1}^{T} g_{t-1}^2 \xrightarrow{d} \int_0^1 \widetilde{W}(r)^2 dr$, then $T^{-1}\sum_{t=1}^{T} g_t^2 \to \infty$, and the NLS estimator γ^T given by $\min_\gamma Q_T(\gamma)$ where $Q_T(\gamma) = T^{-1}\sum_{s=1}^{T} g_t(\gamma)^2$, provides a consistent estimation of γ^*.
Q.E.D.

B.2 Proof of Lemma 3.7

Let us define the $((k+1)\times 1)$ vectors

$$\mathbf{h}_t = \begin{pmatrix} g_{t-1}^* \\ \mathbf{l}_t \end{pmatrix} = \begin{bmatrix} h_{1t} \\ \mathbf{h}_{2t} \end{bmatrix} \quad \text{and} \quad \mathbf{k}_t = \sum_{s=1}^{t}\mathbf{h}_s = \begin{pmatrix} \sum_{s=1}^{t} g_{s-1}^* \\ \mathbf{x}_t \end{pmatrix} = \begin{bmatrix} k_{1t} \\ \mathbf{k}_{2t} \end{bmatrix}.$$

If we apply Theorem 3.5 we obtain the following convergences to the $((k+1)\times(k+1))$ matrices

$$T^{-2}\sum_{t=1}^{T}\mathbf{k}_t\mathbf{k}_t' \xrightarrow{d} \int_0^1 \mathbf{B}(r)\mathbf{B}(r)'dr$$
$$T^{-1}\sum_{t=1}^{T}\mathbf{k}_{t-1}\mathbf{h}_t' \xrightarrow{d} \int_0^1 \mathbf{B}(r)d\mathbf{B}(r)'+\Sigma_1$$

where the $((k+1)\times 1)$ vector $\mathbf{B}(r)$ is given by $\mathbf{B}(r)=[B_1(r),\mathbf{B}_2(r)']'$ and

$$\Sigma_1=\begin{pmatrix}\sigma_1 & \Sigma_{12}' \\ \Sigma_{12} & \Sigma_{22}\end{pmatrix}$$

and an analogous decomposition can be made for Σ. Now we have the following convergences

$$T^{-2}X'X = T^{-2}\sum_{t=1}^{T}\mathbf{x}_t'\mathbf{x}_t = T^{-2}\sum_{t=1}^{T}\mathbf{k}_{2t}'\mathbf{k}_{2t} \xrightarrow{d} \int_0^1 \mathbf{B}_2(r)'\mathbf{B}_2(r)dr$$
$$T^{-1}V'X = T^{-1}\sum_{t=1}^{T}g_t^*\mathbf{x}_t = T^{-1}\sum_{t=1}^{T}h_{1,t+1}\mathbf{k}_{2t} \xrightarrow{d} \int_0^1 \mathbf{B}_2(r)dB_1(r)+\Sigma_{12}$$

and the result follows. Note that $T^{-1}\sum_{t=1}^{T}h_{1,t+1}\mathbf{k}_{2t}=T^{-1}\sum_{t=1}^{T}h_{1t}\mathbf{k}_{2,t-1}+o_p(1)$.
Q.E.D.

C Appendix to Section 4

C.1 Proof of Theorem 4.2

Let us write model [4.2] and [4.3] as

$$\overline{R}-K'B^*-F^T(\gamma_2^*) = \left(F^*(\gamma_2^*)-F^T(\gamma_2^*)\right)+V \quad [4.4]$$
$$G^T(\theta^*) = \left(F^*(\gamma_2^*)-F^T(\gamma_2^*)\right)+V \quad [4.5]$$

For model [4.5] we have

$$\begin{aligned}\lim_{T\to\infty} T^{-1/2}(\theta^T-\theta^*)' &= \lim_{T\to\infty}\left(T^{-1/2}G^T(\theta^*)'G_\theta^T(\theta^*)\right)\left(T^{-1}G_\theta^T(\theta^*)'G_\theta^T(\theta^*)\right)^{-1} \\ &= \lim_{T\to\infty}\left(T^{-1/2}(F^*(\gamma_2^*)-F^T(\gamma_2^*))'G_\theta^T(\theta^*)\right)\left(T^{-1}G_\theta^T(\theta^*)'G_\theta^T(\theta^*)\right)^{-1} \\ &\quad+\lim_{T\to\infty}\left(T^{-1/2}V'G_\theta^T(\theta^*)\right)\left(T^{-1}G_\theta^T(\theta^*)'G_\theta^T(\theta^*)\right)^{-1}\end{aligned}$$

and since we want that

$$\lim_{T\to\infty} T^{-1/2}(\theta^T - \theta^*)' = \lim_{T\to\infty} (T^{-1/2}G^*(\theta^*)'G_\theta^*(\theta^*))(T^{-1}G_\theta^*(\theta^*)'G_\theta^*(\theta^*))^{-1}$$

then Assumption 4.1 is enough.
Q.E.D.

D Appendix to Section 5

D.1 The $\|\cdot\|_S$ Norm

The matrix norm $\|\cdot\|_S$ is defined as follows

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

for M and D_δ being matrices that depend on the matrix A. Analoguosly the asociated vectorial norm is

$$\|Y\|_S \equiv \|(MD_\delta)Y\|_\infty$$

In Mira and Escribano (1995) it is proved that for any matrix A it holds that

$$\|A\|_S \leq \rho(A) + \delta$$

for $\rho(A)$ being the spectral radious of A.

D.2 Proof of Theorem 5.4

Let us define

$$\overline{Z}_t = \begin{cases} F(\overline{Z}_{t-1}) & \text{for} \quad t > 0 \\ 0 & \text{for} \quad t \leq 0 \end{cases}$$

and

$$\tilde{Z}^m_{t,s} = \begin{cases} \Phi\widetilde{W}_{t-1} + F(\tilde{Z}^m_{t-1,s+1}) + U_t & \text{for} \quad s+1 \leq m \\ \overline{Z}_t & \text{for} \quad s+1 > m \end{cases}$$

where $\widetilde{W}_t = E(W_t|A_t, ..., A_{t-m})$, and therefore $E\|W_t - \widetilde{W}_t\|_S^2 \leq \psi_m$ such that $\psi_m \to 0$ when $m \to \infty$. Then it is clear that $\widetilde{Z}_{t,0}^m$ is $\sigma(U_t, \widetilde{W}_{t-1}, ..., U_{t-m+1}, \widetilde{W}_{t-m})$-medible, and then it is $\sigma(U_t, A_{t-1}, ..., U_{t-m+1}, A_{t-m}, ..., A_{t-2m})$-medible,

The difference between Z_t and its predictor $\overline{Z}_t$ is bounded for $t > 0$, because

$$\begin{aligned} \|Z_t - \overline{Z}_t\|_S &= \|\Phi W_{t-1} + F(Z_{t-1}) + U_t - F(\overline{Z}_{t-1})\|_S \\ &\leq \|\Phi W_{t-1} + U_t\|_S + \|F(Z_{t-1}) - F(\overline{Z}_{t-1})\|_S \end{aligned}$$

and by the Mean Value Theorem

$$\begin{aligned} F(Z_t) - F(\overline{Z}_t) &= \begin{pmatrix} F_1(Z_t) - F_1(\overline{Z}_t) \\ \vdots \\ F_r(Z_t) - F_r(\overline{Z}_t) \end{pmatrix} \\ &= \begin{pmatrix} \frac{\partial F_1}{\partial z_1}(\ddot{Z}_t)(z_{1t} - \overline{z}_{1t}) & + & \cdots & + & \frac{\partial F_1}{\partial z_r}(\ddot{Z}_t)(z_{rt} - \overline{z}_{rt}) \\ \vdots & & \ddots & & \vdots \\ \frac{\partial F_r}{\partial z_1}(\ddot{Z}_t)(z_{1t} - \overline{z}_{1t}) & + & \cdots & + & \frac{\partial F_r}{\partial z_r}(\ddot{Z}_t)(z_{rt} - \overline{z}_{rt}) \end{pmatrix} \\ &= \begin{pmatrix} \frac{\partial F_1}{\partial z_1}(\ddot{Z}_t) & \cdots & \frac{\partial F_1}{\partial z_r}(\ddot{Z}_t) \\ \vdots & & \vdots \\ \frac{\partial F_r}{\partial z_1}(\ddot{Z}_t) & \cdots & \frac{\partial F_r}{\partial z_r}(\ddot{Z}_t) \end{pmatrix} \begin{pmatrix} z_{1t} - \overline{z}_{1t} \\ \vdots \\ z_{rt} - \overline{z}_{rt} \end{pmatrix} \\ &= \nabla_Z F(\ddot{Z}_t)(Z_t - \overline{Z}_t). \end{aligned}$$

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

$$\begin{aligned} \|Z_t - \overline{Z}_t\|_S &\leq \|\Phi\|_S\|W_{t-1}\|_S + \|U_t\|_S + \|\nabla_Z F(\ddot{Z}_t)\|_S\|(Z_{t-1} - \overline{Z}_{t-1})\|_S \\ &\leq \delta_{WU,t} + \delta_Z\|(Z_{t-1} - \overline{Z}_{t-1})\|_S \end{aligned}$$

for some $N_{WU,t}$ and since $Z_0 = \overline{Z}_0 = 0$, then by iteration we obtain

$$\begin{aligned} \|Z_t - \overline{Z}_t\|_S &\leq \sum_{j=0}^{t-1} N_{WU,t-j}\delta_Z^j \\ \|Z_t - \overline{Z}_t\|_S^2 &\leq \sum_{j=0}^{t-1} N_{WU,t-j}^2\delta_Z^{2j} + \sum_{j=0}^{t-1}\sum_{i\neq j}^{t-1} N_{WU,t-i}N_{WU,t-j}\delta_Z^{i+j} \\ E\|Z_t - \overline{Z}_t\|_S^2 &\leq \Delta_{Z-\overline{Z}}^{(2)} \end{aligned}$$

for some bound $\Delta_{Z-\overline{Z}}^{(2)}$, because, for instance, $E(N_{WU,t}) = \|\Phi\|_S\Delta_W^{(1)} + \Delta_U^{(1)}$. Now,

$$\begin{aligned} \|Z_t - \widetilde{Z}_{t,0}^m\|_S &= \|\Phi W_{t-1} + F(Z_{t-1}) + U_t - \Phi\widetilde{W}_{t-1} - F(\widetilde{Z}_{t-1,1}^m) - U_t\|_S \\ &\leq \|\Phi(W_{t-1} - \widetilde{W}_{t-1})\|_S + \|F(Z_{t-1}) - F(\widetilde{Z}_{t-1,1}^m)\|_S \end{aligned}$$

and again by the Mean Value Theorem we obtain

$$\|F(Z_{t-1}) - F(\tilde{Z}^m_{t-1,1})\|_S \leq \|\nabla_Z F(Z)\|_S \|Z_{t-1} - \tilde{Z}^m_{t-1,1}\|_S,$$

and since $\|\nabla_Z F(Z)\|_S \leq \delta_Z$ we have

$$\|Z_t - \tilde{Z}^m_{t,0}\|_S \quad \leq \quad \|\Phi\|_S \|W_{t-1} - \widetilde{W}_{t-1}\|_S + \delta_Z \|Z_{t-1} - Z^m_{t-1,1}\|_S$$

and by iteration

$$\|Z_t - \tilde{Z}^m_{t,0}\|_S \leq \sum_{i=0}^{m} \delta^i_Z \|\Phi\|_S \|W_{t-1-i} - \widetilde{W}_{t-1-i}\|_S + \delta^m_Z \|Z_{t-m} - \overline{Z}_{t-m}\|_S$$

and taking expectations

$$\begin{aligned} E\|Z_t - \tilde{Z}^m_{t,0}\|^2_S \quad \leq \quad & E(\sum_{i=0}^{m} \delta^i_Z \|\Phi\|_S \|W_{t-1-i} - \widetilde{W}_{t-1-i}\|_S)^2 + \delta^{2m}_Z E\|Z_{t-m} - \overline{Z}_{t-m}\|^2_S \\ & + 2E((\sum_{i=0}^{m} \delta^i_Z \|\Phi\|_S \|W_{t-1-i} - \widetilde{W}_{t-1-i}\|_S) \times \delta^m_Z \|Z_{t-m} - \overline{Z}_{t-m}\|_S). \end{aligned}$$

If we use for the third term in the summation the Holder inequality with $p = \frac{1}{2} = q$, i.e., $E|YX| \leq E^{1/2}|Y|^2 + E^{1/2}|X|^2$, only remains to work out the following term

$$\begin{aligned} & E(\sum_{i=0}^{m} \|W_{t-1-i} - \widetilde{W}_{t-1-i}\|_S)^2 \\ = \quad & E\sum_{i=0}^{m} \|W_{t-1-i} - \widetilde{W}_{t-1-i}\|^2_S \\ & + E\sum_{i=0}^{m}\sum_{j\neq i}^{m} \|W_{t-1-i} - \widetilde{W}_{t-1-i}\|_S \|W_{t-1-j} - \widetilde{W}_{t-1-j}\|_S \\ \leq \quad & \sum_{i=0}^{m} E\|W_{t-1-i} - \widetilde{W}_{t-1-i}\|^2_S \\ & + \sum_{i=0}^{m}\sum_{j\neq i}^{m} E^{1/2}\|W_{t-1-i} - \widetilde{W}_{t-1-i}\|^2_S E^{1/2}\|W_{t-1-j} - \widetilde{W}_{t-1-j}\|^2_S \end{aligned}$$

and since $E\|W_{t-1-i} - \widetilde{W}_{t-1-i}\|^2_S = \psi_m$ then $E\|Z_t - \tilde{Z}^m_{t,0}\|^2_S$ is bounded by a sumation of terms with ψ_m or terms with δ_Z and since ψ_m goes to zero and $0 < \delta_Z < 1$ we obtain

$$\lim_{m\to\infty} E\|Z_t - \tilde{Z}^m_{t,0}\|_S = 0$$

Now, given $E_{t-2m}(Z_t) \equiv E(Z_t|U_t, A_{t-1}, ..., U_{t-2m+1}, A_{t-2m})$, we can obtain a bound for $\|Z_t - E_{t-2m}(Z_t)\|_{LS}$. Since $\widetilde{Z}^m_{t,0}$ is σ-$(U_t, ..., U_{t-m+1}, A_{t-m}, ...A_{t-2m})$-medible then it is σ-$(U_t, ..., U_{t-2m+1}, A_{t-2m})$-medible so that

$$\begin{aligned} \|Z_t - E_{t-m}(Z_t)\|_{L2} &\leq \delta_K \|Z_t - \widetilde{Z}^m_{t,0}\|_{L2} \\ &= \delta_K E^{1/2} \|Z_t - \widetilde{Z}^m_{t,0}\|^2_S \end{aligned}$$

and since $E\|Z_t - \widetilde{Z}^m_{t,0}\|^2_S \to 0$ at exponential rate then $\{Z_t\}$ is NED on the underlying sequence $\{(U_t, W_t)\}$ of any size. Note that the first inequality is a generalization of the well known fact $E|Z_t - E(Z_t|I_t)|^2 \leq E|Z_t - g(I_t)|^2$ for any function $g(\cdot)$ of the information set I_t and δ_Z is some constant that depends on the norm $\|\cdot\|_S$.

Q.E.D.

D.3 Proof of Theorem 5.5

Apply Proposition 2.8 for parts (i) and (ii). Part (iii) is immediate. For part (iv) apply Teorem 5.4. *Q.E.D.*

D.4 Sketch of the Proof of Theorem 5.6

Let us normalize the $(r \times n)$ matrix, base of the space of cointegration relations, in the following way $\alpha' = [I, -\beta']$ such that $\alpha' X_t = Z_t$, and let us define the $(n \times n)$ matrix M as

$$M = \begin{pmatrix} I & -\beta' \\ 0 & I \end{pmatrix}.$$

Then $MX_t = [Z_t', X_{2t}']'$ for some partition of the vector X_t as $X_t' = [X_{1t}', X_{2t}']$, with X_{1t} of dimension $(r \times 1)$ and X_{2t} of dimension $((n-r) \times 1)$. Given the NEC representation

$$\Delta X_t = \Psi \Delta X_{t-1} + J(\alpha' X_{t-1}) + \varepsilon_t$$

if we multiply by M we obtain the following system

$$\begin{aligned} \alpha' \Delta X_t &= \alpha' \Psi \Delta X_{t-1} + \alpha' J(\alpha' X_{t-1}) + \alpha' \varepsilon_t \\ \Delta X_{2t} &= \Psi_2 \Delta X_{t-1} + J_2(\alpha' X_{t-1}) + \varepsilon_{2t} \end{aligned}$$

for some partition of ε_t, Ψ, and $J(\alpha' X_{t-1})$. Let us represent the vector $[Z'_{t-1}, X'_{2,t-1}]'$ as L_{t-1}, then the system can be rewritten as

$$\begin{aligned} Z_t &= Z_{t-1} + \alpha' \Psi M^{-1} \Delta L_{t-1} + \alpha' J(\alpha' X_{t-1}) + \alpha' \varepsilon_t \\ \Delta X_{2t} &= \Psi_2 M^{-1} \Delta L_{t-1} + J_2(\alpha' X_{t-1}) + \varepsilon_{2t} \end{aligned}$$

or

$$\begin{aligned} Z_t &= Z_{t-1} + P \Delta L_{t-1} + K(Z_{t-1}) + \eta_{1t} \\ \Delta X_{2t} &= \Psi_2 \Delta L_{t-1} + J_2(Z_{t-1}) + \eta_{2t} \end{aligned}$$

that is strightforward to rewrite as in [5.5].
Q.E.D.

D.5 Proof of Theorem 5.7

Apply the proof of Theorem 3.5 of Mira and Escribano (1995), with the caveat that Lemma 3.4 (i) should be modified as in Theorem 3.4.
Q.E.D.

References

Amemiya, T., (1985), *Advanced Econometrics.* Oxford: Basil Blackwell.

Andrews, D. W. K., (1988), "Laws of Large Numbers for Dependent Non-Identically Distributed Random Variables", *Econometric Theory, 4, 458-467.*

Bierens, H. J., (1981), *Robust Methods and Asymptotic Theory in Nonlinear Econometrics.* Lecture Notes in Economics and Mathematical Systems, vol. 192, Berlin: Springer Verlag.

Billingsley, P., (1984), *Convergence of Probability Measures.* New York: Wiley.

Davidson, J., (1994), *Stochastic Limit Theory.* New York: Oxford University Press.

Engle, R. F. and C. W. J. Granger, (1987), "Co-Integration and Error Correction: Representation, Estimation, and Testing", *Econometrica, 55 (2), 251-276.*

Escribano, A., (1986), *Non-Linear Error-Correction: The Case of Money Demand in the U.K. (1878-1970).* Chapter IV. Ph. D. Dissertation. University of California, San Diego.

Escribano, A., (1987), "Error-Correction Systems: Nonlinear Adjustment to Linear Long-Run Relationships", *CORE Discussion Paper 8730. C.O.R.E..*

Gallant, A. R. y H. White, (1988), *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models.* New York: Basil Blackwell.

Granger, C. W. J., (1988), "Models that generate trends", *Journal of Time Series, 9, 329-343.*

Granger, C. W. J., (1995), "Modelling Nonlinear Relationships Between Extended-Memory Variables", *Econometrica, 63 (2), 265-279.*

Granger, C. W. J., and J. Hallman, (1991), "Long Memory Series with Attractors", *Oxford Bulletin of Economics and Statistics, 53, 11-26.*

Granger, C. W. J., and N. Swanson, (1995), "Further developments in the study of cointegrated variables", *Mimeo.*

Granger, C. W. J., and T. Teräsvirta, (1995), *Modelling Nonlinear Economic Relationships.* New York: Oxford University Press.

Hendry, D. F., and N. R. Ericsson, (1991), "An Econometric Analisis of the U.K. 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.*

Heesterman, C. C., and R. D. Gill, (1992), "A Central Limit Theorem for M-estimators by the von Mises Method", *Statistica Neerlandica, 46, 165-177.*

Herrndorf, N., (1984), "A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables", *Annals of Probability, 12, 141-153.*

Johansen, S., (1988), "Statistical Analysis of Cointegration Vectors", *Journal of Economics Dynamic and Control, 12, 231-54.*

Kolmogorov, A. N., and S. V. Fomin, (1978), *Elementos de la Teoría de Funciones y del Análisis Real.* Moscú: Ed. Mir.

Kwiatkowski, D., P. C. B. Phillips, P. Schmidt e Y. Shin, (1992), "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root", *Journal of Econometrics, 54, 159-178.*

Lo, A. W., (1991), "Long-Term Memory in Stock Market Prices", *Econometrica, 59, 1279-1313.*

Mira, S., (1996), *Modelos Econométricos Dinámicos No Lineales con Tendencias Estocásticas.* Ph. D. Dissertation. Universidad Carlos III de Madrid. Madrid.

Mira, S. and A. Escribano, (1995), "Nonlinear Time Series Models: Consistency and Asymptotic Normality of NLS Under New Conditions", *Working Paper 95-42, Universidad Carlos III de Madrid. Madrid.*

Phillips, P. C. B., (1987), "Time Series Regression with a Unit Root", *Econometrica, 55, 277-301.*

Phillips, P. C. B. y Durlauf, S. N., (1986), "Multiple Time Series Regression with Integrated Processes", *Review of Economic Studies, 473-495.*

Potscher, B. M. y I. R. Prucha, (1991a), "Basic Structure of the Asymptotic Theory in Dynamic Nonlinear Econometric Models, Part I: Consistency and Approximation Concepts", *Econometric Reviews, 10 (2), 125-216.*

Sims, C. A., (1972), "Money, Income, and Causality", *The American Economic Review, 62, 540-552.*

Stock, J. H., (1987), "Asymptotic Properties of Least Squares Estimation of Cointegrating Vectors", *Econometrica, 55, 1035-1056.*

Stock, J. H., (1994), "Deciding between I(1) and I(0)", *Journal of Econometrics, 63, 105-131.*

White, H., (1984), *Asymptotic Theory for Econometricians.* New York: Academic Press.

Wooldridge, J. M., and H. White, (1988), "Some Invariance Principles and Central Limit Theorems for Dependent Heterogeneous Processes", *Econometric Theory 4, 210-30.*