ERROR-CORRECTION SYSTEMS : NONLINEAR ADJUSTMENTS TO LINEAR LONG-RUN RELATIONSHIPS.

by

Alvaro ESCRIBANO

CORE DISCUSSION PAPER N° 8730

CENTER FOR OPERATIONS RESEARCH & ECONOMETRICS

UNIVERSITE CATHOLIQUE DE LOUVAIN

Voie du Roman Pays, 34, B-1348 Louvain-la-Neuve (Belgium)

ERROR-CORRECTION SYSTEMS : NONLINEAR ADJUSTMENTS TO LINEAR LONG-RUN RELATIONSHIPS. *

by

Alvaro ESCRIBANO

April 1987

Abstract

In this paper a unifying approach, based on conditional expectations, is used to generate nonlinear time series models. In particular, we show how to derive nonlinear error-correction models that postulate as targets linear long-run relationships and allow for nonlinear short-run adjustments. This framework incorporates both a variety of integrated stochastic processes and departures from long-run relationships with broad time series properties. The analysis is done for system of equations in reduced and structural form. Furthermore, since most of the empirical applications available are based on single equation models, we derive also the corresponding single equation models and discuss the appropriateness of the concept of weak exogeneity in error-correction models.

* This paper is based on my Ph.D. Thesis. I am indebted to my advisor C.W.J. Granger. I wish to thank Rob Engle for his many suggestions. My thanks also go to Herman Bierens, Bernadette Govaerts and Halbert White for their helpful comments. All errors are my sole responsibility.

ISSN. 0771 3894

1. INTRODUCTION

Most macroeconomic variables are nonstationary. This fact was recognized since the early work on time series analysis. The dominant element that generated the non-stationarity was denominated the "trend". The presence of time trends has strong implications either if we are modelling economic relationships or if we are interested in doing statistical inference. Engle and Granger (1985) have shown that the long run properties of the data should be taken into account, even if we are only interested in modelling economic relationships along the so-called "business cycles". However, their approach is restricted to the situation where there exists a transformation of the observations that reduce them, the nonstationary variables, to be stationary. Sometimes, this requirement is too strong for economic variables, see for example White (1984) and Phillips (1987). The concepts of time trends and integrated stochastic processes, introduced by Escribano (1987), are here used in a nonlinear context. This is a difficult task if we work, within the "classical" time series framework, with the usual concept of integration.

In this paper, a unifying approach, based on conditional expectations, is used to generate time series models. The nice feature of this framework is that it can be justified as the solution to an optimization problem. We show how, by using this approach, we can derive nonlinear time series models like : nonlinear vector autoregressions, nonlinear vector autoregressive moving averages and nonlinear vector moving averages. In particular, we derive nonlinear error-correction systems. Since the linear models are nested in this nonlinear extension, we have in addition derived linear error-correction models.

An important advantage of this approach is that it allows us to model economic variables with a broader set of statistical properties than those allowed by the "classical" linear time series approach based on Wold's representation. In fact we can model economic variables that contain general sources of growth and we can allow for different concepts of long-run equilibrium relationships. The reason is that we permit some heterogeneity in the behaviour of the observations.

The paper is organized as follows. In Section 2 we introduce the statistical framework for nonlinear time series models. Section 3 derives and establishes the conditions for a well-defined error-correction system. This is done for systems of equations in reduced and structural form. Special attention is given to the time series properties of nonlinear transformations. Section 4 discusses the single equation error-correction model. And finally section 5 includes some conclusions.

2. A STOCHASTIC FRAMEWORK FOR NONLINEAR TIME SERIES

A probability space is a triple $(\Omega, \mathcal{F}, \mathbb{P})$ where Ω is the sample space, $\mathcal{F}$ is the Borel-σ-field and $\mathbb{P}$ is the probability measure that induces the joint probability distribution of the data. A Borel-σ-field can be thought of as the minimal collection of events of Ω with the property that for all events in $\mathcal{F}$ the probability measure $\mathbb{P}$ is well defined. The translation of events into real numbers is done by $\mathcal{F}$-measurable functions. An $\mathcal{F}$-measurable function $f_i : \Omega \to \mathbb{R}$ assigns for every real number a the set of events $[\omega : f_i(\omega) \leq a] \in \mathcal{F}$. Notice that a random variable x_{jt} is also a function such that $x_{it} : \Omega \to \mathbb{R}$. Finally a random vector x_t is a vector whose elements are random variables $\{x_{it} : i = 0, 1, \ldots, N$ with $x_{it} = 0$ for all $i \leq 0\}$ where t indicates time.

Let x_t be an $N \times 1$ vector with a vector of initial conditions given by x_0. In matrix notation, if we have T observations of each of the N variables,

$$\underline{X}_t = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1T} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{NT} \end{bmatrix} = (x_1, x_2, \ldots, x_T).$$

By introducing the matrix of initial conditions X_0 we form X_T

$$X_T = (X_0, x_1, \ldots, x_T) = (X_0, \underline{X}_T).$$

The matrix X_T contains all data information available in the sample up to time T. In probability theory the way to represent the information set is by the use of the σ-fields generated by those random variables. That is, $\mathcal{F}_0^t \equiv \sigma(X_0, x_1, \ldots, x_t)$ indicates the information generated by the past and present of the vector x_t and $\mathcal{F}_{t+\tau} \equiv \sigma(x_{t+\tau}, \ldots)$ indicates the future information generated by $x_{t+\tau}$ from time $t + \tau$ onwards. In other words, the set of events generated by $X_0, x_1, \ldots, x_{t-1}$ is $\mathcal{F}_0^{t-1}$.

From now on and when there is no risk of confusion we maintain, as it is usual in the econometrics literature, the notation in terms of the sample observations, X_T, instead of the set of events generated by X_T, $\mathcal{F}_0^T$.

Following Hendry and Richard (1983), let the data generating process of $\underline{X}_T$ given X_0, be represented by the joint probability density function

$$\mathbb{P}(\underline{X}_T \mid X_0, \Theta)$$

where Θ is a finite dimensional vector of unknown parameters that characterize the data density.

The joint probability density function can always be factorized in a sequential way as

$$\mathbb{P}(\underline{X}_T \mid X_0, \Theta) = \prod_{t=1}^{T} \mathbb{P}(x_t \mid X_{t-1}, \Theta)$$

where $X_{t-1} = (X_0, x_1, \ldots, x_{t-1})$, or more explicitly,

$$\mathbb{P}(x_1, \ldots, x_T \mid X_0, \Theta) = \prod_{t=1}^{T} \mathbb{P}(x_t \mid X_{t-1}, \Theta).$$

Let $x_t \mid X_{t-1}, \Theta$ have conditional mean and variance given by

$$x_t \mid X_{t-1}, \Theta \sim \{E(x_t \mid X_{t-1}, \Theta), \Sigma_t\} \tag{2.1}$$

The conditional mean

$$E(x_t \mid X_{t-1}, \Theta) = F(X_{t-1}, \Theta) = F(X_0, x_1, \ldots, x_{t-1}, \Theta). \tag{2.2}$$

In general, the functional form F is unknown since all we know is that the conditional expectation is a (Borel-measurable) function of the conditioning set, see Doob (1954). However, it is clear that F depends on the distribution of $x_t \mid X_{t-1}, \Theta$. For example, if $x_t \mid X_{t-1}, \Theta \sim N(\mu_t, \Sigma_t)$ then F is a linear function of the conditioning set.

In this paper we model only the conditional mean. However, we could also model the conditional variance by following the same procedure, getting nonlinear ARCH and nonlinear GARCH. See Engle (1982) and Bollerslev (1986) respectively, for the linear case. We can justify the analysis based on the conditional means by saying that the objective of the econometrician is to minimize the mean square forecast error. In other words, we want to consider those models that give us the "optimal" forecast of x_t conditional on the information set represented by $\mathcal{F}_0^{t-1}$.

It is known that this forecast function is given by $E[x_t \mid \mathcal{F}_0^{t-1}]$, see for example, Granger and Newbold (1977).

If we assume that $\varepsilon_t = x_t - E(x_t \mid X_{t-1}, \Theta)$, so that ε_t is a martingale difference sequence, with zero mean and $E(\varepsilon_t, \varepsilon_t') = \Omega_t$ a matrix with positive and finite main diagonal elements for all t, then the components of the vector ε_t are time trend-free in the second moment although they can be heteroskedastic. From (2.2) we obtain,

$$x_t = F(X_0, x_1, \dots, x_{t-1}, \Theta) + \varepsilon_t. \tag{2.3}$$

Equation (2.3) is a ***nonlinear vector autoregression in reduced form***, NVAR. Notice also that the conditions assumed on the innovations ε_t impose as well some conditions on the function F, namely, that $\varepsilon_t \equiv x_t - F(X_0, x_1, \dots, x_{t-1}, \Theta)$ has to be a martingale difference sequence, time trend-free in the second moment and heteroskedastic (necessary conditions). This is important since the properties of the function F are going to be contingent on the time series properties of the data. For example, if x_t has an unbounded time trend in the second moment, see appendix A, the F cannot be bounded since otherwise ε_t will not be time trend-free in the second moment, which will create a problem of dynamic misspecification in equation (2.3).

Furthermore, since the innovations are also part of the σ-field generated by the x's then they are also implicitly incorporated in the conditioning set. That is $\mathcal{F}_0^{t-1}$ includes those x's and ε's that occur before time t. This fact allows us to obtain a sometimes more parsimonious class of nonlinear systems, that we denominate, ***nonlinear vector autoregressive moving average systems***, NVARMA(p,q),

$$x_t = F_{ab}(x_{t-p}, \dots, x_{t-1}, \varepsilon_{t-q}, \dots, \varepsilon_{t-1}, \Theta) + \varepsilon_t \tag{2.4}$$

where $q < T$ is the maximum order of the moving average components and $p < T$ is the maximum order of the autoregressive components. Particular cases of model (2.4) are the bilinear models discussed in Granger and Andersen (1979) and the bivariate bilinear systems analyzed by Subba Rao (1986).

Furthermore, by recursive substitution on (2.3) we can get a nonlinear system, that we will denominate ***nonlinear vector moving average***, NVMA.

$$x_t = F_b(E_0, \varepsilon_1, \dots, \varepsilon_{t-1}, \Theta) + \varepsilon_t \tag{2.5}$$

where $E_t = (E_0, \varepsilon_1, \varepsilon_2, \dots, \varepsilon_t)$, E_0 is a matrix of initial conditions on the innovations ε_t defined as $\varepsilon_t = x_t - E(x_t \mid X_{t-1}, E_{t-1}, \Theta) = x_t - E(x_t \mid X_{t-1}, \Theta) = x_t - E[x_t \mid \mathcal{F}_0^{t-1}, \Theta]$.

We recognize that a precise specification of models (2.3), (2.4) and (2.5) is needed. For that we have to specify the stability and invertibility conditions of nonlinear systems. However, even if we do not address that question in this section, we hope this discussion highlights the variety of possible models that can be obtained within this stochastic framework. For a very interesting discussion of nonlinear time series models see Priestley (1980), (1981).

We mentioned before that the functional form of F is unknown and so in the next section we are going to make some simplifications so that we can get ***precise nonlinear error-correction models***.

3. NONLINEAR ERROR-CORRECTION SYSTEMS

The purpose of this section is to show how to obtain different error-correction models contingent on the time series properties of the x's. For reasons that will become clear later on, we model variables in deviations about the mean, $x_t^* \equiv x_t - \mu_t$ where μ_t represents the mean of the vector x_t, see also Escribano (1987).

Hendry and Richard (1983) assumed that

$$E[x_t^* \mid X_{t-1}^*, \Theta] = \Pi_L^{\#}(B) x_{t-1}^*$$

where $\Pi_L^{\#}(B)$ is an $N \times N$ matrix of lag polynomials of order $p < T$ and B is the lag operator, $B^k x_t^* = x_{t-k}^*$, and so from here we can obtain linear vector autoregressive systems of order p, VAR(p).

A very simple nonlinear extension that allow us to obtain nonlinear error-correction systems is,

$$E[x_t^* \mid X_{t-1}^*, \Theta] = \Pi_L^{\#}(B) x_{t-1}^* + f^*(\alpha' x_{t-1}^*)$$

where α' is an $r \times N$ matrix of constant parameters. For simplicity we assume that the vector of nonlinear functions $f^*(\alpha' x_{t-1}^*)$ is composed of the following elements,

$$f^*(\alpha' x_{t-1}^*) \equiv \begin{bmatrix} f_{11}^*(\alpha_1' x_{t-1}^*) & + & f_{12}^*(\alpha_2' x_{t-1}^*) & + & \dots & + & f_{1r}^*(\alpha_r' x_{t-1}^*) \\ \vdots & & \vdots & & \vdots & & \vdots \\ f_{N1}^*(\alpha_1' x_{t-1}^*) & + & f_{N2}^*(\alpha_2' x_{t-1}^*) & + & \dots & + & f_{Nr}^*(\alpha_r' x_{t-1}^*) \end{bmatrix}$$

By substituting in equation (2.3) we obtain

$$x_t^* = \Pi_L^{\#}(B)x_{t-1}^* + f^*(\alpha' x_{t-1}^*) + \varepsilon_t. \tag{3.0}$$

This system can be rewritten as

$$\Pi_L(B)x_t^* = f^*(\alpha' x_{t-1}^*) + \varepsilon_t$$

where $\Pi_L(B) = I_N - \Pi_L^{\#}(B)B$, and so it is in reduced form since $\Pi(0) = I_N$, the identity matrix.

If we decompose $\Pi_L(B) = \Pi_L(1)B + (1-B)\Pi_L^*(B)$ and $\Pi_L(1) = \Pi_{1L}\alpha'$ where Π_{1L} is an $N \times r$ matrix and α' is the same $r \times N$ matrix that appears in the $N \times 1$ vector of nonlinear functions $f^*(\alpha' x_{t-1}^*)$. By substituting above

$$\Pi_L^*(B)(1-B)x_t^* = -\Pi_{1L}(\alpha' x_{t-1}^*) + f^*(\alpha' x_{t-1}^*) + \varepsilon_t$$

setting

$$f(\alpha' x_{t-1}^*) \equiv -\Pi_{1L}(\alpha' x_{t-1}^*) + f^*(\alpha' x_{t-1}^*)$$

we obtain

$$\Pi_L^*(B)(1-B)x_t^* = f(\alpha' x_{t-1}^*) + \varepsilon_t \tag{3.1}$$

where $\Pi_L^*(B) = I_N - \Pi_L^{\dagger}(B)B$ so that $\Pi_L^*(0) = I_N$. Later on we will derive systems of equations in structural form.

Notice that when we set the elements of $f^*(\alpha' x_{t-1}^*)$ equal to zero, the model (3.1) has the form of the *error-correction* systems, studied by Granger (1983) and Engle and Granger (1985) and so we have derived those models from an *alternative context*, see also Escribano (1987).

In what follows, theorems 3.1 to 3.4 provide conditions contingent on the time series properties of the vector of economic variables x_t^*, so that model (3.1) is a well specified *reduced form* nonlinear error-correction.

A property used in the following theorems is that of a vector of nonlinear functions $f(\alpha' x_{t-1}^*)$ having components which are *asymptotically linear*. By that we mean that all nonlinear components $f_{ij}(\alpha_j' x_{t-1}^*)$ satisfy

$$\left|\frac{f_{ij}(\alpha_j' x_{t-1}^*)}{(\alpha_j' x_{t-1}^*)}\right| \to a_j, \qquad 0 < a_j < \infty, \quad \text{as} \quad (\alpha_j' x_{t-1}^*) \to \infty, \quad \forall ij.$$

For a discussion of the concepts of time trend, time trend-free, strong asymptotic uncorrelation, integration, co-integration and time co-trending, that we will use from now on, see appendix A and Escribano (1987).

THEOREM 3.1. Let the $N \times 1$ vector x_t^* be generated by

$$\Pi_L^*(B)(1-B)x_t^* = f(\alpha' x_{t-1}^*) + \varepsilon_t \tag{3.1}$$

where

1) *$E[(1-B)x_t^* \mid X_{t-1}^*, \Theta] \equiv \Pi_L^{\dagger}(B)(1-B)x_{t-1}^* - f(\alpha' x_{t-1}^*)$ and $\Pi_L^*(B) = I_N - \Pi_L^{\dagger}(B)B$, so that ε_t form a martingale difference sequence and (3.1) is in reduced form;*
2) *$E[\varepsilon_t\varepsilon_t'] = \Omega_t$ is an $N \times N$ matrix with $\inf_t E(\varepsilon_{it}^2) > 0$ and $\sup_t E(\varepsilon_{it}^2) < \infty$ for $i = 1, \ldots, N$ so that the elements of ε_t are time trend-free in the second moment;*
3) *x_j^* and ε_j are equal to zero for all $j \leq 0$;*
4) *$|\Pi_L^*(\omega)| = 0$ has all roots outside the unit circle, so that $\Pi_L^*(1)$ is of full rank;*
5) *the elements of the $N \times 1$ vector $f(\alpha' x_{t-1}^*)$ are unbounded but finite for finite $(\alpha' x_{t-1}^*)$ and are asymptotically linear,*

   then

   a) x_t^* has a time trend in variance.

   b) x_t^* is heterogeneous almost integrated of order one, HA(1), in variance if and only if $z_t^* \equiv \alpha' x_t^*$ is $HI(0)$ in variance.

Proof. (see appendix B)

This result provides conditions to obtain error-correction models from variables x_t^* that in levels have time trends in the variance and are heterogeneous almost integrated processes of order one in variance since we are not requiring that after differencing the power spectrum of the elements of $(1-B)x_t^*$ be finite at zero frequency (strongly asymptotically uncorrelated), nor that they have constant variance.

The message of Theorem 3.1 part b is that if the observations x_t^* are generated by the nonlinear error-correction of Theorem 3.1 and if differencing once is a time trend removing transformation in variance then heterogeneous almost co-integration in variance occurs, with a heterogeneous almost co-integrating matrix indicated by α'.

An alternative reading is that if the observations x_t^* are generated by the nonlinear error-correction of Theorem 3.1 and heterogeneous almost co-integration in variance occurs, so that $z_t^* \equiv \alpha' x_t^*$ is time trend-free in variance, then differencing once is a time trend removing transformation in variance.

Part a states that if x_t^* is generated by the nonlinear error-correction of Theorem 3.1 then x_t^* has a time trend in variance, or that if x_t^* is time trend-free in variance the nonlinear error-correction of Theorem 3.1 cannot be the model generating process. Observe that the components of model (3.1) are time trend-free in variance but not necessarily stationary nor strongly asymptotically uncorrelated. The characterization of these cases is the purpose of the following theorems.

An important property of the concept of stationarity is that it is preserved against nonlinear transformations. The precise result is given in the following lemma.

LEMMA 3.1. Let the function $f_i : \mathbb{R} \to \mathbb{R}$ with $y_{jt} \equiv f_i(x_{jt})$ be into $\mathbb{R}$ and $\mathcal{F}$-measurable. If $\{x_{jt}\}$ is stationary than $\{y_{jt}\}$ is also stationary.

Proof. (see appendix B).

THEOREM 3.2. Let the $N \times 1$ vector x_t^* be generated by

$$\Pi_L^*(B)(1-B)x_t^* = f(\alpha' x_{t-1}^*) + \varepsilon_t \tag{3.1}$$

where (3.1) satisfies conditions 1), 3), 4), 5) of Theorem 3.1 and

2)′ *ε_t is stationary and $E[\varepsilon_t \varepsilon_t'] = \Omega$ is an $N \times N$ matrix of finite and constant elements, with $\inf_t E(\varepsilon_{it}^2) > 0 \ \forall i$,.*

Then

a) x_t^* has time-trend in variance.

b) x_t^* is an almost integrated process of order one, $AI(1)$, in variance if and only if $z_t^* \equiv \alpha' x_t^*$ is $AI(0)$ in variance.

Proof. (see appendix B).

This theorem tell us that when the ε_t is not only a martingale difference sequence but it is also stationary, x_t^* has the nonlinear error-correction of Theorem

3.2 and almost co-integration in variance occurs, so that z_t^* is almost integrated of order zero, then differencing once is a time trend removing transformation in variance and $(1-B)x_t^*$ is stationary. Alternatively, given that ε_t is a stationary martingale difference sequence if x_t^* has the nonlinear error-correction of Theorem 3.2 and $(1-B)x_t^*$ is stationary and time trend-free in variance then almost co-integration in variance occurs, so that $z_t^* \equiv \alpha' x_t^*$ is stationary and time trend-free in variance with an almost co-integration matrix in variance given by α'.

Before continuing with the other nonlinear error-correction models we have to introduce some useful results. Following Rosenblatt (1978), see also White (1984), we introduce the concept of strong mixing (α-mixing). Let

$$\alpha_x(\mathcal{F}_0^t, \mathcal{F}_{t+\tau}^\infty) \equiv \sup_{[A \in \mathcal{F}_0^t, B \in \mathcal{F}_{t+\tau}^\infty]} |P(AB) - P(A)P(B)|$$

$$\text{and} \qquad \alpha_x(\tau) \equiv \sup_t \alpha_x(\mathcal{F}_0^t, \mathcal{F}_{t+\tau}^\infty).$$

The concept of mixing, measures dependence in probability. It has the nice feature that it captures linear as well as nonlinear relationships and so it is very appropriate for this case.

The following results show first how the order of α-mixing is maintained under nonlinear transformations (Lemma 3.2) and second a relationship between the concepts of strong asymptotical uncorrelation and α-mixing (Proposition 3.1), see also Lemma 3.3 of appendix B.

LEMMA 3.2. Let the function $f_i : \mathbb{R} \to \mathbb{R}$ with $y_{jt} \equiv f_i(x_{jt})$ be into $\mathbb{R}$ and $\mathcal{F}$- measurable. If $\{x_{jt}\}$ is $\alpha_{x_j}(\tau) = 0(\tau^{-\lambda})$ for some $\lambda > 0$ then $\alpha_{y_j}(\tau) = 0(\tau^{-\lambda})$.

Proof. (see appendix B).

The following proposition gives conditions for preserving the strong asymptotic uncorrelation property in the presence of nonlinear transformations and establishes the relationship between the concepts of α-mixing and strong asymptotic uncorrelation.

PROPOSITION 3.1. Let the function $f_i : \mathbb{R} \to \mathbb{R}$ with $y_{jt} \equiv f_i(z_{jt}^*)$ be into $\mathbb{R}$ and $\mathcal{F}$- measurable, and let $\{z_{jt}^*\}$ be α-mixing of order

$$\alpha_{z_j^*}(\tau)^{(1-\frac{2}{\beta})}(4+6c) = 0[(1/\tau)^\delta], \quad \textit{for } \delta > 0,$$

a) if $E|f_i(z^*_{jt})|^\beta < c < \infty$ and $E|f_i(z^*_{jt+\tau})|^\beta < c < \infty$ for $2 < \beta < \infty$, then $f_i(z^*_{jt})$ is strongly asymptotically uncorrelated.

b) if $E|z^*_{jt}|^\beta < c < \infty$ and $E|z^*_{j,t+\tau}| < c < \infty$ for $2 < \beta < \infty$, then z^*_{jt} is also strongly asymptotically uncorrelated.

Proof. (see appendix B).

In general if z^*_{jt} is strongly asymptotically uncorrelated, we do not know if $f_i(z^*_{jt})$ is or not strongly asymptotically uncorrelated. Proposition 3.1 tell us that as long as z^*_{jt} is α-mixing, of an order contingent on the moments of $f_i(z^*_{jt})$ then $f_i(z^*_{jt})$ is also strongly asymptotically uncorrelated. In the following theorems we make use of this result.

THEOREM 3.3. Let the $N \times 1$ vector x^*_t be generated by

$$\Pi^*_L(B)(1-B)x^*_t = f(\alpha' x^*_{t-1}) + \varepsilon_t \qquad (3.1)$$

where (3.1) satisfies the conditions of theorem 3.1 and in addition,

6) the elements of the rows of $\Pi^*_L(B)$ decline at an exponential rate

7) the elements $z^*_{jt} \equiv \alpha'_j x^*_t$ and $f_{ij}(z^*_{jt})$ $\forall\ i = 1 \ldots N$ and $\forall\ j = 1 \ldots r$, satisfy the conditions of Proposition 3.1.

Then

a) x^*_t has a time trend in variance.

b) x^*_t is heterogeneous integrated of order one, $HI(1)$, in variance if and only if z^*_t is $HI(0)$ in variance.

Proof. (see appendix B).

Part a is as in Theorem 3.1. Part b tells us that under some conditions on the speed of decline of the elements of $\Pi^*_L(B)$ and the dependence in probability (α-mixing) of the z^*_t terms, we obtain that $(1-B)x^*_t$ is strongly asymptotically uncorrelated. Then, by definition of heterogeneous integration, we conclude that x^*_t is $HI(1)$ in variance and $(1-B)x^*_t$ is $HI(0)$ in variance. This result can be interpreted as saying that if the observations x^*_t have the nonlinear error-corection of Theorem 3.3 and if $(1-B)x^*_t$ is $HI(0)$ in variance, then heterogeneous co-integration in variance occurs with a heterogeneous co-integrating matrix indicated

by α'. The alternative reading is that if x^*_t is generated by the nonlinear error-correction of theorem 3.3 and z^*_t is $HI(0)$ in variance, then x^*_t is $HI(1)$ in variance.

THEOREM 3.4. Let the $N \times 1$ vector x^*_t be generated by

$$\Pi^*_L(B)(1-B)x^*_t = f(\alpha' x^*_{t-1}) + \varepsilon_t \qquad (3.1)$$

where (3.1) satisfies the conditions of Theorem 3.3 but 2) which is substituted by 2)' of Theorem 3.2, then

a) x^*_t has a time trend in variance;

b) x^*_t is integrated of order one, $I(1)$, in variance if and only if z^*_t is $I(0)$ in variance.

Proof. (see appendix B).

The difference between Theorems 3.3 and 3.4 is that now $(1-B)x^*_t$ and z^*_t are not only time trend-free in variance but also stationary. We conclude then that if x^*_t is generated by the nonlinear error-correction model of Theorem 3.4 and if z^*_t is time trend-free in variance then x^*_t is integrated of order one in variance $I(1)$, and $(1-B)x^*_t$ is integrated of order zero in variance, $I(0)$. Or alternatively if x^*_t is generated by the nonlinear error-correction model of Theorem 3.4 and x^*_t is $I(1)$ in variance then co-integration in variance occurs with co-integrating matrix indicated by α'. Notice that since $f(\alpha' x^*_{t-1}) = \Pi_{1L}(\alpha' x^*_{t-1}) + f^*(\alpha' x^*_{t-1})$ if we set $f^*(\alpha' x^*_{t-1}) = 0$ we obtain a linear error-correction model with integrated variables, in variance, as in Engle and Granger (1985).

Theorem 3.1 gives us necessary and sufficient conditions for z^*_t to be time trend-free in variance when the x_t's are generated by a nonlinear error-correction. Now we want to study the representation of z^*_t.

As we have seen at the beginning of this section

$$\Pi^*_L(B)(1-B)x^*_t = -\Pi_{1L}(\alpha' x^*_{t-1}) + f^*(\alpha' x^*_{t-1}) + \varepsilon_t$$

and since $\Pi^*_L(B) = I_N - \Pi^+_L(B)B$ we can get that

$$x^*_t = x^*_{t-1} - \Pi_{1L}(\alpha' x^*_{t-1}) + \Pi^*_L(B)(1-B)x^*_{t-1} + f^*(\alpha' x^*_{t-1}) + \varepsilon_t$$

premultiplying by α' and rearranging,

$$z^*_t = (I_N - \alpha'\Pi_{1L})z^*_{t-1} + \alpha' f^*(z^*_{t-1}) + \alpha'\Pi^*_L(B)(1-B)x^*_{t-1} + \alpha'\varepsilon_t$$

making $g(z^*_{t-1}) \equiv (I_N - \alpha'\Pi_{1L})z^*_{t-1} + \alpha' f^*(z^*_{t-1})$

$$z^*_t = g(z^*_{t-1}) + \alpha'\Pi^*_L(B)(1-B)x^*_{t-1} + \alpha'\varepsilon_t. \qquad (3.2)$$

Theorem 3.1 tell us that if $(1-B)x^*_t$ is time trend-free in variance then z^*_t is time trend-free in variance. For that to be the case it is sufficient that g be ***globally stable*** (stable everywhere). By that we mean that $|g(z^*_t)| \leq A|z^*_t|$ where A is a $k \times k$ matrix with all eigenvalues being less than one in absolute value. When this is the case z^*_t must be time trend-free in variance since $\alpha'\Pi^*_L(B)(1-B)x^*_{t-1}$ is time trend-free in variance and the vector of nonlinear difference equations is bounded by a stable linear system of difference equations. For more on stability, see Braun (1983) and Caughey (1979). Notice that global stability is not necessary since for time trend-free, for that we only require the function to be stable at the tails. It is for the economic relevance and the interest on the stable points (equilibria) of the system that we want to impose the global stability of the nonlinear adjustment.

It might be worth mentioning that the representation for z^*_t has zero mean because the mean of $\alpha'\Pi^*_L(B)(1-B)x^*_t$ is equal to zero. In this case, even a first order error-correction model is not constantly missing the target, see Kloek (1984), and Pagan (1986) for more in this issue.

It is our purpose now to obtain error-correction ***systems in structural form***, see for example Wallis (1977).

If we assume that $A_0\Pi_{Lj} + A_j = 0$ for all $j = 1, \ldots, p$ and that A_0 is nonsingular, see for example Wallis (1977), we can premultiply equation (2.5) by $A_0 \neq I_N$

$$A_0 x^*_t = A_0 E(x^*_t \mid X^*_{t-1}, \Theta) + A_0\varepsilon_t \qquad (3.3)$$

and substituting for $E(x^*_t \mid X^*_{t-1}, \Theta)$ we get

$$A_0 x^*_t = A_0\Pi^{\#}_L(B)x^*_{t-1} + A_0 f^*(\alpha' x^*_{t-1}) + A_0\varepsilon_t \qquad (3.4)$$

or

$$A(B)x^*_t = f^{\#}(\alpha' x^*_{t-1}) + A_0\varepsilon_t \qquad (3.5)$$

which is a ***structural form*** of a simple ***nonlinear vector autoregression***, NVAR. To obtain the error-correction representation, we can decompose $A(B) = A(1)B + (1-B)A^*(B)$ and $A(1) = \gamma_1\alpha'$ where γ_1 is an $N \times r$ matrix and α' is an $r \times N$. By substituting on (3.5) we obtain a ***nonlinear error-correction*** in ***structural form***.

$$A^*(B)(1-B)x^*_t = f_s(\alpha' x^*_{t-1}) + A_0\varepsilon_t \qquad (3.6)$$

where (3.6) satisfies the same properties of theorems 3.1 to 3.4 with the important differences that now $A^*(0) \neq I_N$, the covariance matrix $E[A_0\varepsilon_t\varepsilon_t' A_0'] = A_0\Omega_t A_0'$ instead of Ω_t, and that $f_s(\alpha' x^*_{t-1}) = -\gamma_1(\alpha' x^*_{t-1}) + f^{\#}(\alpha' x^*_{t-1})$.

For a discussion of the implications of error-correction systems when we are interested in variables that are not in deviations about the mean and we allow the mean to grow through time, see Escribano (1987).

4. SINGLE EQUATION ERROR-CORRECTION MODELS

Most of the empirical literature on error-correction models is based on a single equation models in structural form. See for example Davidson, Hendry, Srbao and Yeo (1978) and Hendry and Ericsson (1987). For this reason we concentrate now on the conditions that allow us to study ***single equation structural form models.***

Consider the following partition of the vector $x_t' = (y_t, q_t')$ where y_t is a scalar time series and q_t is an $(N-1) \times 1$ vector. Factorizing the conditional density introduced in section 2

$$\mathbb{P}(x_t \mid X_{t-1}, \Theta) = \mathbb{P}(y_t \mid q_t, X_{t-1}, \phi_1)\mathbb{P}(q_t \mid X_{t-1}, \phi_2).$$

Let the parameters of interest, ψ, be a function of ϕ_1 only. If ϕ_1 and ϕ_2 are variation-free, then q_t is ***weakly exogenous*** for ψ and so we can make inferences on ϕ_1 based on the conditional model, see Engle, Hendry and Richard (1983). However, when (heterogeneous) co-integration in variance occurs, there are cross equation restrictions through the error-correction terms that enter in the other equations. See the elements of the vector $f^*(\alpha' x_{t-1})$. When this is the case, the partition $\Theta = (\phi_1, \phi_2)$ do not operate a sequential cut and so ϕ_1 and ϕ_2 are not variation-free. Then necessary conditions for q_t to be weakly exogenous are that the error-correction terms do not enter in the other equations, that different error-correction terms enter in each equation or that the common error terms be equal to zero.

A question that deserves further investigation is how serious are the errors committed in practive when q_t is not weakly exogenous just because of the presence of the cross-equation restrictions of some, or all, of the error-correction components.

For the reasons mentioned in the previous section, we are interested in the conditional mean and we will assume that, in terms of deviations about the mean $y_t^* = y_t - \mu_{yt}$,

$$E[y_t^* \mid q_t^*, X_{t-1}^*, \phi_1^*] \equiv A_{0yq} q_t^* + A_y^+(B) x_{t-1}^* + f_y^*(\alpha_y' x_{t-1}^*)$$

where A_{0yq} is an $1 \times (N-1)$ vector, $A_y^+(B)$ is $1 \times N$, and $f_y^*(\alpha_y' x_{t-1}^*)$ is a scalar.

$$f_y^*(\alpha' x_{t-1}^*) = f_{y1}^*(\alpha_{y1}' x_{t-1}^*) + f_{y2}^*(\alpha_{y2}' x_{t-1}^*) + \ldots + f_{yr}^*(\alpha_{yr}' x_{t-1}^*).$$

If we assume that $\varepsilon_{yt} = y_t^* - E[y_t^* \mid q_t^*, X_{t-1}^*, \phi_1^*]$ so that $\{\varepsilon_{yt}\}$ form a martingale difference sequence and that $0 < E(\varepsilon_{yt}^2) = \sigma_{\varepsilon,t}^2 < \infty$ for all t, we obtain a *single equation nonlinear autoregressive distributed lag model*, NARDL.

$$y_t^* = A_{0yq} q_t^* + A_y^+(B) x_{t-1}^* + f^*(\alpha_y' x_{t-1}^*) + \varepsilon_{yt}. \tag{4.1}$$

Expression (4.1) can be simplified to

$$A_y(B) x_t^* = f^*(\alpha_y' x_{t-1}^*) + \varepsilon_{yt} \tag{4.2}$$

where the $1 \times N$ vector $A_y(B) = [(1, -A_{0yq}) + A_y^+(B)B]$. In order to obtain an error-correction representation we can decompose $A_y(B) = A_y(1)B + A_y^*(B)(1-B)$ where $A_y(1) = \gamma_y \alpha_y'$ with γ_y a scalar and α_y' a $1 \times N$ vector. Substituting on (4.2)

$$A_y^*(B)(1-B)x_t^* = -\gamma_y(\alpha_y' x_{t-1}^*) + f^*(\alpha_y' x_{t-1}^*) + \varepsilon_{yt}.$$

Let $f_y(\alpha_y' x_{t-1}^*) \equiv -\gamma_y(\alpha_y' x_{t-1}^*) + f^*(\alpha_y' x_{t-1}^*)$ and we obtain a *single equation structural form nonlinear error-correction model*

$$A_y^*(B)(1-B)x_t^* = f_y(\alpha_y' x_{t-1}^*) + \varepsilon_{yt} \tag{4.3}$$

where $A_y^*(0)$ has the first element equal to 1 which corresponds to the element y_t^* of the vector x_t^*.

Equation (4.3) can be written in a more clear way by setting $(1-B) = \Delta$,

$$\Delta y_t^* = A_{0yq} \Delta q_t^* + \tilde{A}_y(B) \Delta x_{t-1}^* + f_y(\alpha_y' x_{t-1}^*) + \varepsilon_{yt}. \tag{4.4}$$

Notice that if there are r (heterogeneous) co-integrating vectors in variance then, even in a single equation error-correction model, we can have r error-correction components from the $r < N$ possible long-run relationships (equilibria). For (4.3) to be a well-defined error-correction model, it should satisfy the corresponding conditions of Theorems 3.1 to 3.4.

In order to clarify how to decompose the matrix $A_y(B)$ into the sum of the components $A_y(1)B$ and $A_y^*(B)(1-B)$, we are going to discuss a very simple *example*.

Let $x_t^* = (y_t^*, q_t^*)$ be a 1×2 vector so that y_t^* and q_t^* are 1×1.

$$y_t^* = b_0 q_t^* + b_1 q_{t-1}^* + b_2 q_{t-2}^* + a_1 y_{t-1}^* + a_2 y_{t-2}^* + f_y^*(\alpha_y' x_{t-1}^*) + \varepsilon_{yt} \tag{4.5}$$

where now $\alpha_y' x_{t-1}^* = y_{t-1}^* - \alpha_q q_{t-1}^* = (1, -\alpha_q) x_{t-1}^*$. Equation (4.5) can be written as in (4.2)

$$[(1,-b_0) + (-a_1,-b_1)B + (-a_2,-b_2)B^2] \begin{pmatrix} y_t^* \\ q_t^* \end{pmatrix} = f_y^*(y_y^* - \alpha_q q_{t-1}^*) + \varepsilon_{yt}$$

where $A_y(B) = [(1,-b_0) + (-a_1,-b_1)B + (-a_2,-b_2)]B^2]$.

Remember that to obtain a single equation structural form error-correction model, we can decompose $A_y(B) = A_y(1)B + A_y^*(B)(1-B)$. In particular,

$$\begin{aligned} A_y(B) &= [(1,-b_0) + (-a_1,-b_1) + (-a_2,-b_2)]B + [(1,-b_0) + (a_2,b_2)B](1-B) \\ &= (1 - a_1 - a_2, -b_0 - b_1 - b_2)B + [(1,-b_0) + (a_2,b_2)B](1-B) \end{aligned}$$

by substituting above, setting $(1-B) = \Delta$ and rearranging

$$\begin{aligned} \Delta y_t^* = {} & b_0 \Delta q_t^* - a_2 \Delta y_{t-1}^* - b_2 \Delta q_{t-1}^* + (a_1 + a_2 - 1) y_{t-1}^* \\ & + (b_0 + b_1 + b_2) q_{t-1}^* + f_y^*(y_{t-1}^* - \alpha_q q_{t-1}^* + \varepsilon_{yt} \end{aligned}$$

If (heterogeneous) co-integration in variance occurs,

$$b_0 + b_1 + b_2 = -\alpha_q(a_1 + a_2 - 1)$$

so that $A_y(1) = \gamma_y \alpha_y' = (a_1 + a_2 - 1)(1, -\alpha_q)$, we can rewrite the model as in (4.4)

$$\begin{aligned} \Delta y_t^* = {} & b_0 \Delta q_t^* - b_2 \Delta q_{t-1}^* - a_2 \Delta y_{t-1}^* \\ & + f_y(y_{t-1}^* - \alpha_q q_{t-1}^*) + \varepsilon_{yt} \end{aligned} \tag{4.6}$$

where $f_y(y_{t-1}^* - \alpha_q q_{t-1}^*) \equiv -(1 - a_1 - a_2)(y_{t-1}^* - \alpha_q q_{t-1}^*) + f_y^*(y_{t-1}^* - \alpha_q q_{t-1}^*)$. Let $z_t^* \equiv (y_t^* - \alpha_q q_t^*)$ represent the equilibrium error. In terms of the "target" notation, we can say that $y_t^{*T} \equiv \alpha_q q_t^*$ is the target and so $z_t^* = y_t^* - y_t^{*T}$ indicates the deviation from the target. Equation (4.6) is an error-correction model with a linear long-run relationship, represented by $y_t^* = \alpha_q q_t^*$, and a nonlinear adjustment mechanism that compensates for short-run departures from the target.

The time series representation for z_t^* can be derived by substituting $y_t^* = \alpha_q q_t^* + z_t^*$ above

$$\Delta z_t^* = (b_0 - \alpha_q)\Delta q_t^* - b_2 \Delta q_{t-1}^* - a_2 \Delta y_{t-1}^* + f_y(z_{t-1}^*) + \varepsilon_{yt}$$

by adding z_{t-1}^* to both sides and by definition of $f_y(z_{t-1}^*)$ we get

$$\begin{aligned} z_t^* = {} & (b_0 - \alpha_q)\Delta q_t^* - b_2 \Delta q_{t-1}^* - a_2 \Delta y_{t-1}^* \\ & + [1 - (1 - a_1 - a_2)] z_{t-1} + f_y^*(z_{t-1}^*) + \varepsilon_{yt} \end{aligned}$$

and by calling $g_y(z_{t-1}^*) = [1 - (1 - a_1 - a_2)] z_{t-1}^* + f_y^*(z_{t-1}^*)$ we obtain the single equation version of the system (3.2)

$$z_t^* = g_y(z_{t-1}^*) + (b_0 - \alpha_q)\Delta q_t^* - b_2 \Delta q_{t-1}^* - a_2 \Delta y_{t-1}^* + \varepsilon_{yt}. \qquad (4.7)$$

The global stability condition now implies that $|g_y(z_{t-1}^*)| \leq a_y |z_{t-1}^*|$ where $0 < a_y < 1$, which is the well-known stability condition that the function $g_y(z_{t-1}^*)$ should be under the 45° line, in absolute value. See for example Levy and Lessman (1959) and Granger and Weiss (1982). It can be easily seen that if $f_y(z_{t-1}^*)$ is globally stable and has a linear component with a negative sign so that $-(1 - a_1 - a_2) < 0$ then $g_y(z_{t-1}^*)$ has to be also globally stable.

Notice also that (4.7) allows z_t to have a non zero mean, and that even Δq_t and Δy_t do not need to have zero mean (target with a time trend in mean and in variance). Going back to the case where x_t is an $N \times 1$ vector and $x_t = (y_t, q_t)'$ where y_t is 1×1 and q_t is $1 \times (N-1)$, we want to discuss under what conditions we can analyze ***single equations reduced form models*** without any loss of relevant information. We have seen that

$$I\!P(x_t \mid X_{t-1}, \Theta) = I\!P(y_t \mid q_t, X_{t-1}, \phi_1) I\!P(q_t \mid X_{t-1}, \phi_2).$$

then the condition required is that

$$I\!P(y_t \mid q_t, X_{t-1}, \phi_1) = I\!P(y_t \mid X_{t-1}, \phi_1).$$

When this is the case y_t and the elements of q_t are independent conditionallly on the information set $\mathcal{F}_0^{t-1}$, and so we can analyze the $y_t \mid X_{t-1}, \phi_1$ without any

loss of information on the parameters ϕ_1, provided that the sequential cut occurs between ϕ_1 and ϕ_2 so that they are variation-free.

Since we are interested in the conditional mean we can then assume

$$E[y_t^* \mid X_{t-1}^*, \phi_1] = A_y^+(B) x_{t-1}^* + f_y^*(\alpha_y' x_{t-1}^*)$$

and so

$$y_t^* = A_y^+(B) x_{t-1}^* + f_y^*(\alpha_y' x_{t-1}^*) + \varepsilon_{yt}.$$

However, even if by using this approach it is very unlikely that we can justify a single equation ***reduced form*** model it is not that restrictive when we are at the estimation level. To make this point clear, consider the models (3.1) and (3.6) which are in reduced and structural form. The relationship between the errors of both models is clear. The errors from (3.1) are ε_t and from (3.6) are $A_0 \varepsilon_t$ and the corresponding variance-covariance matrices are Ω_t and $A_0 \Omega_t A_0'$. It is clear that since (3.6) uses more information $(A_0 \Omega_t A_0' - \Omega_t)$ should be a positive definite matrix, so that the variance of the disturbance terms in each of the equations of (3.6) should be smaller those of the equations of (3.1). However, if we accept that we can be interested in reduced form models, it is possible that the estimation procedure of the system of equations, in reduced form , coincide with the estimation procedure equation by equation. See Engle and Yoo (1986) for a discussion of the estimation procedures of reduced form linear error-correction systems.

CONCLUSIONS

We have shown how to obtain nonlinear time series models from an optimization point of view when the objective function is the minimum mean square error of the forecast. In addition, we have shown how this approach provides a unifying framework to obtain nonlinear error-correction models and linear ones as a particular case.

Furthermore, we have seen that provided that we have a nonlinear error-correction system like equation (3.1), if the nonlinear adjustment is unbounded, asymptotically linear and under some regularity conditions we conclude that the levels of the variables, x_t^*, are time trending in variance and that there exist long-run equilibrium relationships if and only if the differences of the variables, Δx_t, are time trend-free in variance. If in addition, the nonlinear adjustment is α-mixing of certain order then Δx_t^* is also strongly asymptotically uncorrelated and so x_t^* is an integrated proces of order one, in variance. The integrated process in variance can be heterogeneous or stationary depending on the assumptions made about the behaviour of the innovations.

We have also discussed and derived error-correction systems in structural and reduced form, single equation models in structural and reduced form and provided an example to clarify the results and implications.

The models obtained represent an important extension to the existing literature in the sense that they allow to model economic variables with a broad variety of stochastic and deterministic properties. That is, we can incorporate different sources of growing components on the levels of variables, different concepts of long-run equilibrium relationships and we can incorporate some heterogeneity in the observations that was not allowed in the previous error-correction models.

APPENDIX A

CONCEPTS AND DEFINITIONS

This appendix is based on Escribano (1987).

Let $R[E(x_{jt}^i)] = [a_i, b_i]$ be the range of the deterministic sequence $\{E(x_{jt}^i)\}_{t=0}^{\infty}$, where

$$a_i = \text{Min}[\sup_k \inf\{E(x_{jt}^i)\}_{t=0}^{k}, \sup_k \inf\{E(x_{jt}^i)\}_{t=k}^{\infty}]$$

and

$$b_i = \text{Max}[\inf_k \sup\{E(x_{jt}^i)\}_{t=0}^{k}, \inf_k \sup\{E(x_{jt}^i)\}_{t=k}^{\infty}].$$

Let the indicator function be such that $I[E(x_{jt}^i) \in S] = 0$ if $E(x_{jt}^i) \notin S$ and $I[E(x_{jt}^i) \in S] = 1$ if $E(x_{jt}^i) \in S$. From now on we assume that all relevant i-th order moments exist.

DEFINITION D1. The stochastic process $\{x_{jt}\}$ has a **time trend of order $h_j(i,t)$ in the i-th moment** *if :*

i) $E(x_{jt}^i)$ *is finite for finite* t,

ii) $E(x_{jt}^i) = h_j(i,t)$ *where* $h_j(i,t) : \mathbb{R} \to \mathbb{R}$, *is time dependent, and if*

iii) for **every** *subinterval* $A_i = (c_i, d_i)$, *where* $a_i < c_i < d_i < b_i$, *we have that* $Lim_{t\to\infty}(1/t)\Sigma_{k=0}^{t} I[E(x_{jk}^i) \in A_i] = 0$.

With this definition we can identify the source of the time trend with those moments that are growing. For example, we can have a time trend in mean, variance, etc.

PROPOSITION A. Let the stochastic process $\{x_{jt}\}$ satisfy conditions i) - ii) of Definition D1. If $E(x_{jt}^i) = h_j(i,t)$ is also unbounded as $t \to \infty$ then $\{x_{jt}\}$ has **a time trend in the i-th moment.**

The reason for the above implication is that $E(x_{jt}^i) \to \infty$ as $t \to \infty$ is sufficient for condition iii) of D1. However, observe that is not necessary since we are allowing time trends in the i-th moment to be bounded. In the sequel we make extensive use of this proposition since it is often the case that time trends in the moments are unbounded.

DEFINITION D2. The stochastic process $\{x_{jt}\}$ *is* **time trend-free in the i-th moment** *if* $\{x_{jt}\}$ *has not a time trend in the i-th moment.*

A more general, but less operative, definition of time trend is given in Escribano (1987).

Let the mean of x_{jt} be $E(x_{jt}) \equiv \mu_{jt}$, the variance $E[(x_{jt}-\mu_{jt})^2] \equiv \gamma_j(t,0)$ the autocovariance $E[(x_{jt}-\mu_{jt})(x_{j,t+\tau}-\mu_{j,t+\tau})] \equiv \gamma_j(t,\tau)$ for $\gamma = \pm 1, \pm 2, \ldots$ and the autocorrelation $\rho_{j,t}(\tau) \equiv \{\gamma_j(t,\tau) \,/\, [\gamma_j(t,0)^{\frac{1}{2}}\gamma_j(t+\tau,0)^{\frac{1}{2}}]\}$. In general, the autocovariance and autocorrelations are functions of time if the mean and variance are functions of time.

DEFINITION D3. The stochastic process $\{x_{jt}\}$ *is* **strongly asymptotically uncorrelated** *if :*

i) $0 < \gamma_j(t,0) < \infty \quad \forall\, t$, *and*

ii) there exist a function ρ_τ *for* $\tau \neq 0$ *such that* $\Sigma_{\tau=0}^{\infty}|\rho_\tau| < \infty$ *and* $|\rho_{j,t}(\tau)| \leq |\rho_\tau|$ *for all* $\tau \neq 0$.

Observe, that if $\{x_{it}\}$ is strongly asymptotically uncorrelated then the normalized power spectrum of $\{x_{it}\}$ at zero frequency is finite.

DEFINITION D4. The stochastic process $\{x_{jt}\}$ *is* **integrated of order** $d, I(d), d \in I\!N$, **in variance,** *if :*

i) $\{x_{jt}-\mu_{jt}\}$ *has a time trend in variance, and*

ii) if after differencing **at least** *d times,* $(1-B)^d(x_{jt}-\mu_{jt})$ *is :*

  a) time trend-free in variance,

  b) strongly asymptotically uncorrelated, and

  c) stationary of second order.

Notice that Definition D4 allows $(1-B)^d(x_{jt}-\mu_{jt})$ to have a Wold's representation and hence can be approximated by a finite ARMA model.

When relaxing the stationarity requirement we obtain the following concept of integration.

DEFINITION D5. The stochastic process $\{x_{jt}\}$ *is* **heterogeneous integrated of order** $d, HI(d), d \in I\!N$ **in variance,** *if :*

i) $\{x_{jt}-\mu_{jt}\}$, *has a time trend in variance and,*

ii) if after differencing **at least** *d times,* $(1-B)^d(x_{jt}-\mu_{jt})$ *is :*

  a) time trend-free in variance, and

  b) strongly asymptotically uncorrelated.

This concept allows the structure of $(1-B)^d(x_{jt}-\mu_{jt})$ to be time varying or to have some heteroskedasticity, ARCH see Engle (1982), etc. Examples of this concept are the integrated processes of order one analyzed initially by Phillips (1987).

When relaxing the strong asymptotic uncorrelation condition we get,

DEFINITION D6. The stochastic process $\{x_{jt}\}$ *is* **heterogeneous almost integrated of order** d, $HAI(d)$, $d \in I\!N$ **in variance,** *if :*

i) $\{x_{jt}-\mu_{jt}\}$ *has a time trend in variance, and*

ii) if after differencing **at least** *d times,* $(1-B)^d(x_{jt}-\mu_{jt})$ *is time trend-free in variance.*

Similarly, if we add to Definition D6 the requirement that $(1-B)^d(x_{jt}-\mu_{jt})$ be *stationary of second order* we obtain an **almost integrated process of order** d **in variance,** $AI(d)$.

As is usually the case, the definitions introduced for scalar random variables can be extended to vectors by applying the definitions to each of the component of the vector. We say that the $N \times 1$ *vector* stochastic process $\{x_t\}$ has a time trend in the second moment (time trend-free in the second moment) if *all* the components of the vector $\{x_t\}$ have a time trend in the second moment (time trend-free in the second moment). Similar extension can be applied to the concept of strong asymptotic uncorrelation and so to heterogeneous integration of order d in the second moment, integration of order d in the second moment, etc.

DEFINITION D7. The components, x_{jt} of the vector stochastic process $\{x_t\} = \{x_{jt}, j = 1, \ldots, N\}$ are **co-integrated in variance** *if :*

a) all components are integrated of order d in variance, $I(d)$, $d > 0$ and,

b) there exist a linear combination of them, $z_t^* = \alpha'(x_t - \mu_t)$ which is, $I(d-b)$, $b > 0$, in variance. α' is an $r \times N$ matrix with $r < N$.

Observe, that this is the same as Granger's definition of co-integration. They differ only in that we are using different notions of integrated stochastic processes.

DEFINITION D8. The components, x_{jt} of the vector stochastic process $\{x_t\} = \{x_{jt}, j = 1, \ldots, N\}$ are **heterogeneous co-integrated in variance** *if :*

a) all components are heterogeneous integrated of order d in variance, $HI(d)$, $d > 0$ and,

b) there exist a linear combination of them, $z_t^* = \alpha'(x_t - \mu_t)$ which is $HI(d-b)$, $b > 0$, in variance. α' is an $r \times N$ matrix with $r < N$.

Notice that even when $d = b$ we are not forcing z_t^* to be stationary.

It is also possible to have variables that are not strongly asymptotically uncorrelated.

DEFINITION D9. The components, x_{jt} of the vector stochastic process $\{x_t\} = \{x_{jt}, j = 1, \ldots, N\}$ are **heterogeneous almost co-integrated in variance** *if :*

a) all components are heterogeneous almost integrated of order d in variance, $HAI(d)$, $d \geq 1$ and,

b) there exist a linear combination of them $z_t^* = \alpha'(x_t - \mu_t)$ which is $HAI(d-b)$, $b > 0$, in variance. α' is an $r \times N$ matrix with $r < N$.

As with Definition D8 we can define an **almost co-integrated** relationship by adding to conditions a and b of Definition 9 the requirement of stationarity of second order, so that $HAI(d)$ and $HAI(d-b)$ are substituted by $AI(d)$ and $AI(d-b)$ correspondingly.

The main advantage of those long run relationships is that they are not based on concepts that are *model dependent* and that it is possible to extend them to the *nonlinear* case.

APPENDIX B

Proof of Theorem 3.1

a) The inverse of $\Pi_L^*(B)$ exists by condition 4. Premultiplying model (3.1) by $(1-B)^{-1}\Pi_L^*(B)^{-1}$ and using the initial conditions 3, we obtain

$$x_t^* = -\Pi_L^{*-1}(B)(1-B)^{-1}f(\alpha' x_{t-1}^*) + \Pi_L^{*-1}(B)(1-B)^{-1}\varepsilon_t.$$

From condition 3, $(1-B)^{-1}\varepsilon_t = \Sigma_{j=0}^t \varepsilon_{tj}$, and by 2 $\inf_t E(\varepsilon_{it}^2) > 0$. Then we conclude that the components of x_t^* have time trends in variance.

b) *Necessary Condition* :

Assume that $(1-B)x_t^*$ is time trend-free in variance, generated by (3.1) and allows $z_t^* \equiv \alpha_t' x_t^*$ to be time trending in variance. Then $f(z_t^*)$ is also time trending in variance since the components of $f(z_t^*)$ are asymptotically linear, by condition 5. But by condition 2, ε_t is time trend-free in variance and $\Pi_L^*(B)$ has no unit roots, by condition 4, and so $(1-B)\varepsilon_t^*$ since $f(z_t^*)$ has a time trend in variance then $(1-B)x_t^*$ has to be also time trending in variance, if generated by model (3.1), which contradicts our assumption.

Sufficient Condition

The inverse of $\Pi_L^*(B)$ exists and has all roots outside the unit circle by condition 4. Premultiplying model (3.1) by $\Pi_L^*(B)^{-1}$ we obtain

$$(1-B)x_t^* = -\Pi_L^*(B)^{-1}f(\alpha' x_{t-1}^*) + \Pi_L^*(B)^{-1}\varepsilon_t.$$

ε_t is time trend-free in variance by condition 2, and by 4 the last term represents an invertible moving average and so it is time trend-free in variance. If $z_t^* \equiv \alpha' x_t^*$ is time trend-free in variance, so will then be $f(z_{t-1}^*)$ since, by 5, it is finite for finite z_{t-1}^* and asymptotically linear and so it has asymptotically the same property of z_{t-1}^* and the result follows.

■

Proof of Lemma 3.1

See Stout (1974).

Proof of Theorem 3.2

a) Similar to Theorem 3.1.

b) *Necessary Condition*

The time trend-free part is equivalent to Theorem 3.1. If $z_t^* \equiv \alpha_t' x_t^*$ is not stationary so will not be $f(z_{t-1}^*)$ and by equation (3.1) $(1-B)x_t^*$ will not be stationary either.

Sufficient Condition

The time trend-free part is equivalent to Theorem 3.1. By condition 4 we can write equation (3.1) as

$$(1-B)x_t^* = -\Pi_L^*(B)^{-1} f(\alpha' x_{t-1}^*) + \Pi_L^*(B)^{-1} \varepsilon_t.$$

By 2′, ε_t is stationary and by 4 so is $\Pi_L^*(B)^{-1}\varepsilon_t$. If z_t^* is stationaty so is $f(\alpha' x_{t-1}^*)$, by Lemma 3.1, and so $\Pi_L^*(B)^{-1} f(\alpha' x_{t-1}^*)$. Then from the above system of equations we conclude that $(1-B)x_t^*$ is stationary.

■

Proof of Lemma 3.2

Immediate from White and Domowitz (1984), Lemma 2.1 by allowing τ to be equal to zero.

For the proof of Proposition 3.1 we need to introduce the following important instrumental results.

Lemma 3.3

Let z_{jt}^* be measurable with respect to $\mathcal{F}_{j,0}^t$ and $z_{t+\tau}^*$ with respect to $\mathcal{F}_{j,t+\tau}^{\infty}$. If $E|z_{jt}^*|^{2+\delta} < d_{j,t}$ and $E|z_{j,t+\tau}^*|^{2+\delta} < d_{j,t+\tau}$ are finite for every finite t and τ then for $b = (2+\delta)^{-1}$, $0 < \delta < \infty$,

$$|E(z_{j,t}^* z_{j,t+\tau}^*) - E(z_{j,t}^*)E(z_{j,t+\tau}^*)| \leq \alpha_{z_j^*}(\tau)^{1-2b}[4 + 3(d_{j,t}^b d_{j,t+\tau}^{1-b} + d_{j,t}^{1-b} d_{j,t+\tau}^b)].$$

Proof: (See Ibraginov and Linnik (1971), page 307).

Proof of Proposition 3.1

a) Let

$$E[f_i(z_{jt}^*) f_i(z_{j,t+\tau}^*)] - Ef_i(z_{jt}^*)Ef_i(z_{j,t+\tau}^*) \equiv \gamma_{f_i}(t,\tau)$$

then by applying Lemma 3.3 to $f_i(z_{jt}^*)$ and $f_i(z_{j,t+\tau}^*)$

$$|\gamma_{f_i}(t,\tau)| \leq \alpha_{f_i}(\tau)^{(1-2/\beta)}(4+6c)$$

but from Lemma 3.2 $\alpha_{f_i}(\tau) = \alpha_{z_j^*}(\tau)$ and so

$$|\gamma_{f_i}(t,\tau)| \leq \alpha_{z_j^*}(\tau)^{(1-2\beta)}(4+6c).$$

From the conditions of this lemma we know that

$$\alpha_{z_j^*}(\tau)(4+6c) = 0\left[\left(\frac{1}{\tau}\right)^{\delta}\right] \qquad \delta > 0$$

then if we set

$$|\gamma_\tau| \equiv \alpha_{z_j^*}(\tau)^{(1-2/\beta)}(4+6c),$$

we get that $\Sigma_{\text{all}\tau}|\gamma_\tau| < \infty$ and since $|\gamma_{f_i}(t,\tau)| \leq |\gamma_\tau|$, we conclude that $\Sigma_{\text{all}\tau} \mid \gamma_{f_i}(t,\tau) \mid < \infty$. The variance of $f_i(z_{jt}^*)$ is finite $\forall t$ and so we have that $|\rho_{f_i}(t,\tau)| \leq \mid \rho_\tau \mid$ where $\Sigma_{\tau=0}^{\infty}|\rho_\tau| < \infty$.

■

b) Similar to part a.

Proof of Theorem 3.3

a) Similar to Theorem 3.1.

b) Given part a, for x_t^* to be $HI(1)$ in variance, we have to show first that $(1-B)x_t^*$ is time trend-free in variance iff z_t^* is time trend-free in variance and second, that $(1-B)x_t^*$ is strongly asymptotically uncorrelated, given the conditions of the theorem, since z_t^* is already strongly asymptotically uncorrelated by Proposition 3.1.

Necessary condition for time trend-free in variance of $(1-B)x_t^*$.

Similar to Theorem 3.1. However, observe that condition 7 still allows for the presence of deterministic time trends in variance in z_t^*, in spite of the requirement that z_t^* be α-mixing of certain order..

Sufficient condition for strong asymptotic uncorrelation of $(1-B)x_t^*$.

Remember that

$$(1-B)x_t^* = -\Pi_L^*(B)^{-1}f(z_{t-1}^*) + \Pi_L^*(B)^{-1}\varepsilon_t.$$

By assumption z_t^* and $f(z_{t-1}^*)$ are strongly asymptotically uncorrelated, see Proposition 3.1. All we have to show then is that $\Pi_L^*(B)^{-1}\varepsilon_t$ is strongly asymptotically uncorrelated. This is obtained immediately from condition 6.

■

Proof of Theorem 3.4

Immediate from Theorems 3.2 and 3.3.

REFERENCES

Bollerslev, T.P. (1986), "Generalized Autoregressive Conditional Heteroskedasticity with Applications in Finance," Ph. D. Thesis, Economics Department, University of California San Diego.

Braun, M. (1983), *Differential Equations and their Applications.* 3rd Edition, Applied Mathematical Sciences 15, Springer-Verlag.

Caughey, T.K. (1979), "Nonlinear Ordinary Difference Equations", National Aeronautics and Space Administration, Jet Propulsion Laboratory 79-50.

Davidson, J.E., Hendry, D.F., Srbao, F. and S. Yeo (1978), "Econometric Modelling of the Aggregate Time-Series Relationship Between Consumer's Expenditure and Income in the United Kingdon," *Economic Journal*, 88, 661-692.

Doob, J.L. (1954), *Stochastic Processes.* New York : Wiley.

Engle, R.F. (1982), "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of the United Kingdom Inflation," *Econometrica* 50, 4, 987-1007.

Engle, R.F. and C.W.J. Granger (1985), "Co-integration and Error-Correction : Representation, Estimation and Testing," Discussion Paper, University of California, San Diego.

Engle, R.F. and B.S. Yoo (1985), "Forecasting and Testing in Co-integrated Systems," Department of Economics, Discussion Paper 86-7.

Escribano, A. (1986), "Identification and Modelling of Economic Relationship in a Growing Economy," Ph.D. Dissertation, Department of Economics, University of California San Diego.

Escribano, A. (1987), "Co-integration, Time Co-Trends and Error-Correction Systems : An Alternative Approach ," CORE Discussion Paper No 8715.

Granger, C.W.J. and P. Newbold (1977), *Forecasting Economic Time Series.* Economic Theory and Mathematical Economics. New York : Academmic Press.

Granger, C.W.J. and A.P. Andersen (1979), *An Introduction to Bilinear Time Series Models.* Gottingen : Vandenbroeck and Ruprecht.

Granger, C.W.J. and A.A. Weiss (1982), "Rational Autoregressive Models. An Initial Analysis," UCSD Discussion Paper 82-27.

Granger, C.W.J. and A.A. Weiss (1983), "Time Series Analysis of Error-Correction Models," in S. Karlin, T. Amemiya, and L.A. Goodman (eds.), *Studies in Econometrics, Time Series, and Multivariate Statistics.* New York : Academic Press, 255-278.

Hendry, D.F. and N.R. Ericsson (1987), "Assertion Without Empirical Basis : An Econometric Appraisal of 'Monetary Trends in the United Kingdom' by Milton Friedman and Anna J. Schwartz," Applied Economics Discussion Paper, No 25. March 1987. University of Oxford.

Hendry, D.F. and J.F. Richard (1983), "The Econometric Analysis of Economic Time Series," *International Statistical Review*, 51, 111-163.

Ibragimov, I.A. and Y.V. Linnik (1977), *Independent and Stationary Sequences of Random Variables.* The Netherlands : Wolters-Noodhof.

Kloek, T. (1984), "Dynamic Adjustment When the Target is Nonstationary," *International Economic Review*, 25(2), June 1984, 315-326.

Pagan, A. (1985), "Time Series Behaviour and Dynamic Specification," *Oxford Bulletin and Statistics*, 47, 3, 199-211.

Phillips, P.C.B. (1987), "Time Series Regression with Unit Roots". Forthcoming in *Econometrica.*

Priestley, M.B. (1980), "State-Dependent Models : A General Approach to Nonlinear Time Series Analysis," *Journal of Time Series Analysis*, 1, 45-71.

Priestley, M.B. (1981), "Spectral Analysis and Time Series, " in : *Series on Probability and Mathematical Statistics*, Volumes 1 and 2. New York : Academic Press.

Rosenblatt, M. (1978), "Dependence and Asymptotic Independence for Random Processes", in *Studies in Probability Theory*, (M. Rosenblatt, ed.), Washington D.C., Mathematical Association of America.

Stout, W.F. (1974), *Almost Sure Convergence. Probability and Mathematical Statistics.* A Series of Monographs and Textbooks. New York : Academic Press.

Subba Rao, T. (1986), "Statistical Analysis of Bivariate Bilinear Time Series Models," Technical Report #177, Department of Statistics, University of Manchester, Institute of Science and Technology.

Wallis, K.F. (1977), "Multiple Time Series Analysis and the Final Form of Econometrics Models," *Econometrica*, 45, 1481-1497.

White, H. (1984), " Asymptotic Theory for Econometricians", in : *Series on Economic Theory, Econometrics and Mathematical Economics.* New York : Academic Press.

White, H. and I. Domowitz (1984), "Nonlinear Regression with Dependent Observations," *Econometrica*, 52, 143-162.