Working Paper 93-11
Statistics and Econometrics Series 09
Abril, 1993

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

COINTEGRATION AND COMMON FACTORS

Alvaro Escribano and Daniel Peña*

Abstract

Alternative common factors representations for cointegrated vectors are studied. It is shown that dynamic factor models produce as particular cases the alternative common trend representations for cointegrated variables available in the literature, including the one of Stock and Watson(1988). Furthermore, it is proved that common factor representations with I(1) components imply cointegration. A more efficient procedure for finding the numbers of cointegrated vectors based on this dynamic factors model is suggested.

Key Words
Dynamic Factors Models; Cointegration; Common Factors; Unit Roots; VAR Models.

*Departamento de Estadística y Econometría. Universidad Carlos III de Madrid. We acknowledge financial support from the spanish DGICYT contracts PS-90-0014 and PB-90-0266 and the first author acknowledge as well financial support from the SPES Program of the EEC, contract SPES-CT91-0059. We are very grateful to Antoni Espasa and Jesús Gonzalo for useful discussions.

1. INTRODUCTION

One of the most important contributions of the ARIMA methodology advocated by Box and Jenkins (1976) is to show the advantages of differencing univariate time series to obtain stationary processes. However, in the multivariate case differencing should be made with great care, and some of the problems associated to differencing all the components of a vector of time series when building a multivariate ARIMA model were mentioned by Box and Tiao (1977), when the non stationarity of the vector is due to a small number of non stationary components. Granger (1981, 1986) and Engle and Granger (1987) developed these ideas, proposed the concept of cointegration, and related it to the error correction representation advocated by Sargan (1964). Peña and Box (1984, 1987) showed how to identify common factors in a vector of time series and how to build a simplifying transformation to recover the factors as linear combinations of the original series. Stock and Watson (1988) showed that cointegrated multiple time series must have, at least, one common trend or factor. This work was a first step to clarify the connection between cointegration and common factor analysis, further studied by Peña (1990), Gonzalo and Granger (1991), Johansen (1991), and Reinsel and Ahn (1992).

Tiao and Tsay (1989) proposed finding linear combinations of the observed series which lead to simple multiple time series structures. They showed the importance of exchangeable models in multiple time series and the risks of over parametrization and they developed a canonical correlation analysis to find the simplest representation of a vector of time series. Related work on canonical correlation in time series is found in Akaike (1974), Aoki (1987), Reinsel (1983) and Ahn and Reinsel (1988).

This paper analyzes the importance of taking into account cointegration or common factors structures when building multivariate time series models. The importance of the concept of cointegration in the economic literature is due to the possibility of putting together the information about the long-run equilibrium, coming from economic theory, and the statistical evidence about short-run dynamics in the observed series. However, it is proved in this paper that cointegration is equivalent to a particular type of common factor structure and, therefore, it may be potentially very important in many areas of statistical modeling. In particular, the presence of common factors implies a badly defined vector ARIMA representation, (Peña and Box, 1987) and so it must be taken into account when building multiple time series model with nonstationary variables. Also, at the scalar level, the estimation of single equation models may also be unreliable if cointegration is disregarded, see Stock (1987) and Phillips (1991). Surveys on some of these topics are Escribano(1990), Engle and Yoo(1991), Campbell and Perron(1991), Johansen(1993) and Escribano and Espasa(1993).

The paper is organized as follows. Section 2 presents the notation and the time series representations used in the paper. Section 3 discusses the relationship between cointegration and common factors, shows that all common trends representations proposed in the literature are "equivalent" and that a dynamic factors model nest all common trends (factors) representations. Section 4 critically reviews the different approaches suggested in testing for cointegration (common factors) and in identifying the dimension of the cointegration matrix (number of non-stationary common factors) and suggest a potentially more efficient procedure to determine the number of cointegrating vectors. Finally section 5 incorporate some concluding remarks.

2. ALTERNATIVE MULTIPLE TIME SERIES REPRESENTATIONS: A REVIEW FOR JOINTLY I(1) VARIABLES.

We assume that we have an nx1 time series vector X_t^* and, for simplicity, we will consider all variables in deviations from the mean $X_t = X^*_t - \mu_t$. Given the initial conditions, $X^*_j = 0$ for $j \leq 0$, the mean $\mu_t = E(X^*_t)$ is a column vector of n components which can include constants, deterministic time trends, dummy variables etc.

In general, we assume that the vector X_t follows the vector VARMA representation

$$\Phi(B)X_t = \theta(B)\,\epsilon_t \qquad (2.1)$$

where $\Phi(B) = I - \Phi_1 B - \ldots - \Phi_p B^p$, and $\theta(B) = I - \theta_1 B - \ldots - \theta_q B^q$, Φ_i $(i=1,\ldots p)$ and θ_i $(i=1,\ldots q)$ are nxn square matrices, B is the backshift operator, and ϵ_t is a sequence of uncorrelated variables with zero mean and positive definite covariance matrix Σ.

We assume that $\phi(B)$ and $\theta(B)$ are comprime, that is they do not have roots in common, and that the zeros of the determinantal polynomials are as follows: (i) $|\theta(B)|$ has all zeros outside the unit circle; (ii) $|\phi(B)|$ has zeros on or outside the unit circle.

Assuming that $\phi(B)=(1-B)\phi^*(B)$, where $\phi^*(B)$ has all its roots outside the unit circle, and calling $A(B)=\theta^{-1}(B)\,\phi^*(B)$, model (2.1) can be written

$$A(B)(1-B)X_t = \epsilon_t \qquad (2.2a)$$

where $A(B) = I - \sum_{p}^{\infty} A_i B^i$ has all its roots outside the unit circle. The model can also be written as

$$\Pi(B)X_t = \epsilon_t \qquad (2.2b)$$

where $\Pi(B)=(1-B)A(B)=I-\sum_{i=1}^{\infty}\Pi_i B^i$ and $\Pi(B)$ has n roots on the unit circle. Equations (2.2a) and (2.2b) are called the VAR(∞) representation of the process.

An alternative formulation can be derived easily from (2.2a) by inverting the matrix A(B), and calling $C(B)=A^{-1}(B)$

$$(1-B)\,X_t = C(B)\,\epsilon_t \qquad (2.3)$$

and it will be called the MA(∞) representation. Here $C(B)=\sum_{j=0}^{\infty} C_j B^j$ is an nxn matrix of polynomials in the backward shift operator B, $B^kX_t = X_{t-k}$, with $C_0=I$. The invertibility condition says that the roots of the equation $|C(B)|=0$ are outside the unit circle, and therefore, C(1) has full rank.

Definition 2.1

The time series x_t is *univariate integrated of order d*, **I(d)** d=0,1,2,..., if $(1-B)^d x_t$ follows an invertible stationary MA(∞) process, or if x_t has an ARIMA(p,d,q) representation.

Definition 2.2

The time series vector X_t is *jointly integrated of order d*, **JI(d)** d=0,1,2,..., if:

(i) all components are univariate I(d), and

(ii) $(1-B)^d X_t$ follows an invertible VMA(∞).

We have added the term jointly to distinguish the multivariate I(1) concept from the univariate one.

A jointly I(1) process, JI(1), such as (2.3) can be written as a multiple unobserved components model driven by nx1 linearly independent random walks. This representation, at the univariate level, was obtained by Beveridge and Nelson (1981) by decomposing the matrix C(B) as the sum of the zero frequency (B=1) components, and the rest of the frequency terms,

$$C(B) = C(1) + (1-B)C^{*}(B). \tag{2.4}$$

where $C^{*}(B) = (1-B)^{-1}[C(B)-C(1)]$ and $C_j^{*} = -\sum_{i=j+1}^{\infty} C_i$, under the condition that $\sum_{j=0}^{\infty} j^{1/2} \mid C_j \mid < \infty$. Substituting (2.4) into (2.3) we get,

$$(1-B)X_t = C(1)\epsilon_t + (1-B)C^{*}(B)\epsilon_t \tag{2.5}$$

integrating out this equation under the initial condition that $\epsilon_{t-j} = 0$ for $j \geq t$,

$$X_t = C(1) \sum_{j=0}^{t-1} \epsilon_{t-j} + C^{*}(B)\epsilon_t \tag{2.6}$$

which can be written as a *multiple unobserved components model*, usually called multiple Beveridge and Nelson decomposition,

$$X_t = \tau_t + C^{*}(B)\epsilon_t \tag{2.7a}$$

$$\tau_t = \tau_{t-1} + C(1)\epsilon_t \tag{2.7b}$$

where τ_t is an nx1 vector of *random walks* and C(1) is an nxn matrix of *full rank*, rank[C(1)]=n, and so the nx1 system of equations is driven by nx1 linearly independent random walks.

The three basic representation VARMA, VAR(∞) and VMA(∞) mimic closely the univariate representations of homogeneous time series data. The n-unobserved components model will be related to the common factors model to be discussed in the next section.

3. COINTEGRATION AND COMMON FACTOR MODELS

When all individual series are I(1) but some linear combinations are I(0) Granger (1981) named that *cointegration*. This idea goes back to Box and Tiao (1977) and it is related to the scalar components models analyzed by Tiao and Tsay (1989).

Definition 3.1.

The components of the vector X_t are *cointegrated of order 0, c(1,0), with rank r* if:

a) all individual components are univariate I(1) and

b) there are r linearly independent combinations, $\beta' X_t$, which are JI(0) and rank(β) = r.

It is important to stress that if $\beta' X_t$ is JI(0), R $\beta' X_t$ will also be JI(0) for any non singular R matrix. Calling $T' = R\beta'$ to the cointegration matrix we can always choose R to make T'T = I. To see this, note that given any rxn matrix β of full rank r, we can always take $R = A^{-1}(\beta'\beta)^{-1}$, where A is the square root of the positive definite matrix $\beta'\beta$, that is, it verifies that $(\beta'\beta) = AA'$. Then

$$T'T = A^{-1} (\beta'\beta)^{-1} \beta'\beta(\beta'\beta)^{-1}(A^{-1})' = I$$

Therefore, without lost of generality, we can always assume that the cointegration matrix β is normalized in such a way that $\beta'\beta = I$.

Cointegration and model matrices

In term of the VARMA (2.1) representation cointegration implies that the determinantal polynomial, $|\phi(B)|$ has n-r roots on the unit circle. Let $\phi(B) = E_1(B)\, D(B)\, E_2(B)$, where D(B) is a diagonal matrix that has $(1-B)I_{n-r}$ in the upper left hand corner and I_r in the right hand corner, and $E_1(B)$ and $E_2(B)$ are nonsingular matrices that have all the zeros outside the unit circle. This decomposition is called the Smith-McMillan Lemma and was used by Yoo (1987) and Hylleberg and Mizon (1989).

Differencing all the series is equivalent to multiplying the model by a diagonal matrix $D_2(B)$ that has I in the right hand corner and $(1-B)I_r$ in the upper left hand corner $D(B)D_2(B) = (1-B)I$ and therefore: (i) the matrix A(B) in (2.2a) will contain long terms of the type $(1-B)^{-1}\phi_{ij}(B)$ in order to cancel the extra number of unit roots, and a very complicated and long memory structure may appear when, in fact, a simple one may be suitable for the data; (ii) in the VMA(∞) representation (2.3) $|C(B)|$ has r roots on the unit circle and, therefore, it would not be invertible.

Cointegration impose also restrictions on the rank of the matrices $\Pi(1)$ and C(1) of the VAR y VMA form of the process. Starting from the VAR, let us decompose $\Pi(B)$ as

$$\Pi(B) = \Pi(1)B + (1-B)\, \Pi^{*}(B) \qquad (3.1)$$

where $\Pi^*(0)=\Pi(0)=I$, and $\Pi_j^* = -\sum_{i=1}^{\infty} \Pi_{j+i}$. When $\Pi(B)$ has n roots on the unit circle $\Pi^*(B)=A(B)$ in (2.2a) and $\Pi(1)=0$. However, when $\Pi(B)$ has n-r roots on the unit circle, the VAR representation can be written as

$$\Pi^*(B)\ (1-B)X_t = -\Pi(1)X_{t-1} + \epsilon_t \tag{3.2}$$

and it is called an error correction model. From Granger's representation theorem, see Engle and Granger (1987), we know that cointegration implies certain restrictions on the matrices $\Pi(1)$ and $C(1)$. The matrix $\Pi(1)$ has rank r and can be written as

$$\Pi(1) = \alpha \beta' \tag{3.3}$$

where β is the cointegration matrix introduced in Definition 3.1 and α and β are nxr matrices. Calling $\beta_\perp$ the nx(n-r) matrix that spans the null space of β, ($\beta'\beta_\perp = 0$), we obtain

$$\Pi(1)\ \beta_\perp = 0 \tag{3.4}$$

and the rows of the matrix $\Pi(1)$ span the same space as the rows of the matrix β', which will be called the cointegration space. Also, the columns of the matrix $\Pi(1)$ span the same space as α, and the matrix α verifies

$$C(1)\ \alpha = 0. \tag{3.5}$$

In terms of the VMA(∞) (2.3), cointegration implies that the matrix C(1), which measures the long run impact of shocks in impulse response analysis, has rank n-r. The reason for the reduced rank of C(1) is clear, since premultiplying equation (2.6) by β' we get,

$$\beta' X_t = \beta' C(1) \sum_{j=0}^{t-1} \epsilon_{t-j} + \beta' C^*(B)\epsilon_t \tag{3.6a}$$

and so for $\beta' X_t$ to be JI(0) the condition $\beta' C(1) = 0$ must be satisfied, which reduces equations (3.6a) to

$$\beta' X_t = \beta' C^*(B)\epsilon_t \tag{3.6b}$$

with $|\beta' C^*(B)| = 0$ having all roots outside the unit circle.

Cointegration also implies that scalar components models as defined by Tiao and Tsay (1989) exits. These authors introduced this concept as a way to simplify the structure of vector ARMA models: a scalar component model (SCM) is a linear combination $b'X_t$ such that if ϕ_i and θ_i are the vector ARMA matrices in (2.1) $b'\phi_j=0$ for $j=p_1+1,\ldots,p$ and $b'\theta_j=0$ for $i=q_1+1,\ldots,q$. Suppose, for simplicity, that the elements of X_t is follow an ARIMA (0,1,q) model. Then, according to (3.6) the elements of $\beta' X_t$ are, at most, ARIMA (0,0,q-1), and, therefore, any of the linear combinations $\beta' X_t$ are SCM.

Given these results, an alternative definition of cointegration that takes into account the restrictions on the C(1) and Π(1) matrices is the following.

Definition 3.2.

The components of the vector X_t are *cointegrated of order 0, C(1,0), with rank r* if:

a) all individual components are univariate I(1) and

b) there are r linear independent combinations, $\beta' X_t$, which are JI(0). That is $\beta' C(1)=0$, where β' is the rxn cointegrating matrix, and so rank[C(1)] = n-r, or $\Pi(1)\beta_\perp = 0$ and rank $\Pi[(1)] = r$.

It is possible to allow some of the elements of the vector X_t to be I(0) as long as some of the others are I(d) but this complicates the algebra, see Davidson(1991), and for simplicity we will not consider this possibility here.

Cointegration and Common factors

We will prove in this section that the necessary and sufficient condition for cointegration with rank r is that the series are driven by n-r common factors. Proffs of the result that cointegration implies Stocks and Watson's common trends representations can be found in Stock and Watson(1988), Hylleberg and Mizon(1989), Johansen(1991) and the bellow theorem 3.1. The first two papers start with an infinite VMA while the third one starts with a finite VAR in error correction form. However, none of those papers have proved that a common factors representation with (n-r) JI(1) factors implies cointegration.

Theorem 3.1

The components of the time series vector X_t are *cointegrated* of order 0, C(1,0), with rank r *if and only if*:

(i) X_t can be written as a common trends representation.

(ii) X_t can be written as driven by n-r common factors that are JI(1) and r factors that are JI(0), (observed common factors representations).

Proof. We will prove (i) and (ii) at the same time.

First, we show that cointegration implies a common trends representation. If we decompose the matrix C(1) by using the Jordan canonical form,

$$C(1) = HJH^{-1} \tag{3.7}$$

where H is partitioned as $H=(H_1,H_2)$ with H_1 and H_2, of order nx(n-r) and nxr respectively, containing the right eigenvectors of C(1). The J matrix is block diagonal, contains all eigenvalues of C(1) and since under cointegration the rank of C(1) is n-r, it will have n-r eigenvalues $\neq 0$ in the first block of the diagonal, J_{11}, and zeros in all the other blocks. If H^{-1} is partitioned conformably as $H^{-1} = (H^1,H^2)'$, where H^2 contains the left eigenvectors linked to zero eigenvalues, under cointegration the decomposition (3.7) can be reduced to,

$$C(1) = H_1J_{11}H^1 = H_1H2 \tag{3.8}$$

where $H2 = J_{11}H^1$. Note that as $\beta'C(1)=0$, β' contains the left eigenvectors linked to the zero eigenvalues of C(1) and, therefore, $H^2=\beta'$, and $H_1=\beta_\perp$, where $\beta'_\perp$ is the(n-r)xn matrix such that $\beta'_\perp\beta=0$.

Substituting $C(1) = \beta_\perp H2$ in equation (2.6) and calling the stationary process $C^*(B)\epsilon_t=e_t$ and $H2\epsilon_t = v_t$ we obtain

$$X_t = \beta_\perp\tau^*_t + e_t \tag{3.9a}$$

$$\tau^*_t = \tau^*_{t-1} + v_t \tag{3.9b}$$

where the *common stochastic trend or common factors*, τ^*_t, is a vector of order (n-r)x1 instead of the nx1 vector τ_t of (2.7b). Notice also, that premultiplying in (3.5a) by the cointegrating matrix β' we obtain the jointly I(0) representation (3.2) under the condition that $\beta'\beta_\perp=0$.

Second, we will first show that model (3.9) is a special case of the dynamic common factor model for time series studied by Peña and Box (1984, 1987). These authors proposed the model

$$X_t = A\, f_t + u_t \tag{3.10a}$$

$$\phi(B)\, f_t = \theta(B)\, a_t \tag{3.10b}$$

where A is a nxk factor loading matrix, that without loss of generality can be taken such that $A'A=I$, f_t is a kx1 vector of common factors, u_t is a n-dimensional white noise process and the vector f of $k<n$ common factors follows a vector ARIMA representation. Note that in (3.10) u_t and a_t are independent gaussian process,

whereas in (3.9) both equations are driven by the same white noise sequence ϵ_t. It is straightforward to show that (3.9) is a special case of (3.10). Let us assume that there exists n-r non-stationary factors, that follow the model

$$(1-B)\,\tau^*_t = \theta_1(B)\,a_{1t} \tag{3.11}$$

and s=k-n+r stationary factors that follow the model

$$w_t = \theta_2(B)\,a_{2t} \tag{3.12}$$

where $f'_t=[\tau^{*\prime}_t, w'_t]$ and $\theta_i(B)$ for i = 1,2 have all roots outside the unit circle. Then, partitioning A in (3.10a) as $[A'_1 A'_2]$ the model can be written as

$$X_t = A_1\,\tau^*_t + A_2\,w_t + u_t \tag{3.13}$$

where the first term, $A_1\,\tau^*_t$, contains the non-stationary factors, JI(1), and the second, $A_2 w_t$, the stationary factors, JI(0). We have finished the sufficiency part of the proof of (i) and (ii).

To prove the necessary part, we can write the JI(1) and JI(0) factors of (3.13) as,

$$X_t = A_1\,\tau^*_t + e_t \tag{3.14a}$$

$$\tau^*_t = \tau^*_{t-1} + v_t \tag{3.14b}$$

where $e_t = A_2 w_t + u_t$ and $v_t = \theta_1(B)\,a_{1t}$ are nx1 and (n-r)x1 stationary processes. Note that (3.9) is a particular case of (3.14).

Now we will show that if model (3.10) holds, X_t can be decomposed into a stationary component, using the cointegration relationships, and a JI(1) component. We start with the linear transformation to obtain the common factors suggested by Peña and Box(1987). In the general case (3.10) this transformation is

$$Y_t = M\,X_t = \begin{bmatrix} A_\perp \\ - \\ \bar{A} \end{bmatrix} X_t \tag{3.15}$$

where $\bar{A}$ is the Moore-Penroe generalized inverse of A given by the kxn matrix

$$\bar{A} = (A'A)^{-1}A' = A' \tag{3.16}$$

and $A'_\perp$ is the (n-k)xn orthogonal complement of A such that rank $(A'_\perp)$ = n-k and $A'_\perp A = 0_{n-k}$.

In the particular case (3.14) in which the vector of series is driven by n-r JI(1) common factors and s JI(0) factors, the transformation to recover the JI(1) common factors can be easily built as follows: the orthogonal complement of the factor loading matrix A_1 is β', the cointegration matrix, because $\beta' A_1$ must be zero if $\beta' X_t$ is JI(0) in (3.14a). As the matrix M is partitioned into two orthogonal complements, calling $\beta'_\perp$ the (n-r)xn orthogonal complement of β, the transformation will be

$$Y_t = \begin{bmatrix} \beta'_\perp \\ \beta' \end{bmatrix} X_t \tag{3.17}$$

where $\beta'_\perp$ is also the Moore-Penroe generalized inverse of A_1 and A_1 must be equal to $\beta_\perp$. Therefore, cointegration represents a particular case of the factor analysis model in which the loading matrix, A_1, can be estimated by $\beta_\perp$, the orthogonal complement of the cointegration matrix β.

Transformation (3.17) decompose X_t into two components: the first, $\beta' X_t$ is JI(0), and includes the cointegration relationships; the second, $\beta'_\perp X_t$ is JI(1), and contains the nonstationary common factors driving the observed vector of time series. This decomposition can also be expressed in general terms as

$$X_t = P_1 \beta_\perp' X_t + P_2 \beta' X_t \tag{3.18}$$

and it will be called the observed factors model decomposition of X_t for certain P_1 and P_2 matrices of dimension nx(n-r) and nxr that must satisfy:

$$P_1 \beta_\perp' + P_2 \beta' = I \tag{3.19}$$

Premultiplying equation (3.18) by β' we obtain the conditions $\beta' P_1 = 0$, $\beta' P_2 = I$ and by premultiplying by $\beta'_\perp$ we obtain $\beta'_\perp P_1 = I$, $\beta'_\perp P_2 = 0$. This set of conditions will be satisfied by choosing, for instance, $P_1 = \beta_\perp$, $P_2 = \beta$ and then, a simple case of (3.18) is

$$X_t = \beta_\perp \beta'_\perp X_t + \beta\beta' X_t \tag{3.20}$$

Note that (3.20) is a particular formulation of (3.14a) where the nonstationary JI(1) common factors are given by $\beta'_\perp X_t$, the loading matrix by $\beta_\perp$, and the stationary term e_t by $(\beta\beta')X_t$. ■

Alternative Common Factors Representations

Several authors have suggested common factors models that are particular cases of representation (3.10). For instance, Stock and Watson (1988) define the observed common factors by $\beta'_{\perp}X_t$ and take as factor loading matrix $\beta_{\perp}(\beta'_{\perp}\beta_{\perp})^{-1}$ so that the decomposition (3.18) is

$$X_t=\beta_{\perp}(\beta'_{\perp}\beta_{\perp})^{-1}\beta'_{\perp}X_t+\beta(\beta'\beta)^{-1}\beta'X_t \tag{3.21}$$

which is identical to (3.20) if we impose the normalizing condition $\beta'\beta=I$ and $\beta'_{\perp}\beta_{\perp}=I$. The same representation is also used by Johansen (1991). Kasa (1992) has also suggested the decomposition (3.21) but defining the common factors as $(\beta'_{\perp}\beta_{\perp})^{-1}\beta'_{\perp}X_t$ and loading matrix $\beta_{\perp}$. Of course, the differences on these formulations are due to the identification problem between the loading matrix and the common factors stressed by Peña and Box (1987).

The model for the transformed vector X_t can easily be obtained using the VMA representation. Premultiplying by $(\beta_{\perp}'\beta_{\perp})^{-1}\ \beta_{\perp}'$ in equation (3.9a) and using equation (3.6) we get the corresponding VMA representations for Kasa's components

$$(\beta_{\perp}'\beta_{\perp})^{-1}\ \beta_{\perp}'X_t\ =\ \tau^*_t\ +\ (\beta_{\perp}'\beta_{\perp})^{-1}\beta_{\perp}'C^*(B)\epsilon_t \tag{3.22a}$$

$$\beta'X_t\ =\ 0_r\ +\ \beta'C^*(B)\epsilon_t \tag{3.22b}$$

In this case, the observed common factors, first (n-r)x1 block, perfectly isolates in the first term the common trends, τ^*_t, and so are formed by sum of random walks and jointly JI(0) terms.

Premultiplying equation (3.9a) by $\beta_{\perp}'$ and using again (3.6) we get the VMA representation for the observed common factors and the cointegrating vector of the system derived by Stock and Watson(1988),

$$\beta_{\perp}'X_t\ =\ \beta'_{\perp}\beta_{\perp}\tau^*_t\ +\ \beta_{\perp}'C^*(B)\epsilon_t \tag{3.23a}$$

$$\beta'X_t\ =\ 0_r\ +\ \beta'C^*(B)\epsilon_t \tag{3.23b}$$

where the observed common factors, block (3.23a), contain the sum of random walks and jointly JI(0) variables.

Some authors, (see Gonzalo and Granger, 1992) have suggested a common factor model for cointegrated variables in which the common factors are given by $\alpha'_{\perp}X_t$, where $\alpha_{\perp}$ is the null space of α, the matrix of coefficients in the error correction model that was defined in (3.3). Then, the decomposition is

$$X_t = \beta_\perp(\alpha_\perp'\beta_\perp)^{-1}\alpha'_\perp X_t + \alpha(\beta'\alpha)^{-1}\beta' X_t \tag{3.24}$$

Let us show that the representations (3.21) and (3.24) are equivalent. To prove this, we can substitute equations (3.9a) and (3.6b) into equation (3.24)

$$X_t = \beta_\perp(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'[\beta_\perp\tau_t^* + C^*(B)\epsilon_t] + \alpha(\beta'\alpha)^{-1}\beta' C^*(B)\epsilon_t \tag{3.25}$$

factoring out $C^*(B)\epsilon_t$,

$$X_t = \beta_\perp(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'\beta_\perp\tau_t^* + [\beta_\perp(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp' + \alpha(\beta'\alpha)^{-1}\beta']C^*(B)\epsilon_t \tag{3.26}$$

and since from equations (3.24) we know that $[\beta_\perp(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp' + \alpha(\beta'\alpha)^{-1}\beta'] = I_n$, the above equation simplify to Stock and Watson (1988) common trends model (3.9). Repeating the same steps with (3.21)

$$X_t = \beta_\perp(\beta_\perp'\beta_\perp)^{-1}\beta_\perp'[\beta_\perp\tau_t^* + C^*(B)\epsilon_t] + \beta(\beta'\beta)^{-1}\beta' C^*(B)\epsilon_t \tag{3.27}$$

and rearranging terms,

$$X_t = \beta_\perp(\beta_\perp'\beta_\perp)^{-1}\beta_\perp'\beta_\perp\tau_t^* + [\beta_\perp(\beta_\perp'\beta_\perp)^{-1}\beta_\perp' + \beta(\beta'\beta)^{-1}\beta']C^*(B)\epsilon_t \tag{3.28}$$

which is again simplified to (3.9) because the term in brackets is the identity, $[\beta_\perp (\beta_\perp'\beta_\perp)^{-1}\beta_\perp' + \beta(\beta'\beta)^{-1}\beta'] = I_n$.

From Johansen (1991) we know that $C(1) = \beta_\perp(\alpha_\perp'\Pi^+(1)\beta_\perp)^{-1}\alpha_\perp'$ where $\Pi^+(1)$ comes from $\Pi(B) = \Pi(1) + \Pi^+(B)(1-B)$ and so equations (3.9) can be written as,

$$X_t = \beta_\perp\tau_t^* + C^*(B)\epsilon_t \tag{3.29a}$$

$$\tau_t^* = \tau_{t-1}^* + (\alpha_\perp'\Pi^+(1)\beta_\perp)^{-1}\alpha_\perp'\epsilon_t \tag{3.29b}$$

4. ESTIMATION AND TESTING FOR COINTEGRATION

Identifying the space of cointegration implies determining (i) the dimension of the space, that is the number of linealy independent cointegrating vectors; (ii) an orthonormal basis for this space.

As cointegration implies common factors, the methods developed for choosing the number of factors in a vector of time series can be applied to achieve this goal, although with some modifications, (see for instance the work by Akaike (1974), Aoki (1987), Peña and Box (1987), and Tiao and Tsay (1989)). However, in this section we will review briefly two specific procedures to determine the cointegration space due to Stock and Watson (1988) and Johansen (1991).

Stock and Watson (1988) assumes that the vector X_t follows a VAR(p+1) of known order, and suggested an iterative procedure based on testing the null hypothesis of r_0 cointegration vectors against the alternative hypothesis of $r > r_0$. The procedure is as follows:

Step 1: Compute principal components from the covariance matrix of the data $S = T^{-1}\Sigma X_t X_t' = T^{-1}\Sigma (X_t^*-\mu_t)(X_t^*-\mu_t)'$ where μ_t is the vector of means of the series including all deterministic components and X_t^* are the observed values of the variables. Choose the r_0 smallest principal components of S to form an initial estimate of $\hat{\beta}$, $\hat{\beta}_0$, and the remaining $n-r_0$ to form the vector $\hat{\beta}_\perp$. The value r_0 should be the minimum number of cointegrating vectors and could be zero at the beginning of the process.

Step 2: Compute the vector of common trends by $\hat{\beta}_\perp X_t = W_t$, and calling $\Delta = 1-B$ regress ΔW_t on its p lags as follows:

$$\Delta W_t = \sum_{1}^{p} \Pi_j \Delta W_{t-j} + \eta_t. \qquad (4.1)$$

Step 3: Following the idea of Dickey-Fuller test for unit roots, regress the residual η_t on the summed residuals according to the equation

$$\eta_t = \Phi \sum_{i=0}^{t-1} \eta_{(t-1)-i} + a_t. \qquad (4.2)$$

Step 4: Compute the eigenvalues of Φ and test how many of these values can be assumed to be zero. If it is concluded that all the eigenvalues are zero, the number of cointegration vectors is r_0 and the procedure ended. Otherwise, if the largest eigenvalues of Φ is different from zero take $r_1 = r_0 + 1$, and go to step 2. Iterate steps 2, 3 and 4 until no further cointegrating vectors are found.

We will briefly comment on the justification and weakness of each of the steps previously mentioned.

Step 1: The non-stationary linear combinations formed by $\beta_\perp$ will have larger variance than the stationary combinations $\beta' X_t$ and principal components allows an ordering of the linear combinations according to variance. For instance, if we start from the common factor model (3.10), and compute the moment matrices

$$T^{-2}\Sigma X_t X'_{t-k} = T^{-2} A \Sigma f_t f'_{t-k} A' + T^{-2}\Sigma U_t U'_{t-k} + T^{-2}\Sigma (U_t f'_{t-k} A' + A f_t U'_{t-k}) \qquad (4.3)$$

and taking the probability limit and calling

$$\text{plim } T^{-2}\Sigma X_t X'_{t-k} = M_x(k)$$

$$\text{plim } T^{-2}\Sigma f_t f'_{t-k} = M_f(k)$$

we obtain, for $k=0$,

$$M_x(0) = A\, M_f(0)\, A' \qquad (4.4)$$

where $M_x(0)$ and $M_f(0)$ are random matrices which terms are functionals of a Wiener process (see Phillips and Durlauf (1986) and Chang and Wei (1988)). Therefore, as the matrix $M_f(0)$ will be diagonal, asymptotically $M_x(0)$ will have as eigenvectors the columns of A and as eigenvalues the limiting functional of the diagonal elements of $M_f(0)$.

On the other hand, note that also for $k \geq 1$

$$M_x(k) = A\, M_f(k)\, A' \tag{4.5}$$

and, therefore, the lagged moment matrices of the process could also be used to determine the dimension n-r of $M_x(k)$. As

$$\text{plim } T^{-1} \Sigma\, U_t U'_t = \Sigma$$

but

$$\text{plim } T^{-1} \Sigma\, U_t U'_{t-k} = 0$$

and the other terms

$$\text{plim } T^{-1} \Sigma\, U_t f'_{t-k} = 0$$

go to zero at the same rate, it seems more efficient to determine the rank of $M_x(1)$ than the one of $M_x(0)$.

Step 2: The test is based on the assumption that the order p of the VAR model is known. This is a very strong assumption and more research is needed to understand the effect of errors in the specification of the p on the performance of the test.

Steps 3 y 4: According to H_0, η_t is white noise and, therefore, Φ in (4.2) must be zero. On the other hand, let us assume that it exits a cointegration relationship among the components of W_t and, for instance, w_{1t} is stationary. This means that the first component of η_t, η_{1t}, is not white noise but it follows a non invertible moving average process. Then, the first term of the matrix Φ in (4.2) must be equal to one and all the other terms must be zero. Therefore, the number of non-zero eigenvalues of Φ indicates the additional number of cointegrating vectors that have not been taken into account.

For instance, let us assume that p=1. Then, in (4.1) $\eta_t = \Delta W_t$ and (4.2) reduces to

$$\Delta W_t = \Phi\, W_{t-1} + a_t \tag{4.7}$$

This is a multivariate regression and the matrix ϕ will be estimated by

$$\hat{\Phi} = (\Sigma\, \Delta W_t W'_{t-1})(\Sigma\, W_{t-1} W'_{t-1})^{-1} \quad (4.8)$$

and calling

$$\hat{M}_x^*(1) = T^{-2}\Sigma\, \Delta X_t X'_{t-1}, \quad (4.9)$$

we can write

$$\hat{\Phi} = (\beta'_\perp \hat{M}_x^*(1)\beta_\perp)(\beta'_\perp \hat{M}_x(0)\beta_\perp)^{-1} \quad (4.10)$$

and checking the rank of $\hat{\Phi}$ will indicate the dimension of the cointegration space.

The maximum likelihood procedure developed by Johansen (1988, 1991a) assumes normality, that X_t follows a VAR(k+1) model and can be summarized as follows:

Step 1: Regress ΔX_t on the lagged differences $\Delta X_{t-1},\ldots,\Delta X_{t-k}$ and call R_{0t} the vector of residuals from this regression;

Step 2: Regress X_{t-k} on the lagged differences $\Delta X_{t-1},\ldots, \Delta X_{t-k}$ and call R_{kt} the vector of residuals from this regression

Step 3: Compute S_{00}, the covariance matrix of the residuals R_{0t}, S_{kk} the covariance matrix for R_{kk}, and S_{k0}, the cross covariance matrix, as

$$S_{ij} = T^{-1} \sum_{t=1}^{T} R_{it} R'_{jt} \qquad ij=0,k \quad (4.11)$$

Step 4: Build the matrix

$$M = S_{k0} S_{00}^{-1} S_{0k} S_{kk}^{-1} \quad (4.12)$$

and obtain its eigenvalues λ_i and eigenvectors, v_i. We assume $\lambda_1>\lambda_2>\lambda_3>\ldots>\lambda_n>0$.

Step 5a: Testing the null hypothesis H_0:r=0, is done by computing the likelihood ratio

$$Q(n-r) = -T \sum_{i=r+1}^{n} \log(1-\lambda_i) \text{ for } r=0,1,\ldots,n-1 \quad (4.13)$$

that is called the trace statistics and has been tabulated. If Q(n) is not significant, H_0 is not rejected. Otherwise computed Q(n-1), if it is not significant r=1, otherwise r>1 and compute Q(n-2), and so on. If the last significant value of Q is for Q(n-r) then we conclude that the dimension of the cointegrating space is r+1.

Step 5b: An alternative likelihood ratio test statistic of the hypothesis H_0 (rank (β)=r), versus H_1 (rank (β)=r+1) is given by

$$Q(r \mid r+1) = -T \ln(1-\lambda_{r+1}) \quad (4.14)$$

Step 6: Once the rank $(\beta) = r$ has been established, choose β as the eigenvectors linked to the first r eigenvalues of M in (4.12), and $\beta_\perp$ as the n-r eigenvectors corresponding to the n-r remaining eigenvalues.

This procedure is focused on estimating the matrix $\Pi(1)$ and checking its rank. To justify it let us assume that $k=1$. Then the filtering operations of steps (1) and (2) are not needed, an the model to be estimated is

$$\Delta X_t = \Pi(1) X_{t-1} + \epsilon_t \qquad (4.15)$$

where $\Pi(1) = I - \phi = \alpha\beta'$ must be of rank$=r<n$ if cointegration exits. The estimation of α, assuming β known, is a simple multivariate regression problem which leads to

$$\hat{\alpha} = S_{0k}\beta(\beta' S_{kk}\beta)^{-1} \qquad (4.16)$$

and, therefore, the covariance matrix for the residual ϵ_t is estimated by

$$S_{00} - S_{0k}\beta(\beta' S_{kk}\beta)^{-1}\beta' S_{k0} \qquad (4.17)$$

The minimization of the determinant of (4.17) is well known in the theory of canonical correlations and is achieved by finding the eigenvalues of the matrix M in (4.12).

The rank of the M matrix indicates the dimension of β. Note that the unrestricted least squared estimation of $\hat{\Pi}$ in (4.15) is simple $S_{0k}S_{kk}^{-1}$, and a natural approach to estimate r seems to be estimating $\hat{\Pi}$ and checking its rank. Johansen's procedure is doing that, but using the scale-invariant or standardized matrix M (4.12) that is the product of the unrestricted least squares estimators of ΔX_t on X_t and of X_t on ΔX_t, and, therefore, it will have similar properties to $S_{0k}S_{kk}^{-1}$. It is interesting to point out that (4.10) is approximately $(\beta'_\perp S_{0k}\beta_\perp)(\beta'_\perp S_{kk}\beta_\perp)^{-1}$ and, therefore, the procedures by Stock and Watson and Johansen will lead to similar result in most cases.

5. CONCLUDING REMARKS

In this paper we have shown the relationship between cointegration and common factors, obtained that all common factors models can be seen as particular cases of dynamic factors models and that a common factors representation with JI(1) terms implies cointegration. A dynamic factors model was used by Peña and Box(1984, 1987) to identify common factor in a stationary vector of time series, and we have seen in the paper that the model can immediately be extended to nonstationary time series. In doing so, we have shown the similarities among the "different" common factors representations that have been proposed in the literature for cointegrated vectors. Furthermore, a new procedure for detecting the number of common trends has been suggested. This procedure is based on dynamic moment matrices instead of the contemporaneous one previously proposed in the literature.

REFERENCES

Ahn, S.K. and Reinsel, G.C. (1988) Nested reduced rank autoregressive models for multiple time series. *Journal of the American Statistical Association*, 83, 403, 849-856.

Akaike, H. (1974) A new look at the statistical model identification. *IEEE transactions on Automatic Control*, 19,6, 716-722.

Aoki, M. (1987) *State space modeling of time series*. New York: Springer-Verlag.

Beveridge, S. and Nelson, C.R. (1981) A new approach to the decomposition of economic time series into permanent and transitory components with particular attention to measurement of the business cycle. *Journal of Monetary Economics*, 7, 151-174.

Box, G.E.P. and Jenkins, G.M. (1970) *Time series analysis, forecasting and control.* Holden-Day.

Box, G.E.P. and Tiao (1977) A canonical analysis of multiple time series. *Biometrika*, 64, 355-365.

Campbell, J.Y. & Perron, P. (1991) Pitfalls and opportunities: What macroeconomists should know about unit roots, in Blanchard O. J. and Fisher S. (Eds.). *NBER Macroeconomics Annual 1991*. The MIT Press, Cambridge, Massachussets (1991), 141-201.

Chan, N.H. and Wei, C.Z., (1988) Limiting distribution of least squares estimates of unstable autoregressive processes. *The Annals of Statistics,* 16, 367-401.

Davidson, J.E. (1991) The cointegration properties of vector autoregressions models. *Journal of Time Series Analysis*, Vol. 12, nº 1, 41-62.

Engle, R.F. and Granger, C.W.J. (1987) Co-integration and error correction: representation, estimation and testing. *Econometrica*, 55, 251-276.

Engle, R.F. and Yoo, B.S. (1991) Cointegrated economic time series: A survey with new results. In C.W.J. Granger and R.F. Engle (eds) Log-run economic relations. Readings in cointegration. pp. 237-266. Oxford: Oxford University Press.

Escribano A. (1990) Introducción al tema de cointegración y tendencias. *Cuadernos Económicos de ICE*, 44, 7-42.

Escribano A. and Espasa A. (1993) Single equation dynamic models with cointegrated variables: a critical evaluation. W.P. Universidad Carlos III de Madrid.

Gonzalo, J, and Granger, C. (1992) Estimation of common long-memory components in cointegrated systems. Discussion Paper, University of California, San Diego.

Granger, C.W.J. (1981) Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, 16, 121-130.

Granger, C.W.J. (1986) Developments in the study of cointegrated economic variables, *Oxford Bulletin of Economics and Statistics*, 48, 3, 213-228.

Hylleberg S. and Mizon G. E. (1989) Cointegration and error correction mechanisms. *The Economic Journal*, 99, 113-125.

Johansen, S. (1988) Statistical analysis of co-integration vectors. *Journal of Economic Dynamics and Control*, 12, 231-254.

Johansen, S. (1991) Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, Vol. 59, nº 6, 1551-1580.

Johansen, S. (1992) A representation of vector autoregressive processes integrated of order 2. *Econometric Theory*, 8, 188-202.

Johansen, S. (1992a) Cointegration in partial systems and the efficiency of single-equation analysis. *Journal of Econometrics*, 52, 389-402.

Kasa, K. (1992b) Common stochastic trends in international stock markets. *Journal of Monetary Economics*, 29, 95-124.

Johansen, S. (1993) Estimating systems of trending variables. To appear in *Econometric Reviews*.

Peña, D. (1990) Cointegración y reducción de dimensionalidad en series temporales multivariantes. *Cuadernos Económicos de ICE*, 44, 109-126.

Peña, D. and Box, G.E.P. (1984) Hidden relationships in multivariate time series. *Proc. of Bus. Ec. St. American Statistical Association*, 494-499.

Peña, D. and Box, G.E.P. (1987) Identifying a simplifying structure in time series. *Journal of the American Statistical Association*, 82, 836-843.

Phillips, P.C.B. (1991) Optimal inference in co-integrated systems. *Econometrica*, 59, 283-306.

Phillips, P.C.B. and Durlauf, S.N. (1986) Multiple time series regression with integrated processes. *Review of Economic Studies*, 53, 473-496.

Reinsel, G. (1983) Some results on multivariate autoregressive index models. *Biometrika*, 70, 145-156.

Reinsel, G.C. and Ahn, S.K. (1992) Vector Autoregressive Models with Unit Roots and Reduced Rank Structures, Estimation, Likelihood Ratio Test, and Forecasting. *Journal of Time Series Analysis*, Vol 13, nº 4, 353-375.

Saikkonen, P. (1991) Asymptotically efficient estimation of cointegration regressions. *Econometric Theory*, 7, 1-21.

Sargan, J.D. (1964) Wages and prices in the United Kingdom: A study in econometric methodology, in Hart, P.E., Mills, G., and Whitaker, J.K. (Eds.) *Econometric Analysis for National Planning*. (London: Butterworths).

Stock, J.H. (1987) Asymptotic properties of least squares estimators of cointegrating vectors. *Econometrica*, Vol. 55, nº 5, 1035-1056.

Stock, J.H. and Watson, M.W. (1988) Testing for common trends. *Journal of the American Statistical Association*, 83, 1097-1107.

Tiao, G.C. and Tsay, R.S. (1989) Model specification in multivariate time series. *Journal of the Royal Statistical Society B*, 51, 2, 157-213.

Yoo, B.S. (1987) Co-integrated time series: structure, forecasting and testing. Ph.D. Dissertation, UCSD.