COINTEGRATION

by

Juan J. Dolado [a] , Jesús Gonzalo [b] and Francesc Marmol [b]

a: Department of Economics,

b: Department of Statistics and Econometrics

Universidad Carlos III de Madrid

C/. Madrid 126,

28903 Getafe (Madrid)

SPAIN

Version : 10 February, 1999

1 INTRODUCTION

A substantial part of economic theory generally deals with long-run equilibrium relationships generated by market forces and behavioral rules. Correspondingly, most empirical econometric studies entailing time series can be interpreted as attempts to evaluate such relationships in a dynamic framework.

At one time, conventional wisdom was that in order to apply standard inference procedures in such studies, the variables in the system needed to be stationary since the vast majority of econometric theory is built upon the assumption of stationarity. Consequently, for many years econometricians proceeded as if stationarity could be achieved by simply removing deterministic components (e.g., drifts and trends) from the data. However, stationary series should at least have constant unconditional mean and variance over time, a condition which hardly appears to be satisfied in economics, even after removing those deterministic terms.

Those problems were somehow ignored in applied work until important papers by Granger and Newbold (1974) and Nelson and Plosser (1982) alerted many to the econometric implications of non-stationarity and the dangers of running *nonsense* or *spurious* regressions (see, e.g., Granger, chapter ? in this book). In particular, most of the attention focussed on the implications of dealing with integrated variables which are a specific class of non-stationary variables with important economic and statistical properties. These are derived from the presence of unit roots which give rise to stochastic trends, as opposed to pure deterministic trends, with innovations to an integrated process being permanent rather than transitory.

The presence of, at least, a unit root in economic time series is implied in many economic models. Among them, there are those based on the rational use of available information or the existence of very high adjustment costs in some markets. Interesting examples include future contracts, stock prices, yield curves, exchange rates, money velocity, hysteresis theories of unemployment and, perhaps the most popular, the implications of the permanent income hypothesis for real consumption under rational expectations.

Statisticians, in turn, following the influential approach by Box and Jenkins (1970), had advocated transforming integrated time series into stationary ones by successive differencing of the series before modelization. Therefore, from their viewpoint, removing unit roots through differencing ought to be a pre-requisite for regression analysis. However, some authors, notably Sargan (1964), Hendry and Mizon (1978) and Davidson et al. (1978), inter alia, started to criticized on a number of grounds the specification of dynamic models in terms of differenced variables only, especially because of the difficulties in inferring the long-run equilibrium from the estimated model. After all, if deviations from that equilibrium relationship affect future changes in a set of variables, omitting the former, i.e, estimating a differenced model, should entail a misspecification error. However, for some time it remained to be well understood how both variables in differences and levels could coexist in regression models.

Granger (1981), resting upon the previous ideas, solved the puzzle by pointing out that a vector of variables, all which achieve stationarity after differencing, could have linear combinations which are stationary in levels. Later, Engle and Granger (1987) were the first to formalize the idea of integrated variables sharing an equilibrium relation which turned out to be either stationary or have a lower degree of integration than the original series. They denoted this property by *cointegration*, signifying co-movements among trending variables which could be exploited to test for the existence of equilibrium relationships within a fully dynamic specification framework. In this sense, the basic concept of cointegration applies in a variety of economic models including the relationships between capital and output, real wages and labor productivity, nominal exchange rates and relative prices, consumption and disposable income, long and short-term interest rates, money velocity and interest rates, price of shares and dividends, production and sales, etc. In particular, Campbell and Shiller (1987) have pointed out that a pair of integrated variables that are related through a Present Value Model, as it is often the case in macroeconomics and finance, must be cointegrated.

In view of the strength of these ideas, a burgeoning literature on cointegration has developed over the last decade. In this chapter we will explore the basic conceptual issues and discuss related econometric techniques, with the aim of offering an introductory coverage of the main developments in this new field of research. Section 2 provides some preliminaries on the implications of cointegration and the basic estimation and testing procedures in a single equation framework, when variables have a single unit root. In Section 3, we extend the previous techniques to more general multivariate set-ups, introducing those system-based approaches to cointegration which are now in common use. Section 4, in turn, presents some interesting developments on which the recent research on cointegration has been focusing. Finally, Section 5 draws some concluding remarks.

Nowadays, the interested reader, who wants to deepen beyond the introductory level offered here, could find a number of textbooks (e.g., Banerjee et al., 1993, Johansen, 1995 and Hatanaka, 1996) and surveys (e.g., Engle and Granger, 1991 and Watson, 1994 on cointegration) where more general treatments of the relevant issues covered in this chapter are presented. Likewise, there are now many software packages that support the techniques discussed here (e.g., Gauss-*COINT, E-VIEWS* and *PC-FIML*).

2 PRELIMINARIES : UNIT ROOTS AND COINTEGRATION

A well-known result in time series analysis is Wold's (1938) decomposition theorem which states that a stationary time series process, after removal of any deterministic components, has an infinite moving average (*MA*) representation which, in turn, can be represented by a finite autoregressive moving average (*ARMA*) process.

However, as mentioned in the Introduction, many time series need to be appropriately differenced in order to achieve stationarity. From this comes the definition of integration : a time series is said to be integrated of order d, in short, $I(d)$, if it has a stationary, invertible, non-deterministic *ARMA* representation after differencing d times. A white noise series and a stable first-order autoregressive $AR(1)$ process are well-known examples of $I(0)$ series, a random walk process is an example of an $I(1)$ series, while accumulating a random walk gives rise to an $I(2)$ series, etc.

Consider now two time series y_{1t} and y_{2t} which are both $I(d)$ (i.e., they have compatible long-run properties). In general, any linear combination of y_{1t} and y_{2t} will be also $I(d)$. However, if there exists a vector $(1, -b)'$, such that the linear combination

(1) $\quad z_t = y_{1t} - a - by_{2t}$

is $I(d-b)$, $d \geq b > 0$, then, following Engle and Granger (1987), y_{1t} and y_{2t} are defined as cointegrated of order (d, b), denoted $y_t = (y_{1t}, y_{2t})' \sim CI(d, b)$, with (1, -β)' called the cointegrating vector.

Several features in (1) are noteworthy. *First*, as defined above cointegration refers to a linear combination of nonstationary variables. Although theoretically it is possible that nonlinear relationships may exist among a set of integrated variables, the econometric practice about this more general type of cointegration is less developed (see more on this in Section 4). *Second*, note that the cointegrating vector is not uniquely defined, since for any nonzero value of l, $(l, -lb)'$ is also a cointegrating vector. Thus, a normalization rule needs to be used; for example, $l = 1$ has been chosen in (1). *Third*, all variables must be integrated of the same order to be candidates to form a cointegrating relationship. Notwithstanding, there are extensions of the concept of cointegration, called *multicointegration*, when the number of variables considered is larger than two and where the possibility of having variables with different order of integration can be addressed (see, e.g., Granger and Lee, 1989). For example, in a trivariate system, we may have that y_{1t} and y_{2t} are $I(2)$ and y_{3t} is $I(1)$; if y_{1t} and y_{2t} are $CI(2,1)$, it is possible that the corresponding combination of y_{1t} and y_{2t} which achieves that property be itself cointegrated with y_{3t} giving rise to an $I(0)$ linear combination among the three variables. *Fourth*, and most important, most of the cointegration literature focuses on the case where variables contain a single unit root, since few economic variables prove in practice to be integrated of higher order. If variables have a strong seasonal component, however, there may be unit roots at the seasonal frequencies, a case that we will briefly consider in Section 4; see Ghysels, chapter ? in this book for further details. Hence, the remainder of this chapter will mainly focus on the case of $CI(1,1)$ variables, so that z_t in

(1) is $I(0)$ and the concept of cointegration mimics the existence of a long-run equilibrium to which the system converges over time. If, e.g., economic theory suggests the following long-run relationship between y_{1t} and y_{2t},

(2) $\qquad y_{1t} = a + by_{2t}$,

then z_t can be interpreted as the equilibrium error (i.e., the distance that the system is away from the equilibrium at any point in time). Note that a constant term has been included in (1) in order to allow for the possibility that z_t may have nonzero mean. For example, a standard theory of spatial competition argues that arbitrage will prevent prices of similar products in different locations from moving too far apart even if the prices are nonstationary. However, if there are fixed transportation costs from one location to another, a constant term needs to be included in (1).

At this stage, it is important to point out that a useful way to understand cointegrating relationships is through the observation that $CI(1,1)$ variables must share a set of stochastic trends. Using the example in (1), since y_{1t} and y_{2t} are $I(1)$ variables, they can be decomposed into an $I(1)$ component (say, a random walk) plus an irregular $I(0)$ component (not necessarily white noise). Denoting the first components by m_{it} and the second components by u_{it}, $i = 1, 2$, we can write

(3) $\qquad y_{1t} = m_{1t} + u_{1t}$

(3') $\qquad y_{2t} = m_{2t} + u_{2t}$.

Since the sum of an $I(1)$ process and an $I(0)$ process is always $I(1)$, the previous representation must characterize the individual stochastic properties of y_{1t} and y_{2t}. However, if $y_{1t} - by_{2t}$ is $I(0)$, it must be that $m_{1t} = bm_{2t}$, annihilating the $I(1)$ component in the cointegrating relationship. In other words, if y_{1t} and y_{2t} are $CI(1,1)$ variables, they must share (up to a scalar) the same stochastic trend, say m_t, denoted as *common trend*, so that $m_{1t} = m_t$ and $m_{2t} = bm_t$. As before, notice that if m_t is a common trend for y_{1t} and y_{2t}, lm_t will also be a common trend implying that a normalization rule is needed for identification. Generalizing the previous argument to a vector of cointegration and common trends, then it can be proved that if there are $n - r$ common trends among the n variables, there must be r cointegrating relationships. Note that $0 < r < n$, since $r = 0$ implies that each series in the system is governed by a different stochastic trend and that $r = n$ implies that the series are $I(0)$ instead of $I(1)$. These properties constitute the core of two important dual approaches toward testing for cointegration, namely, one that tests directly for the number of cointegrating vectors (r) and another which tests for the number of common trends $(n-r)$. However, before explaining those approaches in more detail (see Section 3), we now turn to another useful representation of $CI(1,1)$ systems which has proved very popular in practice.

Engle and Granger (1987) have shown that if y_{1t} and y_{2t} are cointegrated $CI(1,1)$, then there must exist a so-called *vector error correction model* (*VECM*) representation of the dynamic system governing the joint behavior of y_{1t} and y_{2t} over time, of the following form

(4) $$\Delta y_{1t} = q_{10} + q_{11} z_{t-1} + \sum_{i=1}^{p_1} q_{12,i} \Delta y_{1,t-i} + \sum_{i=1}^{p_2} q_{13,i} \Delta y_{2,t-i} + e_{1t},$$

(4') $$\Delta y_{2t} = q_{20} + q_{21} z_{t-1} + \sum_{i=1}^{p_3} q_{22,i} \Delta y_{1,t-i} + \sum_{i=1}^{p_4} q_{23,i} \Delta y_{2,t-i} + e_{2t},$$

where Δ denotes the first-order time difference (i.e., $\Delta y_t = y_t - y_{t-1}$) and where the lag lengths $p_i, i=1,\dots,4$ are such that the innovations $e_t = (e_{1t}, e_{2t})'$ are $i.i.d.(0,\Sigma)$. Furthermore, they proved the converse result that a *VECM* generates cointegrated $CI(1,1)$ series as long as the coefficients on z_{t-1} (the so-called *loading or speed of adjustment parameters*) are not simultaneously equal to zero.

Note that the term z_{t-1} in equations (4) and (4') represents the extent of the disequilibrium levels of y_1 and y_2 in the previous period. Thus, the *VECM* representation states that changes in one variable not only depends on changes of the other variables and its own past changes, but also on the extent of the disequilibrium between the levels of y_1 and y_2. For example, if $b = 1$ in (1), as many theories predict when y_{1t} and y_{2t} are taken in logarithmic form, then if y_1 is larger than y_2 in the past $(z_{t-1} > 0)$, then $q_{11} < 0$ and $q_{21} > 0$ will imply that, everything else equal, y_1 would fall and y_2 would rise in the current period, implying that both series adjust toward its long-run equilibrium. Notice that both q_{11} and q_{21} cannot be equal to zero. However, if $q_{11} < 0$ and $q_{21} = 0$, then all of the adjustment falls on y_1, or vice versa if $q_{11} = 0$ and $q_{21} > 0$. Note also that the larger are the speed of adjustment parameters (with the right signs), the greater is the convergence rate toward equilibrium. Of course, at least one of those terms must be nonzero, implying the existence of Granger causality in cointegrated systems in at least one direction. Hence, the appeal of the *VECM* formulation is that it combines flexibility in dynamic specification with desirable long-run properties : it could be seen as capturing the transitional dynamics of the system to the long-run equilibrium suggested by economic theory (see, e.g., Hendry and Richard, 1983). Further, if cointegration exists, the *VECM* representation will generate better forecasts than the corresponding representation in first-differenced form (i.e., with $q_{11} = q_{21} = 0$), particularly over medium and long-run horizons, since under cointegration z_t will have a finite forecast error variance whereas any other linear combination of the forecasts of the individual series in y_t will have infinite variance; see Engle and Yoo (1987) for further details.

Based upon the *VECM* representation, Engle and Granger (1987) suggest a two-step estimation procedure for dynamic modeling which has become very popular in applied research. Assuming that $y_t \sim I(1)$, then the procedure goes as follows :

(*i*) First, in order to test whether the series are cointegrated, the *cointegration regression*

(5) $$y_{1t} = a + by_{2t} + z_t$$

is estimated by ordinary least squares (*OLS*) and it is tested whether the *cointegrating residuals* $\hat{z}_t = y_{1t} - \hat{a} - \hat{b}y_{2t}$ are $I(1)$. To do this, for example, we can perform a Dickey-Fuller test on the residual sequence $\{\hat{z}_t\}$ to determine whether it has a unit root. For this, consider the autoregression of the residuals

(6) $$\Delta\hat{z}_t = r_1\hat{z}_{t-1} + e_t$$

where no intercept term has been included since the $\{\hat{z}_t\}$, being residuals from a regression equation with a constant term, have zero mean. If we can reject the null hypothesis that $r_1 = 0$ against the alternative $r_1 < 0$ at a given significance level, we can conclude that the residual sequence is $I(0)$ and, therefore, that y_{1t} and y_{2t} are $CI(1,1)$. It is noteworthy that for carrying out this test it is not possible to use the Dickey-Fuller tables themselves since $\{\hat{z}_t\}$ are a generated series of residuals from fitting regression (5). The problem is that the *OLS* estimates of a and b are such that they minimize the residual variance in (5) and thus prejudice the testing procedure toward finding stationarity. Hence, larger (in absolute value) critical levels than the standard Dickey-Fuller ones are needed. In this respect, MacKinnon (1991) provides appropriate tables to test the null hypothesis $r_1 = 0$ for any sample size and also when the number of regressors in (5) is expanded from one to several variables. In general, if the $\{\hat{e}_t\}$ sequence exhibits serial correlation, then an augmented Dickey-Fuller (*ADF*) test should be used, based this time on the extended autoregression

(6') $$\Delta\hat{z}_t = r_1\hat{z}_{t-1} + \sum_{i=1}^{p} z_i \Delta\hat{z}_{t-i} + e_t ,$$

where again, if $r_1 < 0$, we can conclude that y_{1t} and y_{2t} are $CI(1,1)$. Alternative versions of the test on $\{\hat{z}_t\}$ being *I*(1) versus *I*(0) can be found in Phillips and Ouliaris (1990). Banerjee et al. (1997), in turn, suggest another class of tests based this time on the direct significance of the loading parameters in (4) and (4') where the b coefficient is estimated alongside the remaining parameters in a single step using nonlinear least squares (*NLS*).

If we reject that $\hat{z}_t$ are *I*(1), Stock (1987) has shown that the *OLS* estimate of b in equation (5) is *super-consistent*, in the sense that the *OLS* estimator $\hat{b}$ converges in probability to its true value b at a rate proportional to the inverse of the sample size, T^{-1}, rather than at $T^{-1/2}$ as is the standard result in the ordinary case where y_{1t} and y_{2t} are *I*(0). Thus, when *T* grows, convergence is much quicker in the $CI(1,1)$ case. The intuition behind this remarkable result can be seen by analyzing the behavior of $\hat{b}$ in (5) (where the constant is omitted for simplicity) in the particular case where $z_t \sim i.i.d.\left(0, s_z^2\right)$, and that $q_{20} = q_{21} = 0$ and $p_3 = p_4 = 0$, so that y_{2t} is assumed to follow a simple random walk

(7) $$\Delta y_{2t} = e_{2t} ,$$

or, integrating (7) backwards with $y_{20} = 0$,

(7') $$y_{2t} = \sum_{i=1}^{t} e_{2i} ,$$

with e_{2t} possibly correlated with z_t. In this case, we get $\mathrm{var}(y_{2t}) = t\,\mathrm{var}(e_{21}) = ts_2^2$, exploding as $T \uparrow \infty$. Nevertheless, it is not difficult to show that $T^{-2}\sum_{t=1}^{T} y_{2t}^2$ converges to a random variable. Similarly, the cross-product $T^{-1/2}\sum_{t=1}^{T} y_{2t} z_t$ will explode, in contrast to the stationary case where a simple application of the Central Limit Theorem implies that it is asymptotically normally distributed. In the *I*(1) case, $T^{-1}\sum_{t=1}^{T} y_{2t} z_t$ converges also to a random variable. Both random variables are functionals of Brownian motions which will be denoted henceforth, in general, as $f(B)$. A Brownian motion is a zero-mean normally distributed continuous (a.s.) process with independent increments, i.e., loosely speaking, the continuous version of the discrete random walk. See Phillips (1987) and Bierens (chapter ? in this book).

Now, from the expression for the *OLS* estimator of b, we obtain

$$(8) \qquad \hat{b} - b = \frac{\sum_{t=1}^{T} y_{2t} z_t}{\sum_{t=1}^{T} y_t^2},$$

and, from the previous discussion, it follows that

$$(9) \qquad T\left(\hat{b} - b\right) = \frac{T^{-1}\sum_{t=1}^{T} y_{2t} z_t}{T^{-2}\sum_{t=1}^{T} y_t^2}$$

is asymptotically (as $T \uparrow \infty$) the ratio of two non-degenerate random variables that in general, is not normally distributed. Thus, in spite of the super-consistency, standard inference cannot be applied to $\hat{b}$ except in some restrictive cases which are discussed below.

(*ii*) After rejecting the null hypothesis that the cointegrating residuals in equation (5) are *I*(1), the $\hat{z}_{t-1}$ term is included in the *VECM* system and the remaining parameters are estimated by *OLS*. Indeed, given the superconsistency of $\hat{b}$, Engle and Granger (1987) show that their asymptotic distributions will be identical to using the true value of b. Now all the variables in (3) and (3') are *I*(0) and conventional modeling strategies (e.g., testing the maximum lag length, residual autocorrelation or whether either q_{11} or q_{21} is zero, etc.) can be applied to assess model adequacy; see Lütkepohl, (chapter ? in this book) for further details.

In spite of the beauty and simplicity of the previous procedure, however, several problems remain. In particular, although $\hat{b}$ is super-consistent, this is an asymptotic result and thus biases could be important in finite samples. For instance, assume that the rates of convergence of two estimators are $T^{-1/2}$ and $10^{10}T^{-1}$. Then, we will need huge sample sizes to have the second estimator dominating the first one. In this sense, Monte Carlo experiments by Banerjee et al. (1993) showed that the biases could be important particularly when z_t and Δy_{2t} are highly serially correlated and they are not independent. Phillips (1991), in turn, has shown analytically that in the case where y_{2t} and z_t are independent at all leads and lags, the distribution in (9) as T grows behaves like a gaussian distribution (technically is a *mixture of normals*) and, hence, the distribution of the *t*-statistic of b is also asymptotically normal. For this reason, Phillips and Hansen (1990) have developed an estimation procedure which corrects for the previous bias while achieves asymptotic normality. The procedure, denoted as a *fully modified ordinary least squares estimator* (*FM-OLS*), is based upon a correction to the *OLS* estimator given in (8) by which the error term z_t is conditioned on the whole process $\{\Delta y_{2t}, t = 0, \pm 1, \ldots\}$ and, hence, orthogonality between regressors and disturbance is achieved by construction. For example, if z_t and e_{2t} in (5) and (7) are correlated white noises with $g = E(z_t e_{2t}) / \mathrm{var}(e_{2t})$, the *FM-OLS* estimator of b, denoted $\hat{b}_{FM}$, is given by

$$(10) \qquad \hat{b}_{FM} = \frac{\sum_{t=1}^{T} y_{2t}\left(y_{1t} - \hat{g}\Delta y_{2t}\right)}{\sum_{t=1}^{T} y_t^2},$$

where $\hat{g}$ is the empirical counterpart of g obtained from regressing the *OLS* residuals $\hat{z}_t$ on Δy_{2t}. When z_t and Δy_{2t} follow more general processes, the *FM-OLS* estimator of b is similar to (10) except that further corrections are needed in its numerator. Alternatively, Saikkonen (1991) and Stock and Watson (1993) have shown that, since $E\left(z_t \middle| \{\Delta y_{2t}\}\right) = h(L)\Delta y_{2t}$, where $h(L)$ is a two-sided filter in the lag operator *L*, regression of y_{1t} on y_{2t} and leads and lags of Δy_{2t} (suitably truncated), using either

OLS or *GLS*, will yield an estimator of b which is asymptotically equivalent to the *FM-OLS* estimator. The resulting estimation approach is known as *dynamic OLS* (respectively *GLS*) or *DOLS* (respectively, *DGLS*).

3 SYSTEM-BASED APPROACHES TO COINTEGRATION

Whereas in the previous section we confined the analysis to the case where there is at most a single cointegrating vector in a bivariate system, this set-up is usually quite restrictive when analyzing the cointegrating properties of an *n*-dimensional vector of *I*(1) variables where several cointegration relationships may arise. For example, when dealing with a trivariate system formed by the logarithms of nominal wages, prices and labor productivity, there may exist two relationships, one determining an employment equation and another determining a wage equation. In this section we survey some of the popular estimation and testing procedures for cointegration in this more general multivariate context, which will be denoted as system-based approaches.

In general, if y_t now represents a vector of *n* *I*(1) variables its Wold representation (assuming again no deterministic terms) is given by

(11) $\Delta y_t = C(L)e_t$,

where now $e_t \sim nid(0, \Sigma)$, Σ being the covariance matrix of e_t and *C*(*L*) an $(n \times n)$ invertible matrix of polynomial lags, where the term "invertible" means that $|C(L) = 0|$ has all its roots strictly larger than unity in absolute value. If there is a cointegrating $(n \times 1)$ vector, $b' = (b_{11}, \ldots, b_{nn})$, then, premultiplying (11) by b' yields

(12) $b'\Delta y_t = b'\left[C(1) + \tilde{C}(L)\Delta\right]e_t$,

where *C*(*L*) has been expanded around $L = 1$ using a first-order Taylor expansion and $\tilde{C}(L)$ can be shown to be an invertible lag matrix. Since the cointegration property implies that $b'y_t$ is *I*(0), then it must be that $b'C(1) = 0$ and hence $\Delta\,(= 1 - L)$ will cancel out on both sides of (12). Moreover, given that *C*(*L*) is invertible, then y_t has a vector autoregressive representation such that

(13) $A(L)y_t = e_t$,

where $A(L)C(L) = \Delta I_n$, I_n being the $(n \times n)$ identity matrix. Hence, we must have that $A(1)C(1) = 0$, implying that $A(1)$ can be written as a linear combination of the elements

b, namely, $A(1) = ab'$, with a being another $(n \times 1)$ vector. In the same manner, if there were r cointegrating vectors $(0 < r < n)$, then $A(1) = B\Gamma'$, where B and Γ are this time $(n \times r)$ matrices which collect the r different a and b vectors. Matrix B is known as the *loading matrix* since its rows determine how many cointegrating relationships enter each of the individual dynamic equations in (13). Testing the rank of $A(1)$ or $C(1)$, which happen to be r and $n - r$, respectively, constitutes the basis of the following two procedures :

(*i*) Johansen (1995) develops a maximum likelihood estimation procedure based on the so-called *reduced rank regression method* that, as the other methods to be later discussed, presents some advantages over the two-step regression procedure described in the previous section. First, it relaxes the assumption that the cointegrating vector is unique, and, secondly, it takes into account the short-run dynamics of the system when estimating the cointegrating vectors. The underlying intuition behing Johansen's testing procedure can be easily explained by means of the following example. Assume that y_t has a $VAR(1)$ representation, that is, $A(L)$ in (13) is such that $A(L) = I_n - A_1 L$. Hence, the $VAR(1)$ process can be reparameterized in the *VECM* representation as

(14) $\Delta y_t = (A_1 - I_n) y_{t-1} + e_t$.

If $A_1 - I_n = -A(1) = 0$, then y_t is $I(1)$ and there are no cointegrating relationships $(r = 0)$, whereas if $rank(A_1 - I_n) = n$, there are n cointegrating relationships among the n series and hence $y_t \sim I(0)$. Thus, testing the null hypothesis that the number of cointegrating vectors (r) is equivalent to testing whether $rank(A_1 - I_n) = r$. Likewise, alternative hypotheses could be designed in different ways, e.g., that the rank is $(r+1)$ or that it is n.

Under the previous considerations, Johansen (1995) deals with the more general case where y_t follows a $VAR(p)$ process of the form

(15) $y_t = A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + e_t$,

which, as in (3) and (3'), can be rewritten in the *ECM* representation

(16) $\Delta y_t = D_1 \Delta y_{t-1} + D_2 \Delta y_{t-2} + \cdots + D_{p-1} \Delta y_{t-p+1} + D y_{t-1} + e_t$.

Where $D_i = -(A_{i+1} + \cdots + A_p)$, $i = 1, 2, \ldots, p-1$, and $D = (A_1 + \cdots + A_p - I_n) = -A(1)$

$= -B\Gamma'$. To estimate B and Γ, we need to estimate D subject to some identification restriction since otherwise B and Γ could not be separately identified. Maximum likelihood estimation of D goes along the same principles of the basic partitioned regression model, namely, the regressand and the regressor of interest $(\Delta y_t$ and $y_{t-1})$ are regressed by *OLS* on the remaining set of regressors $(\Delta y_{t-1},\ldots,\Delta y_{t-p+1})$ giving rise to two matrices of residuals denoted as $\hat{e}_0$ and $\hat{e}_1$ and the regression model $\hat{e}_{ot} = \hat{D}\hat{e}_{1t} + \text{residuals}$. Following the preceding discussion, Johansen (1995) shows that testing for the rank of $\hat{D}$ is equivalent to test for the number of canonical correlations between $\hat{e}_0$ and $\hat{e}_1$ that are different from zero. This can be conducted using either of the following two test statistics

(17) $$l_{tr}(r) = -T\sum_{i=r+1}^{n}\ln\left(1-\hat{l}_i\right)$$

(18) $$l_{max}(r, r+1) = -T\ln\left(1-\hat{l}_{r+1}\right),$$

where the $\hat{l}_i$'s are the eigenvalues of the matrix $S_{10}S_{00}^{-1}S_{01}$ with respect to the matrix S_{11}, ordered in decreasing order $\left(1>\hat{l}_1>\cdots>\hat{l}_n>0\right)$, where $S_{ij} = T^{-1}\sum_{t=1}^{T}\hat{e}_{it}\hat{e}_{jt}^{'}$, $i, j = 0, 1$. These eigenvalues can be obtained as the solution of the determinantal equation

(19) $$\left|lS_{11} - S_{10}S_{00}^{-1}S_{01}\right| = 0.$$

The statistic in (17), known as the *trace statistic*, tests the null hypothesis that the number of cointegrating vectors is less than or equal to *r* against a general alternative. Note that, since $\ln(1)=0$ and $\ln(0)\uparrow -\infty$, it is clear that the trace statistic equals zero when all the $\hat{l}_i$'s are zero, whereas the further the eigenvalues are from zero the more negative is $\ln\left(1-\hat{l}_i\right)$ and the larger is the statistic. Likewise, the statistic in (18), known as the *maximum eigenvalue statistic*, tests a null of *r* cointegrating vectors against the specific alternative of $r+1$. As above, if $\hat{l}_{r+1}$ is close to zero, the statistic will be small. Further, if the null hypothesis is not rejected, the *r* cointegrating vectors contained in matrix Γ can be estimated as the first *r* columns of matrix $\hat{V} = (\hat{v}_1,\ldots,\hat{v}_n)$ which contains the eigenvectors associated to the eigenvalues in (19) computed as

$$\left(l_i S_{11} - S_{10} S_{00}^{-1} S_{01} \right) \hat{v}_i = 0, \qquad i = 1, 2, \ldots, n$$

subject to the length normalization rule $\hat{V}' S_{11} \hat{V} = I_n$. Once Γ has been estimated, estimates of the *B*, D_i and Σ matrices in (16) can be obtained by inserting $\hat{\Gamma}$ in their corresponding *OLS* formulae which will be functions of Γ.

Osterwald-Lenum (1992) has tabulated the critical values for both tests using Monte Carlo simulations, since their asymptotic distributions are multivariate $f(B)$ which depend upon: (*i*) the number of nonstationary components under the null hypothesis $(n-r)$ and (*ii*) the form of the vector of deterministic components, m (e.g., a vector of drift terms), which needs to be included in the estimation of the *ECM* representation where the variables have nonzero means. Since, in order to simplify matters, the inclusion of deterministic components in (16) has not been considered so far, it is worth using a simple example to illustrate the type of interesting statistical problems that may arise when taking them into account. Suppose that $r = 1$ and that the unique cointegrating vector in Γ is normalized to be $b' = \left(1, b_{22}, \ldots, b_{nn}\right)$, while the vector of speed of adjustment parameters, with which the cointegrating vector appears in each of the equations for the n variables, is $a' = \left(a_{11}, \ldots a_{nn}\right)$. If there is a vector of drift terms $m' = \left(m_1, \ldots m_n\right)$ such that they satisfy the restrictions $m_i = a_{ii} m_1$ (with $a_{11} = 1$), it then follows that all Δy_{it} in (16) are expected to be zero when $y_{1,t-1} + b_{22} y_{2,t-1} + \cdots + b_{nn} y_{n,t-1} + m_1 = 0$ and, hence, the general solution for each of the $\left\{ y_{it} \right\}$ processes, when integrated, will not contain a time trend. Many other possibilities, like e.g. allowing for a linear trend in each variable but not in the cointegrating relations, may be considered. In each case, the asymptotic distribution of the cointegration tests given in (17) and (18) will differ, and the corresponding sets of simulated critical values can be found in the reference quoted above. Sometimes, theory will guide the choice of restrictions; for example, if one is considering the relation between short-term and long-term interest rates, it may be wise to impose the restriction that the processes for both interest rates do not have linear trends and that the drift terms are restricted to appear in the cointegrating relationship interpreted as the “term structure”. However, in other instances one may be interested in testing alternative sets of restrictions on the way m enters the system; see, e.g., Lütkepohl, (chapter ? in this book) for further details

In that respect, the Johansen's approach allows to test restrictions on m, B and Γ subject to a given number of cointegrating relationships. The insight to all these tests, which turn out to have asymptotic chi-square distributions, is to compare the number of cointegrating vectors (i.e., the number of eigenvalues which are significantly different from zero) both when the restrictions are imposed and when they are not. Since if the true cointegration rank is *r*, only *r* linear combinations of the variables are stationary, one should find that the number of cointegrating vectors does not diminish if the restrictions are not binding and vice versa. Thus, denoting by $\hat{l}_i$ and l_i^* the set of *r* eigenvalues for the unrestricted and restricted cases, both sets of eigenvalues should be equivalent if the restrictions are valid. For example, a modification of the trace test in the form

$$(20) \qquad T\sum_{i=1}^{r}\left[\ln\left(1-l_i^*\right)-\ln\left(1-\hat{l}_i\right)\right]$$

will be small if the l_i^*'s are similar to the $\hat{l}_i$'s, whereas it will be large if the l_i^*'s are smaller than the $\hat{l}_i$'s. If we impose *s* restrictions, then the above test will reject the null hypothesis if the calculated value of (20) exceeds that in a chi-square table with $r(n-s)$ degrees of freedom.

Most of the existing Monte Carlo studies on the Johansen methodology point out that dimension of the data series for a given sample size may pose particular problems since the number of parameters of the underlying *VAR* models grows very large as the dimension increases. Likewise, difficulties often arise when, for a given *n*, the lag length of the system, *p*, is either over or under-parameterized. In particular, Ho and Sorensen (1996) and Gonzalo and Pitarakis (1998) show by numerical methods that the cointegrating order will tend to be overestimated as the dimension of the system increases relative to the time dimension, while serious size and power distortions arise when choosing too short and too long a lag length, respectively. Although several degrees of freedom adjustments to improve the performance of the test statistics have been advocated (see, e.g., Reinsel and Ahn, 1992), researchers ought to have considerable care when using the Johansen estimator to determine cointegration order in high dimensional systems with small sample sizes. Nonetheless, it is worth noticing that a useful approach to reduce the dimension of the *VAR* system is to rely upon exogeneity arguments to construct smaller conditional systems as suggested by Ericsson (1992) and Johansen (1992a). Equally, if the *VAR* specification is not appropriate, Phillips (1991) and Saikkonen (1992) provide efficient estimation of cointegrating vectors in more general time series settings, including vector *ARMA* processes.

(*ii*) As mentioned above, there is a dual relationship between the number of cointegrating vectors (r) and the number of common trends ($n-r$) in an n-dimensional system. Hence, testing for the dimension of the set of "common trends" provides an alternative approach to testing for the cointegration order in a *VAR//VECM* representation. Stock and Watson (1988) provide a detailed study of this type of methodology based on the use of the so-called Beveridge-Nelson (1981) decomposition. This works from the Wold representation of an $I(1)$ system, which we can write as in expression (11) with $C(L)=\sum_{j=0}^{\infty} C_j L^j$, $C_0 = I_n$. As shown in expression (12), $C(L)$ can be expanded as $C(L)=C(1)+\tilde{C}(L)(1-L)$, so that, by integrating (11), we get

(21) $$y_t = C(1)Y_t + \tilde{w}_t,$$

where $\tilde{w}_t = \tilde{C}(L)e_t$ can be shown to be covariance stationary, and $Y_t = \sum_{i=1}^{t} e_i$ is a latent or unobservable set of random walks which capture the $I(1)$ nature of the data. However, as above mentioned, if the cointegration order is r, there must be an $(r \times n)$ Γ matrix such that $\Gamma' C(1)=0$ since, otherwise, $\Gamma' y_t$ would be $I(1)$ instead of $I(0)$. This means that the $(n \times n)$ $C(1)$ matrix cannot have full rank. Indeed, from standard linear algebra arguments, it is easy to prove that the rank of $C(1)$ is $(n-r)$, implying that there are only $(n-r)$ independent common trends in the system. Hence, there exists the so-called *common trends representation* of a cointegrated system, such that

(22) $$y_t = \Phi y_t^c + \tilde{w}_t,$$

where Φ is an $n \times (n-r)$ matrix of loading coefficients such that $\Gamma' \Phi = 0$ and y_t^c is an $(n-r)$ vector random walk. In other words, y_t can be written as the sum of $(n-r)$ common trends and an $I(0)$ component. Thus, testing for $(n-r)$ common trends in the system is equivalent to testing for r cointegrating vectors. In this sense, Stock and Watson's (1988) testing approach relies upon the observation that, under the null hypothesis, the first-order autoregressive matrix of y_t^c should have $(n-r)$ eigenvalues equal to unity, whereas, under the alternative hypothesis of higher cointegration order, some of those eigenvalues will be less than unity. It is worth noticing that there are other alternative strategies to identify the set of common trends, y_t^c, which do not impose a vector random walk structure. In particular, Gonzalo and Granger (1995), using arguments embedded in the Johansen's approach, suggest identifying y_t^c as linear combinations of y_t which are not caused in the long-run by the cointegration relationships $\Gamma' y_{t-1}$. These linear combinations are the orthogonal complement of matrix B in (16), $y_t^c = B_{\perp} y_t$, where $B_{\perp}$ is an $\left(n \times (n-r)\right)$ full ranked matrix, such that $B' B_{\perp} = 0$, that can be estimated as the last $(n-r)$ eigenvectors of the second moments matrix $S_{01} S_{11}^{-1} S_{10}$ with respect to S_{00}. For instance, when some of the rows of matrix B are zero, the common trends will be linear combinations of those $I(1)$ variables in the system where the cointegrating vectors do not enter into their respective adjustment equations. Since common trends are expressed in terms of observable variables, instead of a latent set of random walks, economic theory can again be quite useful in helping to provide useful interpretation of their role. For example, the rational expectations version of the permanent income hypothesis of consumption states that consumption follows a random walk whilst saving (disposable income minus consumption) is $I(0)$. Thus, if the theory is a valid one, the cointegrating vector in the system formed by consumption and disposable income should be $b' = (1, -1)$ and it would only appears in the second equation (i.e., $a' = (0, a_{22})$), implying that consumption should be the common trend behind the nonstationary behavior of both variables.

To give a simple illustration of the conceptual issues discussed in this section, let us consider the following Wold (MA) representation of the bivariate $I(1)$ process $y_t = (y_{1t}, y_{2t})'$,

$$(1-L)\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = (1-0.2L)^{-1}\begin{pmatrix} 1-0.6L & 0.8L \\ 0.2L & 1-0.6L \end{pmatrix}\begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}.$$

Evaluating $C(L)$ at $L=1$ yields

$$C(1) = \begin{pmatrix} 0.5 & 1 \\ 0.25 & 0.5 \end{pmatrix},$$

so that $rank\, C(1) = 1$. Hence, $y_t \sim CI(1,1)$. Next, inverting $C(L)$, yields the VAR representation

$$\begin{pmatrix} 1-0.6L & -0.8L \\ -0.2L & 1-0.6L \end{pmatrix}\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix},$$

where

$$A(1) = \begin{pmatrix} 0.4 & -0.8 \\ -0.2 & 0.4 \end{pmatrix},$$

so that $rank\ A(1) = 1$ and

$$A(1) = \begin{pmatrix} 0.4 \\ -0.2 \end{pmatrix}(1,-2) = ab'.$$

Hence, having normalized on the first element, the cointegrating vector is $b' = (1,-2)$, leading to the following *VECM* representation of the system

$$(1-L)\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} -0.4 \\ 0.2 \end{pmatrix}(1,-2)\begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}.$$

Next, given $C(1)$ and normalizing again on the first element, it is clear that the common factor is $y_t^c = \sum_{i=1}^{t} e_{1i} + 2\sum_{i=1}^{t} e_{2i}$, whereas the loading vector Φ and the common trend representation would be as follows

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} 0.5 \\ 0.25 \end{pmatrix} y_t^c + \tilde{w}_t .$$

Notice that $b' y_t$ eliminates y_t^c from the linear combination which achieves cointegration. In other words, Φ is the orthogonal complement of b once the normalization criteria has been chosen.

Finally, to examine the effects of drift terms, let us add a vector $m = (m_1, m_2)'$ of drift coefficients to the *VAR* representation. Then, it is easy to prove that y_{1t} and y_{2t} will have a linear trends with slopes equal to $m_1/2 + m_2$ and $m_1/4 + m_2/2$, respectively. When $2m_1 + m_2 \neq 0$ the data will have linear trends, whereas the cointegrating relationship will not have them, since the linear combination in b annihilates the individual trends for any m_1 and m_2.

The interesting case arises when the restriction $2m_1 + m_2 = 0$ holds, since now the linear trend is purged from the system, leading to the restricted *ECM* representation

$$(1-L)\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} -0.4 \\ 0.2 \end{pmatrix} (1,-2,-m_1^*) \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \\ 1 \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix},$$

where $m_1^* = m_1/0.4$.

4 FURTHER RESEARCH ON COINTEGRATION

Although the discussion in the previous sections has been confined to the possibility of cointegration arising from linear combinations of $I(1)$ variables, the literature is currently proceeding in several interesting extensions of this standard set-up. In the sequel we will briefly outline some of those extensions which have drawn a substantial amount of research in the recent past.

(*i*) *Higher Order Cointegrated Systems*

The statistical theory of $I(d)$ systems with $d = 2, 3, \ldots$, is much less developed than the theory for the $I(1)$ model, partly because it is uncommon to find time series, at least in economics, whose degree of integration higher than two, partly because the theory in quite involved as it must deal with possibly multicointegrated cases where, for instance, linear combinations of levels and first differences can achieve stationarity. We refer the reader to Haldrup (1997) for a survey of the statistical treatment of $I(2)$ models, restricting the discussion in this chapter to the basics of the $CI(2,2)$ case.

Assuming, thus, that $y_t \sim CI(2,2)$, with Wold representation given by

(23) $\quad (1-L)^2 y_t = C(L)e_t$,

then, by means of a Taylor expansion, we can write $C(L)$ as

$$C(L) = C(1) - C^*(1)(1-L) + \tilde{C}(L)(1-L)^2,$$

with $C^*(1)$ being the first derivative of $C(L)$ with respect to L, evaluated at $L = 1$. Following the arguments in the previous section, $y_t \sim CI(2,2)$ implies that there exists a set of cointegrating vectors such that $\Gamma' C(1) = \Gamma' C^*(1) = 0$, from which the following *VECM* representation can be derived

(24) $\quad A^*(L)(1-L)^2 y_t = -B_1 \Gamma' y_{t-1} - B_2 \Gamma' \Delta y_{t-1} + e_t$

with $A^*(0)=I_n$. Johansen (1992b) has developed the maximum likelihood estimation of this class of models, which albeit more complicated than in the $CI(1,1)$ case, proceeds along similar lines to those discussed in Section 3.

Likewise, there are systems where the variables have unit roots at the seasonal frequencies. For example, if a seasonally integrated variable is measured every half-a-year, then it will have the following Wold representation

(25) $$\left(1-L^2\right)y_t = C(L)e_t \, .$$

Since $\left(1-L^2\right)=(1-L)(1+L)$, the $\{y_t\}$ process could be cointegrated by obtaining linear combinations which eliminate the unit root at the zero frequency, (1-*L*), and/or at the seasonal frequency, (1+*L*). Assuming that Γ_1 and Γ_2 are sets of cointegrating relationships at each of the two above mentioned frequencies, Hylleberg et al. (1990) have shown that the *VECM* representation of the system this time will be

(26) $$A^*(L)\left(1-L^2\right)y_t = -B_1\Gamma_1^{'}\Delta y_{t-1} - B_2\Gamma_2^{'}\left(y_{t-1}+y_{t-2}\right)+e_t \, ,$$

with $A^*(0)=I_n$. Notice that if there is no cointegration in (1+*L*), $\Gamma_2 = 0$ and the second term in the right hand side of (26) will vanish , whereas lack of cointegration in (1- *L*) implies $\Gamma_1 = 0$ and the first term will disappear. Similar arguments can be used to obtain *VECM* representations for quarterly or monthly data with seasonal difference operators of the form $\left(1-L^4\right)$ and $\left(1-L^{12}\right)$, respectively.

(*ii*) *Fractionally Cointegrated Systems*

As discussed earlier in this chapter, one of the main characteristics of the existence of unit roots in the Wold representation of a time series is that they have “long memory”, in the sense that shocks have permanent effects on the levels of the series so that the variance of the levels of the series explodes. In general, it is known that if the differencing filter $(1-L)^d$, *d* being now a real number, is needed to achieve stationarity, then the coefficient of e_{t-j} in the Wold representation of the *I*(*d*) process has a leading term j^{d-1} (e.g., the coefficient in an *I*(1) process is unity, since *d* = 1) and the process is said to be *fractionally integrated of order d*. In this case, the variance of the series in levels will explode at the rate T^{2d-1} (e.g., at the rate *T* when *d* = 1) and then all that is needed to have this kind of long memory is a degree of differencing $d > 1/2$.

Consequently, it is clear that a wide range of dynamic behavior is ruled out a priori if d is restricted to integer values and that a much broader range of cointegration possibilities are entailed when fractional cases are considered. For example, we could have a pair of series which are $I(d_1)$, $d_1 > 1/2$, which cointegrate to obtain an $I(d_0)$ linear combination such that $0 < d_0 < 1/2$. A further complication arises in this case if the various integration orders are not assumed to be known and need to be estimated for which frequency domain regression methods are normally used. Extensions of least squares and maximum likelihood methods of estimation and testing for cointegration within this more general framework can be found in Jeganathan (1996), Marmol (1998) and Robinson and Marinucci (1998).

(*iii*) *Nearly Cointegrated Systems*

Even when a vector of time series is $I(1)$, the size of the unit root in each of the series could be very different. For example, in terms of the common trend representation of a bivariate system discussed above, it could well be the case that $y_{1t} = f_1 y_t^c + \tilde{w}_{1t}$ and $y_{2t} = f_2 y_t^c + \tilde{w}_{2t}$ are such that f_1 is close to zero and that f_2 is large. Then y_{1t} will not be different from $\tilde{w}_{1t}$ which is an $I(0)$ series while y_{2t} will be clearly $I(1)$. The two series are cointegrated, since they share a common trend. However, if we regress y_{1t} on y_{2t}, i.e., we normalize the cointegrating vector on the coefficient of y_{1t}, the regression will be nearly unbalanced, namely, the regressand is almost $I(0)$ whilst the regressor is $I(1)$. In this case, the estimated coefficient on y_{2t} will converge quickly to zero and the residuals will resemble the properties of y_{1t}, i.e., they will look stationary. Thus, according to the Engle and Granger testing approach, we will often reject the null of no cointegration. By contrast, if we regress y_{2t} on y_{1t}, now the residuals will resemble the $I(1)$ properties of the regressand and we will often reject cointegration. Therefore, normalization plays a crucial role in least squares estimation of cointegrating vectors in nearly cointegrated systems. Consequently, if one uses the static regression approach to estimate the cointegrating vector, it follows from the previous discussion that is better to use the “less integrated” variable as the regressand. Ng and Perron (1997) have shown that these problems remain when the equation are estimated using more efficient methods like *FM*-

OLS and *DOLS*, while the Johansen's methodology provides a better estimation approach, since normalization is only imposed on the length of the eigenvectors.

(*iv*) *Nonlinear Error Correction Models*

When discussing the role of the cointegrating relationship z_t in (3) and (3'), we motivated the *EC* model as the disequilibrium mechanism that leads to the particular equilibrium. However, as a function of an *I*(0) process is generally also *I*(0), an alternative more general *VECM* model has z_{t-1} in (3) and (3') replaced by $g(z_{t-1})$ where $g(z)$ is a function such that $g(0)=0$ and $E[g(z)]$ exists. The function $g(z)$ is such that it can be estimated nonparametrically or by assuming a particular parametric form. For example, one can include $z^{+}=max\{0,z_t\}$ and $z^{-}=min\{0,z_t\}$ separately into the model or large and small values of *z* according to some prespecified threshold in order to deal with possible sign or size asymmetries in the dynamic adjustment. Further examples can be found in Granger and Teräsvirta (1993). The theory of non-linear cointegration models is still fairly incomplete, but nice applications can be found in Gonzalez and Gonzalo (1998) and Balke and Fomby (1997).

(*v*) *Structural Breaks in Cointegrated Systems*

The parameters in the cointegrating regression model (5) may not be constant through time. Gregory and Hansen (1995) developed a test for cointegration allowing for a structural break in the intercept as well as in the slope of model (5). The new regression model now looks like

(27) $$y_{1t}=a_1+a_2D(t_0)+b_1y_{2t}+b_2y_{2t}D(t_0)+z_t,$$

where $D(t_0)$ is a dummy variable such that $D(t_0)=0$ if $0<t\le t_0$ and $D(t_0)=1$ if $t_0<t\le T$. The test for cointegration is conducted by testing for unit roots (for instance, with an *ADF* test) on the residuals $\hat{z}_t$ for each t_0. Gregory and Hansen propose and tabulate the critical values of the test statistic

$$ADF^{*}=\inf_{1<t_0<T}\{ADF(t_0)\}.$$

The null hypothesis of no cointegration and no structural break is rejected if the statistic ADF^{*} is smaller than the corresponding critical value. In this case the structural break will be located at time t^* where the inf of the *ADF* test is obtained. The work of Gregory and Hansen is opening an extensive research on analyzing the stability of the parameters of multivariate possibly cointegrated systems models like the *VECM* in (16). Further work in this direction can be found in Hansen and Johansen (1993), Quintos (1994) and Juhl (1997).

5 CONCLUDING REMARKS

The considerable gap in the past between the economic theorist, who had much to say about equilibrium but relatively less to say about dynamics and the econometrician whose models concentrated on the short-run dynamics disregarding the long-run equilibrium, has been bridged by the concept of cointegration. In addition to allowing the data to determine the short-run dynamics, cointegration suggest that models can be significantly improved by including long-run equilibrium conditions as suggested by economic theory. The generic existence of such long-run relationships, in turn, should be tested using the techniques discussed in this chapter to reduce the risk of finding spurious conclusions.

The literature on cointegration has greatly enhanced the existing methods of dynamic econometric modeling of economic time series and should be consider nowadays as a very valuable part of the practitioner's toolkit.

REFERENCES

Balke, N and T. Fomby (1997), "Threshold cointegration", *International Economic Review* 38, 627-645.

Banerjee, A., J.J. Dolado, J.W. Galbraith and D.F. Hendry, (1993), *Co-integration, error correction and the econometric analysis of non-stationary data*, Oxford University Press, Oxford.

Banerjee, A., J.J. Dolado and R. Mestre, (1997), "Error-correction mechanism tests for cointegration in a single-equation framework", *Journal of Time Series Analysis* 19, 267-284.

Beveridge, S., and C. R. Nelson, (1981), "A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the "Business Cycle"", *Journal of Monetary Economics* 7, 151-74.

Bierens, H. (1999), "Unit roots", chapter ? in this book.

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

Campbell, J.Y., and R. J. Shiller, (1987), “Cointegration and tests of present value models”, *Journal of Political Economy* 95, 1062-88.

Davidson, J.E.H., D. F. Hendry, F. Srba, and S. Yeo, (1978), “Econometric modelling of the aggregate time series relationships between consumer’s expenditure and income in the United Kingdom”, *Economic Journal* 88, 661-92.

Engle, R.F. and C.W.J. Granger (1987), “Co-integration and error correction: Representation, estimation and testing”, *Econometrica* 55, 251-76.

Engle , R.F. and C.W.J. Granger, eds. (1991*), Long-run economic relationships: Readings in cointegration*, Oxford University Press: Oxford.

Engle, R.F. and B.S. Yoo (1987), “Forecasting and testing in cointegrated systems”, *Journal of Econometrics* 35, 143-159.

Ericsson, N.R. (1992), “Cointegration, exogeneity and policy analysis: An overview”, *Journal of Policy Modeling* 14, 424-438.

Ghysels, E. (1999), “Seasonality”, chapter ? in this book.

Gonzalez, M., and J. Gonzalo (1998) “Inference in threshold error correction model”, *Discussion Paper*, Universidad Carlos III de Madrid.

Gonzalo, J., and C. W. J. Granger, (1995), “Estimation of common long memory components in cointegrated systems”, *Journal of Business and Economic Statistics* 13, 27-36.

Gonzalo, J., and J-Y. Pitarakis (1997), “Dimensionality effect in cointegrated systems” to appear in Granger Festschrift edited by R. Engle and H. White. Oxford University Press: Oxford.

Granger, C.W.J., (1981), “Some properties of time series data and their use in econometric model specification”, *Journal of Econometrics* 23, 121-130.

Granger, C.W.J., (1999), “Spurious regressions in econometrics”, chapter ? in this book.

Granger, C.W.J. and T.H. Lee, (1989), “Multicointegration”, *Advances in Econometrics* 8, 71-84.

Granger, C.W.J. and P. Newbold, (1974), “Spurious regressions in econometrics”, *Journal of Econometrics* 2, 111-120.

Granger, C.W.J. and T. Teräsvirta (1993), *Modelling nonlinear economic relationships*, Oxford University Press: Oxford.

Gregory, A.W., and B.E. Hansen (1995), "Residual-based tests for cointegration in models with regime shifts", *Journal of Econometrics* 70, 99-126.

Haldrup, N. (1997), "A review of the econometric analysis of *I*(2) variables", *Working Paper* 97-12, Aarhus University.

Hansen, H., and S. Johansen (1993), "Recursive estimation in cointegrated VAR models", *Preprint* 1, Institute of Mathematical Statistics, University of Copenhagen.

Hatanaka, M. (1996), *Time series-based econometrics*, Oxford University Press: Oxford.

Hendry, D. F., and G. E. Mizon (1978), "Serial correlation as a convenient simplification not a nuisance: A comment on a study of the demand for money by the Bank of England", *Economic Journal* 88, 549-63**.**

Hendry, D. F., and J.-F. Richard (1983), "The econometric analysis of economic time series" (with discussant), *International Statistical Review* 51, 111-63.

Hylleberg, S., R.F. Engle, C.W.J. Granger and B.S. Yoo (1990), "Seasonal integration and cointegration", *Journal of Econometrics* 44, 215-228.

Ho, M. and B. Sorensen (1996), "Finding cointegration rank in high dimensional systems using the Johansen test: An illustration using data based on Monte Carlo simulations", *Review of Economics and Statistics* 78, 726-732.

Jeganathan, P. (1996), "On asymptotic inference in cointegrated time series with fractionally integrated errors", *Discussion Paper*, University of Michigan.

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

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

Johansen, S., (1995), *Likelihood-based inference in cointegrated vector auto-regressive models,* Oxford University Press: Oxford.

Juhl, T., (1997), "Likelihood ratio tests for cointegration in the presence of multiple breaks", *Discussion Paper*, University of Pennsylvania.

Lütkepohl, H., (1999), "Vector autoregressions", chapter ? in this book.

MacKinnon, J.G., (1991), “Critical values for cointegration tests”, in R.F. Engle and C.W.J. Granger, eds., Long-run economic relationships: Readings in cointegration Oxford University Press: Oxford.

Marmol, F. (1998), “Spurious regression theory with nonstationary fractionally integrated processes”, *Journal of Econometrics* 84, 232-250.

Nelson, C.R., and C. I. Plosser, (1982), “Trends and random walks in macroeconomic time series: Some evidence and implications”, *Journal of Monetary Economics* 10, 139-62.

Ng, S., and P. Perron, (1997), “Estimation and inference in nearly unbalanced nearly cointegrated systems”, *Journal of Econometrics* 79, 53-81.

Osterwald-Lenum, M., (1992), “A note with quantiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics”, *Oxford Bulletin of Economics and Statistics* 54, 461-72.

Phillips, P.C.B. (1987), “Time series regression with a unit root”, *Econometrica* 55, 277-301.

Phillips, P.C.B. (1991), “Optimal inference in cointegrated systems”, *Econometrica* 59, 283-306.

Phillips, P.C.B., and B. E. Hansen, (1990), “Statistical inference in instrumental variables regression with *I*(1) processes”, *Review of Economic Studies* 57, 99-125.

Phillips, P.C.B, and S. Ouliaris, (1990), “Asymptotic properties of residual based tests for cointegration”, *Econometrica* 58, 165-94.

Quintos, C. (1994), “Rank constancy tests in cointegrating regressions”, *Discussion Paper*, Yale University.

Reinsel, G.C., and S. K. Ahn, (1992), “Vector autoregressive models with unit roots and reduced rank structure: Estimation, likelihood ratio test, and forecasting”, *Journal of Time Series Analysis* 13, 353-375.

Robinson, P., and D. Marinucci (1998), “Semiparametric frequency domain analysis of fractional cointegration”, *Discussion Paper*, London School of Economics.

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

Saikkonen, P. (1992), “Estimation and testing of cointegrated systems by an autoregressive approximation”, *Econometric Theory* 8, 1-27.

Sargan, J. D. (1964), “Wages and prices in the United Kingdom: A study in Econometric methodology”, repr. in D. F. Hendry and K. F. Wallis (ed), *Econometrics and Quantitative Economics*, Blackwell: Oxford.

Stock, J.H. (1987), “Asymptotic properties of least squares estimators of cointegrating vectors”, *Econometrica* 55, 277-302.

Stock, J.H., and M. W. Watson, (1988), “Testing for common trends”, *Journal of the American Statistical Association* 83, 1097-107.

Stock, J.H., and M. W. Watson, (1993), “A simple estimator of cointegrating vectors in higher order integrated systems”, *Econometrica* 61, 783-820.

Watson, M.W. (1994), “Vector autoregressions and cointegration”, in Engle, R.F. and D.L. McFadden, eds., *Handbook of Econometrics* IV, Elsevier: New York.

Wold, H. (1938), *A study in the analysis of stationary time series,* Almqvist and Wiksell Stockholm.