LWATI: A Journal of Contemporary Research, 8(1), 194-213, 2011
ISSN: 1813-2227

# Cointegration and Econometric Analysis of Non-Stationary Data in Nigeria: An Empirical Evidence

**Iyeli I. Iyeli, Anthony Imoisi and Aham Kelvin Uko**
*Department of Economics, University of Port Harcourt , Nigeria*

## ABSTRACT

This study discussed the concept of cointegration and the econometric analysis of non-stationary time- series in the context of Nigeria data. The outcome reveals a high degree of non-stationarity of Nigeria data and also show along –run relationship between corporation income tax and the variables identified as its determinants, namely, market capitalization and value of stock market transactions. This is in conformity with the philosophy underlying the cointegration theory. Therefore, ignoring cointegration in non-stationary time series variables could lead to misspecification of the underlying process in the determination of corporate income tax in Nigeria. Thus, the study conclude that cointegration is greatly enhanced the existing dynamic econometric modeling of economic time series and should be considers nowadays as a very valuable part of the applied econometrician tool kit in analyzing economic problems for effective decision making.

## INTRODUCTION

Greater portion of economic theory deals with long-run equilibrium relationships generated by market forces. In the same vain, most empirical econometric studies using time series can be interpreted as attempts to emphasis such relationships in a dynamic framework. Conventional wisdom suggest 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. In line with this argument econometricians proceeded as if stationarity could be achieved by simply removing deterministic components of 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 economic theory even after removing those deterministic terms. Those problem were somehow disregarded in applied econometric work until Granger and Newbold (1974) and Nelson and Plosser (1982) call the attention of many to the econometric implications of non-stationarity and the dangers of running spurious regressions. In particular, most of the attention focused 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 integration process being permanent rather than transitory.

The presence of at least one or more unit roots 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 stock prices, exchange rates, money velocity, interest rate, corporate income tax, etc and, perhaps the most popular, the implications of the permanent income hypothesis for real consumption under rational expectations.

Following the influential approach by Box and Jenkins (1970), statisticians had advocated transforming integrated time series into stationary ones by successive differencing of the series before modelization. Therefore, in their perspective removing unit roots through differencing ought to be a pre-requisite for regression analysis. However, Sargan (1964), Hendry and Mizon (1978) and Davidson et al. (1978), among others, 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 deviation 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), relying on the previous ideas, clarified the confusion 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 the variety of economic models including the relationships between interest rate policy and domestic investment capital and output, real wages and labour productivity, nominal exchange rates and relative prices, consumption and disposable income, money velocity, etc.

In view of the strength of these ideas, substantial literature on cointegration and its application to solving economic problems has developed over the last decade. In this section we will explore the basic conceptual issues and discuss related economietric techniques, there in. section 2 provides some implications of cointegration and the basic estimationand testing procedures in a multivariate equation framework.

In Section 3, we present an empirical evidence using Nigerian data. Finally, Section 4 draws some concluding remarks.

**Unit Roots and Cointegration**

World (1938) in his decomposition theorem 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*(I) series, while accumulating a random walk gives rise to an *I(2)* series, etc.

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

$z_t = x_{1t} - \alpha - \beta_{x_{2t}}$ ------------------------------------------------------------------------------(1)

Is *I(d – b), d ≥ b > 0,* then, following Engle and Granger (1987), $x_{1t}$ and $x_{2t}$ are defined as cointegrated of order *(d,b)* $x_t = (x_{1t}, x_{2t})' \sim CI(d,b)$, with $(1,\beta)'$ call 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 integrated variables, the econometric practice about this more general type of cointegration is less developed. *Second,* note that the cointegrating vector is not uniquely defined, since for any nonzero value of Ψ, (Ψ,- Ψβ)' is also a cointegrating vector. Thus, a normalization rule needs to be used; for example Ψ-1 has been chosen in (1). *Third,* all variable must be integrated of the same order to be qualified 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 (Granger and Lee, 1989). For example, in a trivariate system, we may have that $x_{1t}$ and $x_{2t}$ are *I(2)* and $x_{3t}$ is *I(1);* if $x_{1t}$ and $x_{2t}$ are CI(2,1), it is possible that the corresponding combination of $x_{1t}$ and $x_{2t}$ which achieves that property be itself cointegrated with $x_{3t}$ giving rise to an *I(o)* 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,. Hence, the remainder of this section will mainly focus on the case of CI (I,I) variable, so that $z_t$ in (1) is *I(0)* and the concept of cointegraion 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 $x_{1t}$ and $x_{2t}$,

$x_{1t} = \alpha + \beta x_{2t}$’-------------------------------------------------------------------(2)

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 example, a standard theory of spatial competition argues that arbitrage will prevent prices of similar products in different location from moving too far apart 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 (I).

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

$x_{1t} = \mu_{1t} + u_{1t}$ -------------------------------------------------------------------(3)

$x_{2t} = \mu_{2t} + u_{2t}$ ------------------------------------------------------------------ (4)

Since the sum of an I(I) process and an I()) process is always I(I), the previous representation must characterize the individual stochastic properties of $x_{1t}$ and $x_{2t}$. However, if $x_{1t}$ and $\beta x_{2t}$ is I(0), it must be that $\mu_{1t} = \beta\mu_{2t}$, annihilating the I(I) component in the cointegrating relationship. In other words, if $x_{1t}$ and $x_{2t}$ are CI(1,1) variables, they must share (up to a scalar) the same stochastic trends, $\mu_t$ denoted as common trend, so that $\mu_{1t} = \mu_t$ and $\mu_{2t} = \beta\mu_t$. As before, notice that if $\mu_t$ is a common trend for $x_{1t}$ and $x_{2t}$, $\Psi\mu_t$, will 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 stochaistic trend and that r = n implies that the I(0) instead of I(I). 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 test for the number of common trend (n – r).

Engle and Granger (1987) have shown that if $x_{1t}$ and $x_{2t}$ are cointegrated CI(I,I), ten there must exist a so-called vector error correction model (VECM) representation of the dynamic system governing the join behaviour of $x_{1t}$ and $x_{2t}$ over time of the following form:

$$\Delta x_{1t} = \theta_{10} + \theta_{11} z_{t-1} + \sum_{i=1}^{p1} \theta_{12,i}\, \Delta x_{1t-i} + \sum_{i=1}^{p2} \theta\Delta_{13,i}\, \Delta x_{2,t-i} + \varepsilon_{1t}{}' \text{-----------------} (5)$$

$$\Delta x_{1t} = \theta_{20} + \theta_{21} z_{t-1} + \sum_{i=1}^{p3} \theta_{22,i} + \sum_{i=1}^{p4} \theta\Delta_{23,i}\, \Delta x_{2,t-i} + \varepsilon_{2t}' \text{-----------------} (6)$$

Where $\Delta$ denotes the first-order time difference (i.e., $\Delta x_t = x_t - x_{t-1}$) ad where the lag lengths *L<sub></sub>*$L_{i,}I$ *= 1, ..., 4* are such that the innovations $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})'$ 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 $z_{t-1}$ in (the so-called loading or speed of adjustment parameters) are not simultaneously equal to zero.

Note that the term $z_{t-1}$ in equation (5) and (6) represents the extent of the disequilibrium levels of $x_1$ *and* $x_2$ 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 $x_1$ *and* $x_2$. For example, if $\beta$ =1 in (1), as many theories predict when $x_{1t}$ *and* $x_{2t}$ are taken in logarithmic form, then if $x_1$ *is larger than* $x_2$ in the past $(z_{t-1} > 0)$, then $\theta_{11} < 0$ *and* $\theta_{21} > 0$ 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 $\theta_{11}$ *and* $\theta_{21}$ cannot be equal to zero. However, if $\theta_{11} < 0$ *and* $\theta_{21} = 0$, then all of the adjustment falls on $x_1$, or vice versa if $\theta_{11} < 0$ *and* $\theta_{21} > 0$. 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., wit $\theta_{11} < 0$ *and* $\theta_{21} = 0$), 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 $x_t$ will have infinite variable; (Engle and Yoo (1987).

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 $x_t \sim I(1)$, then the procedure goes as follows:

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

$$x_{1t} = \alpha + \beta x_{2t} + z_t \text{ ----------------------------------------------------------------------(7)}$$

is estimated by ordinary least squares (OLS) and it is tested whether the cointegrating residuals $\hat{z}_t = x_{1t} - \hat{\alpha} - \hat{\beta} x_{2t}$ are $I(1)$. To do this, we performa 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.

$$\Delta\hat{z}_t = \lambda_1 \hat{z}_{t-1} + \varepsilon_t \text{ ---------------------------------------------------------------------(8)}$$

where no intercept term has been included since then $\{\hat{z}_t\}$, being residuals from a regression equation with a constant term, have zero mean. If we can reject the null hypothesis that $\lambda_1 = 0$ against the alternative $\lambda_1 < 0$ at a given significance level, we can conclude that the residual sequence is I(0) and, therefore, that $x_{1t}$ and $x_{2t}$ are CI (1,1). It is noteworthy tat for carrying out this test it is not possible to use the Dickey-Fuller table themselves since $\{\hat{z}_t\}$ are a generated series of residuals from fitting regression (7). The problem is that the OLS estimates of $\alpha$ and $\beta$ are such that they minimize the residual variance in (7) 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 $\rho_1 = 0$ for any sample size and also when the number of regressors in (7) is expanded from one to several variables. In general, if the $\{\hat{z}_t\}$ sequence exhibits serial correlation, then an augmented Dickey-Fuller (ADF) test should be used, based this time on the extended autoregression.

$$\Delta\hat{z}_t = \lambda_1 \hat{z}_{t-1} + \sum_{i=1}^{p} \xi_i \Delta\hat{z}_{t-i} + \varepsilon_t \text{ -----------------------------(9)}$$

where again, if $\lambda_1 < 0$, we can conclude that $y_{1t}$ and $y_{2t}$ are CI(1,1). Alternative version s of the test on $\{\hat{z}_t\}$ being I(I) versus I(0) can be found in Philips and Ouliaris (1990). Banerje et al. (1997), in turn, suggest another class of tests based this time on the direct significance of the loading parameters in (5) and (6) where the β 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 β in equation (7) is super-consistent, in the sense that the OLS estimator $\hat{\beta}$ converges in probability to its true value β at a rate proportional to the inverse of the sample size, $T^{-1}$, rather that at $T^{-1/2}$ as is the standard result in the ordinary case where $x_{1t}$ and $x_{2t}$ 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 behaviour of $\hat{\beta}$ in (7) (where the constant is omitted for simplicity) in the particular case where $z_t \sim i.i.d.(0, \sigma_z^2)$, and that $\theta_{20} = \theta_{21} = 0$ and $L_3 = L_4 = 0$, so that $x_{2t}$ is assumed to follow a simple random walk

$\Delta y_{2t} = \varepsilon_{2t}$' ---------------------------------------------------------------------------(10)

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

$$x_{2t} = \sum_{i=1}^{t} \varepsilon_{2i} \text{ ----------------------------------------(11)}$$

with $\varepsilon_{2i}$ possible correlated with $z_t$. In this case, we get var $(x_{2t})$ = t var $(\varepsilon_{2t}) = t\sigma_2^2$ exploding as $T \uparrow \infty$. Nevertheless, it is not difficult to show that $T^{-2}\sum_{t=1}^{T} x_{2t}^2$ converges to a random variable. Similarly, the cross-product $T^{-1/2}\sum_{t=1}^{T} x_{2t}\varepsilon_t$ 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(I) case, $T^{-1}\sum_{t=1}^{T} x_{2t}\varepsilon_t$ converges also to a random variable. Both random variables are functionals of Brownian motions which will be donated 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 Philips 1987).

Now, from the expression for the OLS estimator of $\beta$, we obtain

$$\hat{\beta} - \beta = \frac{T^{-1}\sum_{t=1}^{T} x_{2t}z_t}{\sum x_{2t}z_t} \text{ ---------------------------------- (12)}$$

and, from the previous discussion, it follows that

$$T(\hat{\beta} - \beta) = \frac{T^{-2}\sum^{T} x_t^2}{T^{-2}\sum^{T} x_t^2} \text{ ----------------------------------------------(13)}$$

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

(ii) After rejecting the null hypothesis that the cointegrating residuals in equation (7) are I(1), the $\hat{z}_{t-1}$ term is included in the VECM system and the remaining parameter are estimated by OLS. Indeed, given the superconsistency of $\hat{\beta}$, Engle and Granger (1987) show that their asymptotic distributions will be identical to using the true value of β. Now all the variables in (3) and (4), are *I(0)* and conventional modeling strategies (e.g., testing the maximum lag length, residual autocorrelation or whether either

$\theta_{11}$ or $\theta_{12}$ is zero, etc) can be applied to assess model adequacy; (see Lütkepohl, 1999).

In spite of the beauty and simplicity of the procedure, however, several problems remain. In particular, although $\hat{\beta}$ is supper-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 $\Delta x_{2t}$ are highly serially correlated and they are not independent. Phillips (1991), in turn, has shown analytically that in the case where $x_{2t}$ and $z_t$ are independent at all leads and lags, the distribution in (3) as T grows behaves like a Gaussian distribution (technical is a *mixture of normal's*) and, hence, the distribution of the t-statistic of β 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 (12) by which the error term $z_t$ is conditioned on the whole process $\{\Delta x_{2t}, t = 0, \pm 1, \dots\}$ and, hence, othogonality between regressors and disturbance is achieved by construction. For example, if $z_t$ and $\varepsilon_{2t}$ in (7) and (10) are correlated white noises with $\gamma = E(z_t \varepsilon_{2t})/var(\varepsilon_{2t})$, the FM-OLS estimator of β, denoted $\hat{\beta}_{FM}$, is given by

$$\hat{\beta}_{FM} = \sum_{t=1}^{T} x_t^2 \quad \sum^{T} x_{2t}(x_{1t} - \hat{\gamma}\Delta x_{2t}) \qquad \text{------(14)}$$

where $\hat{\gamma}$ is the empirical counterpart of $\gamma$ obtained from regressing the OLS residuals $\hat{z}_t$ on $\Delta y_{2t}$. When $z_t$ *and* $\Delta y_{2t}$ follow more general process, the FM-OLS estimator of β is similar to (14) except that further corrections are needed in its numerator. Alternatively, Saikkonen (1991) and Stock and Watson (1993) have shown that, since $E(z_t | \{\Delta x_{2t}\}) = h(L)$ is a two-sided filter in the lag operator L, regression of $x_{1t}$ on $x_{2t}$ and leads and lags of $\Delta x_{2t}$ (suitably truncated), using either OLS or GLS, will yield an estimator of β which is asymptomatically equivalent to the FM-OLS estimator. The resulting estimation approach is known as dynamic OLS or, DGLS.

In the beginning of this section we confined the analysis to the case where there is at most a single cointegrating vector in a bivaiate 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 aggregate corporate income tax, market capitalization and value of stock transaction, there may exist two relationships, one determining corporate income tax equation and another determining capital market equation. In furtherance to this, we comment on some of the popular estimation and testing procedures for cointegration in this general multivariate context which will be denoted as system-based approaches has presented in the table below.

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

$\Delta x_t = C(L)\varepsilon_t$ -------------------------------------------------------------------(15)

where now $\varepsilon_t \sim nid(0, \Sigma)$, $\Sigma$ being the covariace matrix of $\varepsilon_t$ and C(L) an *( n x n)* invertible matix of polynomial lags, where the term "invetible" means that $|C(L) = 0|$ has all its strictly larger than unity in absolute value. If there is a cointegrating *(n x 1)* vector, $\beta' = (\beta_{11}, \ldots, \beta_{nn})$, then, premultiplying (11) $\beta'$ yields

$\beta'\Delta x_t = \beta'[C(1) + C(L)\Delta]\varepsilon_t$,---------------------------------------------------(16)

where C(L) has been expanded around L = 1 using a first-order Taylor expansion and $C(L)$ can be shown to be an invertible lag matrix. Since the cointegration property implies that $\beta' x_t$ is *I(0)*, then it must be that $\beta' C(I)$ =0 and hence Δ (= 1 – L) will cancel out on other sides of (16). Moreover, given that C(L) is invertible, then $y_t$ has a vector autoregressive representation such that

$A(L)x_t = \varepsilon_t'$ --------------------------------------------------------------------- (17)

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 $\beta$, namely, $A(1) = \alpha\beta'$, with $\propto$ 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 $\Gamma$ are this time $(n \times r)$ matrices which collect the $r$ and $\beta$ vectors. Matrix $B$ is known as the *loading matrix* since its rows determine how man cointegrating relationships enter each of the individual dynamic equations in (17). 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. Fist, 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 tat $y_t$ has a $VAR(1)$ representation, that is, $A(L)$ IN (17) is such that $A(L) = I_n - A_1 L$. Hence, the $VAR(1)$ process can be reparameterized in the $VECM$ representation as

$\Delta X_t = (A_1 - I_n)X_{t-1} + \varepsilon_t$,-----------------------------------------------------(18)

If $A_1 - I_n - A(1) = 0$, then $X_t$ is $I(1)$ and there are no cointegrating relationship $(r = 0)$, whereas if 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 $X_t$ follows a VAR(ρ) process of the form

$X_t = A_1X_{t-1} + A_2X_{t-2} + \cdots + A_pX_{t-p} + \varepsilon_t$ -------------------------------(19)

Which, as in (3) and (4), can rewritten in the ECM representation

$\Delta X_t = D_1\Delta X_{t-1} + D_2\Delta X_{t-2} + \cdots + D_p\Delta X_{t-p+1} + DX_{t-1}\varepsilon_t$ ------------ (20)

Where $D_i = -(A_{i+1} + \cdots + A_p), i = 1,2, \ldots, L-1,$ *and*

$D = (A_1 + \cdots + A_pI_n) = A(1)$

In that respect, the Johansen's approach allows to test restrictions on $\mu$, B and $\Gamma$ 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 coiintegrating vectors does not diminish if the restrictions are not binding and vice versa. Thus, denoting by $\hat{\Psi}$ and $\Psi_i^*$ 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.

$$T\sum_{i=1}^{r}\left[\ln(1-\Psi_i^*) - \ln(1-\hat{\Psi}_i)\right] \text{-----------------------------}(24)$$

will be small if the $\hat{\Psi}_i^*$'s are similar to the $\hat{\Psi}_i$ 's whereas it be large if the $\hat{\Psi}_i^*$'s are smaller than the $\hat{\Psi}_i$'s. if e impose $s$ restrictions, then the above test will reject the null hypothesis if the calculated value of (24) exceeds that in a chi-square table with $r(n - s)$ degree 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. (Reinsel and Ahn, 1992), researchers ought to have considerable care when using the Johansen estimator to determine cointegration order in high dimensional system 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 erogeneity 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 estamtion 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 World representation of an $I(1)$ system, which we can write as in expression (15) with $C(L) = \sum_{j=0}^{\infty} C_j = L^j, C_0 I_n$. As shown in expression (16), $C(L)$ can be expanded as $C(L) = C(1) + C(L)(1-L)$, so that, by integrating (15),

we get $X_t = C(1)X_t + \tilde{v}_t$, ------------------------------------------------------(25)

where $\tilde{v}_t = C(L)\varepsilon_t$ can be shown to be covariance stationary, and $X = \sum_{i=1}^{t} \varepsilon_i$ is a latent or unobservable set of random walks which capture the $I(1)$ nature of the data. However, as mentioned above I the cointegration order is $r$, there must be an $(r \times n)\,\Gamma$ matrix such that $\Gamma' C(1) = 0$ since, otherwise, $\Gamma' X_t$ 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 $C(1)(n \times r)$, implying that there are only $(n \times r)$ independent common trends in the system. Hence, there exists the so-called *common trends representation* of a cointegrated system, such that

$X_t = \Phi X_t^c + \tilde{v}_t$,------------------------------------------------------------------(26)

Where $\Phi$ is an $n \times (n-r)$ matrix of loading coefficients such that $\Gamma'\Phi = 0$ and $X_t^c$ is an $(n-r)$ vector random walk. In other words, $X_t$ can be written as the sum of $(n-r)$ common trends and an $I(0)$ component. Thus, testing for $(n-r)$ a common trend 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 authoregressive matrix of $X_t^c$ 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 unit. It is worth noticing that there are other alternative strategies to identify the set of common trends, $X_t^c$, 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 $X_t^c$ as linear combinations of $X_t$ which are not caused in the long-run by the cointegration relationship $\Gamma' X_{t-1}$, these linear combinations are the orthogonal complement of matrix $B$ in (20), $X_t^c = B_\perp X_t'$ where $B_\perp$ is an $(n \times (n-r))$ 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 $S_{00}$. Or 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 no 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 stations that consumption follows a random walk whilst saving (disposable income minus consumption) is $I(0)$. Hus, if the theory is a valid one, the cointegrating vector in the system formed by consumption and disposable income should be $\beta' = (1, -1)$ and it would only appears in the second equation (i.e $\alpha' = (0, \alpha_{12})$), implying that consumption should be the common trend behind the nonstationary behaviour of both variables……….. 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 (20). Further work in this direction can be found in Hansen and Johansen (1993), Quintos (1994) and Juhl (1997).

## Empirical Evidence from Nigerian Data

### Results of Stationary Tests

Taking the Nigerian data into consideration, we present our estimated result of the model for the determinants of corporate income tax to enable us evaluate the results and then X-ray the findings in the context of cointegration theory in other for us to determine the extent of the non-stationarity of the Nigerian data and its policy implications. The results of the unit root tests show that the three time series are non-stationary in levels (or log levels) but are stationary in first differences (or log differences). This is deduced from the fact that for levels of the variables, the absolute value of the, ADF test statistics are less than the absolute value of the critical values of the ADF at 1%, 5% and 10% significance levels respectively. However for the first differences of the variables the ADF test statistic of each is greater than the 5% and 10% critical value of the ADF statistic in absolute values.

**Table 1:** Test for Unit Root

(A) At ordinary levels

| Variables | ADF | Order of Integration | Decision Rules |
|---|---|---|---|
| LN (CIT) | -2.208534 | 1(0) | Non stationary |
| LN (VSM) | -1.347236 | 1(0) | Non stationary |
| LN (MKCAP) | 1.801015 | 1(0) | Non stationary |

Rejection of Null hypothesis of unit root at
1% = *
5% = **
10% = ****
The null hypothesis are all accepted at both 1, 5 & 10% levels of significance

**Source**: Authors computation

(B) At first differences

| Variables | ADF | Order of Integration | Decision Rule |
|---|---|---|---|
| ΔLN (CIT) | -3.200652 ** (***) | 1(1) | Stationary |
| ΔLN(VSM) | -5.770307 * (** (***) | 1(1) | Stationary |
| ΔLN (MKCAP) | -3.331303 ** (***) | 1(1) | Stationary |

Rejection of Null hypothesis of unit root at
* = 1%
** = 5%
*** = 10%

**Source**: Authors computation

Therefore, the ADF test presented above justify the test for cointegration in the equation. The presence of cointegration makes it possible to estimate an error correction mechanism (ECM), which is a solution to the problem of spurious result associated with estimating equations involving time series variables. Based on the revelation of the presence of cointegration we reject the null hypothesis of no cointegration and conclude that the variables are cointegrated. The Johansen (1991) cointegration result presented in table (2) bellow was carried out on the residuals of the static cointegrating regression involving the non-stationary variable identified as the determinants of corporate income tax. The test results indicate that there are at most 3 cointegrating vectors at 5% levels of significance. Following the decision rule for the Johansen cointegration tests, we reject the null hypothesis of no cointegration in the variables at this level of significance. Since there are at most 3 cointegrating vectors and 3 linear combinations of the variables that are stationary then all other linear combinations are non –stationary. The results therefore suggest the presence of cointegration in the time series variables implying that the normalized cointegrating coefficient gives the long-run relationship in the variables. This is given by the solved static long-run equation in table 2 above. This solved static long-run Corporate Income Tax (CIT) equation agrees with the ECM long-run run solution because the variables and its coefficients are the same in the equation. From our cointegrating test, it was seen that there are 3 cointegrating relationship and following Johansen (1988) and Johnnsen & Juelius (1990, 1992) representation theorem, the residual from the cointegrating regression is a valid error correction and the speed of adjustment, ρ in the equation is determined as the co-efficient of the error correction variable. If ρ is statistically equal to zero, the change in corporate Income tax does not respond at all to the deviation from the long-run equilibrium in $\text{period}_{t-1}$ (Taylor and Sarno, 1997).

## Results from Cointegration and Error Correction Models

Following our findings on the unit root test above that all the variables are integrated of order one 1(1) (or non-stationary), we therefore test for possible cointegration or non-stationarity among the variables. The result is presented in table 2 below. Adopting the general to specific framework, as discussed in the literature, we proceeded to estimate an over-parameterized error correction model of the determinant of corporate income tax from where a parsimonious (preferred) error correction model would be obtained. The attractiveness of the ECM is that it provides a framework for establishing links between the short-run and long run approaches to econometric modeling. Thus with the ECM, no information associated with variable first differencing is lost because the modeling technique incorporates both the short-run dynamics (i.e the first differences) and the long run information through the error correction term.

**Table 2:** Johansen Integration Tests Results.

Test assumption: linear deterministic trend in the data

Series: D(LOG(CIT),2) D(LOG(VSM),2)D(LOG(MKCAP),2)

Lags interval: 1 to 1

| Eigenvalue | Likelihood Ratio | 5 percent Critical Value | 1 Percent Critical Value | Hypothesized No. of CE(s) |
|---|---|---|---|---|
| 0.931974 | 206.0351 | 29.68 | 35.65 | None * |
| 0.863020 | 117.3357 | 15.41 | 20.04 | At most 1** |
| 0.791478 | 51.73439 | 3.76 | 6.65 | At most 2** |

* (**) denotes rejection of the hypothesis at 5% (1%) significance level

L.R. test indicates 3 cointegrating equation(s) at 5% significance level

CIT = -0.02 – 0.30D (LOG(VSM)) + 0.51D(LOG(MKCAP)

SEE = (0.04) (0.13)

ECM = CIT + 0.30 D (LOG(VM) – 0.51D (LOG(MKCAP) + 0.02

**Source:** Authors' computation.

**Table 3: Overparameterized Error Correction Model Result**

Dependent Variable D(LOG(CIT))

Method: Least Squares

Sample (adjusted): 1979 2007

Included observation: 29 after adjusting endpoints

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 1.482935 | 0.446505 | 3.3321204 | 0.1862 |
| D(LOG(CIT(-1))) | -1917052 | 1.011150 | -1.895912 | 0.3090 |
| D(LOG(CIT(-2))) | 2.544865 | 1.455205 | 1.748802 | 0.3307 |
| D(LOG(CIT(-3))) | 0.659098 | 0.651703 | 1.011347 | 0.4964 |
| D(LOG(CIT(-4))) | -2.154549 | 1.048301 | -2.055278 | 0.2883 |
| D(LOG(CIT(-5))) | -2.928627 | 1.306852 | -2.240979 | 0.2672 |
| D(LOG(CIT(-6))) | 0.550164 | 0.844774 | 0.651257 | 0.6325 |
| D(LOG(CIT(-7))) | -0.722709 | 1.145259 | -0.631044 | 0.6416 |
| D(LOG(CIT(-8))) | -1.935066 | 0.778936 | -2.484241 | 0.2436 |
| D(LOG(VSM)) | -0.036488 | 0.270592 | -0.134846 | 0.9147 |
| D(LOG(VSM(-1))) | -0.713130 | 0.235272 | -3.032375 | 0.2028 |
| D(LOG(VSM(-2))) | -1.170589 | 0.475373 | -2.462465 | 0.2456 |
| D(LOG(VSM(-3))) | -0.372936 | 0.223877 | -1.665810 | 0.3442 |
| D(LOG(VSM(-4))) | -0.445566 | 0.343046 | -1.298852 | 0.4177 |
| D(LOG(VSM(-5))) | 0.730112 | 0.334581 | 2.1821168 | 0.2736 |
| D(LOG(VSM(-6))) | 0.629595 | 0.521879 | 1.206398 | 0.4406 |
| D(LOG(VSM(-7))) | 0.236526 | 0.417841 | 0.566066 | 0.6721 |
| D(LOG(VSM(-8))) | -0.799015 | 0.649376 | -1.230435 | 0.4345 |
| D(LOG(MKCAP)) | 2.065457 | 1.541821 | 1.339622 | 0.4082 |
| D(LOG(MKCAP(-1)) | -1.079449 | 0.414267 | -2.605684 | 0.2333 |
| D(LOG(MKCAP(-2))) | 5.357992 | 1.882877 | 2.845641 | 0.2151 |
| D(LOG(MKCAP(-3))) | 2.108849 | 1.528389 | 1.379785 | 0.3993 |
| D(LOG(MKCAP(-4))) | 5.698041 | 2.144568 | 2.65964 | 0.2292 |
| D(LOG(MKCAP(-5))) | 3.228460 | 1.580172 | 2.043107 | 0.2898 |
| D(LOG(MKCAP(-6))) | -0.435148 | 0.902916 | -0.481936 | 0.7141 |
| D(LOG(MKCAP(-7))) | -0.888087 | 1.075159 | -0.826005 | 0.5605 |
| D(LOG(MKCAP(-8))) | 2.363000 | 2.150089 | 1.099024 | 0.4700 |
| ECM(-1) | -0754984 | 0.403659 | -1.870351 | 0.3126 |
| R-squared | 0.984622 | Mean dependent var | | 0.101369 |
| Adjusted R-squared | 0.569426 | S.D dependent var | | 0.219055 |
| S.E of regression | 0.143740 | Akaike info criterion | | -2.477883 |
| Sum of squared resid | 0.020661 | Schwarz criterion | | -1.157735 |
| Log likelihood | 63.92930 | F-statistic | | 2.371462 |
| Durbin-Watson stat | 3,301777 | Prob (F-statistic) | | 0.478414 |

**Source**: Author's Computation 2009.

**Table 4: Parsimonious Error Correction Model Result**

Dependent Variable D(LOG(CIT))
Method: Least Squares
Sample (adjusted): 1979 2007
Included observation: 29 after adjusting endpoints

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 1.396254 | 0.186949 | 7.468645 | 0.0017 |
| D(LOG(CIT(-1))) | -2.182790 | 0.428826 | -5.090150 | 0.0070 |
| D(LOG(CIT(-2))) | 3.111721 | 0.544682 | 5.712908 | 0.0046 |
| D(LOG(CIT(-3))) | 0.405551 | 0.145569 | 2.785980 | 0.0495 |
| D(LOG(CIT(-4))) | -2.599082 | 0.047257 | -6.381919 | 0.0031 |
| D(LOG(CIT(-5))) | -3.117954 | 0.559287 | -5.574876 | 0.0051 |
| D(LOG(CIT(-6))) | 1.027257 | 0.211089 | 4.866465 | 0.0082 |
| D(LOG(CIT(-8))) | -7.733443 | 0.270556 | -6.406963 | 0.0030 |
| D(LOG(VSM)) | -0.190096 | 0.039995 | -4.752957 | 0.0090 |
| D(LOG(VSM(-1))) | -0.673377 | 0.093995 | -7.164001 | 0.0020 |
| D(LOG(VSM(-2))) | -1.193267 | 0.167918 | -7.106241 | 0.0021 |
| D(LOG(VSM(-3))) | -0.309789 | 0.059075 | -5.244014 | 0.0063 |
| D(LOG(VSM(-4))) | -0.561731 | 0.081364 | -6.903942 | 0.0023 |
| D(LOG(VSM(-5))) | 0.585121 | 0.11321 | 4.945180 | 0.0078 |
| D(LOG(VSM(-6))) | 0.345155 | 0.079380 | 4.348110 | 0.0122 |
| D(LOG(VSM(-8))) | -1.129662 | 0.188154 | -6.003914 | 0.0039 |
| D(LOG(MKCAP)) | 2.652045 | 0.585785 | 4.527333 | 0.0106 |
| D(LOG(MKCAP(-1)) | -1.024841 | 0.150633 | -6.803557 | 0.0024 |
| D(LOG(MKCAP(-2))) | 5.159129 | 0.784427 | 6.576943 | 0.0028 |
| D(LOG(MKCAP(-3))) | 1.244508 | 0.297988 | 4.176365 | 0.0140 |
| D(LOG(MKCAP(-4))) | 5.388727 | 0.780010 | 6.908537 | 0.0023 |
| D(LOG(MKCAP(-5))) | 2.689161 | 0.380213 | 7.072779 | 0.0021 |
| D(LOG(MKCAP(-7))) | -1.379423 | 0.370887 | -3.719255 | 0.0205 |
| D(LOG(MKCAP(-8))) | 3.397042 | 0.584581 | 5.811068 | 0.0044 |
| ECM(-1) | -0.792024 | 0.147420 | -5.372555 | 0.0058 |
| R-squared | 0.978420 | Mean dependent var | | 0.101369 |
| Adjusted R-squared | 0.848940 | S.D dependent var | | 0.219055 |
| S.E of regression | 0.085139 | Akaike info criterion | | -2.345925 |
| Sum of squared resid | 0.028995 | Schwarz criterion | | -1.167221 |
| Log likelihood | 59.01591 | F-statistic | | 7.556512 |
| Durbin-Watson stat | 3.309724 | Prob (F-statistic) | | 0.031018 |

**Source:** Authors Computation, 2009.

As earlier on mentioned, the presence of cointegration makes it possible to estimate an error correction model, which is a solution to the problem of spurious result often associated with estimating equations involving time series variables. Therefore based on the result in table 3 above, we proceeded to re-estimate the over-parameterized model, adopting the general to specific approach and the summary of the result showing the parsimonious model is presented in table 4 above. An examination of the Parsimonious model in table 4 above shows that market capitalization and value of stock market transaction explains 98% of the variations in corporate income tax. (This is shown by the $R^2$ value of 0.98). The overall regression was significant (This is shown by the F-ratio of 7.6) and the absence of serial correlation of 3.31. All the variables have the expected signs and are all significant at 1%, 5% and 10% level of significance respectively.

The coefficient of the error correction term is statistically significant and carries the expected negative sign at both 1%, 5% and 10% level of significance. However, the speed of adjustment is rapid (fast) at 79.2% of the adjustment to equilibrium corporate income tax implying that ignoring cointegration in non-stationary time series variables could lead to misspecification of the underlying process in the determination of corporation income tax in Nigeria. Finally, the result show along-run relationship between corporate income tax and the variables identified as its determinants namely, market capitalization and value of stock market transaction by firms in Nigeria.

## CONCLUDING REMARKS

The wide 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 ignoring 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 Nigerian data applied in this study exhibit the features of cointgration theory therefore the generic existence of such long-run relationships, in turn, should be tested using the techniques discussed in this study to reduce the risk of finding spurious conclusions associated with Nigeria data. The method of cointegration has greatly enhanced the existing methods of dynamic econometric modeling of economic time series as we can see from the Nigeria data applied here, and should be consider nowadays as a very valuable part of the practitioner's toolkit in analyzing economic problems towards decision making.

## REFERENCES

Balke, N and T. Fomby (1997), “Threshold cointegration”, *International Economic Review* 38, 627 – 642.

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.

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

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.

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 modeling 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, 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 Policy Modeling* 14, 424 – 438.

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

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 system”, *Journal of Business and Economic Statistics* 13, 27 – 36.

Gonzalo, J., and J-Y. Pitarakis (1997), “Dimensionality effect in cointegrated systems” to appear in Grander 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 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.

Granger, C.W.J. and T.H. Lee, (1989), "Multicointegration", *Advances in econometrics* 8, 71 – 84.

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 Statistical, University of Copehagen.

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.

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.

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

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 authoegressive models,* Oxford University Press: Oxford.

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

Lütkepohl, H., (1999),, "Vector autoegressions", chapter 7.

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.

Mormol, 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., 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.

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

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, forecasting", *Journal of Time Series Analysis* 13, 353-375.

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

Saikkonen, P. (1991), "Asymptotically efficient estimation of cointegrated regression", *Econometric Theory* 7, 1-21.

Saikkonen, P. (1992), "Estimation and testing of cointegration 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., 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. McFafdden, eds., *Handbook of Econometrics* IV, Elsevier: New York.

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