Differencing and Cointegration

Class Notes by Prof. Vinod:

Differencing

It is common to observe that many economic variables have trends which cannot be described by simple deterministic functions. Sometimes the trends are removed by considering logarithmic, Box-Cox or other transformations where the time variable is used, and sometimes differencing is needed. Differencing of order 1 means replacing y_t by $\Delta y_t = (y_t - y_{t-1}) = (1-L)\, y_t$ in terms of the familiar difference operator Δ and the lag operator L. Differencing of order d replaces y_t by $(1-L)^d\, y_t$. The converse of differencing (undifferencing) is the operator $(1-L)^{-1} = 1 + L + L^2 \cdots$, which leads to accumulation or integration. Integration of order d is defined in terms of stationarity property as follows.

DEFINITION of **Integration of order d** If a nondeterministic time series has a stationary invertible ARMA representation after differencing d times, it is said to be integrated of order d, denoted by:

$$y_t \sim I(d) \text{ if } (1-L)^d y_t \text{ is stationary}$$

According to the above definition, when $d = 0$, the series y_t is stationary in levels of its values, and when $d = 1$ it is the change in the levels from one time period to the next that is stationary. A simple example of I(0) series is white noise or a time series sample generated from the normal distribution $N(0,\sigma^2)$. An example of I(1) model is the random walk model:

$$y_t = y_{t-1} + a_t \tag{1}$$

where the error term is $N(0,\sigma^2)$, or integrated of order zero. The differencing operation often removes certain (linear) trends in the series, and one obtains a stationary series. As an exercise, the reader can verify that for the above model also the means and variances of the differenced series are finite and not time-dependent, innovations have only a temporary effect, the autocorrelations eventually decrease steadily and add up to a finite number. Let us start the process in (1) at t=0, where the initial value y_0 is assumed to be known. Now, $y_1 = y_0 + a_1$, $y_2 = y_1 + a_2 = y_0 + a_1 + a_2$, $y_3 = y_2 + a_3 = y_0 + a_1 + a_2 + a_3$, and for an arbitrary t, the random walk series y_t is written as:

$$y_t = y_0 + \sum_{i=1}^{t} a_i \tag{2}$$

where old shocks have the same importance as the new shocks in determining random walk series. That is, <u>random walk process exhibits long memory</u>, contrasted with AR(1) with parameter $|\rho|<1$ has short memory since the coefficient on the t-period old past shock is ρ^t which becomes smaller and smaller as t increases.

Now the first difference $\Delta y_2 = y_2 - y_1 = a_2$, and $\Delta y_3 = a_3$. In general,

$$\Delta y_t = y_t - y_{t-1} = a_t \sim N(0, \sigma^2) = I(0) \tag{3}$$

Thus first differencing of $y_t \sim I(1)$ from a random walk (integrated of order 1) series yields $a_t \sim I(0)$. Observe that, for $d \geq 1$, we have $\Delta^d \equiv (1-L)^d$, a polynomial in the lag operator L whose root is unity – that is the polynomial is zero when L=1, the root. Thus differencing removes unit roots, and the resulting polynomial in L is of lower order, since the largest power of L is reduced by d. Common sense suggests that it might be simpler, parsimonious or efficient to work with lower order processes. Hence it is not surprising that removal of unit roots is often recommended before processing of a series. For example, if $y_t \sim I(1)$, is integrated of order 1 and we do not take the first differences, the series is non-stationary, with infinite variance. Any innovation then has a permanent effect and the autocorrelation coefficients of order k denoted by $\rho_k \to 1$ for all k as $t \to \infty$.

Regarding linear combinations, let us note a a few results.

If $y_t \sim I(1)$ and if $x_t \sim I(0)$, then the linear combination $z_t = x_t + y_t \sim I(1)$, (4)

If $y_t \sim I(d)$ and a and b are constants, then the linear transformation $(a\, y_t + b) \sim I(d)$, (5)

To verify (4) note that the infinite variance of I(1) series will eventually dominate. We would also expect that:

If $y_t \sim I(1)$ and if $x_t \sim I(1)$, then the linear combination $z_t = a\, x_t + b\, y_t \sim I(1)$. (6)

If $y_t \sim I(d)$ and if $x_t \sim I(d)$, then the linear combination $z_t = a\, x_t + b\, y_t \sim I(d)$. (7)

Since deterministic trends are often difficult to assess, the order of integration is also difficult to estimate

Co-integration

If two or more variables seem to trend together there is an interest in studying their joint trending. Familiar examples are: short and long term interest rates, household incomes and expenditures, commodity prices (gold) in geographically separated markets, capital appropriations and expenditures by business. Sometimes Economic theory suggests that certain theoretically related variables have an equilibrium relation between them and the equilibrating process forces them to trend together. For example, the quantity that the buyers are willing to demand (q_d) and quantity that the sellers are willing to supply (q_s) cannot drift too far apart. They may drift apart in the short run perhaps due to seasonal effects, but if they can drift apart without bound, there is no equilibrium. The equilibrium error, $z_t = q_d - q_s$ is excess demand in this case. Adam Smith showed that the invisible hand of the market forces will tend to force z_t to approach zero in the dynamic sense implied by $z_t \sim I(0)$.

In general, z_t is a linear combination of two or more series. Economic theory suggests that equilibrium error z_t should not diverge over time, but rather should be stationary. When z_t=0, the equilibrium is achieved. If $z_t \sim N(0, \sigma^2)$, it can achieve the equilibrium several times, whenever there is a crossing from positive to negative values, or vice versa. Hence having z_t become I(0) is a result supportive of economic theory. Cointegration is a statistical description of some equilibria.

Co-integration is a study of the behavior of linear combinations of time series, and looks for results which test economic theory. If two or more I(d) series are trend together, the order of some particular linear combination z_t –usually representing the equilibrium error– should be smaller than d. The joint trending of the two variables may mean that the long-run components cancel each other in some sense. Hence we have the formal definition due to Granger(1981), for a vector y_t of two or more time series.

DEFINITION of **Co-integration**: The components of a $k \times 1$ vector y_t with $k \geq 2$ time series are said to be co-integrated of order (d, b) denoted by $y_t \sim CI(d,b)$ if there exist $r \geq 1$ "co-integrating vectors" α_i ($\neq 0$) of dimension $k \times 1$ defining r linear combinations which are integrated of order $d-b$, or

$$z_{it} = \alpha_i{}' y_{it} \sim I(d-b), \quad b>0,\ i = 1,2,\cdots, r. \tag{1}$$

where the elements of y_t are denoted by y_{jt} with $j=1,2,\cdots, k$.

Clearly, some reduction (b>0) in the order of integration is necessary as a result of the linear combination by the co-integrating vectors α_i. For the equilibrium example above, the quantities demanded and supplied $y_{1t} = q_d$ and $y_{2t} = q_s$ are I(1), and we say that they are co-integrated if excess demand $q_d - q_s$ defines a linear combination which is I(0). Of course, the units of measurement of q_s and q_d must be comparable for their difference to be zero.

Multivariate Case:

The reduction in the order of integration of equation (1) may be stated in terms of multivariate time series analysis of the elements of the vector y_t. At each lag, the multivariate theory allows for possible relations between the elements of y_t. The reduction in the order of integration from d to d–b can be stated as a reduction in the rank of a so-called co-integration matrix G(L) whose elements are polynomials in the lag operator L, defined as follows. Let $\Delta y_t = (1-L)y_t$ denote a difference operator applied to each element of the y_t vector. Multivariate ARMA models for y_t can be used to write the following matrix equation.

$$\Delta\, y_t = G(L)\, a_t \tag{2}$$

where a_t denotes a white noise process. Engle and Granger(1987) show that when the elements of y_t are co-integrated, the matrix G(1) obtained by simply replacing the operator L by unity does not have a full rank, that is G(1) is singular. To understand why we have polynomials in the lag operator, let us first consider the simple univariate case ($k = 1$, $d = 1$). Recall that I(1) means that the elements of y_t are integrated of order 1, which means that $(1-L)\ y_t = a_t$ is the white noise process. Now divide both sides by $(1-L)$ to yield $y_t = (1-L)^{-1} a_t$, and observe that $(1-L)^{-1}$ equals the infinite series $1+L+L^2 \cdots$. Hence the following multivariate (Wold) decomposition is plausible:

$$y_t = \sum_{j=0}^{\infty} G_j\ a_{t-j}\,, \qquad G_0 = I \tag{3}$$

where G_j are called Green's function matrices of dimension $k \times k$ each. It is shown in the literature that the representation (3) is unique under certain regularity conditions. The novel feature in the present context of co-integration is that we have Δy_t instead of y_t on the left side of (3).

VAR representation and Co-integration:

Sims (1980, Econometrica) argued that economic theory does not really justify the “structural equations" of the traditional simultaneous equations models (SEM). Since the vector autoregressive (VAR) models widely used in Econometrics are a special case of multivariate ARMA models when the MA terms are absent, there is no new theory needed for considering the VAR models. Assume that the elements of the $k \times 1$ vector y_t of time series have a vector autoregressive representation, VAR(p) of order p, with each of the Φ_j coefficient matrices of order $k \times k$. Denote by I_k an identity matrix of order k.

$$\Phi(L)\, y_t = (\, I_k - \Sigma_{j=1}^{p} \Phi_j L^j)\, y_t = a_t \tag{8}$$

Observe that $\Phi(L)$ is also a $k \times k$ matrix whose elements are polynomials in the lag operator L. According to Sims, these matrices are identified without any constraints from economic theory. Of course, economic theory imposes some overidentifying restrictions, which can be tested.

Matrix algebra of Polynomial Matrices, a Digression:

Matrix algebra is available for such polynomial matrices. Finding the eigenvalues of a matrix A involves solving a characteristic equation $|\,A - \lambda I| = 0$. The expression $A - \lambda I$ is a polynomial in λ. Thus polynomial matrices are commonplace. The determinant of $\Phi(L)$ is a scalar polynomial, its inverse is an infinite order matrix, which is well defined only if the series converges. (See Banerjee et al p. 141). Eigenvalues of polynomial matrices are of interest ina analyzing dynamics and cointegration. Let $A(L)y_t = A_0 y_t + A_1 y_{t-1} \cdots A_k y_{t-k} = 0$, where $A_0 = I$ for normalization and represent this k-lag equation in a companion form which reduces it to AR(1) form in matrix polynomials.

$$Y_t = \begin{bmatrix} y_t \\ y_{t-1} \\ . \\ . \\ y_{t-k+1} \end{bmatrix}, \Phi = \begin{bmatrix} A_1 & A_2 & \cdots & A_{k-1} & A_k \\ -I & 0 & \cdots & 0 & 0 \\ 0 & -I & \cdots & 0 & 0 \\ & & & & \\ 0 & 0 & 0 & -I & 0 \end{bmatrix}, \ A(L)=0 \Rightarrow Y_t + \Phi Y_{t-1} = 0$$

The companion form is obtained by simply inserting identities of the form $y_{t-1} \equiv y_{t-1}$. Note that $Y_t + \Phi Y_{t-1} = 0$ is an AR(1) form in Y_t which is a vector instead of y_t which is a scalar. If $A_1, \cdots A_k$ are matrices, we have a VAR system. Thus a k-th order VAR(k) can be reduced to VAR(1) using more complicated matrices. If a polynomial matrix is defined in terms of z instead of L and we consider its characteristic polynomial, $|\det \Phi(z)|=0$, its roots can be assumed to be =1 or >1. If the $n \times n$ matrix has a reduced rank (=r), it can be expressed as a product of two $n \times r$ matrices. Many results in cointegration theory can be understood in terms of matrix algebra results.

A general formulation (Banerjee et al p. 179, Sims et al 1990, Etrica) is

$Y_t = A \ Y_{t-1} + G\Omega^{1/2} \eta_t$

where A is $k \times k$ matrix, G is $k \times N$ selection matrix (mostly zeros and ones), Ω is covariance matrix of errors, η_t is $N \times 1$ vector of martingale difference sequence (MDS). If k_1 eigenvalues of A are less than unity and $k - k_1$ are exactly equal to unity. Since $T^{-p} \Sigma_t Y_t Y_t'$ converges to a singular limit for a suitably chosen p, we must work with a transformed set $Z_t = DY_t$ of variables, where D is nonsingular. Using D, Z can be decomposed between stochastic and nonstochastic components. As long as the stochastic components are not trended, regression coefficients are asymptotically normal.

<u>Back to Cointegration after the digression:</u>

If the elements of y_t are co-integrated, the $\Phi(1)$ matrix, obtained by simply replacing L by 1 is singular. Assuming that the rank of $\Phi(1)$ is r which is strictly positive, but less than k.

When y_t data are first differenced by the operator $\Delta = 1 - L$, we have Δy_t. If we attempt the VAR representation of the differenced data we have

$$\Phi^*(L)\,(1-L)\,y_t \ = \ a_t \tag{9}$$

where $\Phi^*(L) \ = (\, I_k - \Sigma_{j=1}^{p-1} \Phi_j L^j)$ and $\Phi_j^* = - \ \Sigma_{m=j+1}^{p} \Phi_m$

See Ahn and Reinsel (1990, JASA p813) for algebra of Jordan canonical forms in this context. Note that $\Phi^*(0)=\Phi(0)= I_k$. Does the singularity of $\Phi(L)$ imply the singularity of $\Phi^*(L)$? singularity of $\Phi^*(L)\,(1-L)$? singularity of both ?

Engle and Granger (1987) suggest that the VAR estimates derived from <u>differenced data are misspecified</u> if the elements of y_t are cointegrated. They prove their result based on error correction interpretation of co-integration. By analogy, MacDonald and Kearney (1987) show that if one is interested in causality testing, and if the variables are co-integrated, the models based on differenced data are misspecified.

Estimation of Cointegration

In any estimation of the cointegration, one must recognize 3 elements, nonsymmetric distribution and dependence on nuisance parameters.

(i) our prior knowledge regarding existence of unit roots should be used. The median bias should be eliminated, if possible. This will eliminate the nonsymmetric part of the dependence on nuisance parameters and will increase efficiency.

(ii) multivariate aspect of the problem (use simultaneous system equation estimation, if possible to eliminate simultaneous equations bias)

(iii) flexible enough to capture the dynamics of the system.

Stock (1987, Etrica p1035) has shown a useful result that if two variables are co-integrated OLS estimates of the slope coefficients are consistent and highly efficient. The cointegrating regression $y_t = \hat{\alpha} x_t + z_t$, where z_t are residuals and x_t and y_t are I(1) series (random walks) with infinite variance. Only the one linear combination has finite variance, which is reliably found by the OLS. Stock shows that the OLS estimate converges to the true value at a rate T^{-1} rather than the usual $T^{-1/2}$, because all other linear combinations have infinite variance. This suggests that OLS is good no matter which is the dependent variable and despite the possible existence of simultaneous equations bias or serial correlation among errors OLS is consistent. However, in some cases there are problems of bias and nonnormality associated with the OLS. The usual t-ratios produced by the OLS are not appropriate for inference.

According to Stock consumption and income are cointegrated and simple regression gives consistent estimator. Haavelmo (1943) argued that OLS regression of consumption on income gives rise to inconsistent estimate of MPC because of the simultaneous equations bias. Stock(p.1042) claims that the simultaneous equations bias may be important in small samples, but not asymptotically. Harvey (1989, p.294) explains these problems concisely. Imagine two economic time series, y_t and x_t, both of which are nonstationary, I(1) say. If economic theory suggests that the two cannot drift too far apart, they must satisfy a stationary (steady state) relationship between them. Then we are often able to find a linear combination $y_t - a_0 - a_1 x_t$ which is I(0). One may interpret a_1 as a long run elasticity of y with respect to x. If we take first differences and expectation of both sides of $y_t = a_0 + a_1 x_t + u_t$ it can be seen that the ROG(y)=a_1ROG(x), where ROG denotes a rate of growth. This illustrates the case where x_t and y_t are CI(1,1).

Gonzalo (1994, J of Etr. Vol. 60, p. 203) compares five methods of estimating cointegration: OLS, NLS, Maximum Likelihood in an Error Correction model (MLECM), Principal Components (PC) and Canonical Correlation (CC). He finds that Johansen's MLECM is the best. His Table 1 gives estimates of a simple model by the five methods. He explains each method in detail, including their asymptotic properties. Asymptotic efficiency of MLECM is proved by Saikkonen (1991). His Table 2 describes various methods. Hatanaka (1996,Time Series Based Econometrics, Oxford, p.117) expresses scepticism about the usefulness of Johansen's methods for macroeconomics. In order to discuss these results, we need to discuss error correction models.

Error Correction Models (ECM's) of Co-integration:

One of the appealing aspects of co-integration to the economist is its link with economic equilibrium. In dynamic economic models often give rise to the so-called equilibrium steady state. For a two variable system a typical error correction model (ECM) relates Δy_t, the change in one variable is a function of past equilibrium errors (u_t) and past changes in both variables:

$$\Delta x_t = \gamma_1 (y_{t-1} - \alpha x_{t-1}) + \text{p– lags of } \Delta x_t \text{ and } \Delta y_t + \text{white noise errors}$$

$$\Delta y_t = \gamma_2 (y_{t-1} - \alpha x_{t-1}) + \text{p– lags of } \Delta x_t \text{ and } \Delta y_t + \text{white noise errors}$$

where the errors may be correlated. The terms $(y_{t-1} - \alpha x_{t-1})$ are called error correction terms, which are I(0) if the variables are cointegrated. Thus all variables in the above two equations are I(0). Thus cointegration implies an ECM and vice versa, which is called a Granger representation theorem. A direct estimation strategy is to first obtain a superconsistent estimate of α, compute the error correction terms and then estimate the two equations by the seemingly unrelated regressions (SUR) system of equations.

More generally, Engle and Granger (1987, Etrica p254) define that a (multivariate) vector of time series y_t has an error correction interpretation if it has at least one unit root, in the sense that it can be expressed as:

$$A(L)\,(1–L)\,y_t = -\gamma z_{t-1} + u_t,$$

where u_t is stationary multivariate disturbance, and there are exactly r cointegrating relations (equilibria) satisfying $z_r = \alpha y_r$. This is a multivariate representation without assuming a subset of variables to be (weakly) exogenous. Also, α need not be a set of constants arising from economic theory. It is treated as a parameter. Another way of thinking about this is in terms of a special case:

$$\Delta y_t = a_0 + a_1 \Delta y_{t-1} + b_0 \Delta x_t + b_1 \Delta x_{t-1} + (\alpha - 1)\{ y_{t-1} - \frac{\beta}{1-\alpha} x_{t-1} \} + \epsilon_t$$

using Harvey's 1990 book (p. 290). The term in the curly brackets {.} represents the long run relation where the level variables are involved. If α is close to 1, one may concentrate on the short run and ignore the long run term {.}. When α<1 the level terms play a crucial role of error correction as follows. If due to a shock y_{t-1} becomes too big for equilibrium, the {.} term becomes positive, but the presence of $(\alpha - 1)$ which is negative for α<1 will bring down the left side Δy_t and slow down the short run growth in y.

Phillips (1991, Etrica p. 283) calls the above ECM as autoregressive because of the presence of Δy_{t-1} and suggests that a triangular system format has the advantage that its error structure does not depend on the parameters involved in cointegrating relations. In Phillips' framework, y_t is an $n \times 1$ vector of I(1) variables, u_t is an $n \times 1$ vector stationary time series. We partition y_t into y_{1t} and y_{2t} which are $n_1 \times 1$ and $n_2 \times 1$ vectors respectively so that $n=n_1+n_2$. The linear long run equilibrium between the two sets is given by $y_{1t}=By_{2t}$. After allowing for randomness we have:

$y_{1t}=By_{2t}+u_{1t}$, and $\Delta y_{2t} = u_{2t}$,

where B is an $n_1 \times n_2$ matrix, u_{1t} are stationary deviations from the equilibrium. The resulting ECM is $\Delta y_t = -EAy_{t-1} + v_t$, $E=\begin{pmatrix} I \\ 0 \end{pmatrix}$, $A=[I, -B]$ and $v_t=\begin{pmatrix} I & B \\ 0 & I \end{pmatrix} u_t$, where asymptotics are convenient.

Engle-Granger Two Step Estimation and Testing:

This is a simple OLS based technique for handling cointegration.

Step 1: OLS on static regression: $y_t=\alpha x_t + \epsilon_t$ to get $\hat{\alpha}_{ols}$.

It can be shown that $(\hat{\alpha}_{ols} - \alpha) \sim O_p(T^{-1})$, or super consistent.

Step 2: Substitute this $\hat{\alpha}_{ols}$ in the error correction model

$\Delta y_t = \gamma_2(y_{t-1} - \hat{\alpha}_{ols}\, x_{t-1})$ + (p– lags of Δx_t and Δy_t) + white noise errors

Engle-Granger theorem shows that the usual OLS standard errors in the second step will provide consistent estimates of true standard errors (See Banerjee et al p. 159)

Econometricians in 1970's started worrying about static time series regressions in levels y_t on x_t, due to spurious regression results. Then came models with general specifications followed by simplifications dictated by statistical tests, e.g. Hendry, Mizon, Richard, etc. Several variables and their lags are included and chances of having cointegrated subset of regressors is thereby enhanced. Engle-Granger theorem reinstated the faith in static regression in late 1980's. Thus the literature seems to have come a full circle. However, some problems with regressions with nonstationary variables remain: (i) Bias is present in OLS estimation of cointegrating parameter, and poor performance in some situations is documented by simulations. Mankiw and Shapiro (1986, JME p 165) document size distortions in t tests. (ii) The coefficients follow non-normal distributions as functionals of Wiener process. (iii) The OLS error process may not be a martingale difference sequence (MDS). (iv) If more than one cointegrating vector exists, there will be problems due to a failure of weak exogeneity.

Park-Phillips Fully Modified Estimation (2nd order bias removal):

See Park and Phillips (1988, ET p.468) and Phillips and Hansen (1990, R Eco Stud, p.99), Banerjee et al (1993, p.). In the system $y_{1t}=\beta\, y_{2t}+u_{1t}$, $y_{2t}=y_{2,t-1}+u_{2t}$

where u_{1t} and u_{2t} are intercorrelated and autocorrelated, OLS can be biased. Hence full system ML estimation of cointegrated system is recommended by Phillips and others. The idea is to obtain median-unbiased and asymptotically normal estimates. Assume that $u_t=(u_{1t}, u_{2t})'$ is weakly stationary with zero mean and cov. matrix $\Omega = V + \Gamma + \Gamma'$ where $V=Eu_0\, u_0'$ is the cov. matrix for lag 0, $\Gamma=\Sigma_{k=1}^{\infty} E\, u_0\, u_k'$. The fully modified estimator replaces

$(X'X)^{-1}X'y$ by $(X'X)^{-1}[\,X'y - T\hat{\delta}^{+}]$

where $X'X$ is sum of squares of y_{2t}. $X'y$ is $\Sigma_{t=1}^{T} y_{1t}^{+}\, y_{2t}$, where correction for long-run simultaneity is

$y^{+}_{1t} = y_{1t} - \hat{w}_{12}\hat{w}^{-1}_{22}\Delta\, y_{2t}$, where w_{ij} are partitioned 2×2 elements of the Ω matrix. The hat denotes a consistent estimate. The 2nd order bias correction term $\hat{\delta}^{+}$ to $X\prime y$ is as follows. $\hat{\delta}^{+} = \hat{\Lambda}\begin{pmatrix} 1 \\ -\hat{w}^{-1}_{22}\hat{w}_{21} \end{pmatrix}$ where $\Lambda = \Sigma^{\infty}_{k=1} Eu_{20}\, u_{k}\prime$. The fully modified standard error s^{+} is given by

$(s^{+})^{2} = [\hat{w}_{11} - \hat{w}^{2}_{21}\ \hat{w}^{-1}_{22}]\,\{\Sigma^{T}_{t=1} y^{2}_{2t}\}^{-1}$. Fully modified $(\hat{\beta}_{fm} - \beta)/s^{+} \Rightarrow N(0,1)$

The simulation in Phillips-Hansen (1990) does not show unequivocal superiority of fully modified estimator of β, but does show superior estimate of t-value for $\hat{\beta}$. Bartlett triangular spectral window with lag length 5 and OLS residuals $\hat{u}_{it}$ are used in practice to estimate Λ and Ω in frequency domain.

FIVE Estimation Methods:

Now we return to various estimators summarized by Gonzalo (1994, J Etr Vol. 60, p. 203) OLS, NLS, MLECM, Principal Components (PC) and Canonical Correlation (CC). The model is:

$$y_t = \beta x_t + z_t\ ,\ z_t = \rho z_{t-1} + e_{zt}\,,\ \Delta x_t = e_{xt}\ \text{and}\ \begin{pmatrix} e_{zt} \\ e_{xt} \end{pmatrix} \sim \text{iid } N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma^{2}_{1} & \theta\sigma_1\sigma_2 \\ \theta\sigma_1\sigma_2 & \sigma^{2}_{2} \end{bmatrix}\right]$$

where both y_t and x_t are I(1) but there is a linear combination or cointegrating vector $\alpha = (1, -\beta)\prime$ that is I(0).

A very useful Error correction form is:

$$\begin{pmatrix} \Delta y_t \\ \Delta x_t \end{pmatrix} = \begin{pmatrix} \rho - 1 \\ 0 \end{pmatrix}(y_{t-1} - \beta\, x_{t-1}) + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \text{ where } u_{1t} = \beta\, e_{xt} + e_{zt} \text{ and } u_{2t} = u_{zt}$$

Since $\Delta\, x_t$ si weakly exogenous wrt β (for any θ), we can estimate the cointegrating vector $\alpha = (1, -\beta)\prime$ from the conditional model:

$$\Delta y_t = \alpha_1 \Delta x_t + (\rho - 1)(y_{t-1} - \beta\, x_{t-1}) + u_{1.2t} \text{ where } u_{1.2t} = \Delta y_t - E(\Delta y_t | \Delta x_t,\ z_{t-1})$$

or $u_{1.2t} = e_{zt} - \theta\frac{\sigma_1}{\sigma_2}\, e_{xt}$

The parameter α_1 is called short-run or impact multiplier. In general it involves lags of Δy_t and Δx_t and plays a key role in derivations. This was first noted by Johansen (1991. Etrica p 1551) in his representation of the Granger representation theorem. Paruolo (1997, ET 13, p79) develops theory for inference based on the impact multiplier from a dural MA representation.

<u>OLS:</u> $y_t = \beta\, x_t + z_t$ and estimate $\hat{\beta}_{ols}$

The asymptotic distribution has 3 terms (i) a mixture of normals (ii) unit root term which makes the distribution asymmetric and (iii) a simultaneous equations bias caused by long-run correlation between x_t and errors z_t.

<u>NLS</u> <u>(q):</u> q denotes the number of lags included in the summation.

Estimate by OLS: $\Delta y_t = \pi_{11} y_{t-1} + \pi_{12} x_{t-1} + \Sigma^{q}_{i=1}\theta_i \Delta y_{t-i} + \Sigma^{q}_{i=1}\delta_i \Delta x_{t-i} + u_{1t}$

$\hat{\beta}_{nls} = -\hat{\pi}_{12}/\hat{\pi}_{11}$

<u>MLECM(q):</u> Vector autoregression (VAR) is used with lag order q

Denote $H_t = (y_t\,,\ x_t)\prime$. $\Delta H_t = -\gamma\alpha\prime\, H_{t-1} + \Sigma^{q}_{i=1}\Gamma_i \Delta H_{t-i} + u_t$ where $u_t \sim \text{iid } N(0, \Lambda)$

Regress ΔH_t on $\Delta H_{t-1}, \cdots, \Delta H_{t-q}$ and save residuals as R_{ot}

Regress H_{t-1} on $\Delta H_{t-1}, \cdots, \Delta H_{t-q}$ and save residuals as R_{qt}

Define $S_{jk} = (1/T) \Sigma_{t=1}^{T} R_{jt} R'_{kt}$ based on residual matrices for j and k =o,q

Solve the eigenproblem: $(S_{qo} S_{oo}^{-1} S_{oq}) \hat{\alpha}_i = \nu_i S_{qq} \alpha_i$ for i=1,2.

In practice, decompose $S_{qq}=FF'$ and solve the symmetric eigenproblem:
$[(F')^{-1} S_{qo} S_{oo}^{-1} S_{oq} (F)^{-1} - \nu_i I] F\hat{\alpha}_i = 0$ Rank the eigenvalues in descending order and use

$$\hat{\beta}_{mlecm} = -\hat{\alpha}_{12}/\hat{\alpha}_{11}$$

The asymptotic distribution is a mixture of normals and standard asymptotic χ^2 tests can be used. The bias terms are absent, and asymptotic theory suggests that this estimator is most efficient provided the no. of lags is known and errors are Gaussian.

Principal Components: (PC) This method is designed to find the linear combination of y_t and x_t with minimum variance. If Σ denotes the cov. matrix of H_t then PC minimizes teh quadratic form $p'\Sigma p$ subject to $p'p=1$. The solution arises from the first order conditions.
Define $M = \Sigma_{t=1}^{T} H_t H'_t$. Solve the eigenproblem: $M \hat{p}_i = \mu_i \hat{p}_i$ for i=1,2
Rank the eigenvalues in descending order and compute $\hat{\beta}_{pc} = -\hat{p}_{22}/\hat{p}_{11}$
This method is very sensitive to units of measurement, since the normalization $p'p=1$ is often incorrect. The asymptotic distribution is very similar to OLS.

Canonical Correlations: (CC) seeks maximally correlated linear combinations of H_t and H_{t-1}
Define $M_{jk} = \Sigma_{t=1}^{T} H_{t-j} H'_{t-k}$ for j,k = 0,1.

Solve the eigenproblem: $(M_{01} M_{11}^{-1} M_{10}) \hat{c}_i = \delta_i M_{00} \hat{c}_i$ for i=1,2, or its symmetric equivalent.

Rank the eigenvalues in descending order and compute $\hat{\beta}_{cc} = -\hat{c}_{22}/\hat{c}_{11}$
This method does not incorporate information about unit roots. MLECM also performs canonical correlation between ΔH_t and H_{t-1}. This slight difference makes a big difference to asymptotic distribution. Bossaerts JEDC 1988, 12, 347-364. Ahn and Reinsel (1990, JASA p813) propose reduced rank estimation and involves Jordan canonical forms. They propose a two-step estimator superior to Engle and Granger's. However their method would need additional software.

Saikkonen-type Estimation Methods

Saikkonen (1991, Econometric Theory, pp. 1-21) reports a new simple time domain correction to the OLS regression
$y_t = \alpha + \beta x_t + \Sigma_{j=-K}^{K} \eta_j \Delta x_{t-j}$ when x and y are cointegrated. Note that in the above expression the subscript starts with negative values of j, interpreted as leads (not lags). This is proved to be efficient asymptotically. Phillips (1991, Etrica, p283) suggests a triangular simultaneous equations type estimator which is shown to be optimal in some sense. A three-step estimator proposed by Engle and Yoo (1990) which approximate Johansen's maximum likelihood estimator has useful t ratios whose limiting distribution is normal.

Saikkonen (1992, Econometric Theory, pp. 1-27 and 1993, pp 19-35) also makes an important contribution to estimation of cointegrations. His 1993 paper has won an ET award in 1997. It suggests making the cointegration regressions as the “Reduced form of a simultaneous equations model." The most appealing aspect of his theoretical results is that usual exogeneity assumptions is NOT NEEDED. The asymptotic distributions are shown to be mixed normal, so usual Wald type tests with asymptotic χ^2 distributions are available. After the coefficients of the cointegrating regressions are estimated, only GLS is required.

Saikkonen (1993, ET p.19) starts with
$By_t=Cx_t + e_t$ where the dimensions are $B(n \times n)$, $C(n \times m)$, $x_t(m\times 1)$, $y_t(n \times 1)$ (1)
$x_t=x_{t-1}+\nu_t$, $t=1,2,\cdots$ (2)
Denote w_t=combined vector of (e_t and ν_t) of dimension $(n+m) \times 1$. If w_t is strictly stationary, ergodic with zero mean, finite covariance matrix, and continuous spectral density matrix $f_{ww}(\lambda)$ with a positive definite $\Omega=2\pi f_{ww}(0)$, evaluating the spectral density at $\lambda=0$ frequency, i.e., for the long-run component. These assumptions imply that y_t and x_t are cointegrated with exactly n cointegrating relations defined by (1). Note that if there is no cointegration, B in $By_t=Cx_t + e_t$ is singular. Assuming nonsingular B, premultiply by B^{-1} to write: $y_t=B^{-1}C x_t+B^{-1}e_t=Ax_t+u_t$. The traditional view does not have the B matrix on the left side and the matrix A is treated as coefficient matrices of cointegrating regressions. Clearly $A=B^{-1}C$, by definition, where B and C may not be identifiable, without identifying restrictions. Saikkonen develops a theory by imposing a rank condition on the combined matrix [B C], which is similar to the rank consition from econometrics of simultaneous equation models.

The basic steps may be reviewed before giving the details. Any suitable estimation of A in $y_t=Ax_t+u_t$ is the starting point of this theory. To get B and C, assume that all diagonals of B are unity (normalization) to obtain uniqueness (decide which variable to regress on which others). Now write $A=B^{-1}C$ (by definition) as $a_i=H_i\delta_i$ set of $i=1,2,\cdots n$ equations, where H_i are known numbers and δ_i are then estimated by a regression of a_i on H_i defined below. If A is estiamted by OLS, the estimation of δ_i can be viewed as the standard 2SLS estimator. However, it is not necessary to assume that x_t are exogenous, all we need is that the components of x_t are not cointegrated among themselves. They must be all nonstationary, which assumption can be tested by using Saikkonen (1992), when AR(∞) representation is available for u_t and ν_t. Identification restrictions can be simple exclusion restrictions as follows.
β_i=i-th row of $(B - I_n)$ of dimension $(n_i \times 1)$, where we subtract identity since diagonals of B are unity.
γ_i=i-th row of C of dimension $(m_i \times 1)$.
Selection matrix J_i has ones and zeros, it is $(n \times n_i)$ such that $J_i\beta_i$ is the i-th row of $B - I_n$.
Selection matrix K_i has ones and zeros, is $(m \times m_i)$ such that $K_i\gamma_i$ is the i-th row of C.
Define $A'=[a_1, a_2, \cdots a_n]$ from rows of $A=B^{-1}C$, and denote $G_i=A'J_i$.

$$a_i=H_i\delta_i=[A'J_i\ ,\ K_i] \begin{pmatrix} -\beta'_i \\ \gamma_i \end{pmatrix} \quad (3)$$

Assuming that i-th equation of (1) satisfies the usual rank condition for identification, is equivalent to assuming that H_i has full column rank. Thus parameters of interest β_i and γ_i can be uniquely obtained by solving (3).

Recall that we first find suitable estimator of A in $y_t=Ax_t+u_t$. Now, compute A' and get a_i from its rows. Construct estimators of $H_i=[A'J_i\ ,\ K_i]$. Now regress a_i on H_i in an auxiliary regression by using the GLS method. The components of x_t may be interpreted as common trends which describe the nonstationarity in y_t. The common trends method decomposes a nonstationary series into two components: stationary and stochastically trended component.

Godambe-Durbin Estimating Function (EF) Methods

Maekawa, K. T. Yamamoto , Y. takeuchi and M. Hatanaka paper in J of Statistical Planning and Inference 49 (1996) pp. 279-303 uses Durbin two-step estimator. They fully work out a model like ADL(1,0), derive the asymptotics, compare the results to the older Maddala-Rao (1973 Etrica p.761).
Since Durbin 2-step is optimal in the sense of reaching Cramer-Rao lower bound, according to newer theory of “Godambe-Durbin Estimating Functions in Econometrics," to appear in a paper by H. D. Vinod. It is not surprising that Maekawa et al found that the 2-step estimator works very well.

If x_t and y_t are I(1) and there is linear combination $x_t - cy_t \sim I(0)$, then x_t and y_t are cointegrated.

Consider AR(1) model $y_t= \alpha y_{t-1} +u_t$. If α=1, the limiting distribution of OLS estimator of α is nonstandard. It is a functional of the Wiener process W(r) on [0,1], let $\Rightarrow$ denote weak convergence.

$T(\hat{\alpha} - 1) \Rightarrow \frac{0.5\{W(1)^2 - (\sigma_u/\sigma^2)\}}{\int_0^1 W(r)^2 dr}$, where $\sigma_u^2 = \lim_{t\to\infty}(1/T)\,\Sigma_{i=1}^T E(u_t^2)$

$\sigma^2 = \lim_{t\to\infty} E(S_T^2/T)$, where $S_T = \Sigma_{i=1}^T Eu_t$ is a partial sum. Note that expression having squared W(1) and sum of squares of u_t is in the numerator and integral of $W(r)^2$ is in the denom. Note there is T on the left side of this expression and the results do not depend on initial values.

This means that $\hat{\alpha} \Rightarrow 1$ at a fast rate of T for the AR(1) case, rather than the usual rate $\surd$ T (in prob.). This property is called "super consistency of OLS." Before the recent theory for integrated processes, the asymptotic theory for dynamic regressions was well developed, e.g., Maddala-Rao (1973, Etrica, p.761). This theory showed that if lagged-dependent variables and autocorrelated errors are present, OLS is biased and inconsistent. There remains some confusion in the literature. Park and Phillips (1989, ET, p.95) showed that if the regressors are I(1), OLS loses the super-consistency, it become only $\surd$ T consistent. (still not too bad).

Now consider 3 equation model having cointegrated I(1) regressor z_t:
(1) $y_t = \alpha y_{t-1} + \beta z_t + u_t$, $|\alpha|<1$.
(2) $u_t = \rho u_{t-1} + v_t$, $|\rho|<1$, v_t is iid $N(0, \sigma_v^2)$
(3) $z_t = \lambda z_{t-1} + \epsilon_t$, ϵ_t is iid $N(0, \sigma_\epsilon^2)$ and indep of v_t.

Remember that α is the coeff. of y_{t-1}, the lagged dep. varible, β is the coeff. of z_t, I(1) regressor (when λ=1) and ρ denotes the coeff. of autoregressive errors u_t. In Maddala-Rao $|\lambda|<1$ was assumed and λ=1 is what is new now. Eq. (1) is ADL(1,0) with autocorrelated errors. If (3) is substituted in (1) it becomes ADL(1,1) with autocorrelated errors.

Denote the asymptotic biases: $(\hat{\alpha}-\alpha) \to \gamma$, $(\hat{\beta}-\beta) \to \delta$ and reconsider equation (1): $y_t = \alpha y_{t-1} + \beta z_t + u_t$, by replacing $\alpha = \hat{\alpha} - \gamma$ and $\beta = \hat{\beta} - \delta$. Note that the residual of this regression (1) is $e_t = u_t - (\gamma y_{t-1} + \delta z_t)$. After considerable algebra, it can be shown that y_{t-1} and z_t are cointegrated. Write
$y_{t-1} = (\beta z_t)/(1-\alpha) + a_t + \Delta$, where $\Delta = \alpha^t y_0 - (\beta\alpha^t/(1-\beta))z_0$, the effect of initial values and

$a_t = \Sigma_{s=0}^{t-2} \alpha^S u_{t-1-s} - \beta/(1-\alpha)\,\Sigma_{s=1}^{t-1} \alpha^S \epsilon_{t-s}$. More conveniently, $a_t = u_{t-1} - (\beta\epsilon_t)/(1-\alpha) + \alpha\, a_{t-1}$.
The true error of regression (1) can be written as $e_t = u_t - \gamma\, a_t - o_p(1)$. Using explicit expressions for sums of squares of e_t, sums of products of e_t and e_{t-1}, etc. one can write the usual OLS test statistics and their plims to make important conclusions.

The presence of asymptotic bias for OLS estimates of ADL(1,1) from (1) obviously means that OLS is inconsistent. Unlike the case of integrated regressors, where lagged dependent variables are absent, the biases are not random terms with Wiener integrals in them. Here the biases γ and δ are nonrandom, although the OLS estimates $\hat{\alpha}$ and $\hat{\beta}$ do remain biased and inconsistent. It can be shown that the t-statistics converge to infinity. For example, plim $t_{\hat{\alpha}} \to \infty$ unless the null hypothesis is $\alpha_0 = \alpha + \gamma$. Also, plim $t_{\hat{\beta}} \to \infty$ unless the null hypothesis is $\beta_0 = \beta + \delta$, which includes the bias δ.

Infinite t-statistic means that observed t-stats will be too large always rejecting the null hypotheis, hence misleading. Sometimes this is related to spurious regression problem (See Hamilton p.560). Also plim $\hat{\rho} = \rho - \gamma$ means that the Cochrane-Orcutt will not be able to correct for autocorrelated errors and provide consistgent estiamtes of α and β. The estimated $\hat{\rho}$ will not be the true ρ even asymptotically. Since (1) has lagged dep. variables, Durbin's h statistic for α is given by

$$h = \left(1 - \frac{DW}{2}\right)\left[\frac{T}{1 - T[se(\hat{\alpha})]^2}\right]^{1/2},$$ where DW =Durbin-Watson statistic and se= standard error.

Unfortunately, this statistic is not reliable here, since plim $h \to \infty$. By contrast, DW is better according to the simulation by Maekawa et al, although strictly speaking DW is not applicable when there are lagged dep. variables.

Now we show that Durbin (1960, JRSS-B, p. 139) **2-step estimator** can provide consistent estimates of α and β. To understand why, recall that (1) is $y_t = \alpha y_{t-1} + \beta z_t + u_t$. Now just change t to t-1 without loss of generality and write:
$y_{t-1} = \alpha y_{t-2} + \beta z_{t-1} + u_{t-1}$. Now multiply this equation by ρ and subtract from (1) to write
$[y_t - \rho y_{t-1}] = \alpha(y_{t-1} - \rho y_{t-2}) + \beta(z_t - \rho z_{t-1}) + u_t - \rho u_{t-1}$. Now rearrange this and write:
$y_t = (\alpha+\rho) y_{t-1} - \alpha\rho y_{t-2} + \beta z_t - \beta\rho z_{t-1} + v_t$, where v_t is from (2) above.

Step 1: Regress ADL(2,1) model instead of (1).
$y_t = a_1 y_{t-1} + a_2 y_{t-2} + a_3 z_t + a_4 z_{t-1} + v_t$
define $\tilde{\rho} = -\hat{a}_4/\hat{a}_3$, as a consistent estimator of ρ since $a_4 = -\beta\rho$ and $a_3 = \beta$ and in taking the ratio, β cancels. Inserting the extra lagged dependent variable is important.
Step 2: using the $\tilde{\rho}$ consistent estimate of ρ construct pseudo first differences as:
$Y_t = y_t - \tilde{\rho} y_{t-1}$
$Z_t = z_t - \tilde{\rho} z_{t-1}$ and $U_t = u_t - \tilde{\rho} u_{t-1}$. Now run ADL(1,0) regression similar to (1):
$Y_t = \alpha Y_{t-1} + \beta Z_t + U_t$. It is possible to show that plim $(\hat{\alpha} - \alpha) = 0$ and plim $(\hat{\beta} - \beta) = 0$

Pretest bias: Sims et al (1990, Etrica p 17) If nonstationarity itself is not the center of interest or if its form and degree is unknown, the discontinuity of asymptotic theory near unit roots raises serious problems of pretesting bias. Testing for presence of nonstationarity and then testing for its nature is needed before Granger causation test. But, the test results for nonstationarity and cointegration may be correlated. These problems are absent for the EF approach.

Cointegration Tests

The statistical problems of testing for cointegration are non-standard. There are seven estimators used for cointegration and corresponding tests given by Engle and Granger(1987, Econometrica, p.268). Stock and Watson (1988b) variable trend interpretation:

Before the advent of cointegration econometrics it was customary to separate the (long term) trend part from the (short term) cyclical part arising from the business cycles. Unfortunately the distinction is often artificial because there are important interactions between the two. Any shifts in the long run prospects for important economic variables can be expected to change short run behavior as well. The trends themselves are often stochastic and changing over time. Wrong conclusions regarding how the economy works and/or wrong forecasts are possible if regression methods are applied to time series data without proper understanding of the appropriate econometric theory. For example, due to the oil shock the trend in US GNP changed between 1960's and 1970's. One way to avoid the pitfalls is to include level variables along with differenced variables in the analysis.

Recall that cointegration is often between two I(1) variables. To test for I(1) one needs unit root tests. The usual AR(1) model $y_t = \rho\, y_{t-1} + \epsilon_t$ can be rewritten as $\Delta y_t = (\rho - 1)\, y_{t-1} + \epsilon_t$. Here we test whether the coefficient of y_{t-1} is zero against the alternative that it is NEGATIVE. This t-ratio is the most familiar version of the Dickey-Fuller (DF) test and the t-ratio does not have limiting normal distribution. The DF distribution is skewed to the left having a considerable mass in the negative region because of the bias in the presence of lagged dependent variables. The presence of an intercept alters the distribution and the errors ϵ_t may not be white noise. A parametric correction to this problem is the augmented Dickey-Fuller (ADF) test extended by Engle and Yoo (1990) from the regression:

$$\Delta y_t = (\rho - 1)\, y_{t-1} + \Sigma_{j=1}^{K} \eta_j\, \Delta y_{t-j} + \text{intercept} + \text{trend terms} + \epsilon_t$$

where the sampling distribution is the same as that of DF and does not depend on η_j parameters. This method requires the investigator to choose the order K using some criterion as AIC.

For cointegration tests we wish to make sure that the x_t and y_t are I(1) by using the above equation. The next step is to test for NON- COINTEGRATION. This means checking whether there is a unit root in the residuals of the cointegrating regression (y_t on x_t). If the series are cointegrated the residuals are stationary. Five percent critical values on ρ (first order autoregressive coefficient) are –1.64 for normal, –2.86 for DF and –3.34 for the case when the intercept and trend terms are present.

If one concludes that there is cointegration, the model should include both differences and error correction terms. If the trend term is included in the regression, critical value becomes –3.78. After all we are testing that the two series are not cointegrated even after extracting a trend. Sims, Stock and Watson (1990, Etrica) show if regressions involve I(1) variables and we are interested in θ, the coeff. of an I(0) variable with zero mean, then $\sqrt{T}(\hat{\theta} - \theta)$ is asymptotically normal. In some cases standard asymptotics applies to all the coefficients of interest to the economists. Individual coefficient tests are appropriate, but not joint tests. Although the two-step procedure is consistent for α it is not efficient. By contrast, Saikkonnen's method is consistent and efficient. Johansen's (1991) approach estimates a VAR by maximum likelihood subject to a rank constraint.

SEVEN tests Listed by Engle-Granger:

1. The Co-integration Regression Durbin Watson (CRDW test): $y_t = \alpha x_t + c + u_t$

ξ_1 = DW. The null hypothesis here is that DW = 0, (usually the null is that DW=2 or autocorrelation is zero). The interest is in testing whether the errors are stationary or not. It can be shown that if errors are non-stationary, DW $\rightarrow$ 0. Hence, DW can indicate nonstationarity. This is not the same as the traditional use of the DW statistic, where we are interested in checking whether the errors are serially uncorrelated (e.g. if they are iid), where a desired value of DW is about 2. Here, DW>0.386 gives a good indication of the existence of cointegration according a simulation reported by Engle and Granger.

2. Dickey Fuller (DF) Regression on residuals: $\Delta u_t = -\phi u_{t-1} + \epsilon_t$.

$\xi_2 = \tau$: the t statistic for ϕ. Again the idea is to estimate the long-run parameters by OLS regression in level variables and make sure that the residuals are not random walk. If the test suggests that the residuals are random walk, we reject the null hypothesis of no cointegration.

3. Augmented DF Regression on residuals: $\Delta u_t = -\phi u_{t-1} + b_1 \Delta u_{t-1} + \ \ldots \ + b_i \Delta u_{t-i} + \epsilon_t$.

$\xi_3 = \tau$ or t statistic for ϕ.

4. Restricted VAR: $\Delta y_t = \beta_1 u_{t-1} + \epsilon_{1t}$, $\Delta x_t = \beta_2 u_{t-1} + \gamma \Delta y_t + \epsilon_{2t}$.

$\xi_4 = \tau^2_{\beta 1} + \tau^2_{\beta 2}$.

5. Augmented VAR : Same as (4) but with p lags of Δy_t and Δx_t in each equation.

$\xi_5 = \tau^2_{\beta 1} + \tau^2_{\beta 2}$.

6. Unrestricted VAR : $\Delta y_t = \beta_1 y_{t-1} + \beta_2 x_{t-1} + c_1 + \epsilon_{1t}$, $\Delta x_t = \beta_3 y_{t-1} + \beta_4 x_{t-1} + \gamma \Delta y_t + c_2 + \epsilon_{2t}$.

$\xi_6 = 2[F_1 + F_2]$ where F_1 is the F statistic for testing β_1 and β_2 both equal to zero in the first equation, and F_2 is the comparable statistic in the second.

7. Augmented Unrestricted VAR: The same as (6) except for p lags of Δx_t and Δy_t in each equation.

$\xi_7 = 2[F_1 + F_2]$.

Suppose that x_t is stationary in first differences. In other words it is integrated of order one, that is $x_t \sim$ I(1), and it can be expressed as

$$\Delta x_t = g_x + \eta_t \tag{1}$$

where η_t is a stationary process with a mean of zero. If x_t is in logarithms, as is usually the case, then g_x is the average growth rate. Suppose also that y_t is I(1), but mean of g_y. Models of the form above are not unreasonable for many macroeconomic time series as they are capable of reproducing the kind of evolutionary behaviour often observed in practice. Now suppose that there is a relationship between x_t and y_t, so that although they are both non-stationary they tend to keep together in the long run. More specifically, they obey an equation of the form

$$y_t = v_0 + v x_t + u_t \tag{2}$$

where v_0 and v are parameters and u_t is a zero mean stationary disturbance term. Again if x_t and y_t are in logarithms, v can be interpreted as the long run elasticity of y with respect to x. Furthermore on taking first differences in (2), followed by expectations, the following relationship is seen to hold between the growth rates in the two series:

$$g_y = vg_x \qquad (3)$$

Finally if capital letters are used to denote antilogarithms the underlying long run relationship in (2) may be thought of as

$$Y = KX^v \text{ with } K = \exp(v_0) \qquad (4)$$

<u>Spurious Regression:</u> See Granger-Newbold (1974), Phillips (1986), Hamilton (1994, p. 557), Banerjee (1993, p.71). Let $y_t = y_{t-1} + u_t$ and $x_t = x_{t-1} + v_t$ both are I(1) Random Walks. Errors are uncorrelated. Yet in a regression of y_t on x_t the coefficients DO NOT tend to zero: $y_t = \beta_0 + \beta_1 x_t$ does not have $\beta_0 = 0$ and $\beta_1 = 1$. Note that if $\beta_1 = 0$, then $y_t = \beta_0 + x_t +$ error. This error cannot be I(0) because y_t is random walk I(1). This leads to internal <u>inconsistency in hypothesis testing with t of F tests.</u>

The estimate $\hat{\beta}_0/\surd T$ converges in law to $\sigma_u h_1$. (coeff diverges unless divided by the square root T)

The estimate $\hat{\beta}_1$ converges in law to $(\sigma_u/\sigma_v)\, h_2$. (coeff is inconsistent)

$$\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = \begin{bmatrix} 1 & \int_0^1 W_2 \\ \int_0^1 W_2 & \int_0^1 W_2 W_2{\prime} \end{bmatrix}^{-1} \begin{bmatrix} \int_0^1 W_1 \\ \int_0^1 W_2 W_1 \end{bmatrix}$$

where W_1 denotes a scalar standard Brownian motion and W_2 denotes vector valued standard Brownian motion, Hamilton, p. 559. The conventionally calculated F tests and t tests are also wrong. If a regression has $R^2 >$ DW, it suggests spurious regression.

Solutions for curing the spurious regr. problem: (Hamilton, p. 560).
1) Include lagged dep. and indep. variables as regressors.
2) Difference the data before regressing, use Δ rather than levels.
Differencing avoids spurious regr. problem, as well as, the problem of nonstandard distributions, but routine use of this method is not recommended, since the variables may be cointegrated. Also, if the data are AR(1) with coeff. close to 1, they may appear to be nonstationary. Differencing such data is not appropriate.
3) Use Cochrane-Orcutt or similar correction for serial correlation, since this correction is asymptotically like differencing the data, this also solves the problem.

Spurious regr problems occur if I(2) variable is regressed on I(1) or vice versa. It is not as bad if I(1) is regressed on I(0). Banerjee et al 1993, p.80. Sims et al (1990, Etrica, p.136) note that the practice of transforming data to stationarity is unnecessary in many cases. When are they necessary? Certain generalized cointegrating vectors reduce Y_t to a stationary process with mean zero. These have no problem. The problems arise for those “forbidden" linear combinations which orthogonal to the above cointegrating vectors. Coefficients of only that variable which is nonstationary AND which does not appear in ANY stationary linear combination, which has problems.

A test for cointegration is a useful method of distinguishing between equilibrium relations and spurious relations.

<u>Near integrated processes</u> (Banerjee, p 95): $y_t = \rho y_{t-1} + u_t$, where u_t is N(0,1), $|\rho|<1$ and the initial value is $y_0 \sim N(0, \sigma^2/(1-\rho^2))$. There appears to be a discontinuity near $\rho=1$ as the constant unconditional variance $\sigma^2/(1-\rho^2)$ which is not a function of time becomes $t\sigma^2$, which is a function and hence nonstationary. However if we let $\rho = 1+\epsilon$, such that $\epsilon<0$. then $\mathrm{var}(y_t) = \sigma^2(1+z^2+\cdots z^{2t-2})$ with $z=(1+\epsilon)$. Hence the $\mathrm{Var}(y_t) = \sigma^2[1+\epsilon(t-1)]$ to $O(\epsilon)$ In finite samples if ϵ is close to zero, it may be a better approximation to treat it as nonstationary I(1) rather than I(0).
If $\rho = \exp(\epsilon/T)$ then as $\epsilon=0$ this $\rho = \exp(0) = 1$. Define the functional

$$K_\epsilon(r)=W(r)+\epsilon\int_0^r \exp[\epsilon(r-s)]W(s)ds$$

known as Ornstein-Uhlenbeck process. For fixed r it is $N(0, (1/2)\epsilon^{-1}[\exp(2r\epsilon)-1]$. This is a first order diffusion process closely related to W(r). It is similar to an error correction process generated by the stochastic differential equation: $dK_\epsilon(s)=\epsilon K_\epsilon(s) + dW(s)$. Note that ϵ is like a noncentrality parameter and if it is zero, $K_\epsilon(r)=W(r)$

What about time trends?. If DGP is random walk but mistaken for $y_t = c+\gamma t+u_t$, a linear trend, Durlauf and Phillips (1988) show that under the null of $c=0=\gamma$, the distribution of $\hat{c}$ diverges, and the distribution of $\hat{\gamma}$ is such that the estimate does converge to zero, but the inference can be unreliable, since the t and F stats do NOT converge to zero. They are asymptotically unbounded with prob. 1. (Banerjee et al p.83)

Rethinking about Multicollinearity in the presence of I(1) variables:

In the past, it was common to delete multicollinear variables. For I(1) data such elimination jeopardizes the cointegration possibility.

$X'X$ is $O(T^2)$

$\alpha' X' X\alpha$ is O(T) if x is I(1) but $\alpha' x$ is I(0) due to cointegration.

If the dependent variable is also I(1), then elimination is ok. But Wiener based critical values need to be used. Economists need no longer assume long-run equilibrium relations, but can test for them.

Constancy of consumption - income ratio, capital-output ratio, wage share is total income, etc. was observed by economists long before, cointegration was developed. Once cointegration is tested and established, economists can move from I(1) space to I(0) space, where the well established tools are appropriate.

Hatanaka Scepticism:

Hatanaka (1996,Time Series Based Econometrics, Oxford, p.117) expresses scepticism about the usefulness of Johansen's methods for macroeconomics. He makes three remarks (pages 150-151).

[1] Consider GNP $\sim$ I(1), M (money supply) $\sim$ I(1), and INT (interest rate) $\sim$ I(0) is stationary. The variables are assembled in a vector (GNP, M, INT). The coefficients (0,0,1) yield a cointegrating vector, since INT is I(0). If money demand equation holds, M = f(GNP and INT), the coefficients (1, –1, –b) yield ANOTHER cointegrating vector, which alone is economically meaningful. The vectors (1, –1, 0) and (0, 0, 1) have no economic meaning.

[2] It is often claimed that cointegration is related to long-run equilibrium. However, in economic theory, long-run refers to the LENGTH of TIME needed by the system to restore equilibrium when the system is perturbed. By contrast, in cointegration analysis, the order of integration I(1) means long-run and I(0) means short-run. Short run relation may well provide cointegration in economics, e.g. many studies of uncovered interest parity (UIP).

[3] To be cointegrated from the viewpoint of economic theory, cointegration space should nullify both deterministic and stochastic trends. Stock and Watson (1989) investigate neutrality of money only in the context of stochastic trends. Hence their conclusions may not be valid.

Highlights of cointegration in Hatanaka (1996, p 161 and p. 201) summarize the advanced theory from his viewpoint. It involves VAR and VMA representations, perpendicular matrices, ranks, etc.