Cointegration in Single Equation Models

Artur Silva Lopes, April 2022 (rev. ed.)

ISEG–ULisboa

1 Introduction

We begin by studying a set of small useful properties and a first definition of cointegration.

1.1 Basic properties

These properties are almost axioms. Consider two constants different from zero, a and b. Then:

i. if $x_t \sim I(0) \Rightarrow (a + b\,x_t) \sim I(0)$;
   if $x_t \sim I(1) \Rightarrow (a + b\,x_t) \sim I(1)$.

ii. if x_t and y_t are both $I(0) \Rightarrow (a\,x_t + b\,y_t) \sim I(0)$;

iii. if $x_t \sim I(1)$ but $y_t \sim I(0) \Rightarrow (a\,x_t + b\,y_t) \sim I(1)$;

iv. *in general*, if $x_t \sim I(1)$ and $y_t \sim I(1) \Rightarrow (a\, x_t + b\, y_t) \sim I(1)$; *in general* (but not always), any linear combination of $I(1)$ series is still $I(1)$.

Actually, this last one is not really a property.

1.2 Cointegration: first definition

Definition: if $x_t \sim I(1)$ and $y_t \sim I(1)$, but there is a linear combination, $(y_t - \beta x_t) = u_t \sim I(0)$, then x_t and y_t are said to be cointegrated, $(x_t, y_t) \sim CI(1,1)$, and the vector $[1 \quad -\beta]'$ is the cointegration vector.

It is only demanded that the linear combination is I(0); it can have a non zero mean or it can even have a trend. But, if besides being $I(0)$, we want that u_t has a zero mean, we may need to introduce a constant and/or a linear trend term in the relation, as in

$$y_t - \gamma_0 - \gamma_1 t - \beta x_t = u_t \sim I(0).$$

So, a cointegration relation is special because it is an exception to the general rule.

If y_t and x_t are both $I(1)$, then, by the BN decomposition, both have a stochastic trend. But if $(y_t, x_t) \sim CI(1,1)$, i.e., if $y_t - \beta x_t = u_t \sim I(0)$, then that stochastic trend is annihilated; therefore, it is common to both series.

That is, although they appear to evolve in an arbitrary way, wandering stochastically, without any attraction for any fixed value or without any systematic behaviour, their long run behaviour is driven or guided by the same stochastic trend; in the long run they covary jointly, in a way to satisfy that (linear) relation. They evolve tending to follow or to keep up with each other.

Example: the paths of a drunk and his dog leaving a bar and returning home at 6am (Murray, 1994).

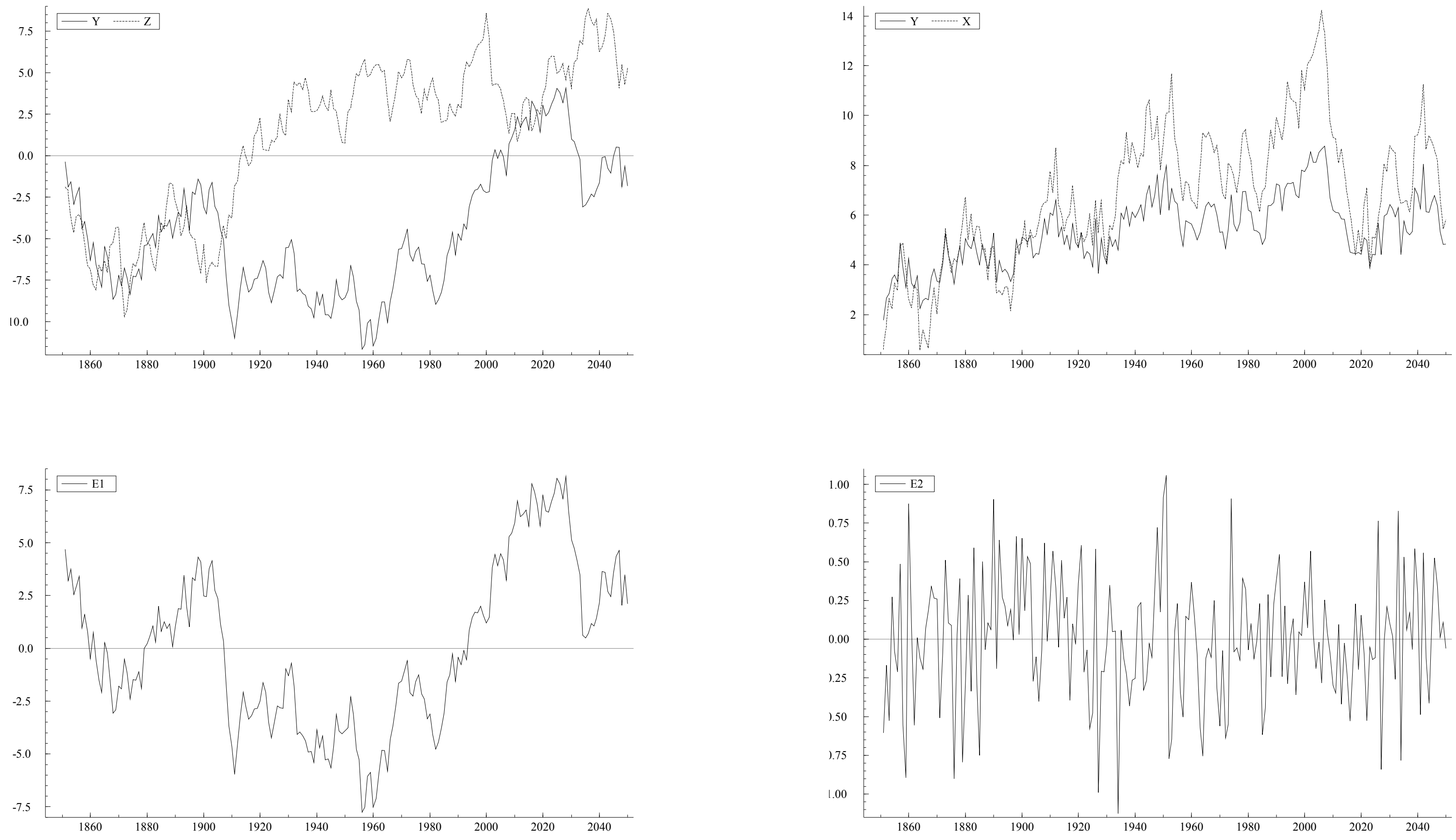


Figure 1: a) in the left upper panel the two series are not cointegrated, they are two independent random walks; in the lower panel, e_{1t} represents the OLS residuals from the regression of y_t on z_t. b) in the upper right panel the two series are cointegrated and below them the e_{2t} series is the series of OLS residuals OLS from the regression of y_t on x_t.

2 Cointegration and economic theory

Economic theory often suggests the existence of relations between variables, which may be interpreted as long-run equilibrium relations. This means that their behaviour in short time stretches may show only disequilibria, situations where the relation does not hold. But in the long run variables will tend to move in order to satisfy, approximately, that relation, with stationary deviations or errors, that do not tend to diverge through time, that show no tendency to accumulate.

It is as if in cases of large disequilibria or after some disequilibrium periods, some forces come into action that try to restore equilibrium. A long run equilibrium relation *tends* to hold *approximately*, *on average*, *in the long run*.

Cointegration reflects a long run equilibrium relationship, of a systematic long-run co-movement between variables.

The condition $y_t - \beta\, x_t = u_t \equiv 0, \forall t$ is highly demanding, it cannot be empirically observed.

But if despite individually each series being $I(1)$, $y_t - \beta\, x_t = u_t \sim I(0)$, preferably with zero mean, then u_t may be interpreted as the equilibrium error or deviation, as the disequilibrium, because:

a) the behaviour of *mean reversion* or *mean regression* ensures that it frequently returns to that zero mean (to equilibrium);

b) that it is merely temporary or transitory, without exhibiting any tendency to diverge through time;

c) since it has a finite and constant variance, it varies always inside certain bounds.

Example 1: permanent income hypothesis of consumption. A simple version implies that household consumption has two components,

$$c_t = c_t^P + c_t^T,$$

(permanent or long run component + transitory). Assuming that $c_t^T \sim I(0)$ and since, $c_t^P = \beta\, y_t^P$, with y_t^P permanent income, we have

$$c_t = \beta\, y_t^P + c_t^T,$$

which, provided $y_t^P \sim I(1)$, is a cointegration equation, with c_t^T the stationary equilibrium error.

Example 2: the purchasing power parity hypothesis (PPP). With p_t the log of the internal price level and p_t^* the one of external level, and with e_t the log of price of national currency in external currency, i.e., the exchange rate, it must hold that

$$e_t = p_t - p_t^* + u_t,$$

where the error represents the deviations from equilibrium (transportation costs, market frictions, custom taxes, etc.). However, since that error must be stationary and since e_t, p_t and p_t^* are typically $I(1)$, the equation

$$e_t - p_t + p_t^* = u_t \sim I(0)$$

represents a cointegration relationship and in this case economic theory goes even further because it even specifies the cointegration vector: $[1 \quad -1 \quad 1]$.

Example 3: stable money demand function. Let m_t denote the log of nominal money *stock* (e.g., M1) and p_t the log of internal price level. Often they are both I(2), but if the difference between them (i.e., the log of the quotient) is I(1), this is also a case of cointegration: a linear combination of I(2) series produces a series with a lower order of integration, $I(1)$, the log of real money.

Real money demand is considered to be a function of the volume of transactions in the economy, proxied by $\log(GDP) = y_t$ — demand for transactions motive

—, and also of some interest rate, r_t — demand for speculative motive. Then

$$\underbrace{m_t - p_t}_{I(1)} = \beta_1 + \beta_2\, y_t + \beta_3\, r_t + u_t,$$

which is a cointegration equation if both y_t and r_t are also $I(1)$ but $u_t \sim I(0)$.

There are many other examples because the definition of cointegration is not specially demanding.

Two observations:

a) since any linear combination of $I(0)$ series is still (trivially) $I(0)$, there is no cointegration between this type of series;

b) a spurious regression equation between $I(1)$ series is an equation where all series, including the one of the errors, are $I(1)$, and this reflects the absence of a cointegration relationship:

$$\underbrace{y_t}_{I(1)} = \alpha + \beta \underbrace{x_t}_{I(1)} + \underbrace{u_t}_{I(1)}.$$

3 Generalization

The general definition of cointegration is owed to Engle and Granger (1987).

Definition: consider a vector $\mathbf{y_t}$ of m series or variables: $\mathbf{y_t}' = [y_{1t}\ y_{2t}\ \ldots y_{mt}]'$. It is said to be cointegrated of orders (d, b), with d and b positive integers, and we write $\mathbf{y_t} \sim CI(d, b)$, if:

i. each component of $\mathbf{y_t}$ (each series) is $I(d)$: $y_{it} \sim I(d), i = 1, 2, \ldots, m$;

ii. there is a vector $\boldsymbol{\beta} \neq 0$ such that

$$\boldsymbol{\beta}'\mathbf{y_t} = \beta_1 y_{1t} + \beta_2 y_{2t} + \ldots + \beta_m y_{mt} = u_t \sim I(d-b),$$

with $d \geq b$; $\boldsymbol{\beta}$ is called the cointegration vector.

The definition demands that the linear combination of series integrated with the same order originates a series with a lower order of integration.

a) The most important case in economics is the $CI(1,1)$: because often the series "are" $I(1)$ but also because it translates economic relationships into long-run equilibrium relationships.

b) When only two $I(1)$ series are involved there is a single cointegration vector which is linearly independent.
Proof: by contradiction. Consider the cointegration vector $[1 \;\; \beta_1]'$:

$$y_{1t} + \beta_1 y_{2t} = u_{1t} \sim I(0),$$

with y_{1t} and y_{2t} both $I(1)$. Contradictorily, admit the existence of another cointegration vector, different from the first, $[1 \;\; \beta_2]'$, with $\beta_2 \neq \beta_1$ in order to be linearly independent:

$$y_{1t} + \beta_2 y_{2t} = u_{2t} \sim I(0).$$

Subtracting member by member:

$$\beta_1 y_{2t} - \beta_2 y_{2t} = \underbrace{u_{1t} - u_{2t}}_{I(0)},$$

because u_{1t} and u_{2t} are both $I(0)$. But the left member, equal to $(\beta_1 - \beta_2)y_{2t}$ can be $I(0)$ only in the trivial case when $\beta_1 = \beta_2$, contrarily to the initial assumption. Therefore, the second vector must be equal to the first.

Any linear combination of a cointegration vector is still a cointegration vector: if $[1 \;\; \beta_1]'$ is a cointegration vector $\Rightarrow$ $[k \;\; k\beta_1]'$ is too $(k \neq 0)$. Therefore, the number of cointegration vector is infinity. What is only one is the number of cointegration vectors that are *linearly independent*.

c) But with m series that are $I(1)$ there can exist, at most, $m-1$ cointegration vectors which are linearly independent. Indeed, there cannot exist m.

Let $\boldsymbol{\beta}_i$ denote each cointegration vector, $\boldsymbol{\beta}_i'\mathbf{y}_t \sim I(0)$ and $\mathbf{B}$ the matrix whose columns are the cointegration vectors, with r columns:

$$\mathbf{B} = [\boldsymbol{\beta}_1 \, \boldsymbol{\beta}_2 \;\; \dots \;\; \boldsymbol{\beta}_r] \, .$$

Begin by assuming, contradictorily, that there are m linearly independent cointegration vectors. Then, the matrix $\mathbf{B}$ is $(m \times m)$ and regular and $\mathbf{B}'\mathbf{y}_t$

is a vector of m $I(0)$ variables. Then,

$$(\mathbf{B}')^{-1}\underbrace{\mathbf{B}'\mathbf{y}_t}_{I(0)} = \mathbf{I}\,\mathbf{y}_t = \mathbf{y}_t$$

would also have to be, after all, a vector of $I(0)$ series, contradicting the initial assumption. Therefore, there cannot exist m linearly independent cointegration vectors .

d) The number of linearly independent cointegration vectors is the *cointegration rank* and it is represented with r: $r \leq m - 1$.

The possible existence of several cointegration vectors poses an economic interpretation problem, that is, one of identification.

For instance, if we augment the set of variables of the money demand function with inflation, it may occur that the number of cointegration relationships increases to two.

The possibility that $r > 1$ appears only in the framework of multi-equational or system modeling, which will not be addressed here.

e) The previous definition is the one of *stochastic cointegration* because the linear combinations of variables that eliminate unit roots are allowed to have linear trends. That is, each linear combination annihilates only the stochastic trend.

Deterministic cointegration is stronger because it demands that the same vectors $\boldsymbol{\beta}_i$ that eliminate unit roots also eliminate the deterministic trends that series might have. That is, based on the BN decomposition, for each series we have

$$y_{it} = \mathsf{DT}_{it} + z_{it},$$

where DT can be only a constant (or may even not exist), and z_{it} is the "noise function" of the series, that is, the stochastic component, which contains the stochastic trend. To allow that all variables have deterministic non zero trends:

$$\mathbf{DT}_t = \boldsymbol{\gamma}_0 + \boldsymbol{\gamma}_1 t,$$

with $\mathbf{DT}_t$ and $\boldsymbol{\gamma}_0$ and $\boldsymbol{\gamma}_1$ $(m \times 1)$ vectors.

Deterministic cointegration demands not only that $\boldsymbol{\beta}'\mathbf{z}_t \sim I(0)$ but also that the vector $\boldsymbol{\beta}$ satisfies $\boldsymbol{\beta}'\mathbf{DT}_t =$ constant. That is, that deterministic trends are also annihilated by the same vector that annihilates the stochastic trend; clearly, this condition is very demanding and it is rarely satisfied.

f) Admitting that cointegration is "only" stochastic, for the linear combination to have zero mean — making the interpretation of errors as equilibrium errors easier — it is often necessary to include an intercept in the equation and a deterministic trend term (and possibly seasonal dummies as well).

Assuming a single cointegration vector, we have

$$u_t = \boldsymbol{\beta}'\mathbf{y}_t - (\gamma_0 + \gamma_1 t),$$

that is, in matrix form:

$$\mathbf{u} = \mathbf{Y}\,\boldsymbol{\beta} - \;\mathbf{D}\,\boldsymbol{\gamma},$$

with $\mathbf{D}$ the matrix of observations of deterministic regressors and $\boldsymbol{\gamma}$ its vector of coefficients.

g) It may not exist any cointegration relationship with only two $I(1)$ series, but including a third (or a third and a fourth, etc.) might lead to a cointegration relationship. For instance, in the case of PPP theory, there are three series involved.

Problems to handle:

i) how to estimate the cointegration vector (assumed to be just one)?

ii) How to test the existence of cointegration?

The estimation methods assume that there is cointegration and, hence, they are preceded by tests. But we will begin by studying the first ones.

Estimation methods can be single equation or system or multi-equation. The first assume that $r = 1$ and allow estimating a single cointegration vector. The second not only allow estimating the possible different cointegration vectors but their number as well, that is, r. The most important of these is Johansen's method.

4 Estimating the Cointegration Vector

Three single equation methods will be presented: SOLS, DOLS and the estimator based in a dynamic model (ADL).

4.1 SOLS Estimation

Return to $u_t = \boldsymbol{\beta}'\mathbf{y}_t - (\gamma_0 + \gamma_1 t)$ and normalize the coefficient of the first variable, y_{1t}, that is, make $\beta_1 = 1$:

$$y_{1t} = \mathbf{d}_\mathbf{t}'\boldsymbol{\gamma} + \mathbf{y}_{2t}'\boldsymbol{\beta}^* + u_t,$$

with $\boldsymbol{\beta}^* = [-\beta_2^* \ \ldots \ -\beta_m^*]'$, with β_j^*, $j = 2, \ldots, m$ the remaining coefficients resulting from the normalization. Statistically, this normalization is purely arbitrary, that is, it could be different. In general, it is economic theory that indicates the regressand and the regressors.

The SOLS estimator, suggested by Engle and Granger, is simply the OLS estimator of this static equation — *static* OLS. The good news first.

If the variables are cointegrated, then it is likely that they are generated in a joint and simultaneous way $\Rightarrow$ endogeneity problem. But OLS is not inconsistent!

Suppose that the DGP is

$$\begin{cases} \lambda_1 y_{1t} - y_{2t} = u_{1t}, & \text{with } (1-\rho_1 L)u_{1t} = \epsilon_{1t},\ \epsilon_{1t} \sim iid(0,\sigma_1^2), \\ y_{1t} - \lambda_2 y_{2t} = u_{2t}, & \text{with } (1-\rho_2 L)u_{2t} = \epsilon_{2t},\ \epsilon_{2t} \sim iid(0,\sigma_2^2), \end{cases}$$

which is driven by the ρ_1 and ρ_2 parameters:

a) if $|\rho_1| < 1$ and $|\rho_2| < 1$, y_{1t} and y_{2t} are both $I(0)$;

b) if $\rho_1 = \rho_2 = 1$, y_{1t} and y_{2t} are both $I(1)$ and they are not cointegrated;

c) if, for instance, $\rho_1 = 1$ and $|\rho_2| < 1$, $(y_{1t}, y_{2t}) \sim CI(1,1)$ with cointegration vector $[1 \quad -\lambda_2]'$.

Indeed for this case, solving the first equation with respect to y_{2t} and replacing in the second:

$$y_{1t} = \underbrace{\frac{1}{1-\lambda_2\lambda_1} u_{2t}}_{I(0)} - \lambda_2 \underbrace{\frac{1}{1-\lambda_2\lambda_1} u_{1t}}_{I(1)} \sim I(1),$$

assuming that $\lambda_2\lambda_1 \neq 1$. Now, since $\rho_1 = 1$ the process u_{1t} is $I(1)$ because it is a random walk; and since $|\rho_2| < 1$, $u_{2t} \sim I(0)$ because it follows a stationary AR(1). Then, y_{1t} is $I(1)$.

Substituting again in the equation for y_{2t}

$$y_{2t} = \underbrace{\frac{\lambda_1}{1-\lambda_2\lambda_1}u_{2t}}_{I(0)} - \underbrace{\frac{1}{1-\lambda_2\lambda_1}u_{1t}}_{I(1)} \sim I(1).$$

Allowing us to conclude that the second equation is a cointegration one, with the mentioned cointegration vector. Write it as a regression equation:

$$\begin{aligned} y_{1t} &= \lambda_2 y_{2t} + u_{2t} \\ &= \lambda_2 y_{2t} + \rho_2 u_{2,t-1} + \epsilon_{2t} \\ &= \lambda_2 y_{2t} + \rho_2(y_{1,t-1} - \lambda_2 y_{2,t-1}) + \epsilon_{2t}, \end{aligned}$$

where the last equation derives from the second equation in the system. The error of the static cointegration equation (u_{2t}) has two components, both strongly correlated with the regressor: the first because it is a lag of a variable which is cointegrated with it and the second because it is a lag of itself, which is an $I(1)$ variable and hence strongly and positively autocorrelated. Therefore, the regressor of the static equation is endogenous.

But (as we will see) the existence of cointegration, including the nature of the regressor, $I(1)$, is sufficient to ensure that the endogeneity problem disappears asymptotically, and hence it does not affect the consistency of OLS estimation. Asymptotically, the variability of y_{2t} makes the endogeneity problem negligible.

Is there inconsistency of the spurious regression type involving $I(1)$ variables? No, in the case of cointegration OLS will tend to select the true parameter value or a closer one even more often than usual. Remember that OLS minimizes the estimated variance of the errors and, for any other λ_2 value the errors will be $I(1)$ and therefore they will have a variance which goes to ∞ when $T \to \infty$.

With cointegration the OLS estimator is even **super-consistent**: instead of converging to the true parameter value at rate $\sqrt{T}$ (as usual), i.e., instead of being $O_p(T^{-1/2})$, the convergence is at rate T, the estimator is $O_p(T^{-1})$.

Actually

$$\widehat{\lambda}_2 = \lambda_2 + \frac{\sum y_{2t} u_{2t}}{\sum y_{2t}^2},$$

where the sampling error denotes the usual expression $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$ for this case.

But the numerator is the sum of products of an $I(1)$ series by an $I(0)$, which diverges at speed or rate T, it is $O_p(T)$; and the denominator is a sum of squares of an $I(1)$ series, which also diverges but at rate T^2, i.e., it is $O_p(T^2)$.

Then, since it is the quotient of a $O_p(T)$ magnitude by one which is $O_p(T^2)$, $(\widehat{\lambda}_2 - \lambda_2)$ is $O_p(T^{-1})$, that is, to get a non-degenerate limit distribution it is necessary to scale the sampling error with T; it is not sufficient to do it with $\sqrt{T}$, as is usual.

It is $T(\widehat{\lambda}_2 - \lambda_2)$ that is limited and that has a defined distribution and not $\sqrt{T}(\widehat{\lambda}_2 - \lambda_2)$: the OLS estimator is consistent at rate T, it is *super-consistent*.

Intuitively, the large variability of the regressor dominates its correlation with the error, making endogeneity asymptotically negligible.

The bad news are in a larger number but only is very serious.
1. The endogeneity problem is asymptotically negligible but in small samples it can be significant: the OLS estimator of the static cointegration regressions is generally biased ($\mathsf{E}(\widehat{\lambda}_2) \neq \lambda_2$) and the bias can be large.

2. Generally, the OLS estimator is not asymptotically efficient. In the example, OLS is estimating a dynamically misspecified equation; since it neglects information it cannot be efficient.

3. In general, asymptotic inference methods are still invalid, as in spurious regressions. Why? Due to the usual regressor endogeneity and error autocorrelation typical of static regressions, which make appear nuisance parameters in asymptotic distributions, invalidating the employment of usual test statistics. In general, neither the t nor the F statistics can be validly employed to do inferences about cointegration parameters.

There is, however, an exception where:

a) the regression errors are normal and *iid*;

b) the regressors of the cointegration regression are strictly exogenous.

Returning to the general case of a vector of m variables, this is the situation of a *triangular system*:

$$\begin{cases} y_{1t} = \alpha + \boldsymbol{\beta}'\mathbf{y}_{2t} + u_{1t} \\ \mathbf{y}_{2t} = \mathbf{y}_{2,t-1} + \mathbf{u}_{2t}, \end{cases} \tag{1}$$

with a *recursive determination*, with $\mathbf{y}_{2t}$ determined first by the lower equations (similar to random walks), and thereafter determining y_{1t} through the first (cointegration) equation.

Additionally, it must also hold that

$$\begin{bmatrix} u_{1t} \\ \mathbf{u}_{2t} \end{bmatrix} \sim iid\mathcal{N}\left(\begin{bmatrix} 0 \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_{22} \end{bmatrix}\right),$$

in particular,

$$\mathsf{E}(u_{1t}\mathbf{u}_{2s}) = \mathbf{0} \Leftrightarrow \mathsf{E}(u_{1t}\Delta\mathbf{y}_{2s}) = \mathbf{0}, \forall t, s.$$

Hence, besides cointegration errors *iid* and normal, the regressors are strictly exogenous. In this case, exact distribution theory is valid: the usual t and F statistics have *exact* t and F distributions, respectively. But this is not a surprising result: the assumptions of the classical model hold, and so its results are valid, there is no novelty.

It is true that the intervening variables are neither stationary nor ergodic, but the classical model does not require any condition about that.

If the normality assumption is relaxed and the errors are only *iid*, not necessarily normal, the distributions of test statistics are valid only asymptotically. Moreover, OLS is asymptotically efficient. But this is also not much realistic: not only there cannot exist any *feedback* effect but also the errors of a static equation, with no dynamics, cannot be autocorrelated.

The solutions to overcome the general inefficiency of the OLS estimator and to allow making inferences about the cointegration parameters are (besides those that we will study):

a) the FIML estimator by Johansen, in the multi-equation framework;

b) the 3 steps method by Engle and Yoo, simple but rarely used;

c) the FM-OLS (*fully modified*) estimator by Phillips and Hansen, where two non parametric corrections are applied to the OLS estimator.

And if, given m variables $I(1)$ there are $m-1>1$ cointegration vectors linearly independent (violating a crucial assumption in our approach)? In that case, the OLS estimator may be estimating a linear combination of those $m-1$ vectors from the base of the cointegration space.

4.2 DOLS estimator

The DOLS (*dynamic* OLS) estimator was proposed at the beginning of the 90's, separately and independently, by Saikkonen, Phillips and Loretan and Stock and Watson. It has good properties, despite its relative simplicity, and it should be used more frequently.

Consider the previous system of equations, (1), but suppose now that the regressors $\mathbf{y}_{2t}$ are endogenous, Cov$(u_{1t}, \mathbf{u}_{2t}) \neq 0$, violating a crucial assumption of both the classical model and the model with pre-determined regressors. Consider the linear projection of u_{1t} on $\mathbf{u}_{2,t-p}, \mathbf{u}_{2,t-p+1}, \ldots,$ $\mathbf{u}_{2,t-1}$, $\mathbf{u}_{2t}$,$\mathbf{u}_{2,t+1}, \ldots \mathbf{u}_{2,t+p}$ and let $\tilde{u}_t$ denote the respective error:

$$u_{1t} = \sum_{s=-p}^{p} \boldsymbol{\gamma}_s' \mathbf{u}_{2,t-s} + \tilde{u}_t.$$

Therefore, by definition of linear projection, i.e., by construction, $\tilde{u}_t$ is not correlated with $u_{2,t-s}$ for $s = -p, -p+1, \ldots, p$. But since $u_{2t} \equiv \Delta y_{2t}$, the cointegration equation can now be written as

$$y_{1t} = \alpha + \boldsymbol{\beta}' \mathbf{y}_{2t} + \sum_{s=-p}^{p} \boldsymbol{\gamma}_s' \boldsymbol{\Delta} \mathbf{y}_{2,t-s} + \tilde{u}_t,$$

where there is no correlation between the regressors and the error provided that $\text{Cov}(u_{1t}, u_{2,t-s}) = 0$ for $|s| > p$.

That is, provided there is no correlation between the errors of the system equations distanced more than p time periods, the regressors of the cointegration equation augmented with the lags and the future values are now strictly exogenous; the initial correlation is presumed to be totally captured by the introduced *lags* and *leads*.

Hence, the parametric correction of OLS asymptotically eliminates the endogeneity effect on the distribution of the OLS estimator: all that is needed is to add to the static regression leads and lags of the differenced regressors.

How to choose p? We can use asymptotic distribution theory to eliminate non significant regressors, adopting a modeling strategy from general to specific (GTS): start the process with a number of leads and lags that appears to be sufficient to capture any correlation between the original regressors and the errors and "test down", eliminating regressors in first differences that are irrelevant.

If $\tilde{u}_t$ is not autocorrelated, the DOLS estimator is asymptotically efficient and the t- and F-statistics on the cointegration parameters are asymptotically valid. Including *leads* and *lags* is often sufficient to "whitten" the errors: "dynamizing" the model that way may remove the autocorrelation of original errors rendering unnecessary any further correction.

But if the $\tilde{u}_t$s are autocorrelated, additional corrections are necessary. The different methods diverge with respect to this: the additional correction may be parametric or non-parametric. Phillips and Loretan: including lags of the equilibrium error in the cointegration equation, as in the first equation from the system example of the previous subsection:

$$y_{1t} = \alpha + \boldsymbol{\beta}'\mathbf{y}_{2t} + \sum_{s=-p}^{p} \boldsymbol{\gamma}_s' \boldsymbol{\Delta}\mathbf{y}_{2,t-s} + \sum_{s=1}^{p} \theta_s (y_{1,t-s} - \boldsymbol{\beta}'\mathbf{y}_{2,t-s}) + \epsilon_t,$$

which demands NLLS estimation.

Another method: correcting the test statistics using a parametric estimate of the long-run variance of the errors from the dynamic equation.

4.3 A small example

Consider again the example of the consumption function for the portuguese economy. Assuming that there is cointegration, SOLS provides the following estimate of the long-run relationship

$$\widehat{LCP} = 0.099 + 0.023A75 + 0.935LRD + 0.192LSR - 0.004INF$$

where $A75$ represents an impulse dummy variable, equal to 1 only in 1975, and where the usual standard errors have been omitted because the t-ratios are unlikely to be valid (not only due to the likely autocorrelation of the errors — strongly indicated by the usual statistics — but also due to the much plausible regressor endogeneity and, in particular, of LRD).

Still assuming that there is cointegration, DOLS was initiated with a model with leads and lags until order 2. With some simplifications one arrives at:

$$\begin{aligned}\widehat{LCP}_t = \; & -0.037 + 0.105 A75_t + 0.976 LRD_t - 0.012 INF_t + 0.160 LSR_t + 0.427 \Delta LRD_{t-1} \\ & + 0.002 \Delta INF_t - 0.007 \Delta INF_{t+1} - 0.009 \Delta INF_{t+2} - 0.714 \Delta LSR_t - 0.619 \Delta LSR_{t+2},\end{aligned}$$

whose standard errors for the estimators of the cointegration coefficient are, respectively, 0.124, 0.030, 0.021, 0.0015 and 0.062.

ΔINF_t is not statistically significant at the usual levels (p-value of 0.106) but its presence seems important to ensure the absence of symptoms of error autocorrelation; since these are non-existent (p-values for the BG(1) and BG(3) of 0.867 and 0.143, respectively), usual inference methods appear to be valid. For instance, the test of $H_0 : \beta_3 = 1$ $vs.$ $H_1 : \beta_3 \neq 1$(unit consumption-income elasticity), can now be based in the t-statistic $= \frac{0.976-1}{0.021} = -1.143$, insignificant at the usual levels.

Can the errors of the equation estimated with DOLS be employed to analyse the existence of cointegration? No, because they are different from the errors of the static cointegration relation; they are not the entity represented here with u_t. DOLS assumes the existence of cointegration; it is not useful to test whether it exists.

4.4 The Estimator from the ADL Model — Review

ADL models also allow obtaining easily the estimates for the long-run multipliers, which are really the cointegration parameters, i.e., the coefficients of the long-run equilibrium relationships. The preferred form is the Bardsen one.

Considering an ADL(r,s) with k explanatory variables assumed as (weakly) exogenous, we have

$$\Delta y_t = \mu - A(1) y_{t-1} + \sum_{i=1}^{r-1} \delta_i \Delta y_{t-1} + \mathbf{B}(1)' \mathbf{x}_{t-1} + \sum_{j=0}^{s-1} \boldsymbol{\gamma}_i' \boldsymbol{\Delta} \mathbf{x}_{t-j} + \epsilon_t,$$

where $A(L)$ denotes the autoregressive polynomial, and $\mathbf{B}(L)$ the vector of polynomials of distributed lags for the exogenous variables $(B_j(L) = \beta_{0j} + \beta_{1j}L + \beta_{2j}L^2 + \ldots + \beta_{s_j,j}L^{s_j}, j = 1, \ldots, k)$.

Excluding the constant, the cointegration parameters are given by

$$\begin{aligned} \boldsymbol{\lambda} = &\ [\boldsymbol{\lambda}_1 \ \ \boldsymbol{\lambda}_2 \ldots \ \boldsymbol{\lambda}_k]' \\ = &\ A(1)^{-1} \mathbf{B}(1)'. \end{aligned}$$

If one wishes to include a constant in the cointegration relationship, one just makes $\lambda_0 = \mu / A(1)$.

4.5 Revisiting the example

Recall that starting with an ADL(3,3) and without any *dummy* variable I have got*:

$$\widehat{LCP} = 0.183 + 0.914LRD + 0.358LSR - 0.00034INF.$$

In the table summarizing estimation results notice: 1) the coherence of the sign for the coefficient of inflation produced by the different methods; 2) the plausibility of the signs and the magnitudes of the estimated coefficients; the only exception is the estimate for the coefficient of LRD with DOLS, which appears to be very large. However, it is also with DOLS that the influence from the wage variable on consumption appears weaker. On the contrary, this effect appears

*When the *dummy* is included but LSR is excluded, one gets:

$$\widehat{LCP} = 0.175 + 0.221A75 + 0.958LRD - 0.0099INF.$$

Table 1. Estimates for the cointegration parameters

variable	OLS	DOLS	ADL
C	0.099	−0.037	0.183
$A75$	0.023	0.105	—
LRD	0.935	0.976	0.914
LSR	0.192	0.160	0.358
INF	−0.004	−0.012	−0.0003

more important in the case of the ADL, which is also the case where the long-run elasticity consumption-income is smaller.

5 Testing for Cointegration

Consider again the **potential** cointegration equation, possibly with deterministic regressors:

$$y_{1t} = \mathbf{d}_t'\boldsymbol{\gamma} + \mathbf{y}_{2t}'\boldsymbol{\beta} + u_t.$$

If the variables are cointegrated $\Rightarrow$ u_t is the error of the long-run equilibrium relationship and hence $u_t \sim I(0)$; on the contrary, if they are not cointegrated $\Rightarrow u_t \sim I(1)$.

Therefore, a test procedure consists of testing

$$H_0 : u_t \sim I(1) \Leftrightarrow \text{non cointegration , vs.}$$
$$H_1 : u_t \sim I(0) \Leftrightarrow \text{cointegration.}$$

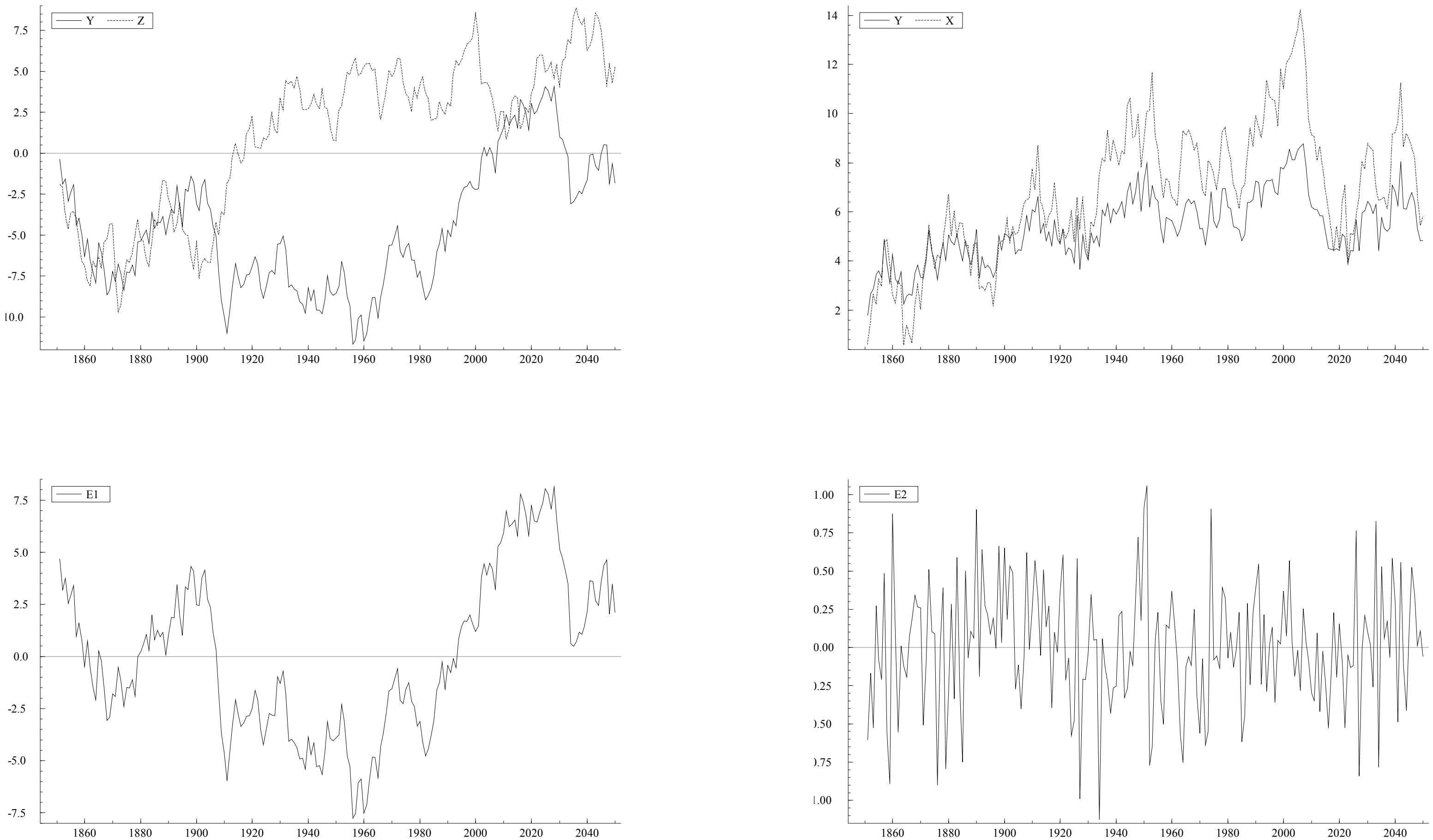
As the most popular unit root tests are those where this assumption is the null hypothesis and since there are many of these tests, they are the most common. If the cointegration vector is known (provided by economic theory) it is imposed on data and a test for cointegration is easy to do performing a vulgar UR test on u_t.

5.1 Engle-Granger Tests

Generally, however, the cointegration vector is not known. Hence, u_t is not observable and the previous tests cannot be performed. Engle and Granger proposed replacing the errors u_t, unobserved, with OLS residuals, e_t (or $\widehat{u}_t$). That is: 1) the potential cointegration equation is estimated and e_t are obtained; 2) the test of

$$H_0 : e_t \sim I(1) \Leftrightarrow \text{non cointegration, vs.}$$
$$H_1 : e_t \sim I(0) \Leftrightarrow \text{cointegration,}$$

substitute of the test above, is performed with a Dickey-Fuller regression.



Returning to figure 1, in the left panel the series y_t and z_t are $I(1)$ and non-cointegrated: they are two independent random walks, generated with independent shocks. In the lower graph the series of OLS residuals from the regression between the two series has a typical shape of an $I(1)$ series: although the presence of the constant ensures that residuals have zero mean, they rarely return to that mean, and their behaviour is very smooth.

The two series from the upper right panel are ($I(1)$ and) cointegrated: although sometimes they evolve in opposed ways in general they tend to vary jointly, moving together in the long-run. Therefore, the series of OLS residuals from the regression between the two has a clearly stationary appearance: very oscillating, frequently crossing the x-axis, etc.

The drawback is that since OLS chooses the cointegration vector so as to minimize the residual variance, it forces them to appear the most stationary as possible (even when they are really not). Therefore, the DF distribution is not valid to perform DF tests on the residuals: it is the Engle and Granger (EG) distribution, with even larger (in absolute value) critical values.

Generally, the test regressions are of the form

$$\Delta e_t = \phi e_{t-1} + \sum_{i=1}^{k} \gamma_i \Delta e_{t-i} + \epsilon_t,$$

where e_t are the OLS residuals from the potential cointegration equation, $\phi = \rho - 1$ and k is large enough (but not too much) to "whiten" the errors. That is, generally the constant is not included in this regression because it must have been already included in the potential cointegration equation. A trend term is not included as well because the variables included in that regression must already include it (implicitly or explicitly).

The distributions depend on the number of involved variables: the more variables are included the more the residuals are "made up" or "masked" by OLS; hence, the critical values must be also more demanding. These are presented in Table 2 and, in general, they are from Phillips and Ouliaris (1990).

There are 3 cases to consider:

1. $\mathsf{E}(\Delta y_{1t}) = 0$ and $\mathsf{E}(\Delta \mathbf{y}_{2t}) = 0$, i. e., none of the series has drift. This is the simplest case and the critical values are those from the upper panel of the table.

2. $\mathsf{E}(\Delta \mathbf{y}_{2t}) \neq 0$ and $\mathsf{E}(\Delta y_{1t})$ may be 0 or not: at least one of the regressors has drift (and the dependent variable may have it or not). Even if there are several regressors with drift, their linear trends can be combined in only one and the test regression may be written as the regression of y_{1t} on a constant, $g - 1$ driftless regressors and one regressor with drift. Since the asymptotic statistical behaviour of the regressors with drift is dominated by their drift $\Rightarrow$ the asymptotic behaviour of the $I(1)$ regressor with drift is asymptotically equivalent to a linear trend (t). Hence, asymptotically the residuals also behave as those of a regression with constant, $g - 1$ regressors with drift and a trend. In this case we must use the lower panel of the table.

3. $\mathsf{E}(\Delta \mathbf{y}_{2t}) = 0$ and $\mathsf{E}(\Delta y_{1t}) \neq 0$: y_{1t} has drift but $\mathbf{y}_{2t}$ does not (example: explaining the behaviour of investment, growing with time, on the basis of

Table 2. Asymptotic critical

values for the EG tests (DF on the residuals)

g	1%	5%	10%
	caso A: driftlesss regressors		
1	−3.96	−3.37	−3.07
2	−4.31	−3.77	−3.45
3	−4.73	−4.11	−3.83
4	−5.07	−4.45	−4.16
5	−5.28	−4.71	−4.43
	caso B: regressors with drift		
1	−3.96	−3.41	−3.13
2	−4.36	−3.80	−3.52
3	−4.65	−4.16	−3.84
4	−5.04	−4.49	−4.20
5	−5.36	−4.74	−4.46

Souces: Phillips and Ouliaris (1990), tables IIb) and IIc), p. 190, except the first line of the lower panel, taken from Fuller (1996), table 10.A.2. g denotes the number of $I(1)$ regressors.

an interest rate). To remove the trend from the errors we must include the t term in the regression. Then, the statistic obtained from the regression with constant, trend and g $I(1)$ regressors has the distribution provided by the lower panel of the table but with $g+1$ regressors; for instance, if $g=3$ the 5% critical value is -4.49.

Question: should one include t in the regressions? This is because in cases 1 and 2 we can include the trend term provided we use the lower panel with $g+1$ regressors. This type of practice would have the benefit of reducing the cases for the critical values to a single one. However, simulation studies show that including t in the test regressions, even when there is a justification, reduces their power, and therefore it should be avoided (in the test regressions only, to be clear).

And what must be the order for the autoregression? These "AEG" tests were not studied the same way the ADF were. We will continue to privilege the GTS t-sig method and the use of the AIC criteria.

Although asymptotically equivalent, in finite samples the test results can be sensitive to the variable that is chosen as dependent.

Ng and Perron (1997): if the first difference of a series has negative autocorrelation, the estimates based on the equation where that series is the dependent variable always have better properties than those that are normalized with another dependent variable. Since UR tests reject the null hypothesis more often for those series $\Rightarrow$ the "less integrated" variable must be used as dependent.

Intuitively: it is better that the regressors have much variability to identify their effects precisely; now, variables with more variability should be the "most integrated".

If AEG tests do not provide statistical evidence for cointegration, that indicates that the analysed static regression, after all, seems to be a spurious regression (although it may not be one in the full sense of the word).

5.2 Revisiting the example

Question: but is there really cointegration between LCP_t, LRD_t, LSR_t and INF_t (including also a constant and the *dummy*)?

With AEG test regressions, starting with $k_{MAX} = 3$ and using GTS to determine the optimal lag number I arrived at $\widehat{k} = 0$ and obtained $t_\phi = AEG(0) = -3.08$ ($\phi = \rho - 1$). Since this corresponds to case 2 because LRD_t has drift and since $g = 3$, the asymptotic critical value at 5% is -4.16:

$$CR = \{t_\phi : t_\phi < -4.16\}.$$

We are far from obtaining evidence for the existence of cointegration. When the roles of variables are changed and LRD_t is used as dependent: $\widehat{k} = 0$ and $t_\phi = AEG(0) = -3.29$, which is a bit better but still insufficient to reject the null for non-cointegration. Therefore, although not spurious in the usual negative sense, the previous regression appears to have no value, not corresponding to a long-run equilibrium relationship. But the small T and the nature of the test may be impeding a more positive conclusion.

6 Cointegration and the ECM

Even in the most favourable case, the one of cointegration, **in general**, usual inference methods are not valid.

Possible solution: use the first differenced series to be able to resort to conventional methods. Problem: if there is cointegration, the information about the long-run (involving the levels of variables) is being neglected. Simple short-run dynamic relationships are used that ignore the restriction on the existence of a long-run equilibrium relationship.

6.1 The Granger Representation Theorem

To overcome this problem we must use the error correction model (ECM), that reconciles short-run dynamic adjustment with the long-run equilibrium relationship. Justification: the **Granger Representation Theorem** (*light* version):

If $(y_t, x_t) \sim CI(1,1)$ with cointegration parameter λ, then the variables can be represented through the vector ECM:

$$\begin{cases} \Delta y_t = \delta_1 + \alpha_1(y_{t-1} - \lambda x_{t-1}) + \sum \gamma_{1i} \Delta y_{t-i} + \sum \delta_{1j} \Delta x_{t-j} + \epsilon_{1t} \\ \Delta x_t = \delta_2 + \alpha_2(y_{t-1} - \lambda x_{t-1}) + \sum \gamma_{2i} \Delta y_{t-i} + \sum \delta_{2j} \Delta x_{t-j} + \epsilon_{2t} \end{cases}$$

where at least one of the αs is different from zero ($\alpha_1 \neq 0 \vee \alpha_2 \neq 0$). The reverse is also true: if the variables are $I(1)$ and they admit a vector ECM representation, then they are cointegrated.

1. The vector ECM is a VAR model in the first differences of the variables augmented with the error correction term.

2. Since all involved variables are in the $I(0)$ form, usual inference methods should be valid.

3. The indexes for the summations are omitted on purpose because there is no restriction for this issue.

4. The interest lies almost always only in the first equation, which represents the distribution of y_t conditional on x_t and on the lags of both. Now, estimation and inference methods based on that equation only, neglecting the second one, are valid and efficient provided the restriction $\alpha_2 = 0$ holds. x_t is said to be weakly exogenous[†] for the parameters (of the first equation) of

[†]Let $f(y, x; \theta)$ denote the joint density of variables Y and X. It can always be factorized as

$$f(y, x; \theta) = g(y|x; \phi_1)h(x; \phi_2).$$

Suppose that interest lies on a certain parameter vector, ψ. Then: a) if ψ is a function only of ϕ_1, $\psi = l(\phi_1)$; b) and if ϕ_1 and ϕ_2 are "variation free", we say that X is weakly exogenous for ψ and inference based in $g(y|x)$ is valid and totally efficient.

On the notion of "variation free": consider the parameter spaces of ϕ_1 and ϕ_2, $\phi_1 \in \mathbf{\Phi}_1$ and $\phi_2 \in \mathbf{\Phi}_2$; to be variation free means that $(\phi_1, \phi_2) \in \mathbf{\Phi}_1 \times \mathbf{\Phi}_2$, that is, the space $\mathbf{\Phi}_1$ is not a function of ϕ_2 and the space $\mathbf{\Phi}_2$ is not a function of ϕ_1, that is, knowledge about the value of a parameter does not provide any information about the potential range of values for the other.

the ECM and, in that case, the second equation is seen as representing the marginal distribution of x_t.

But if $\alpha_2 \neq 0$ the isolated estimation of the first equation is not efficient because the parameter λ appears in both equations. The analysis should begin estimating jointly the two equations and testing $H_0 : \alpha_2 = 0$, that is, doing a test for weak exogeneity. It seems to be absurd but one should initiate the analysis estimating jointly the two equations to know whether that estimation can be skipped.

Since we are restricted to single equation methods, it will be assumed that $\alpha_2 = 0$ and attention will be focused on the conditional ECM.

6.2 The t-ECM test

The Granger Representation Theorem suggests an alternative way to test the existence of cointegration: testing the presence of an error correction mechanism, with the t-ECM test (that we have seen in Cha. 3).

In the first equation of the system, we want to test

$$H_1 : \alpha_1 = 0 \quad vs. \quad H_1 : \alpha_1 < 0.$$

Still assuming weak exogeneity of the regressor(s), the test can be performed in ADL models, as in the ADL(1,1), for instance,

$$y_t = \mu + \alpha y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \epsilon_t.$$

Written in the Bardsen form,

$$\Delta y_t = \mu + (\alpha - 1) y_{t-1} + \beta_0 \Delta x_t + (\beta_0 + \beta_1) x_{t-1} + \epsilon_t,$$

it is about testing the necessary (but not sufficient) condition for stability

$$\begin{aligned} H_0 &: \alpha = 1 \Leftrightarrow \ -A(1) = 0, \ \ vs. \\ H_1 &: \alpha < 1 \Leftrightarrow \ -A(1) < 0. \end{aligned}$$

And similarly for the second (in)equality for ADL models with more regressors.

But under H_0 the tested coefficient is from an $I(1)$ variable; even written in the ECM form, the coefficient cannot be written as one of an $I(0)$ variable with zero mean, and so the result from Sims, Stock and Watson cannot be employed. The distribution of the test statistic is *non-standard*.

A coarse critical value, approximate, for 5% level tests is, from the tables by Ericsson and MacKinnon (2002):

$$CV(0.05) \approx -3.0 - 0.2m - 0.3(nd - 1),$$

where m is the total number of $I(1)$ involved variables $(g+1)$ and nd the number of deterministic terms.

Often the EG tests do not provide evidence for cointegration but the t-ECM test does. It is a problem of low power from EG tests. Why? Because there is a common factor restriction that, when it does not hold, reduces the power of tests.

Example: consider the ADL(1,1), written in the ECM form:

$$\Delta y_t = (\alpha - 1)(y_{t-1} - \lambda x_{t-1}) + \beta_0 \Delta x_t + \epsilon_t,$$

where $\lambda = (\beta_0 + \beta_1)/(1 - \alpha)$. Subtracting $\lambda \Delta x_t$ to both members:

$$\underbrace{\Delta(y_t - \lambda x_t)}_{\Delta u_t} = \underbrace{(\alpha - 1)}_{\phi} \underbrace{(y_{t-1} - \lambda x_{t-1})}_{u_{t-1}} + \underbrace{(\beta_0 - \lambda)\Delta x_t + \epsilon_t}_{v_t},$$

where v_t is white noise only when $\beta_0 = \lambda$, that is, if the short and long-run multipliers are identical, which is a very strong dynamic restriction, typical of static models with autocorrelated errors; generally this restriction is invalid, and so the estimation of the EG test equation will be inefficient and the test will have low power.

In the example, if the ADL includes only the level variables LCP, LRD and INF and the impulse *dummy* $A75$ (but not LSR), one gets $t_{MCE} = -4.26$. Since

$$VC(0.05) \approx -3 - 0.2 \times 3 - 0.3(1 - 1) = -3.6,$$

there is strong evidence against the null, that is, strong evidence for cointegration (contradicting the EG test). And one must choose the inference provided by the t-ECM model.

7 Estimation of the conditional ECM

7.1 One step estimation

There is nothing new: it consists of estimating the ADL model, preferably in the Bardsen form, the estimation of the cointegration (long-run) parameters and the short-run dynamics made simultaneously.

Again the example from ADL models but the version that includes the $A75$ *dummy* (and excludes LSR):

$$\begin{aligned}\widehat{\Delta LC_t} = \ & 0.061 + 0.062A75 - 0.271LCP_{t-1} + 0.258LRD_{t-1} - 0.003INF_{t-1} \\ & +0.403\Delta LC_{t-1} + 0.508\Delta LR_t - 0.267\Delta LR_{t-2}.\end{aligned}$$

7.2 Two steps method

This was the estimation method proposed at an early stage by Engle and Granger and it dominated empirical applications for several years.

In the first step the cointegration vector is estimated (with OLS) using the static regression. Objective: to obtain the OLS residuals, e_t or $\widehat{u}_t$, which are estimates of equilibrium errors.

In the second step, those estimates are inserted in the ECM equation, as if they are the true values of those errors:

$$\Delta y_{1,t} = \delta + \phi e_{t-1} + \sum \theta_i \Delta y_{1,t-i} + \sum \boldsymbol{\gamma}_j' \mathbf{y}_{2,t-j} + \epsilon_{1t}.$$

Since OLS is superconsistent in the first step, substituting cointegration parameters by their estimators does not affect, asymptotically, the estimation properties and the validity of inference methods. But simulation studies have shown that the one step estimator appears to have better unbiasedness and efficiency properties.

The $t - ECM$ test may be also based in this model. However, due to reduced efficiency in small samples, the power will tend to be low as well.

It is not mandatory that in the first step the estimates of equilibrium errors are obtained with OLS applied on the static regression. Any other cointegration vector estimator may be employed in the first step: for instance, the DOLS estimator or even Johansen's (although this one does not make much sense for this).