COINTEGRATION AND COMMON TRENDS

Econometrics C ♦ Lecture Note 6
Heino Bohn Nielsen
February 6, 2012

In this note we discuss some important issues in regression models for non-stationary time series. It is illustrated how linear combinations of non-stationary time series are non-stationary in general, and *cointegration* is defined as the special case where a linear combination is stationary. We emphasize that relations between non-stationary variables can only be interpreted as defining an equilibrium if the variables cointegrate, and we discuss *error-correction* as the force that sustain the equilibrium relation. We then present some single-equation tools for cointegration analysis, e.g. the so-called Engle-Granger two-step procedure and cointegration analysis based on unrestricted ADL models. We show how to estimate the cointegrating parameters and how to test the hypothesis of no-cointegration. Towards the end of the note we discuss some limitations of the single-equation approach.

Outline

1 Unit-Root Time-Series and Cointegration

In this section we look at linear combinations of unit root non-stationary time series and define the concept of cointegration. To simplify the notation we consider the case of $p = 2$ variables in most of the presentation below, but the discussion is easily extended to more variables.

Let x_{1t} and x_{2t} be two time series that are integrated of first order, I(1). We can write the two processes on the form

$$\begin{aligned} x_{1t} &= \tau_{1t} + \text{stationary process} + \text{initial value} \quad (1) \\ x_{2t} &= \tau_{2t} + \text{stationary process} + \text{initial value}, \quad (2) \end{aligned}$$

where τ_{1t} and τ_{2t} are random walk components generated by unit roots. We often refer to τ_{it} as the *stochastic trend* of x_i.

Next define the linear combination, $z_t := \beta' x_t$, where x_t is a vector of variables, and β is a vector of weights in the linear combination, i.e.

$$x_t = \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} \quad \text{and} \quad \beta = \begin{pmatrix} 1 \\ -\beta_2 \end{pmatrix}.$$

Inserting (1) and (2), we can write the linear combination as

$$\begin{aligned} z_t &= \beta' x_t = \begin{pmatrix} 1 & -\beta_2 \end{pmatrix} \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = x_{1t} - \beta_2 x_{2t} \\ &= \tau_{1t} - \beta_2 \tau_{2t} + \text{stationary process} + \text{initial value}. \quad (3) \end{aligned}$$

We note that z_t contains the random walk component, $\tau_{1t} - \beta_2 \tau_{2t}$, and in most cases z_t will also be I(1). The result that a combination of I(1) variables is *in general* I(1) can easily be extended to higher order of integration, and a combination of variables integrated of order 2 (say), will also be I(2) in general.

An exception from this result is if there exist a vector, β, so that z_t defined in (3) is a stationary process. This property is denoted *cointegration* and the vector β is called a *cointegration vector*. For cointegration we need τ_{it} to be *common*, that is generated by the same underlying random walk, τ_t^*, i.e.

$$\tau_{1t} = \theta_1 \tau_t^* \quad \text{and} \quad \tau_{2t} = \theta_2 \tau_t^*.$$

If we choose $\beta_2 = \theta_1 / \theta_2$ we have from (3) that

$$z_t = \underbrace{\theta_1 \tau_t^* - (\theta_1/\theta_2)\, \theta_2 \tau_t^*}_{=0} + \text{stationary process} + \text{initial value}.$$

The *common stochastic trends* cancel and z_t is a stationary process.

Example 1 (cointegrated processes): As an example of a data generating process (DGP) that generates cointegrated variables, consider the following system:

$$
\begin{aligned}
\Delta x_{2t} &= \epsilon_{2t} && (4)\\
x_{1t} &= \beta_2 x_{2t} + \epsilon_{1t}, && (5)
\end{aligned}
$$

where ϵ_{1t} and ϵ_{2t} are IID and uncorrelated error processes. We solve for the levels to find

$$
\begin{aligned}
x_{2t} &= \tau_{2t} + \text{initial value} &&= \sum\nolimits_{i=1}^{t} \epsilon_{2i} + x_{20}\\
x_{1t} &= \tau_{1t} + \text{initial value} + \text{stationary process} &&= \beta_2 \sum\nolimits_{i=1}^{t} \epsilon_{2i} + \beta_2 x_{20} + \epsilon_{1t}.
\end{aligned}
$$

Here $\tau_{1t} = \beta_2 \tau_{2t}$ is a common stochastic trend and the processes cointegrate with cointegration vector $\beta = (1 : -\beta_2)'$. In particular we find

$$
z_t = \beta' x_t = \epsilon_{1t},
$$

which is stationary. An economic example could be that income (x_{2t}) develops as a random walk process according to (4), while consumption (x_{21}) according to (5) is a linear function of income plus a stationary noise term. The dynamics of both equations could of course be more complicated. ♦

If we consider a regression type formulation,

$$
x_{1t} = \mu + \beta_2 x_{2t} + u_t, \tag{6}
$$

where μ is the mean of $z_t = x_{1t} - \beta_2 x_{2t}$, then cointegration implies that the deviation, u_t, is a (mean zero) stationary process. It is important to realize, however, that if $z_t = \beta' x_t = x_{1t} - \beta_2 x_{2t}$ is a stationary process then so is $\widetilde{z}_t = b z_t = b x_{1t} - b\beta_2 x_{2t}$. This means that for all values $b \neq 0$, both

$$
\beta = \begin{pmatrix} 1 \\ -\beta_2 \end{pmatrix} \quad \text{and} \quad \widetilde{\beta} = \begin{pmatrix} \widetilde{\beta}_1 \\ \widetilde{\beta}_2 \end{pmatrix} = \begin{pmatrix} b \\ -b\beta_2 \end{pmatrix}
$$

are cointegration vectors for the variables in x_t. In the first case, β, we have imposed a *normalization* on the first coefficient, $\beta_1 = 1$. This normalization is natural if we have a relation of the form (6) in mind, but we could equally well have chosen a different normalization, e.g. $\widetilde{\beta} = \left(-\beta_2^{-1}, 1\right)'$, corresponding to an equation with x_{2t} on the left hand side.

The definition of cointegration is easily extended to more variables. In particular, let $x_t = (x_{1t}, x_{2t}, x_{3t}, ..., x_{pt})'$ be a $p-$dimensional vector of variables. Then a vector $\beta = \left(1, -\beta_2, ..., -\beta_p\right)'$ is a cointegration vector if

$$
z_t = \beta' x_t = x_{1t} - \beta_2 x_{2t} - ... - \beta_p x_{pt}
$$

is a stationary process. Note that with p I(1) variables there can be several (at most $p-1$) different cointegration vectors. This is not a problem for the theory, but the single-equation tools presented in this note are only appropriate for the existence of a single cointegration vector. As the number of variables, p, increases it becomes less and less likely that there is only one stationary combination.

1.1 Cointegration and Economic Equilibrium

Consider again a DGP as in Example 1:

$$\begin{aligned} \Delta x_{2t} &= \epsilon_{2t} & (7) \\ x_{1t} &= \mu + \beta_2 x_{2t} + u_t, & (8) \end{aligned}$$

where ϵ_{2t} is IID and u_t is a stationary process uncorrelated with ϵ_{2t}. Notice that the individual variables, x_{1t} and x_{2t}, are I(1) non-stationary, while $\beta' x_t$ is a stationary process. An implication is that the shock ϵ_{2t} has permanent effects on the levels of both variables but only transitory effects on $\beta' x_t$.

That makes it natural to think of the cointegrating relation (8) as defining an *economic equilibrium*: The variables themselves wander arbitrarily far up and down due to the accumulation of shocks to x_{2t}, but they never deviate too much from equilibrium. When the variables cointegrate, we can define $x_{1t}^* = \mu + \beta_2 x_{2t}$, and we will refer to x_{1t}^* as the equilibrium value of x_{1t}, and $u_t = x_{1t} - x_{1t}^*$ is the deviation from equilibrium. The equilibrium value can be interpreted as the value at which there is no inherent tendency for x_{1t} to move away, but it is important to realize that because the economy is continuously hit by shocks, the system will never settle down at x_{1t}^*, and x_{1t} will not converge to x_{1t}^* in any sense.

Example 2 (purchasing power parity): Let $x_{1t} = \log(E_t)$ denote the log of the bilateral exchange rate between Dollar and Euro (denominated as Dollar per Euro), and let $x_{2t} = \log(P_t^{US}) - \log(P_t^{EU})$ denote the corresponding difference between the logs of the consumer prices. Then

$$z_t = x_{1t} - x_{2t} = \log(E_t) - \big(\log(P_t^{US}) - \log(P_t^{EU})\big) = \log\left(\frac{E_t \cdot P_t^{EU}}{P_t^{US}}\right)$$

is the relative deviation from purchasing power parity (PPP) between the US and the Euro area. For most countries consumer prices and exchange rates appear non-stationary, and if the deviation from PPP is stationary we can think of PPP as a valid equilibrium relation for parity between US and the Euro area. In this case $\beta = (1, -1)'$ would be a cointegrating vector for $x_t = (x_{1t}, x_{2t})'$. If, on the other hand, the deviation, z_t, is non-stationary, it means that the price differential can wander arbitrarily far from the PPP value and there is no equilibrium interpretation of the PPP. ♦

Example 3 (prices on the orange market): As an empirical example, Figure 1 (A) illustrates the price of organic and regular oranges, p_t^{org} and p_t^{reg}, in pence per lb., while graph (B) illustrates the price differential, $p_t^{\text{org}} - p_t^{\text{reg}}$. The individual prices in graph (A) are obviously non-stationary and a possible interpretation is that the non-stationarity is driven by stochastic trends. The prices show strong co-movements, however, and the price differential looks much more stable and could be a sample path from a stationary process.

This suggests that the relation

$$p_t^{\text{org}} = \mu + p_t^{\text{reg}} + u_t,$$

defines an equilibrium for the orange market, where μ is the additional price of organic oranges in equilibrium, and u_t is the deviation from equilibrium in period t. Note again, that $p_t^{\text{org}} - p_t^{\text{reg}}$ will not equal μ in any specific period and $p_t^{\text{org}} - p_t^{\text{reg}}$ will not approach μ as $T \to \infty$. The equilibrium concept refers to the fact that fluctuations of $p_t^{\text{org}} - p_t^{\text{reg}}$ around μ will be stationary as suggested by graph (B). ♦

Example 4 (private consumption): Similarly, Figure 1 (C) illustrates the log of real private consumption in Denmark, c_t, the log of real disposable income, y_t, and the log of real private wealth including the value of owner occupied housing, w_t (we have subtracted 2 from w_t in the graph to make the levels comparable). All three time series are clearly trending. The series for consumption and income have many similarities and co-move in some periods. Deviations from this pattern seem to occur primarily when there are large fluctuations in private wealth. People familiar with the Danish business cycle will recognize the peak in private wealth in 1986 as the result of a boom in the housing market, which apparently drove up the consumption-to-income ratio. The time series behavior, as well as simple economic theory, suggest that consumption depends on both income and wealth, and graph (D) depicts the deviation, $u_t = c_t - c_t^*$, from a simple consumption function

$$c_t = -0.404 + 0.364 \cdot y_t + 0.516 \cdot w_t + u_t.$$

We note that the deviation, u_t, looks much more stable than the variables themselves, suggesting that $\beta = (1, -0.364, -0.516)'$ may be a cointegrating vector for $x_t = (c_t, y_t, w_t)'$. Whether the deviation, u_t, actually corresponds to a stationary process is a testable hypothesis to which we return in §3. ♦

Example 5 (money demand): To estimate a long-run money demand relation we may consider the variables $x_t = (m_t, y_t, r_t, b_t)'$, where m_t is the real money stock (in logs), y_t is real income (in logs), r_t is the short interest rate as a measure of the yield of holding money, while b_t is the bond rate measuring the yield on holdings alternative to money. Some theories suggest that in the long run the demand for money is given by

$$m_t = y_t - \omega\left(b_t - r_t\right),$$

so that money demand increases with the amount of transactions, measured by y_t, and decreases with the opportunity cost of holding money, $b_t - r_t$. This suggests that $\beta = (1, -1, \omega, -\omega)'$ could be a cointegrating vector for the variables in x_t.

Alternatively, theories for the determination of interest rates would suggest that two interest rates with different maturities should be cointegrated, and also the velocity, $y_t -$

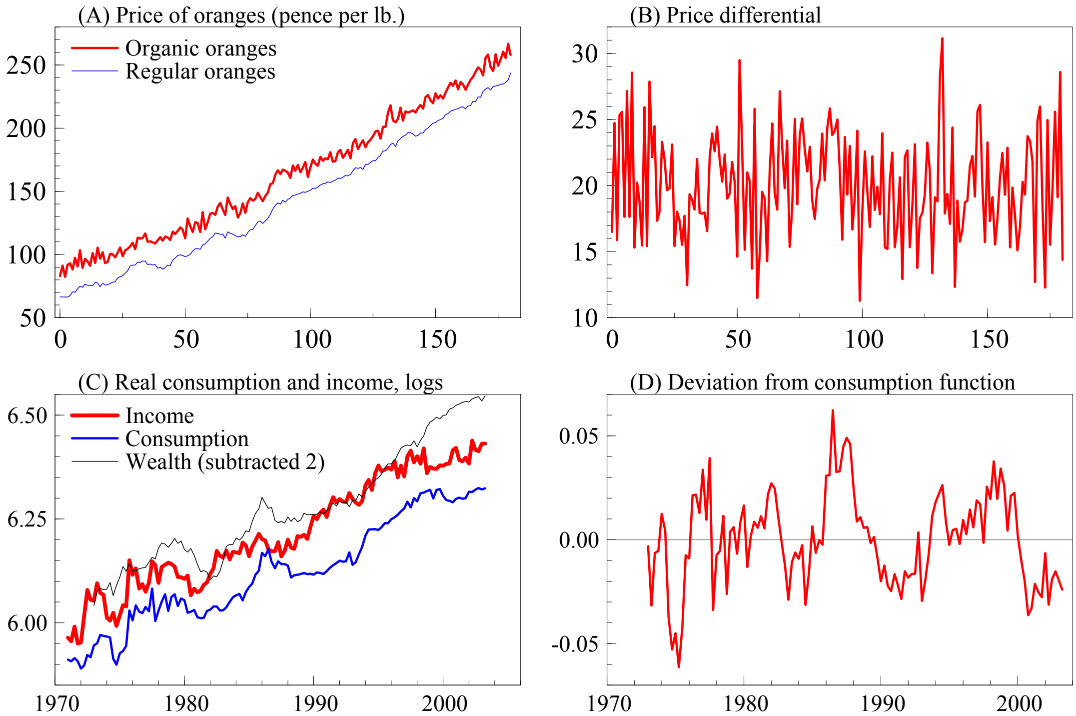

Figure 1: *Examples of some possibly cointegrated series. (A): Price of organic oranges, p_t^{org}, and regular oranges, p_t^{reg}, measured in pence per lb. (B): The price differential, $p_t^{org} - p_t^{reg}$. (C): Real aggregate consumption, c_t, disposable income, y_t, and private wealth, w_t, in logs. (D): The linear combination, $u_t = c_t - 0.364 \cdot y_t - 0.516 \cdot w_t + 0.404$.*

m_t, may be stationary. That suggests a different scenario with two cointegration relations:

$$\beta_1' x_t = (0, 0, -1, 1) \begin{pmatrix} m_t \\ y_t \\ r_t \\ b_t \end{pmatrix} = b_t - r_t \quad \text{and} \quad \beta_2' x_t = (1, -1, 0, 0) \begin{pmatrix} m_t \\ y_t \\ r_t \\ b_t \end{pmatrix} = m_t - y_t.$$

It is an empirical question, which of the scenarios (if any) that characterizes a data set, but the single-equation tools presented in this note are only appropriate in the scenario with one cointegrating relation. ♦

1.2 Deterministic Terms

In the definition of cointegration above we have assumed that the variables, x_{1t} and x_{2t}, are I(1) and that $z_t = \beta' x_t$ is a stationary variable with mean $E[z_t] = \mu$. The theory of cointegration can easily be extended to other specifications of deterministic variables.

As an example we might believe that z_t is stationary around a deterministic linear trend. This would be the case if x_{1t} and x_{2t} contain both deterministic and stochastic trends, and that the linear combination, $\beta' x_t$, cancels the stochastic trends but *not* the deterministic trends. To model this case we can extend (6) with a deterministic trend term, e.g.

$$x_{1t} = \mu + \mu_1 t + \beta_2 x_{2t} + u_t.$$

The interpretation is that $z_t = x_{1t} - \beta_2 x_{2t}$ is *trend-stationary*, i.e. stationary around the linear trend, $\mu + \mu_1 t$. The deviation, u_t, is a mean zero stationary process.

Similarly, linear combinations could be stationary around other deterministic components, e.g. level shifts.

1.3 How is the Equilibrium Sustained?

In the previous section we defined cointegration of variables, $x_t = (x_{1t}, x_{2t})'$, as the existence of a vector β so that the combination, $z_t = \beta' x_t$, is a stationary process, and we interpreted the relation as defining an equilibrium for the variables. Logically, an equilibrium requires the existence of some forces in the DGP which ensures that the non-stationary variables, x_{1t} and x_{2t}, do not move too far away from equilibrium. In this section we present *error-correction* as a way of describing these forces, and we discuss how cointegration and error-correction are two complementary ways of characterizing the same phenomenon.

There exists a famous representation theorem, due to Engle and Granger (1987), stating that x_{1t} and x_{2t} cointegrate if and only if there exist an *error correction model* for either x_{1t}, x_{2t} or both. To illustrate the link, let

$$x_{1t} = \mu + \beta_2 x_{2t} + u_t$$

be an equilibrium relation between two I(1) variables. Since u_t is a stationary mean zero variable, there exist a stationary ARMA model for u_t. Assume for simplicity that it is an AR(2),

$$u_t = \theta_1 u_{t-1} + \theta_2 u_{t-2} + \epsilon_t,$$

where $\theta(1) = 1 - \theta_1 - \theta_2 > 0$ from stationarity. Inserting the definition of u_t, this is equivalent to

$$(x_{1t} - \mu - \beta_2 x_{2t}) = \theta_1 (x_{1t-1} - \mu - \beta_2 x_{2t-1}) + \theta_2 (x_{1t-2} - \mu - \beta_2 x_{2t-2}) + \epsilon_t,$$

or collecting terms:

$$x_{1t} = (1 - \theta_1 - \theta_2)\mu + \theta_1 x_{1t-1} + \theta_2 x_{1t-2} + \beta_2 x_{2t} - \theta_1 \beta_2 x_{2t-1} - \theta_2 \beta_2 x_{2t-2} + \epsilon_t,$$

which is an autoregressive distributed lag model, ADL(2,2). This can also be written as the error-correction model

$$\Delta x_{1t} = \beta_2 \Delta x_{2t} + \theta_2 \beta_2 \Delta x_{2t-1} - \theta_2 \Delta x_{1t-1} - (1 - \theta_1 - \theta_2) \{x_{1t-1} - \mu - \beta_2 x_{2t-1}\} + \epsilon_t, \quad (9)$$

where the long-run solution is the lagged deviation from the cointegrating relation, u_{t-1}, and the error-correction parameter $-(1-\theta_1-\theta_2)<0$ ensures that deviations from the equilibrium are eliminated[1].

To intuitively understand the link between cointegration and error correction, notice that under the maintained assumptions, Δx_{1t}, Δx_{1t-1}, Δx_{2t}, Δx_{2t-1}, and ϵ_t are all stationary terms. Since x_{1t} and x_{2t} are assumed to be I(1), the equation in (9) is only balanced in terms of the order of integration if the combination $x_{1t-1}-\beta_2 x_{2t-1}$ is stationary, i.e. if the variables cointegrate. If the variables do not cointegrate, the only way to balance the equation is to exclude the levels from the equation by setting $(1-\theta_1-\theta_2)=0$.

The link between cointegration and error correction also emphasizes that cointegration is essentially a system property; and from the result of the representation theorem we do not know whether x_{1t} or x_{2t} or both variables error correct. This suggests that a general formulation of the error-correction model consists of an equation for each variable, in our case

$$\begin{aligned}\Delta x_{1t} &= \delta_1+\Gamma_{11}\Delta x_{1t-1}+\Gamma_{12}\Delta x_{2t-1}+\alpha_1\left(x_{1t-1}-\beta_2 x_{2t-1}\right)+\epsilon_{1t}\\ \Delta x_{2t} &= \delta_2+\Gamma_{21}\Delta x_{1t-1}+\Gamma_{22}\Delta x_{2t-1}+\alpha_2\left(x_{1t-1}-\beta_2 x_{2t-1}\right)+\epsilon_{2t},\end{aligned}$$

where one lag of each first difference has been included. Stacking the equations we may write the model as the so-called *vector error correction model*,

$$\begin{pmatrix}\Delta x_{1t}\\ \Delta x_{2t}\end{pmatrix}=\begin{pmatrix}\delta_1\\ \delta_2\end{pmatrix}+\begin{pmatrix}\Gamma_{11} & \Gamma_{12}\\ \Gamma_{21} & \Gamma_{22}\end{pmatrix}\begin{pmatrix}\Delta x_{1t-1}\\ \Delta x_{2t-1}\end{pmatrix}+\begin{pmatrix}\alpha_1\\ \alpha_2\end{pmatrix}(x_{1t-1}-\beta_2 x_{2t-1})+\begin{pmatrix}\epsilon_{1t}\\ \epsilon_{2t}\end{pmatrix},$$

or

$$\Delta x_t=\delta+\Gamma\Delta x_{t-1}+\alpha\beta' x_{t-1}+\epsilon_t,$$

where we have used the definitions

$$\delta=\begin{pmatrix}\delta_1\\ \delta_2\end{pmatrix},\quad \Gamma=\begin{pmatrix}\Gamma_{11} & \Gamma_{12}\\ \Gamma_{21} & \Gamma_{22}\end{pmatrix},\quad \alpha=\begin{pmatrix}\alpha_1\\ \alpha_2\end{pmatrix},\quad \text{and}\quad \beta=\begin{pmatrix}1\\ -\beta_2\end{pmatrix}.$$

We note that the lagged deviation from the cointegrating relation, $\beta' x_{t-1}=x_{1t-1}-\beta_2 x_{2t-1}$, appears as an explanatory variable in both equations. For x_{1t} to error correct we need $\alpha_1<0$. To see this, imagine that x_{1t-1} is above equilibrium so that $x_{1t-1}-\beta_2 x_{2t-1}$ is positive. For x_{1t} to move towards the equilibrium we need $\Delta x_{1t}<0$, which requires $\alpha_1<0$. If x_{1t} error corrects, the magnitude of α_1 measures the proportion of the deviation that is corrected each period, and α_1 is sometimes referred to as the *speed of adjustment*. As an example, a value of $\alpha_1=-0.4$ would indicate that 40% of a deviation from equilibrium is removed each period. Using the same line of arguments, $\alpha_2>0$ is consistent with error correction of x_{2t}.

[1] The simple assumptions used in the present derivation impose a common factor restriction on (9) but that is not necessarily true in practice.

To illustrate the graphical implications of cointegration and error correction we consider a simple model for two cointegrated variables,

$$\begin{pmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{pmatrix} = \begin{pmatrix} -0.2 \\ 0.1 \end{pmatrix} (x_{1t-1} - x_{2t-1}) + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}, \tag{10}$$

where ϵ_{1t} and ϵ_{2t} are independent standard normals, $N(0,1)$. Here $\beta = (1,-1)'$ is a cointegrating vector and both variables error correct, with speeds of adjustment given by $\alpha = (-0.2, 0.1)'$. One realization of x_{1t} and x_{2t} ($t = 1, 2, ..., 100$) generated from the DGP in (10) is illustrated in Figure 2 (A). Notice the strong co-movement between the variables, which reflects that they have the same stochastic trend. Graph (B) depicts the deviation from the long-run relation,

$$z_t = \beta' x_t = x_{1t} - x_{2t}.$$

The series z_t is relatively persistent and is often above or below equilibrium for longer periods of time. This illustrates the moderately slow error-correction in (10). In graph (C) we illustrate the speed of adjustment. We consider a large deviation $z_t = \beta' x_t = 10$ in a particular period and show the adjustment towards equilibrium in a situation where no shocks hit the system. In the present case the deviation from $\beta' x_t$ is visible for approximately 10 periods and the convergence is exponential. It is the equilibrating force in graph (C) that ensures that the levels in graph (A) do not move to far apart. Finally graph (D) depicts a cross plot of x_{1t} on x_{2t}. The variables are non-stationary and will wander arbitrarily on the real axis. Cointegration (i.e. the force implied by error-correction) implies that the observations will never move to far from the equilibrium defined by the straight line. Finally, observe that the most recent observation is far from equilibrium, $\beta' x_{100} > 0$. If we were to make an out-of-sample forecast of the series, $x_{101}, x_{102}, ...$, then we would conjecture that x_t would be drawn towards equilibrium, i.e. that either x_{1t} would decrease or that x_{2t} would increase to close the gap.

Example 6 (prices on the orange market, continued): For the case of the organic and regular oranges, an estimation yields the two error-correction equations

$$\begin{aligned} \Delta p_t^{\text{org}} &= \underset{(1.665)}{22.864} - \underset{(0.081)}{1.090} \cdot \left(p_{t-1}^{\text{org}} - p_{t-1}^{\text{reg}}\right) + \hat{\epsilon}_t^{\text{org}} \\ \Delta p_t^{\text{reg}} &= \underset{(0.634)}{1.147} - \underset{(0.031)}{0.008} \cdot \left(p_{t-1}^{\text{org}} - p_{t-1}^{\text{reg}}\right) + \hat{\epsilon}_t^{\text{reg}}, \end{aligned}$$

where the numbers in parentheses are standard errors of the estimated coefficients. We can write the system as a vector error correction model,

$$\begin{pmatrix} \Delta p_t^{\text{org}} \\ \Delta p_t^{\text{reg}} \end{pmatrix} = \begin{pmatrix} 22.864 \\ 1.147 \end{pmatrix} - \begin{pmatrix} 1.090 \\ 0.008 \end{pmatrix} \left(p_{t-1}^{\text{org}} - p_{t-1}^{\text{reg}}\right) + \begin{pmatrix} \hat{\epsilon}_t^{\text{org}} \\ \hat{\epsilon}_t^{\text{reg}} \end{pmatrix},$$

where

$$\beta' x_{t-1} = (1, -1) \begin{pmatrix} p_{t-1}^{\text{org}} \\ p_{t-1}^{\text{reg}} \end{pmatrix} = p_{t-1}^{\text{org}} - p_{t-1}^{\text{reg}}$$

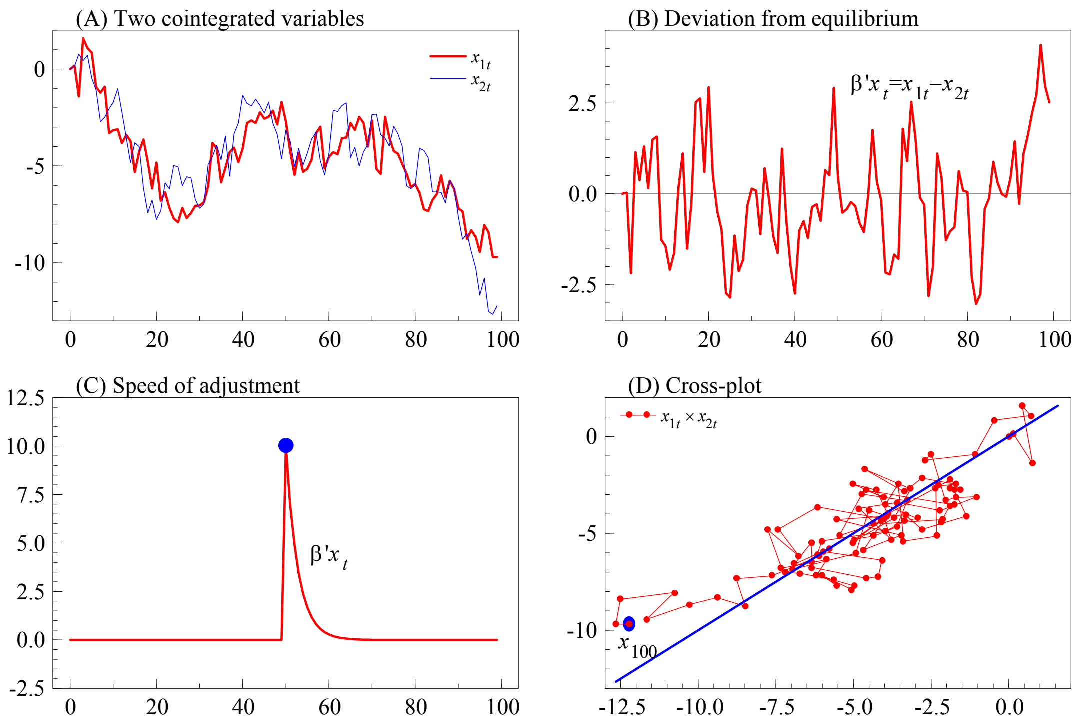

Figure 2: *Simulated series to illustrate cointegration and error-correction.*

is the cointegrating relation, and $\alpha = (-1.090, -0.008)'$ characterizes the speed of adjustment towards equilibrium.

The organic orange price seems to error correct very strongly, removing the entire disequilibrium each month. The regular orange price, on the other hand, does not seem to error correct. The coefficient is negative, indicating a movement *away* from equilibrium, but it is very small and not significantly different from zero. A simple interpretation of this result is that the orange price is essentially determined on the large market for regular oranges. The price of organic oranges has to follow the price of regular oranges, with an additional premium of approximately 23 pence per lb. Note that changes in the price of regular oranges (i.e. a shock to ϵ_t^{reg} *ceteris paribus*) will be fully transmitted to the price of organic oranges after one month, while changes to the price of organic oranges (i.e. a shock to ϵ_t^{org} *ceteris paribus*) will not be transmitted to the market for regular oranges. ♦

2 Estimation and Inference

Above, the concepts of cointegration and error correction was introduced. In this section we discuss how the parameters in the cointegrating vector, $\beta = (1, -\beta_2, ..., -\beta_p)'$, can be estimated and how inference on β can be conducted.

2.1 The Engle-Granger Two-Step Approach

Recall, that if a set of variables, x_{1t} and x_{2t}, cointegrate then there exists coefficients, μ and β_2, so that

$$x_{1t} = \mu + \beta_2 x_{2t} + u_t \tag{11}$$

defines an equilibrium. It is natural to try to estimate β_2 in the static regression (11) and this is the approach suggested in the seminal paper of Engle and Granger (1987).

It can be shown that if x_{1t} and x_{2t} are I(1) and cointegrated then the OLS estimator from (11), $\widehat{\beta}_2$, is consistent for the true parameter, β_2. We do not postulate that the model in (11) is the DGP that generated the data, and it turns out that consistency of $\widehat{\beta}_2$ holds even if the estimation model is misspecified relative to the DGP–as long as the misspecification only relates to stationary terms. The reason is that the stochastic trends will dominate asymptotically, so for $T \to \infty$ any misspecification of stationary terms will not affect the estimator. As an example the static regression in (11) will produce consistent estimators even if the true DGP is dynamic. This is discussed in some detail in Box 1. This result is in contrast to the stationary case, where consistency is normally only obtained if the DGP is contained in the estimation model.

Consistency of the estimator tells you that $\widehat{\beta}_2$ converges to β_2 as T diverges. It turns out that the non-stationarity of the variables in x_t affects the so-called *rate of convergence*, i.e. the speed at which the variance of $\widehat{\beta}_2$ go to zero. If x_{1t} and x_{2t} are stationary variables, we know that under usual conditions,

$$\sqrt{T}\left(\widehat{\beta}_2 - \beta_2\right) \to N(0, V),$$

where V is the asymptotic variance of $\widehat{\beta}_2$. The interpretation is that the variance of $\widehat{\beta}_2$ is $T^{-1}V$, which approaches zero at the rate of T^{-1}. For cointegrated I(1) series, the variance of $\widehat{\beta}_2$ approaches zero at a faster rate of T^{-2}, known as *super consistency* of $\widehat{\beta}_2$. To illustrate this phenomenon, graph (A) and (B) in Figure 3 show the distributions of the estimator $\widehat{\beta}_2$ of a true value $\beta_2 = 1$ from the static regression (11). In graph (A) x_{2t} is generated as an IID variable, while x_{1t} is x_{2t} plus an IID error term. In graph (B) we use the same setup but now x_{2t} is I(1), generated as a random walk. In both cases the estimators are consistent and the distributions collapse around the true value, $\beta_2 = 1$. In the cointegrated I(1) case, however, convergence is much faster, and the distributions are much less dispersed.

Whereas the specification of stationary terms is not important asymptotically, it might nevertheless be important in finite samples, and some authors suggest that the super-consistent OLS estimator $\widehat{\beta}_2$ can be severely biased in finite samples. We return to an alternative estimator and an illustration of the bias in §2.2

In a cointegration analysis, the static regression (11) is sometimes referred to as the first step of an *Engle-Granger two-step procedure*; where the second step is a description of the dynamic adjustment towards equilibrium. Given the estimated cointegration parameters,

Box 1: Static Regression when the DGP is Dynamic

In most cases we believe that the DGP, generating the observed data in the economy, is dynamic. In this case the static regression (11) is misspecified; but the misspecification is related only to stationary terms and the obtained estimator, $\widehat{\beta}_2$, is still consistent.

As an example, consider a simple dynamic DGP given by

$$\begin{aligned} x_{1t} &= \delta + \theta_1 x_{1t-1} + \phi_0 x_{2t} + \phi_1 x_{2t-1} + \epsilon_{1t} && \text{(B1-1)} \\ x_{2t} &= x_{2t-1} + \epsilon_{2t}, && \text{(B1-2)} \end{aligned}$$

where ϵ_{1t} and ϵ_{2t} are IID error processes. Here, x_{2t} is a random walk, while x_{1t} is generated as an autoregressive distributed lag model ADL(1,1). The equation in (B1-1) can be rewritten as

$$(1-\theta_1)\, x_{1t} = \delta + (\phi_0 + \phi_1)\, x_{2t} - \theta_1(x_{1t} - x_{1t-1}) - \phi_1(x_{2t} - x_{2t-1}) + \epsilon_{1t}$$

or

$$x_{1t} = \mu + \beta_2 x_{2t} + a_1 \Delta x_{1t} + a_2 \Delta x_{2t} + \widetilde{\epsilon}_{1t}, \quad \text{(B1-3)}$$

where we have defined

$$\mu = \frac{\delta}{1-\theta_1}, \quad \beta_2 = \frac{\phi_0+\phi_1}{1-\theta_1}, \quad a_1 = -\frac{\theta_1}{1-\theta_1}, \quad a_2 = -\frac{\phi_1}{1-\theta_1}, \quad \text{and} \quad \widetilde{\epsilon}_{1t} = \frac{\epsilon_{1t}}{1-\theta_1}.$$

Comparing the expressions in (11) and (B1-3), we note that the static regression is a simplified version of the DGP, obtained by excluding the stationary terms, Δx_{1t} and Δx_{2t}. Since the misspecification is related to only stationary terms, the estimator from the static regression is still consistent.

From a first look it seems natural to use the model (B1-3) for estimating the parameters. Note, however, that $\Delta x_{1t} = x_{1t} - x_{1t-1}$ is correlated with $\widetilde{\epsilon}_{1t}$, so the OLS estimator of a_1 is not consistent. Asymptotically this will not affect the estimator of β_2, and the OLS estimator in (B1-3) is consistent. Alternatively we can use x_{1t-1} as an instrument for Δx_{1t} and estimate the parameters using instrumental variables (IV). This IV estimator is numerically equivalent to the estimator obtained from applying OLS to the ADL model (B1-1). The IV estimator may sound complicated compared to the ADL model, but depending on the software system you use, the IV estimator is sometimes a convenient way to get the estimated standard errors for the cointegrating parameter, $\widehat{\beta}_2$, see also in §2.4.

---

we may define the so-called *error correction term* as the deviation from equilibrium,

$$\widehat{u}_t = x_{1t} - \widehat{\mu} - \widehat{\beta}_2 x_{2t}.$$

Under cointegration $\widehat{u}_t$ is a stationary process, and since the estimators converge to the true values very fast we can include $\widehat{u}_{t-1}$ as a fixed regressor in a dynamic model. The second step of the Engle-Granger procedure is therefore to estimate an error-correction model *given* $\widehat{u}_{t-1}$, e.g.

$$\Delta x_{1t} = \delta + \lambda_1 \Delta x_{1t-1} + \kappa_0 \Delta x_{2t} + \kappa_1 \Delta x_{2t-1} + \alpha \widehat{u}_{t-1} + \epsilon_t,$$

where we have assumed one lag in the first differences and have conditioned on the contemporaneous change Δx_{2t}. All terms in the error correction models are stationary and standard inference procedures apply to all parameters, in the sense that $t-$ratios will follow standard normal distributions, $N(0,1)$, asymptotically.

2.2 Dynamic Regression Models

An alternative to the estimator obtained by OLS in the static regression (11) is to construct a dynamic model, which is believed to be a better approximation of the DGP, and derive the estimator of the cointegrating coefficients from this model. One possibility is to construct the best possible description of the auto-covariance structure of the data by estimating an appropriate autoregressive distributed lag (ADL) model, and derive estimators of the cointegrating parameters from the long-run solution. In particular we could estimate by OLS the unrestricted ADL model, where the lag-lengths are set to eliminate residual autocorrelation, e.g. an ADL(2,2) model,

$$x_{1t} = \delta + \theta_1 x_{1t-1} + \theta_2 x_{1t-2} + \phi_0 x_{2t} + \phi_1 x_{2t-1} + \phi_2 x_{2t-2} + \epsilon_t. \tag{12}$$

Recall that the unrestricted ADL model can be written as an error-correction model. In particular we can use the reformulations

$$\begin{aligned} x_{1t} - \theta_1 x_{1t-1} - \theta_2 x_{1t-2} &= \Delta x_{1t} + \theta_2 \Delta x_{1t-1} - (\theta_1 + \theta_2 - 1)\, x_{1t-1} \\ \phi_0 x_{2t} + \phi_1 x_{2t-1} + \phi_2 x_{2t-2} &= \phi_0 \Delta x_{2t} - \phi_2 \Delta x_{2t-1} + (\phi_0 + \phi_1 + \phi_2)\, x_{2t-1}, \end{aligned}$$

to obtain the ECM form

$$\Delta x_{1t} = \delta + \lambda_1 \Delta x_{1t-1} + \kappa_0 \Delta x_{2t} + \kappa_1 \Delta x_{2t-1} + \gamma_1 x_{1t-1} + \gamma_2 x_{2t-1} + \epsilon_t, \tag{13}$$

where $\lambda_1 = -\theta_2$, $\kappa_0 = \phi_0$, $\kappa_1 = -\phi_2$, $\gamma_1 = (\theta_1 + \theta_2 - 1)$, and $\gamma_2 = (\phi_0 + \phi_1 + \phi_2)$. For both (12) and (13) the estimator of the cointegrating coefficient is given by the long-run solution, i.e.

$$\widehat{\beta}_2 = \frac{\widehat{\phi}_0 + \widehat{\phi}_1 + \widehat{\phi}_2}{1 - \widehat{\theta}_1 - \widehat{\theta}_2} = -\frac{\widehat{\gamma}_2}{\widehat{\gamma}_1}. \tag{14}$$

The model in (13) is often referred to as the unrestricted ECM form. Recall, that we may also write the model with the long-run solution explicit as

$$\Delta x_{1t} = \lambda_1 \Delta x_{1t-1} + \kappa_0 \Delta x_{2t} + \kappa_1 \Delta x_{2t-1} + \gamma_1 (x_{1t-1} - \mu - \beta_2 x_{2t-1}) + \epsilon_t. \tag{15}$$

These formulations are equivalent but (13) can be estimated with OLS while (15) is non-linear in the parameters and requires a more elaborate estimation procedure (e.g. maximum likelihood).

Compared to the estimator from the static regression, the estimator derived from a dynamic model has the advantage of being based on a well-specified model. The main problem in empirical applications is that the DGP is not known, so the precise form of (12) has to be determined from the data. The usual approach is to start with a general ADL(p,q), where p and q are large enough to eliminate residual autocorrelation. From this model insignificant lags can be removed.

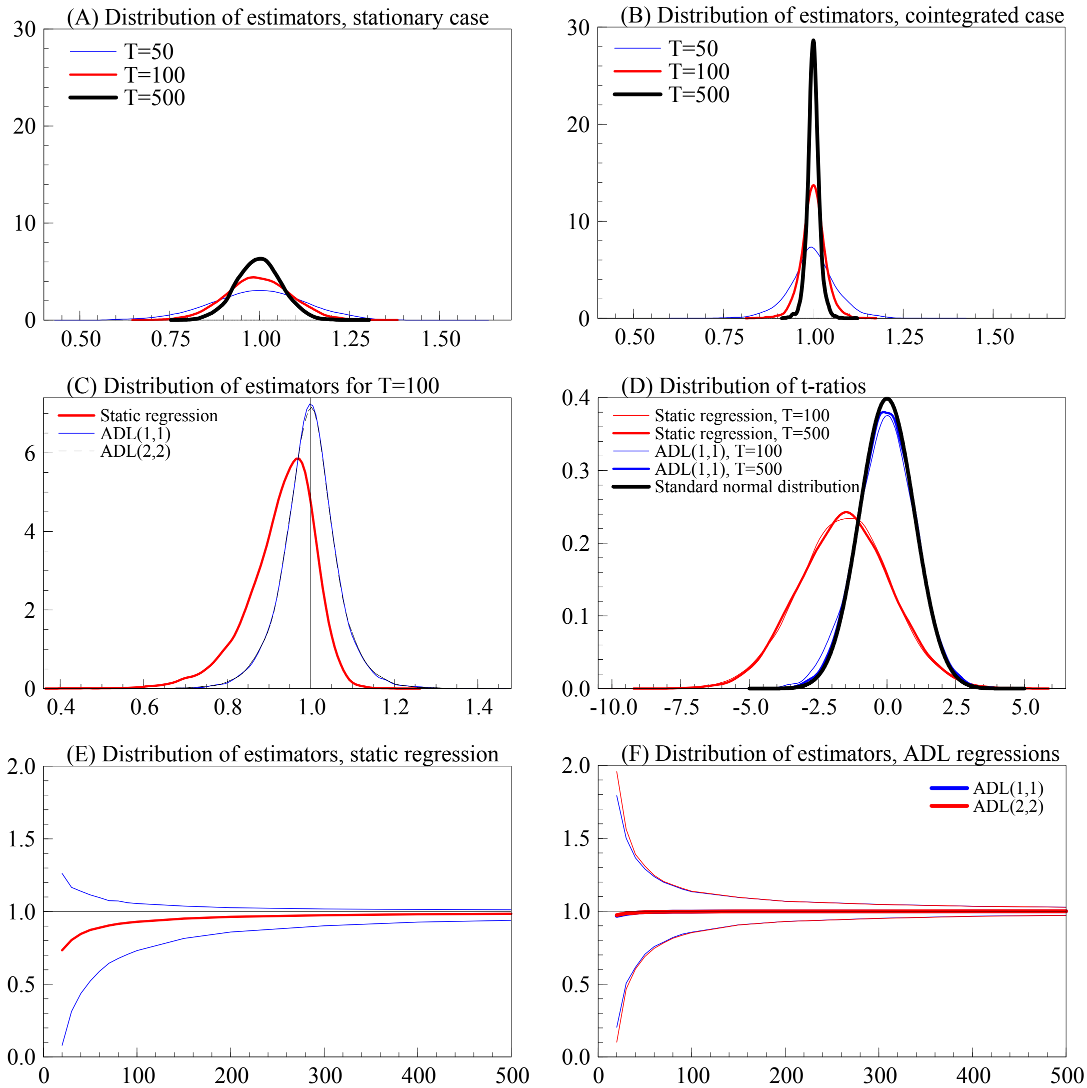

Figure 3: *(A): Consistency of the estimated parameter in a static regression for stationary variables. (B): Superconsistency for cointegrated I(1) variables. (C): Distributions of the estimated cointegration parameter based on a static and a dynamic regression. (D): Distributions of the t−ratios under a true null hypothesis for* $T = 100$ *and* $T = 500$ *based on the static and dynamic regressions. (E)-(F): Mean and 95% confidence bands of the distributions of the estimated cointegration parameter for different sample lengths* $T = 20, 30, ..., 500$. *The Monte Carlo simulations are based on* $10,000$ *replications.*

Box 2: Inference on Coefficients in an ADL Model

Consider an ADL(2,2) model given by

$$x_{1t} = \delta + \theta_1 x_{1t-1} + \theta_2 x_{1t-2} + \phi_0 x_{2t} + \phi_1 x_{2t-1} + \phi_2 x_{2t-2} + \epsilon_t, \tag{B2-1}$$

where ϵ_t is an IID process. Given that the variables in (B2-1) are all I(1), it is interesting to ask if any of the estimators obtained by applying OLS to equation (B2-1) follow standard distributions, so that inference based on the standard normal distribution applies.

The answer to this question is given in Sims, Stock, and Watson (1990). They give the general result that an estimated parameter follow a normal distribution asymptotically if it can be written as the coefficient to a mean zero stationary variable–possibly after a linear transformation of the model. This means that if the model can be reformulated so that e.g. the parameter α is the coefficient to a stationary variable with mean zero, then the distribution of $\widehat{\alpha}$ is asymptotically normal.

Again we may rewrite the ADL model in ECM form as

$$\Delta x_{1t} = \lambda_1 \Delta x_{1t-1} + \kappa_0 \Delta x_{2t} + \kappa_1 \Delta x_{2t-1} + \alpha \left\{ x_{1t-1} - \mu - \beta_2 x_{2t-1} \right\} + \epsilon_t, \tag{B2-2}$$

where $\lambda_1 = -\theta_2$, $\kappa_0 = \phi_0$, $\kappa_1 = -\phi_2$, $\alpha = (\theta_1 + \theta_2 - 1)$, $\mu = \delta / (1 - \theta_1 - \theta_2)$, and $\beta_2 = (\phi_0 + \phi_1 + \phi_2) / (1 - \theta_1 - \theta_2)$.

Note that Δx_{1t-1}, Δx_{2t}, and Δx_{2t-1} are stationary variables with mean zero, so estimators of the corresponding parameters: λ_1, κ_0, and κ_1 will follow a normal distribution. *Given cointegration*, the term $x_{1t-1} - \mu - \beta_2 x_{2t-1}$ is also stationary with mean zero, so also the estimator of α will follow a normal distribution.

Unfortunately there is no way to rewrite the model so that β_2 is the coefficient to a stationary mean zero term, so this argument cannot be used to show that the estimator of the cointegrating coefficient, β_2, has a normal distribution. If x_{1t} is the only variable that error corrects, however, then all information on β_2 is present in the equation (B2-1) and the single equation OLS estimator is identical to the maximum likelihood estimator in the vector error-correction model. It follows that $\widehat{\beta}_2$ is asymptotically efficient and *asymptotically normal*. This is strong assumption to make, however, and inference on cointegrating parameters should always be done with some caution.

2.3 Comparison in a Monte Carlo Simulation

To compare the two approaches and illustrate the practical importance of the bias in a static regression we set up a small Monte Carlo simulation. As the DGP we consider a specific model

$$\begin{aligned} x_{1t} &= 0.30 \cdot x_{2t} + 0.20 \cdot x_{2t-1} + 0.50 \cdot x_{1t-1} + \epsilon_{1t} \\ x_{2t} &= x_{2t-1} + \epsilon_{2t}, \end{aligned}$$

for $t = 1, 2, ..., T$, where the innovations, ϵ_{1t} and ϵ_{1t}, are assumed $N(0, 1)$ and uncorrelated. The DGP implies a long-run solution with a cointegrating coefficient of $\beta_2 = \frac{0.30+0.20}{1-0.50} = 1$. Based on $10,000$ data sets from this DGP we look at the properties of the estimators obtained from the static regression (11) and from the ADL(1,1) model. We note that the used ADL model is identical to the DGP and we expect it to perform better than the static regression. To illustrate the effect of choosing an estimation model which is more general than the DGP we also consider the estimates obtained from an ADL(2,2), which estimates a redundant lag for both variables. This setup amounts to using one regression model that coincides with the DGP (an ADL(1,1)), one that is too general (an ADL(2,2)), and one that is too restricted (a static regression).

Figure 3 (C) illustrates the distributions of the estimated parameters for the three cases for $T = 100$ observations. We note that the distributions for the ADL(1,1) and ADL(2,2) almost coincide and are symmetric and nicely centered around the true value. This indicates that estimating a redundant lag will only marginally affect the estimators. The distribution of the estimates from the static regression is shifted to the left, reflecting the bias of the estimator. The mean of the estimates is 0.93, which is significantly smaller than unity. We also note that the distribution is skewed, with a long left tail. Graph (E) and (F) illustrate the mean and the 5% and 95% quantiles of the distributions of the estimates for different sample lengths $T = 20, 30, ..., 500$. We see that the estimator from the static regression is consistent, but it is severely biased in small samples and the distribution is clearly asymmetric. The estimator from the dynamic regression has the correct expectation for all considered sample lengths. We also note that the cost of the two redundant regressors in the ADL(2,2) is only visible for very short sample lengths, and even for $T = 20$, the difference between the estimates from an ADL(1,1) and an ADL(2,2) is small.

The results for this specific DGP thus seem to suggest that estimators derived from a dynamic regression model are clearly preferable to the two-step Engle-Granger estimators. This seems to be confirmed for more general classes of DGPs in the literature.

2.4 Inference on Cointegrating Parameters

Besides getting estimates of the parameters we are often interested in testing specific hypotheses on the cointegrating coefficients, which may link the statistical model to economic theory. This requires that we know the distribution of $\widehat{\beta}_2$. Unfortunately, it turns out that the estimator obtained from the static regression (11) is not normal and in general the distribution depends on unknown parameters, which invalidates standard inference. As a consequence, we can only use the static regression to estimate the parameters, while the estimated standard errors cannot be used for inference in general.

In the dynamic regression, (12) or equivalently (13) or (15), the situation is a bit more promising, and *given cointegration* $t-$ratios constructed from the estimated standard errors follow standard normal distributions under the null. Assuming cointegration this result implies that we can make inference on the cointegration coefficients derived as the long-run solution from an ADL or ECM model. As an example we can test hypotheses on the cointegrating coefficient using the standard $t-$ratio

$$\widehat{t}_{\beta_2=b} = \frac{\widehat{\beta}_2 - b}{\mathsf{se}(\widehat{\beta}_2)},$$

which will follow a standard normal distribution asymptotically. A more theoretical discussion of the inference on the parameters of the ADL model for I(1) variables is given in Box 2.

The only complication is that $\widehat{\beta}_2$ is a non-linear function of the estimated parameters, cf. (14), and the standard error of $\widehat{\beta}_2$ is a complicated function of the covariance matrix for the estimated parameters in (12). The software package PcGive has a procedure to calculate the static long-run solution and supply the derived standard errors. In other software packages it is sometimes more convenient to use the alternative (but numerically equivalent) IV estimator mentioned in Box 1 or the non-linear etimation of (15) since they automatically produce standard errors to $\widehat{\beta}_2$.

The distributions of the $t-$ratios in the Monte Carlo simulation are reported in graph 3 (D). The $t-$ratios from a static regression have a distribution which is skewed to the left, and inference based on a standard normal would be very misleading. The $t-$ratios from the dynamic regression, on the other hand, seem to be close to a standard normal–making it possible to test hypotheses on the parameters.

Example 7 (private consumption, continued): To illustrate estimation and inference on cointegrating coefficients, consider again the Danish quarterly consumption data: $x_t = (c_t, y_t, w_t)'$. Applying OLS to a static regression model for the 122 observations, $1973:1 - 2003:2$, yields

$$c_t = -\underset{(0.129)}{0.404} + \underset{(0.049)}{0.364} \cdot y_t + \underset{(0.044)}{0.516} \cdot w_t + \widehat{u}_t, \tag{16}$$

where the numbers in parentheses are standard errors. The estimates seem consistent with a simple consumption function in which consumption depends positively on income and wealth. We may note that a one per cent increase in income and wealth give less than a one per cent increase in consumption as $0.364 + 0.516 = 0.88$. Consequently, the consumption-income ratio will not be constant in a steady state, which may be regarded as unsatisfactory from an economic point of view. Note that $t-$ratios constructed from the reported standard errors in (16) do not follow a standard normal distribution and they are not suitable for testing. For example we cannot test if the sum of coefficients, 0.88, is significantly different from zero.

Based on the estimates of the static regression (16) we may define the error-correction term

$$\widehat{u}_t = c_t - 0.404 - 0.364 \cdot y_t - 0.516 \cdot w_t,$$

which is interpretable as the deviation from equilibrium. The term $\widehat{u}_t$ may be used in the construction of error-correcting models to characterize the dynamic properties of the data as suggested by the Engle-Granger approach. In principle there may exist error correction models of Δc_t, Δy_t, and Δw_t, and starting with a model with two lags in the first differences and deleting insignificant lags, produces the three equations:

$$\begin{aligned}
\Delta\widehat{c}_t &= \underset{(0.002)}{0.001} - \underset{(0.077)}{0.195} \cdot \Delta c_{t-1} + \underset{(0.057)}{0.229} \cdot \Delta y_t + \underset{(0.117)}{0.426} \cdot \Delta w_t - \underset{(0.064)}{0.250} \cdot \widehat{u}_{t-1} \\
\Delta\widehat{y}_t &= \underset{(0.002)}{0.002} + \underset{(0.118)}{0.433} \cdot \Delta c_t + \underset{(0.115)}{0.387} \cdot \Delta c_{t-1} - \underset{(0.087)}{0.353} \cdot \Delta y_{t-1} + \underset{(0.099)}{0.066} \cdot \widehat{u}_{t-1} \\
\Delta\widehat{w}_t &= \underset{(0.001)}{0.003} + \underset{(0.060)}{0.232} \cdot \Delta c_t - \underset{(0.050)}{0.030} \cdot \widehat{u}_{t-1}.
\end{aligned}$$

Note that only consumption corrects deviations form the long-run relation, with a speed of adjustment of -0.25, while Δy_t and Δw_t do not adjust significantly when the variables are out of equilibrium.

An alternative estimator of the cointegrating coefficients can be derived from a conditional ADL model for consumption. Assuming at most three lags and deleting insignificant terms lead to the preferred ADL model

$$\begin{aligned}
c_t &= \underset{(0.093)}{-0.080} + \underset{(0.092)}{0.544} \cdot c_{t-1} + \underset{(0.079)}{0.204} \cdot c_{t-2} + \underset{(0.060)}{0.240} \cdot y_t - \underset{(0.065)}{0.125} \cdot y_{t-1} \\
&\quad + \underset{(0.124)}{0.401} \cdot w_t - \underset{(0.129)}{0.291} \cdot w_{t-1} + \widehat{\epsilon}_t. \qquad (17)
\end{aligned}$$

According to misspecification tests, the model seems relatively well-behaved. No-autocorrelation of order 1 to 5 is not rejected with a $p-$value of 0.64; and no-ARCH of order 1 to 4 is not rejected with a $p-$value of 0.21. Solving equation (17) for the static long-run solution yields

$$c_t = \underset{(0.357)}{-0.320} + \underset{(0.146)}{0.458} \cdot y_t + \underset{(0.130)}{0.436} \cdot w_t + \widehat{u}_t, \qquad (18)$$

where the long-run coefficients are derived as

$$\frac{0.240 - 0.125}{1 - 0.544 - 0.204} = 0.458 \quad \text{and} \quad \frac{0.401 - 0.291}{1 - 0.544 - 0.204} = 0.436,$$

and where the standard errors to the cointegrating coefficients are complicated functions of the covariance matrix of the estimated parameters. Compared to the static regression, the estimated coefficient to income is somewhat higher, whereas the coefficient to private wealth is lower. We also note that the standard errors in (18), which can be used for testing hypotheses on the cointegrating coefficients, are much larger than the standard errors in (16).

Based on the dynamic model, the sum of the coefficients is still below unity, $0.458 + 0.436 = 0.894$, but now we can test the hypothesis that it is actual unity. A Wald test for this hypothesis gives a test statistic of 5.26, corresponding to a $p-$value of 0.022 in a $\chi^2(1)$ distribution. We therefore reject the hypothesis and conclude that the sum of the coefficients seems to be significantly smaller than unity.

To illustrate the dynamic properties of the estimated cointegration model Figure 4 shows the impulse-response functions for a permanent change in income and wealth, i.e. the cumulated values of $\partial c_{t+i}/\partial y_t$ and $\partial c_{t+i}/\partial w_t$ $(i = 0, 1, ..., 40)$. For disposable income

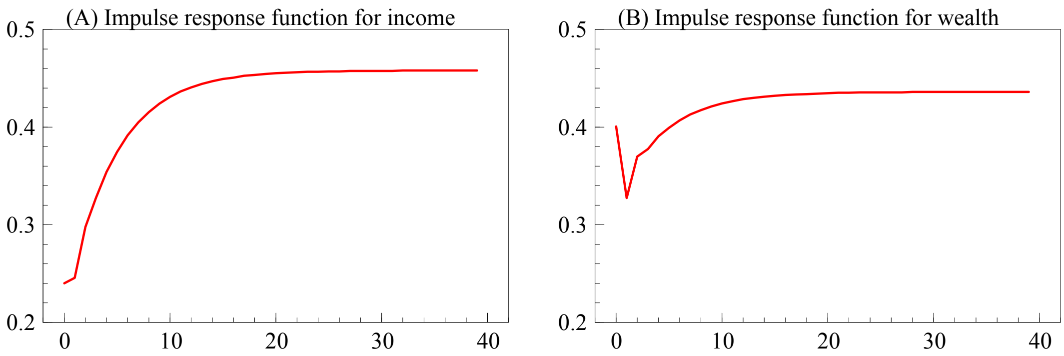

Figure 4: *Impulse-response functions for a permanent change in income and wealth, i.e. the accumulated values of* $\partial c_{t+i}/\partial y_t$ *and* $\partial c_{t+i}/\partial w_t$ $(i = 0, 1, ..., 40)$.

the contemporaneous impact is 0.240, and there is a smooth convergence to the long-run impact of 0.458. A permanent change in the private wealth have a contemporaneous effect on consumption of 0.401, which is not far from the long-run impact of 0.436. The convergence is not monotone, however, and the large contemporaneous impact is followed by a decrease in the next period and then a gradual convergence.

Notice that the results obtained in the estimation of (17) can also be obtained by estimating the equivalent unrestricted error-correction model, i.e.

$$\begin{aligned}\Delta c_t \;&=\; \underset{(0.093)}{-0.080} - \underset{(0.079)}{0.204} \cdot \Delta c_{t-1} + \underset{(0.060)}{0.240} \cdot \Delta y_t + \underset{(0.124)}{0.401} \cdot \Delta w_t \\ &\quad \underset{(0.065)}{-0.251} \cdot c_{t-1} + \underset{(0.044)}{0.115} \cdot y_{t-1} + \underset{(0.046)}{0.110} \cdot w_{t-1} + \widehat{\epsilon}_t. \qquad (19)\end{aligned}$$

Here the cointegrating coefficients can be found as the long-run solution, $0.115/0.251 = 0.458$ and $0.110/0.251 = 0.436$, which are identical to the results in (18). ♦

2.5 What if Variables do Not Cointegrate?

Recall that cointegration is the special case where the stochastic trends in the individual variables cancel. From a logical point of view this is an exception, and it is interesting to ask for the properties of regression models with I(1) variables that do not cointegrate.

To discuss this case assume that x_{1t} and x_{2t} are two unrelated I(1) variables. Both variables contain stochastic trends, but they are unrelated and do not cointegrate. Ideally we would like the static regression

$$x_{1t} = \mu + \beta_2 x_{2t} + u_t, \qquad (20)$$

to reveal that $\beta_2 = 0$, at least asymptotically. This turns out *not* to be the case, however, and the estimator $\widehat{\beta}_2$ is not consistent. Moreover, as $T \to \infty$ the $t-$ratio, $\widehat{t}_{\beta_2=0}$, will indicate a significant relation between x_{1t} and x_{2t}. This is known as the *spurious regression* result. The problem is that when the variables do not cointegrate, u_t is an $I(1)$ process and standard results do not hold.

Example 8 (spurious regression): As an example of a spurious regression, consider two presumably unrelated I(1) variables, namely yearly data covering $1980 - 2000$ for the log of real private consumption in Denmark, cons_t, and the log of the number of breeding cormorants in Denmark, bird_t. We estimate a static regression:

$$\mathsf{cons}_t = \underset{(0.150)}{12.145} + \underset{(0.015)}{0.095} \cdot \mathsf{bird}_t + \widehat{u}_t.$$

The $t-$ratio for the hypothesis that there is no relation, $\beta_2 = 0$, is given by

$$\widehat{t}_{\beta_2=0} = \frac{0.095}{0.015} = 6.30,$$

which seems highly significant in a $N(0,1)$ distribution, apparently suggesting a clear positive relation between the number of birds and aggregate consumption! Furthermore, R^2 in the equation is 0.69, indicating that the number of breeding birds can account for large proportion of the variation in consumption. These results are of cause totally spurious–a simple consequence of the variables being I(1). ♦

To illustrate the spurious regression problem we set up a simple Monte Carlo simulation. As a comparison we first reproduce the standard results for a stationary regression. We generate data series as independent IID variables,

$$x_{1t} = \epsilon_{1t} \quad \text{and} \quad x_{2t} = \epsilon_{2t}, \quad t = 1, 2, ..., T,$$

where ϵ_{1t} and ϵ_{2t} are independent drawings from a $N(0,1)$. The results from the regression model (20) are reported in Figure 5 (A) and (B) for sample lengths $T = 50, 100, 500$. We note in graph (A) that the distributions of $\widehat{\beta}_2$ are centered around the true value and the convergence of the estimator $\widehat{\beta}_2$ implies that the variance decreases as $T \to \infty$. In graph (B) we consider the distributions of the $t-$ratio, $\widehat{t}_{\beta_2=0}$. The distribution is close to the asymptotic $N(0,1)$ for all considered sample lengths.

Next consider the spurious regression between I(1) variables. Here we generate data as independent random walks, i.e.

$$x_{1t} = x_{1t-1} + \epsilon_{1t} \quad \text{and} \quad x_{2t} = x_{2t-1} + \epsilon_{2t}, \quad t = 1, 2, ..., T.$$

The results are reported in graph (C) and (D). In graph (C) we note that the distribution of $\widehat{\beta}_2$ is centered around the true value, but it does not collapse as $T \to \infty$. This reflects that the estimator is unbiased but not consistent. In graph (D) we notice that the distributions of the $t-$ratios get increasingly dispersed as T increases, and the distributions are far from a standard normal. As $T \to \infty$ this implies that using the conventional critical values of ± 1.96, we would always reject the true hypothesis that $\beta_2 = 0$.

2.5.1 Dynamic Regression Models

One way to explain the spurious regression result is to note that the regression model in (20) is logically inconsistent if the variables do not cointegrate. Since there is no genuine relation between x_{1t} and x_{2t}, the true value of the parameter is zero, $\beta_2 = 0$; so a model with $\beta_2 \neq 0$ is necessarily false. Note, however, that the model with $\beta_2 = 0$ is also false. If $\beta_2 = 0$, then the only way to balance the equation is if u_t is I(1), but that is not consistent with the assumptions of the regression model. One problem with the spurious regression is therefore that the actual DGP can not be contained in the estimation model; and the test for $\beta_2 = 0$ against $\beta_2 \neq 0$ compares two false models.

Based on this insight it is easy to suggest a modification of the static regression, which circumvents some of the problems of the spurious regression. As an example, consider the simple ADL(1,0) model, where the lagged value of x_{1t} is included in the regression:

$$x_{1t} = \mu + \theta_1 x_{1t-1} + \phi_0 x_{2t} + \epsilon_t. \tag{21}$$

In this model the simple DGP is obtained if

$$\theta_1 = 1 \quad \text{and} \quad \phi_0 = \mu = 0,$$

which is consistent with the assumption of a stationary error term.

To analyze how the dynamic regression model behaves with unrelated I(1) variables, we redo the Monte Carlo simulation using now the dynamic regression in (21). The distributions of the estimator, $\widehat{\phi}_0$, and the $t-$ratio, $\widehat{t}_{\phi_0=0}$, are reported in graph (E) and (F), respectively. In graph (E) we note that the estimator is consistent, and comparing with graph (A) we also note that the rate of convergence is faster than for the stationary case. It is remarkable, that simply augmenting the static regression with the lagged left hand side variable eliminates the inconsistency of the estimator. The variables, x_{1t} and x_{2t}, are still I(1), however, and standard results for hypothesis testing do not automatically apply. In graph (F) we note that the distributions are fixed for different sample length as in the stationary case, but the distribution is not an $N(0, 1)$. This fact can be explained using the argument in Box 2.

3 Testing for No-Cointegration

The discussion of spurious regression makes it obvious, that it is important to the able to test whether a set of p variables, $x_{1t}, x_{2t}, ..., x_{pt}$, are cointegrated or not. If the variables are cointegrated we can use the methods suggested above for estimation and inference on the equilibrium relation. If the variables do not cointegrate, the regression model is useless and should be disregarded or changed to obtain cointegration.

In this section we discuss how the hypothesis that a set of variables are not cointegrated can be tested. We consider two different approaches. One approach is based on the deviation from a proposed cointegrating relation or on the residual from a static regression. This is the approach implemented in the Engle-Granger methodology. The second approach is based on the equivalence between cointegration and error correction and it is actually a test for no-error-correction in an unrestricted dynamic model.

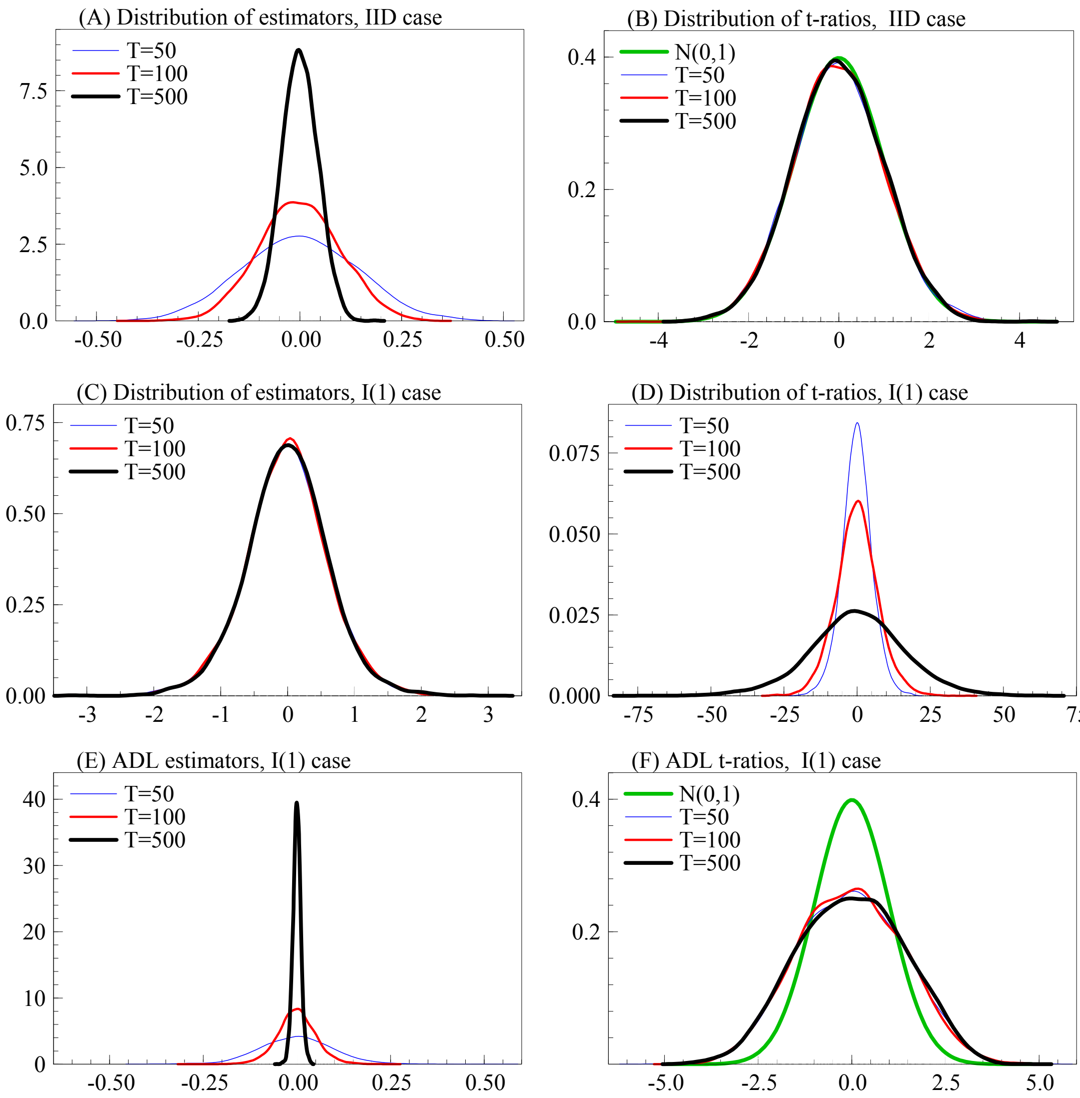

Figure 5: *Monte Carlo results for a static regression for stationary variables and for a spurious regression.*

3.1 Residual-Based Tests

Following the definition, a set of variables $x_1, x_2, ..., x_p$ cointegrate with cointegration vector $\beta = \left(1, -\beta_2, -\beta_3, ..., -\beta_p\right)'$, if the linear combination

$$z_t = \beta' x_t = x_{1t} - \beta_2 x_{2t} - \beta_3 x_{3t} - ... - \beta_p x_{pt}$$

is stationary. It follows that the null hypothesis of *no-integration* can be translated into the hypothesis of a unit root in z_t. This hypothesis can be tested using a conventional augmented Dickey-Fuller (ADF) test. Allowing z_t to have a mean different from zero but no deterministic linear trend, the hypothesis of no-cointegration can be tested as the hypothesis $\mathcal{H}_0 : \pi = 0$ in the ADF regression with a constant term,

$$\Delta z_t = \delta + \sum_{i=1}^{k-1} c_i \Delta z_{t-i} + \pi z_{t-1} + \eta_t, \tag{22}$$

where η_t is an IID error term. The alternative to a unit root is stationarity, $\mathcal{H}_A : -2 < \pi < 0$, and under the null of a unit root the ADF $t-$test statistic,

$$\widehat{\tau}_c = \frac{\widehat{\pi}}{\mathsf{se}(\widehat{\pi})},$$

follows a DF distribution. Critical values for the DF distribution are reproduced in part (A) of Table 1 in the row with zero estimated parameters.

If the relevant alternative to a unit root is trend-stationarity, the ADF regression (22) may be augmented with a linear trend term, and the test for no-cointegration is the ADF test with a linear trend, τ_l.

Example 9 (prices on the orange market, continued): As an example of a unit root test where the potential cointegration vector is known, reconsider the prices from the orange market. The potential stationary variable is the price differential

$$z_t = p_t^{\text{org}} - p_t^{\text{reg}},$$

implying a cointegration vector $\beta = (1, -1)'$. To test the hypothesis of no-cointegration, we test for a unit root in z_t. Setting up an ADF regression with a constant term and 5 lags in Δz_t and deleting insignificant lags, lead to the simple DF regression

$$\Delta z_t = \underset{(1.534)}{21.718} - \underset{(0.0750)}{1.082}\, z_{t-1} + \widehat{\eta}_t.$$

The Dickey-Fuller test is given by the $t-$ratio,

$$\widehat{\tau}_c = \frac{-1.082}{0.075} = -14.43.$$

The 5% critical value for the case of a constant is -2.86, so we can easily reject the null of no-cointegration. Also recall from Figure 1 (B) that the price differential looks extremely stable. ♦

3.1.1 Estimated Cointegration Vector

In many cases the cointegration vector, β, is unknown and the above approach is not feasible. The test procedure can easily be modified, however, to the case where β is estimated as in the Engle-Granger procedure. Recall that if the cointegration coefficients, $\beta_2, ..., \beta_p$, are unknown, they can be super-consistently estimated in the static regression

$$x_{1t} = \mu + \beta_2 x_{2t} + ... + \beta_p x_{pt} + u_t, \tag{23}$$

or from the long-run solution of a dynamic model. The regression (23) corresponds to a cointegrating relation if the deviation from the relation, u_t, is a stationary process. Estimating the parameters we can test for *no-cointegration* by testing whether the estimated residual, $\widehat{u}_t$, contains a unit root. This test is translated into the hypothesis $\mathcal{H}_0 : \pi = 0$ in the ADF regression

$$\Delta\widehat{u}_t = \sum_{i=1}^{k-1} c_i \Delta\widehat{u}_{t-i} + \pi\widehat{u}_{t-1} + \eta_t. \tag{24}$$

We note that since the estimated residual, $\widehat{u}_t$, has a mean of zero, there is no constant term in the ADF regression (24). Nonetheless, the critical values for the ADF test depend on the deterministic specification of the static regression, e.g. whether (23) contains a constant or a linear trend.

The fact that the cointegrating vector $\widehat{\beta}$ is estimated also changes the critical values for the ADF test, and the estimation uncertainty has to be taken into account. The intuition is that OLS applied to the static regression (23) will minimize the variance of $\widehat{u}_t$, and graphically the estimated residuals will look as 'stationary as possible'. And the more explanatory variables we include in (23), i.e. the more parameters we estimate, the smaller is the variance of $\widehat{u}_t$, and the more stationary it will look. In the test procedure we will have to account for that, and the critical values depend on the number of I(1) regressors in (23). The asymptotic distributions of tests for no-cointegration are illustrated in Figure 6 (A). As the number of regressors in the static regression increases, the distribution of the ADF test statistic moves to the left. This reflects that the OLS procedure makes the variance of the estimated residual smaller and smaller. The critical values of the residual based test are reproduced in Table 1 (A).

Example 10 (private consumption, continued): To test whether the static regression of a consumption function in (16) corresponds to a cointegrating relation, we construct the estimated residual

$$\widehat{u}_t = c_t - 0.404 - 0.364 \cdot y_t - 0.516 \cdot w_t,$$

which is depicted in Figure 1 (D). To test for no-cointegration we use an ADF regression without deterministic terms. In the present case one lag is needed,

$$\Delta\widehat{u}_t = -\underset{(0.089)}{0.223}\Delta\widehat{u}_{t-1} - \underset{(0.068)}{0.221}\widehat{u}_{t-1} + \widehat{\eta}_t,$$

and the test statistic is given by

$$\widehat{\tau}_c = \frac{-0.221}{0.068} = -3.27.$$

The 5% and 10% critical values for the case of a constant term and two estimated parameters in the static regression are given by -3.74 and is -3.45, respectively, so we cannot reject the hypothesis of no-cointegration. This is reflected in Figure 1 (D), where the deviations from the relation are relatively persistent. The deviations seem to be related

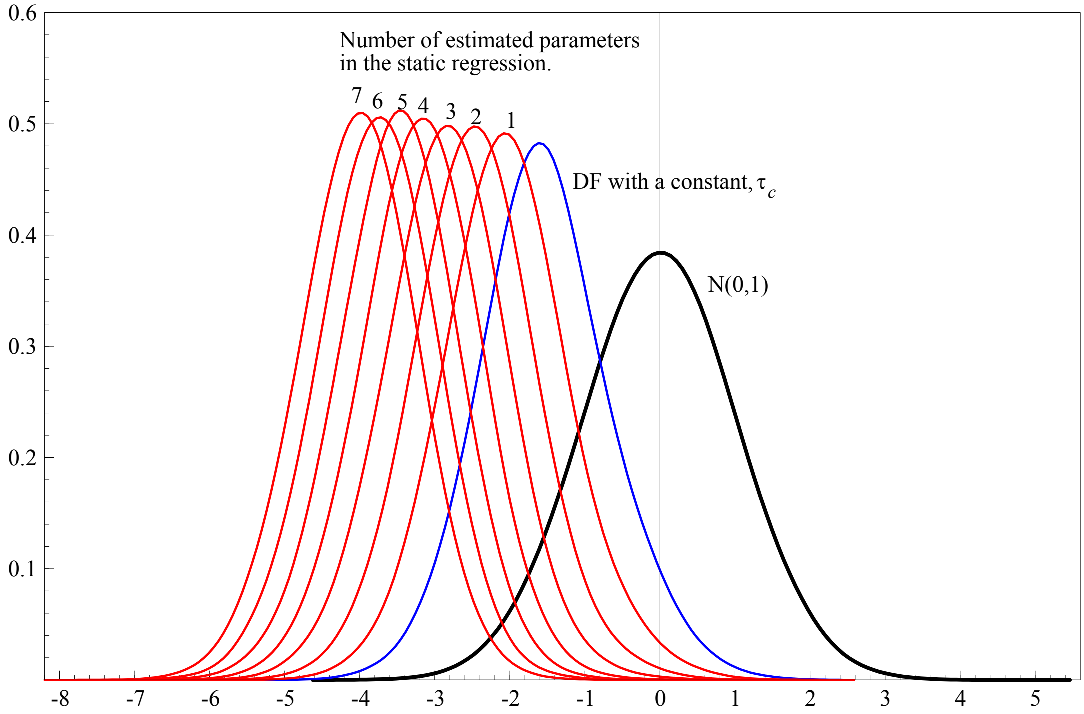

Figure 6: *Asymptotic distributions of the residual-based test for no-cointegration.*

to the business cycle, suggesting that the consumption-income ratio is pro-cyclical besides the wealth effects. To obtain stronger evidence of cointegration one possible solution is to augment the model with a measure of the business cycle, e.g. a variable measuring the effects of unemployment. ♦

3.2 Testing for No-Cointegration in the ECM

Due to the representation theorem discussed in §1.3, the null hypothesis of *no-cointegration* corresponds to the null of *no-error-correction*. This observation has been used to construct several tests for whether variables cointegrate. The most convenient is based on the unrestricted error-correction model, e.g.

$$\Delta x_{1t} = \delta + \lambda_1 \Delta x_{1t-1} + \kappa_0 \Delta x_{2t} + \kappa_1 \Delta x_{2t-1} + \gamma_1 x_{1t-1} + \gamma_2 x_{2t-1} + \epsilon_t. \tag{25}$$

Here we can test the hypothesis that x_{1t} do not error correct, i.e. $\mathcal{H}_0 : \gamma_1 = 0$ against the cointegrating alternative, $\mathcal{H}_A : \gamma_1 < 0$. The test statistic is just the conventional $t-$ratio, given by

$$\widehat{t}_{\gamma_1=0} = \frac{\widehat{\gamma}_1}{\mathsf{se}(\widehat{\gamma}_1)}.$$

As for the residual based test, the distribution of $\widehat{t}_{\gamma_1=0}$ depends on the deterministic terms in the regression (25) as well as the number of I(1) variables in x_t. The asymptotic critical

(A) Residual-based (ADF) test for no-cointegration

Number of estimated parameters	Constant in (22)			Constant and trend in (22)		
	1%	5%	10%	1%	5%	10%
0	−3.43	−2.86	−2.57	−3.96	−3.41	−3.13
1	−3.90	−3.34	−3.04	−4.32	−3.78	−3.50
2	−4.29	−3.74	−3.45	−4.66	−4.12	−3.84
3	−4.64	−4.10	−3.81	−4.97	−4.43	−4.15
4	−4.96	−4.42	−4.13	−5.25	−4.72	−4.43

(B) PcGive test for no-cointegration

Number of variables in x_t (p)	Constant in (25)			Constant and trend in (25)		
	1%	5%	10%	1%	5%	10%
2	−3.79	−3.21	−2.91	−4.25	−3.69	−3.39
3	−4.09	−3.51	−3.19	−4.50	−3.93	−3.62
4	−4.36	−3.76	−3.44	−4.72	−4.14	−3.83
5	−4.59	−3.99	−3.66	−4.93	−4.34	−4.03

Table 1: *Asymptotic critical values for tests of no-cointegration. Reproduced from Davidson and MacKinnon (1993).*

values are reproduced in part (B) of Table 1. This test appeared very early in the PcGive software package and is often referred to as the *PcGive test for no-cointegration*.

Comparing the residual-based test for no-cointegration with the test for no-error-correction in the dynamic model three things are worth noting. First, the test for no-error-correction is based on the assumption that x_{1t} is the only variable which error corrects to the potential cointegrating relation. This implies that we should test for no-cointegration in the 'correct' error-correction model; in the present case that is the model for Δx_{1t} and not the model for Δx_{2t}. In most cases, prior knowledge from economic theory suggests which equation to consider.

Secondly, the test for no-error-correction of Δx_{1t} is parallel to a test for no-cointegration for a relation involving x_{1t}. Even if we cannot reject the hypothesis of no-error-correction of Δx_{1t}, the other right hand side variables in levels, $x_{2t}, ..., x_{pt}$, may still cointegrate in a relation *not involving* x_{1t}.

Thirdly, a comparison of (25) with the ADF test (24) shows, that the latter imposes a common factor restriction on the dynamics when the hypothesis of a unit root is tested. There is no *a priori* reason to believe that the data obey a common factor restriction, and the test may be negatively affected by imposing the restriction. The relation between (25) and (24) is explored in more details in Box 3.

Example 11 (private consumption, continued): To test whether the unrestricted

Box 3: ADF Tests and Common Factor Restrictions

Consider a potential cointegrating relation between two I(1) variables

$$x_{1t} = \mu + \beta_2 x_{2t} + u_t.$$

To test for no-cointegration we use the residual, $u_t = x_{1t} - \mu - \beta_2 x_{2t}$, and consider an ADF regression. Assume for simplicity that only one lag of Δu_t is needed, i.e.

$$\Delta u_t = c_1 \Delta u_{t-1} + \pi u_{t-1} + \eta_t.$$

Inserting the definition of u_t and collecting terms yields a model

$$\Delta x_{1t} = -\pi\mu + c_1 \Delta x_{1t-1} + \beta_2 \Delta x_{2t} - c_1 \beta_2 \Delta x_{2t-1} + \pi x_{1t-1} - \pi \beta_2 x_{2t-1} + \eta_t.$$

This is an ECM model, but subject to a number of common factor restrictions. We have 6 regressors on the right hand side, but only 4 parameters to be estimated: π, μ, c_1, and β_2, and hence two restrictions.

The restrictions imply e.g. that the contemporaneous impact of x_{2t} on x_{1t} is β_2, which is identical to the long-run impact. There is no compelling reason to believe that this is true in practice, see also Example 7.

error correction model in (19) suggests cointegration we test for no-error-correction using the $t-$ratio,

$$\widehat{t}_{\gamma_1=0} = \frac{-0.251}{0.065} = -3.86.$$

The 5% critical value is given in part (B) of Table 1 as -3.51, so we can borderline reject no-cointegration.

The different conclusions from the residual-based test and the PcGive test for no-cointegration could be related to the fact that the common factor restrictions imposed on the ADF test are not in line with the data. The test statistic for the two common factor restrictions is 10.41, which is highly significant according to the asymptotic $\chi^2(2)$ distribution, so the common factors are easily rejected. Rejection of the common factor restrictions and our knowledge that consumption is the only error-correcting variable suggest that the PcGive test for no-cointegration is probably preferable in the present case. ♦

4 Limitations of the Single-Equation Approach

So far we have presented some single-equation tools for cointegration analysis. Although these methods are powerful in many situations, it is important to know the assumptions that are implicitly made in the analysis and the drawbacks and limitations of the methods. Below we focus on the analysis based on an unrestricted ADL or error correction model and outline some important limitations.

The starting point for the discussion is an ECM model for three variables,

$$\Delta c_t = \delta + \kappa_0 \Delta y_t + \kappa_1 \Delta w_t + \gamma_1 c_{t-1} + \gamma_2 y_{t-1} + \gamma_3 w_{t-1} + \epsilon_t, \tag{26}$$

and to make the discussion less abstract we think of a consumption function and use the notation c_t, y_t, and w_t for the three variables. In the analysis of the equation (26) we implicitly make three sets of assumptions:

(1) Cointegration is a system property and in principle there exist error correction equations for all variables: Δc_t, Δy_t, and Δw_t. We only consider the equation for Δc_t.

(2) We assume that there is only one cointegrating relation between the variables, given by the long-run solution.

$$c_t = -\frac{\gamma_2}{\gamma_1} \cdot y_t - \frac{\gamma_3}{\gamma_1} \cdot w_t.$$

We mention in §1 that this is not necessarily true; and as the number of variables in the model increases it actually becomes less and less likely.

(3) To condition equation (26) on the contemporaneous changes, Δy_t and Δw_t, we assume that they are predetermined, i.e. there is no feedback from Δc_t to Δy_t and Δw_t. This is not necessarily true in practice, where several variables may be simultaneously determined.

Below we discuss these three issues in turn.

4.1 A Vector Error Correction Model

We consider the case where the variables, $x_t = (c_t, y_t, w_t)'$, are cointegrated with cointegration vector $\beta = (1, -\beta_2, -\beta_3)'$, so that $\beta' x_t$ is a stationary process. Assume that we are mainly interested in estimating the long-run parameters, β_2 and β_3.

We consider the three error correction models:

$$\begin{aligned}
\Delta c_t &= \delta_1 + \alpha_1 (c_{t-1} - \beta_1 y_{t-1} - \beta_2 w_{t-1}) + \text{dynamics} + \epsilon_{1t} \\
\Delta y_t &= \delta_2 + \alpha_2 (c_{t-1} - \beta_1 y_{t-1} - \beta_2 w_{t-1}) + \text{dynamics} + \epsilon_{2t} \\
\Delta w_t &= \delta_3 + \alpha_3 (c_{t-1} - \beta_1 y_{t-1} - \beta_2 w_{t-1}) + \text{dynamics} + \epsilon_{3t}
\end{aligned}$$

where 'dynamics' represents lagged values of first differences. Cointegration implies the existence of error correction, so one or more of the three coefficients, α_1, α_2, or α_3, have to be significantly different from zero. We note that the cointegrating parameters, β_1 and β_2, appear in all equations, so if we want the best possible (or efficient) estimators, we have to use the information in all three equations and not just the equation for Δc_t. Remember that we can stack the three equations in a vector error correction model (VECM):

$$\begin{pmatrix} \Delta c_t \\ \Delta y_t \\ \Delta w_t \end{pmatrix} = \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{pmatrix} + \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} \begin{pmatrix} 1 & -\beta_1 & -\beta_2 \end{pmatrix} \begin{pmatrix} c_{t-1} \\ y_{t-1} \\ w_{t-1} \end{pmatrix} + ... + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \end{pmatrix},$$

where we have left out the dynamics. The parameters of this model can be estimated using maximum likelihood, but that is beyond the scope of the present note.

In the special case where $\alpha_2 = \alpha_3 = 0$ it is sufficient to consider the first equation Δc_t, and the single equation analysis will be efficient. This assumption is implicitly imposed by the single equation model.

4.2 More Cointegrating Relations

Now assume that there actually exists two cointegrating relations between the variables in x_t, e.g.

$$c_t - \beta_1 y_t \sim I(0) \quad \text{and} \quad c_t - \beta_2 w_t \sim I(0).$$

The first one represent a consumption-income ratio if $\beta_1 = 1$ and the second one show the propensity to consume out of private wealth. The long-run relations can be written as

$$\begin{pmatrix} c_t - \beta_1 y_t \\ c_t - \beta_2 w_t \end{pmatrix} = \begin{pmatrix} 1 & -\beta_1 & 0 \\ 1 & 0 & -\beta_2 \end{pmatrix} \begin{pmatrix} c_t \\ y_t \\ w_t \end{pmatrix} = \beta' z_t.$$

We can write the VECM as

$$\begin{pmatrix} \Delta c_t \\ \Delta y_t \\ \Delta w_t \end{pmatrix} = \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{pmatrix} + \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{pmatrix} \begin{pmatrix} 1 & -\beta_1 & 0 \\ 1 & 0 & -\beta_2 \end{pmatrix} \begin{pmatrix} c_{t-1} \\ y_{t-1} \\ w_{t-1} \end{pmatrix} + ... + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \end{pmatrix}.$$

Here α_{11} measures how Δc_t is affected by deviations from the first long-run relation, $c_{t-1} - \beta_1 y_{t-1}$, while α_{12} measures how Δc_t is affected by deviations from the second long-run relation, $c_{t-1} - \beta_2 w_{t-1}$. Second row in α measures how Δy_t is affected by deviations from equilibrium etc.

The parameters of this model can again be estimated using ML, but we will not discuss that here. Instead we just note that if we only consider the first equation, then we will estimate the first row of the model, i.e.

$$\Delta c_t = \delta_1 + (\alpha_{11} + \alpha_{12})\, c_{t-1} - \beta_1 \alpha_{11} y_{t-1} - \beta_2 \alpha_{12} w_{t-1} + \epsilon_{1t},$$

which contains a combination of the two stationary relations. Since a combination of stationary relations will also be stationary, the equation is still balanced in terms of the order of integration, but we will not be able to separately interpret the two equilibrium relations.

4.3 Exogeneity

To be able to condition on Δy_t and Δw_t in the single-equation cointegration analysis we have to assume that they are *predetermined*, i.e.

$$E[\epsilon_{1t} \Delta y_t] = 0 \quad \text{and} \quad E[\epsilon_{1t} \Delta w_t] = 0.$$

This requirement states that there can be *no feedback* from Δc_t to Δy_t and Δw_t. In the present case consumption is a main component in the gross domestic product, which is the (national accounts) basis for defining disposable income. This link may suggest that consumption and income are simultaneously determined in a given quarter.

If the regressor is not predetermined we may exclude it, focussing on the reduced form with no contemporaneous effects. Alternatively we may be able to find good instruments for Δy_t, and estimate the model using an instrumental variables estimator.

A third possibility is again to estimate the vector error correction model directly. In this setting the variables, Δc_t, Δy_t, and Δw_t, are treated on equal footing, without imposing *a priori* restrictions of exogeneity.

5 Concluding Remarks

This note has illustrated that regression models for unit-root non-stationary time series give unreliable results, and the usual tools will not be able to distinguish between genuine relationships and spurious regressions. This suggests that for non-stationary time series we should *always* think in terms of cointegration. Logically a relationship can only be interpreted as defining an economic equilibrium if the variables cointegrate; and if they don't–there is no interpretable relationship between the variables.

We have presented a number of single-equation tools for cointegration. The conceptually simplest approach is the Engle-Granger two-step estimation, but for practical purposes the cointegration analysis based on unrestricted ADL or ECM models are probably preferable. This also fits within the general-to-specific framework, in which we first find an appropriate statistical description of the data (the unrestricted ADL model), and afterwards test hypotheses to link the statical model to economic theory (testing for cointegration and interpreting the long-run relationship).

5.1 Further Readings

The literature on cointegration analysis is huge, and most references are far more technical that the present note. An accessible introduction is Hendry and Juselius (2000). Alternative presentations of time series econometrics, including sections on single equation cointegration analysis, are given in Patterson (2000) and Enders (2004). A classic reference on cointegration analysis based on the ADL model is the book by Banerjee, Dolado, Gailbraith, and Hendry (1993). Maddala and Kim (1998) give a review of the literature on unit roots and cointegration. A specific reference for the test for no-error-correction (with references to the earlier literature) is Ericsson and MacKinnon (2002). The classic reference for time series analysis in general, which includes rather technical sections on cointegration models is Hamilton (1994). An introduction to vector error correction models and the analysis of cointegration in a VAR model is given in Hendry and Juselius (2001) and Juselius (2007), while the (very technical) theory is given in Johansen

(1996).

References

Banerjee, A., J. Dolado, J. W. Gailbraith, and D. Hendry (1993): *Co-Integration, Error-Correction, and the Econometric Analysis of Non-Stationary Data*. Oxford University Press, Oxford.

Davidson, R., and J. G. MacKinnon (1993): *Estimation and Inference in Econometrics*. Oxford University Press, Oxford.

Enders, W. (2004): *Applied Econometric Time Series*. John Wiley & Sons, 2nd edn.

Engle, R., and C. Granger (1987): "Co-Integration and Error Correction: Representation, Estimation and Testing," *Econometrica*, 55, 251–276.

Ericsson, N. R., and J. G. MacKinnon (2002): "Distributions of Error Correction Tests for Cointegration," *The Econometrics Journal*, 5, 285–318.

Hamilton, J. D. (1994): *Time Series Analysis*. Princeton University Press, Princeton.

Hendry, D. F., and K. Juselius (2000): "Explaining Cointegration Analysis: Part I," *Energy Journal*, 21(1), 1–42.

——— (2001): "Explaining Cointegration Analysis: Part II," *Energy Journal*, 22(1), 75–120.

Johansen, S. (1996): *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, Oxford, 2nd edn.

Juselius, K. (2007): *The Cointegrated VAR model: Econometric Methodology and Macroeconomic Applications*. Oxford University press, Oxford.

Maddala, G. S., and I.-M. Kim (1998): *Unit Roots, Cointegration, and Structural Change*. Cambridge University Press, Cambridge.

Patterson, K. (2000): *An Introduction to Applied Econometrics. A Time Series Approach*. Palgrave MacMillan, New York.

Sims, C. A., J. H. Stock, and M. W. Watson (1990): "Inference in Linear Time Series Models with Some Unit Roots," *Econometrica*, 58(1), 113–144.