Cointegration

- The VAR models discussed so fare are appropriate for modeling $I(0)$ data, like asset returns or growth rates of macroeconomic time series.

- Economic theory, however, often implies equilibrium relationships between the levels of time series variables that are best described as being $I(1)$.

- Similarly, arbitrage arguments imply that the $I(1)$ prices of certain financial time series are linked.

- The statistical concept of cointegration is required to make sense of regression models and VAR models with $I(1)$ data.

Spurious Regression

If some or all of the variables in a regression are $I(1)$ then the usual statistical results may or may not hold. One important case in which the usual statistical results do not hold is *spurious regression,* when all the regressors are $I(1)$ and not cointegrated. That is, there is no linear combination of the variables that is $I(0)$.

Example: Granger-Newbold JOE 1974

Consider two independent and not cointegrated $I(1)$ processes y_{1t} and y_{2t}

$$\begin{aligned} y_{it} &= y_{it-1} + \varepsilon_{it}, \\ \varepsilon_{it} &\sim \text{GWN}(0,1),\ i = 1,2 \end{aligned}$$

Estimated levels and differences regressions

$$\begin{aligned} y_1 &= \underset{(0.39)}{6.74} + \underset{(0.05)}{0.40} \cdot y_2,\ R^2 = 0.21 \\ \Delta y_1 &= \underset{(0.07)}{-0.06} + \underset{(0.06)}{0.03} \cdot \Delta y_2,\ R^2 = 0.00 \end{aligned}$$

Statistical Implications of Spurious Regression

Let $\mathbf{Y}_t = (y_{1t}, \ldots, y_{nt})'$ denote an $(n \times 1)$ vector of $I(1)$ time series that are not cointegrated. Write

$$\mathbf{Y}_t = (y_{1t}, \mathbf{Y}_{2t}')',$$

and consider regressing of y_{1t} on $\mathbf{Y}_{2t}$ giving

$$y_{1t} = \hat{\boldsymbol{\beta}}_2' \mathbf{Y}_{2t} + \hat{u}_t$$

Since y_{1t} is not cointegrated with $\mathbf{Y}_{2t}$

- true value of $\boldsymbol{\beta}_2$ is zero
- The above is a *spurious regression* and $\hat{u}_t \sim I(1)$.

The following results about the behavior of $\hat{\boldsymbol{\beta}}_2$ in the spurious regression are due to Phillips (1986):

- $\hat{\boldsymbol{\beta}}_2$ does not converge in probability to zero but instead converges in distribution to a non-normal random variable not necessarily centered at zero. This is the spurious regression phenomenon.

- The usual OLS t-statistics for testing that the elements of $\boldsymbol{\beta}_2$ are zero diverge to $\pm\infty$ as $T \to \infty$. Hence, with a large enough sample it will appear that $\mathbf{Y}_t$ is cointegrated when it is not if the usual asymptotic normal inference is used.

- The usual R^2 from the regression converges to unity as $T \to \infty$ so that the model will appear to fit well even though it is misspecified.

- Regression with $I(1)$ data only makes sense when the data are cointegrated.

Intuition

Recall, with $I(1)$ data sample moments converge to functions of Brownian motion. Consider two independent and not cointegrated $I(1)$ processes y_{1t} and y_{2t} :

$$y_{it} = y_{it-1} + \varepsilon_{it},\ \varepsilon_{it} \sim \mathsf{WN}(0, \sigma_i^2),\ i = 1, 2$$

Then

$$\begin{aligned} T^{-2} \sum_{t=1}^{T} y_{it}^2 &\Rightarrow \sigma_i^2 \int_0^1 W_i(r)^2 dr,\ i = 1, 2 \\ T^{-2} \sum_{t=1}^{T} y_{1t} y_{2t} &\Rightarrow \sigma_1 \sigma_2 \int_0^1 W_1(r) W_2(r) dr \end{aligned}$$

where $W_1(r)$ and $W_2(r)$ are independent Wiener processes.

In the regression

$$y_{1t} = \hat{\beta} y_{2t} + \hat{u}_t$$

Phillips derived the following convergence result using the FCLT and the CMT:

$$\begin{aligned}
\hat{\beta} &= \left(T^{-2} \sum_{t=1}^{T} y_{2t}^2 \right)^{-1} T^{-2} \sum_{t=1}^{T} y_{1t} y_{2t} \\
&\Rightarrow \left(\sigma_2^2 \int_0^1 W_2(r)^2 dr \right)^{-1} \sigma_1 \sigma_2 \int_0^1 W_1(r) W_2(r) dr
\end{aligned}$$

Therefore,

$$\hat{\beta} \overset{p}{\nrightarrow} 0$$

$$\hat{\beta} \Rightarrow \text{random variable}$$

Cointegration

Let $\mathbf{Y}_t = (y_{1t}, \ldots, y_{nt})'$ denote an $(n \times 1)$ vector of $I(1)$ time series. $\mathbf{Y}_t$ is *cointegrated* if there exists an $(n \times 1)$ vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)'$ such that

$$\boldsymbol{\beta}'\mathbf{Y}_t = \beta_1 y_{1t} + \cdots + \beta_n y_{nt} \sim I(0)$$

In words, the nonstationary time series in $\mathbf{Y}_t$ are cointegrated if there is a linear combination of them that is stationary or $I(0)$.

- The linear combination $\boldsymbol{\beta}'\mathbf{Y}_t$ is often motivated by economic theory and referred to as a *long-run equilibrium* relationship.

- Intuition: $I(1)$ time series with a long-run equilibrium relationship cannot drift too far apart from the equilibrium because economic forces will act to restore the equilibrium relationship.

Normalization

The cointegration vector $\boldsymbol{\beta}$ is not unique since for any scalar c

$$c \cdot \boldsymbol{\beta}'\mathbf{Y}_t = \boldsymbol{\beta}^{*\prime}\mathbf{Y}_t \sim I(0)$$

Hence, some *normalization* assumption is required to uniquely identify $\boldsymbol{\beta}$. A typical normalization is

$$\boldsymbol{\beta} = (1, -\beta_2, \ldots, -\beta_n)'$$

so that

$$\boldsymbol{\beta}'\mathbf{Y}_t = y_{1t} - \beta_2 y_{2t} - \cdots - \beta_n y_{nt} \sim I(0)$$

or

$$\begin{aligned} y_{1t} &= \beta_2 y_{2t} + \cdots + \beta_n y_{nt} + u_t \\ u_t &\sim I(0) = \text{cointegrating residual} \end{aligned}$$

In long-run equilibrium, $u_t = 0$ and the long-run equilibrium relationship is

$$y_{1t} = \beta_2 y_{2t} + \cdots + \beta_n y_{nt}$$

Multiple Cointegrating Relationships

If the $(n \times 1)$ vector $\mathbf{Y}_t$ is cointegrated there may be $0 < r < n$ linearly independent cointegrating vectors. For example, let $n = 3$ and suppose there are $r = 2$ cointegrating vectors

$$\begin{aligned} \boldsymbol{\beta}_1 &= (\beta_{11}, \beta_{12}, \beta_{13})' \\ \boldsymbol{\beta}_2 &= (\beta_{21}, \beta_{22}, \beta_{23})' \end{aligned}$$

Then

$$\begin{aligned} \boldsymbol{\beta}_1'\mathbf{Y}_t &= \beta_{11}y_{1t} + \beta_{12}y_{2t} + \beta_{13}y_{3t} \sim I(0) \\ \boldsymbol{\beta}_2'\mathbf{Y}_t &= \beta_{21}y_{1t} + \beta_{22}y_{2t} + \beta_{23}y_{3t} \sim I(0) \end{aligned}$$

and the (3×2) matrix

$$\mathbf{B}' = \begin{pmatrix} \boldsymbol{\beta}_1' \\ \boldsymbol{\beta}_2' \end{pmatrix} = \begin{pmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{33} \end{pmatrix}$$

forms a *basis* for the space of cointegrating vectors. The linearly independent vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ in the cointegrating basis $\mathbf{B}$ are not unique unless some normalization assumptions are made. Furthermore, any linear combination of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, e.g. $\boldsymbol{\beta}_3 = c_1\boldsymbol{\beta}_1 + c_2\boldsymbol{\beta}_2$ where c_1 and c_2 are constants, is also a cointegrating vector.

Examples of Cointegration and Common Trends in Economics and Finance

Cointegration naturally arises in economics and finance. In economics, cointegration is most often associated with economic theories that imply equilibrium relationships between time series variables:

- The permanent income model implies cointegration between consumption and income, with consumption being the common trend.
- Money demand models imply cointegration between money, income, prices and interest rates.
- Growth theory models imply cointegration between income, consumption and investment, with productivity being the common trend.

- Purchasing power parity implies cointegration between the nominal exchange rate and foreign and domestic prices.

- Covered interest rate parity implies cointegration between forward and spot exchange rates.

- The Fisher equation implies cointegration between nominal interest rates and inflation.

- The expectations hypothesis of the term structure implies cointegration between nominal interest rates at different maturities.

- The present value model of stock prices states that a stock's price is an expected discounted present value of its expected future dividends or earnings.

Remarks:

- The equilibrium relationships implied by these economic theories are referred to as *long-run equilibrium* relationships, because the economic forces that act in response to deviations from equilibrium may take a long time to restore equilibrium. As a result, cointegration is modeled using long spans of low frequency time series data measured monthly, quarterly or annually.

- In finance, cointegration may be a high frequency relationship or a low frequency relationship. Cointegration at a high frequency is motivated by arbitrage arguments.

  - The *Law of One Price* implies that identical assets must sell for the same price to avoid arbitrage opportunities. This implies cointegration between the prices of the same asset trading on different markets, for example.

- Similar arbitrage arguments imply cointegration between spot and futures prices, and spot and forward prices, and bid and ask prices.

Here the terminology long-run equilibrium relationship is somewhat misleading because the economic forces acting to eliminate arbitrage opportunities work very quickly. Cointegration is appropriately modeled using short spans of high frequency data in seconds, minutes, hours or days.

Cointegration and Common Trends

If the $(n \times 1)$ vector time series $\mathbf{Y}_t$ is cointegrated with $0 < r < n$ cointegrating vectors then there are $n - r$ *common* $I(1)$ *stochastic trends.*

For example, let

$$\begin{aligned}
\mathbf{Y}_t &= (y_{1t}, y_{2t})' \sim I(1) \\
\boldsymbol{\varepsilon}_t &= (\varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t})' \sim I(0)
\end{aligned}$$

and suppose that $\mathbf{Y}_t$ is cointegrated with cointegrating vector $\boldsymbol{\beta} = (1, -\beta_2)'$. This cointegration relationship may be represented as

$$\begin{aligned}
y_{1t} &= \beta_2 \sum_{s=1}^{t} \varepsilon_{1s} + \varepsilon_{3t} \\
y_{2t} &= \sum_{s=1}^{t} \varepsilon_{1s} + \varepsilon_{2t}
\end{aligned}$$

The common stochastic trend is $\sum_{s=1}^{t} \varepsilon_{1s}$.

Notice that the cointegrating relationship annihilates the common stochastic trend:

$$\begin{aligned}
\boldsymbol{\beta}'\mathbf{Y}_t &= y_{1t} - \beta_2 y_{2t} \\
&= \beta_2 \sum_{s=1}^{t} \varepsilon_{1s} + \varepsilon_{3t} - \beta_2 \left(\sum_{s=1}^{t} \varepsilon_{1s} + \varepsilon_{2t} \right) \\
&= \varepsilon_{3t} - \beta_2 \varepsilon_{2t} \sim I(0).
\end{aligned}$$

Some Simulated Cointegrated Systems

Cointegrated systems may be conveniently simulated using Phillips' (1991) *triangular representation*. For example, consider a bivariate cointegrated system for $\mathbf{Y}_t = (y_{1t}, y_{2t})'$ with cointegrating vector $\boldsymbol{\beta} = (1, -\beta_2)'$. A triangular representation has the form

$$\begin{aligned} y_{1t} &= \beta_2 y_{2t} + u_t, \text{ where } u_t \sim I(0) \\ y_{2t} &= y_{2t-1} + v_t, \text{ where } v_t \sim I(0) \end{aligned}$$

- The first equation describes the long-run equilibrium relationship with an $I(0)$ disequilibrium error u_t.

- The second equation specifies y_{2t} as the common stochastic trend with innovation v_t:

$$y_{2t} = y_{20} + \sum_{j=1}^{t} v_j.$$

- In general, the innovations u_t and v_t may be contemporaneously and serially correlated. The time series structure of these innovations characterizes the short-run dynamics of the cointegrated system.

- The system with $\beta_2 = 1$, for example, might be used to model the behavior of the logarithm of spot and forward prices, spot and futures prices, stock prices and dividends, or consumption and income.

Example: Bivariate system with $\boldsymbol{\beta} = (1, -1)'$

$$\begin{aligned} y_{1t} &= y_{2t} + u_t \\ y_{2t} &= y_{2t-1} + v_t \\ u_t &= 0.75 u_{t-1} + \varepsilon_t, \\ \varepsilon_t &\sim \text{iid } N(0, (0.5)^2), \\ v_t &\sim \text{iid } N(0, (0.5)^2) \end{aligned}$$

Note: y_{2t} defines the common trend

Trivariate cointegrated system with 1 cointegrating vector $\boldsymbol{\beta} = (1, -\beta_1, \beta_2)'$

$$\begin{aligned} y_{1t} &= \beta_1 y_{2t} + \beta_2 y_{3t} + u_t,\ u_t \sim I(0) \\ y_{2t} &= y_{2t-1} + v_t,\ v_t \sim I(0) \\ y_{3t} &= y_{3t-1} + w_t,\ w_t \sim I(0) \end{aligned}$$

An example of a trivariate cointegrated system with one cointegrating vector is a system of nominal exchange rates, home country price indices and foreign country price indices. A cointegrating vector $\boldsymbol{\beta} = (1, -1, -1)'$ implies that the real exchange rate is stationary.

Example: Trivariate cointegrated system $\boldsymbol{\beta} = (1, -0.5, -0.5)'$

$$\begin{aligned} y_{1t} &= 0.5 y_{2t} + 0.5 y_{3t} + u_t, \\ u_t &= 0.75 u_{t-1} + \varepsilon_t, \varepsilon_t \sim \text{iid } N(0, (0.5)^2) \\ y_{2t} &= y_{2t-1} + v_t,\ v_t \sim \text{iid } N(0, (0.5)^2) \\ y_{3t} &= y_{3t-1} + w_t,\ w_t \sim \text{iid } N(0, (0.5)^2) \end{aligned}$$

Note: y_{2t} and y_{3t} are the common trends

Simulated trivariate cointegrated system with 2 cointegrating vectors

A triangular representation for this system with cointegrating vectors $\boldsymbol{\beta}_1 = (1, 0, -\beta_{13})'$ and $\boldsymbol{\beta}_2 = (0, 1, -\beta_{23})'$ is

$$\begin{aligned} y_{1t} &= \beta_{13} y_{3t} + u_t, \;\; u_t \sim I(0) \\ y_{2t} &= \beta_{23} y_{3t} + v_t, \;\; v_t \sim I(0) \\ y_{3t} &= y_{3t-1} + w_t, \;\; w_t \sim I(0) \end{aligned}$$

An example in finance of such a system is the term structure of interest rates where y_3 represents the short rate and y_1 and y_2 represent two different long rates. The cointegrating relationships would indicate that the spreads between the long and short rates are stationary.

Example: Trivariate system with $\boldsymbol{\beta}_1 = (1, 0, -1)'$, $\beta_2 = (0, 1, -1)'$

$$\begin{aligned} y_{1t} &= y_{3t} + u_t, \\ u_t &= 0.75u_{t-1} + \varepsilon_t, \varepsilon_t \sim \text{iid } N(0, (0.5)^2) \\ y_{2t} &= y_{3t} + v_t, \\ v_t &= 0.75v_{t-1} + \eta_t, \eta_t \sim \text{iid } N(0, (0.5)^2) \\ y_{3t} &= y_{3t-1} + w_t,\ w_t \sim \text{iid } N(0, (0.5)^2) \end{aligned}$$

Note: y_{3t} defines the common trend.

Cointegration and Error Correction Models

Consider a bivariate $I(1)$ vector $\mathbf{Y}_t = (y_{1t}, y_{2t})'$ and assume that $\mathbf{Y}_t$ is cointegrated with cointegrating vector $\boldsymbol{\beta} = (1, -\beta_2)'$ so that $\boldsymbol{\beta}'\mathbf{Y}_t = y_{1t} - \beta_2 y_{2t}$ is $I(0)$. Engle and Granger's famous (1987) *Econometrica* paper showed that cointegration implies the existence of an *error correction model* (ECM) of the form

$$\begin{aligned}
\Delta y_{1t} &= c_1 + \alpha_1(y_{1t-1} - \beta_2 y_{2t-1}) \\
&\quad + \sum_j \psi_{11}^j \Delta y_{1t-j} + \sum_j \psi_{12}^j \Delta y_{2t-j} + \varepsilon_{1t} \\
\Delta y_{2t} &= c_2 + \alpha_2(y_{1t-1} - \beta_2 y_{2t-1}) \\
&\quad + \sum_j \psi_{21}^j \Delta y_{1t-j} + \sum_j \psi_{22}^j \Delta y_{2t-j} + \varepsilon_{2t}
\end{aligned}$$

The ECM links the long-run equilibrium relationship implied by cointegration with the short-run dynamic adjustment mechanism that describes how the variables react when they move out of long-run equilibrium.

Let y_t denote the log of real income and c_t denote the log of consumption and assume that $\mathbf{Y}_t = (y_t, c_t)'$ is $I(1)$. The *Permanent Income Hypothesis* implies that income and consumption are cointegrated with $\boldsymbol{\beta} = (1, -1)'$:

$$\begin{aligned} c_t &= \mu + y_t + u_t \\ \mu &= E[c_t - y_t] \\ & u_t \sim I(0) \end{aligned}$$

Suppose the ECM has the form

$$\begin{aligned} \Delta y_t &= \gamma_y + \alpha_y(c_{t-1} - y_{t-1} - \mu) + \varepsilon_{yt} \\ \Delta c_t &= \gamma_c + \alpha_c(c_{t-1} - y_{t-1} - \mu) + \varepsilon_{ct} \end{aligned}$$

The first equation relates the growth rate of income to the lagged disequilibrium error $c_{t-1} - y_{t-1} - \mu$, and the second equation relates the growth rate of consumption to the lagged disequilibrium as well. The reactions of y_t and c_t to the disequilibrium error are captured by the *adjustment coefficients* α_y and α_c.

Consider the special case

$$\begin{aligned}\Delta y_t &= \gamma_y + 0.5(c_{t-1} - y_{t-1} - \mu) + \varepsilon_{yt}, \\ \Delta c_t &= \gamma_c + \varepsilon_{ct}.\end{aligned}$$

Consider three situations:

1. $c_{t-1} - y_{t-1} - \mu = 0$. Then

$$\begin{aligned}E[\Delta y_t|\mathbf{Y}_{t-1}] &= \gamma_y \\ E[\Delta c_t|\mathbf{Y}_{t-1}] &= \gamma_d\end{aligned}$$

2. $c_{t-1} - y_{t-1} - \mu > 0$. Then

$$E[\Delta y_t|\mathbf{Y}_{t-1}] = \gamma_y + 0.5(c_{t-1} - y_{t-1} - \mu) > \gamma_y$$

Here the consumption has increased above its long-run mean (positive disequilibrium error) and the ECM predicts that y_t will grow faster than its long-run rate to restore the consumption-income ratio its long-run mean.

3. $c_{t-1} - y_{t-1} - \mu < 0$. Then

$$E[\Delta y_t | \mathbf{Y}_{t-1}] = \gamma_y + 0.5(c_{t-1} - y_{t-1} - \mu) < c_y$$

Here consumption-income ratio has decreased below its long-run mean (negative disequilibrium error) and the ECM predicts that y_t will grow more slowly than its long-run rate to restore the consumption-income ratio to its long-run mean.

Tests for Cointegration

Let the $(n \times 1)$ vector $\mathbf{Y}_t$ be $I(1)$. Recall, $\mathbf{Y}_t$ is cointegrated with $0 < r < n$ cointegrating vectors if there exists an $(r \times n)$ matrix $\mathbf{B}'$ such that

$$\mathbf{B}'\mathbf{Y}_t = \begin{pmatrix} \boldsymbol{\beta}_1'\mathbf{Y}_t \\ \vdots \\ \boldsymbol{\beta}_r'\mathbf{Y}_t \end{pmatrix} = \begin{pmatrix} u_{1t} \\ \vdots \\ u_{rt} \end{pmatrix} \sim I(0)$$

Testing for cointegration may be thought of as testing for the existence of long-run equilibria among the elements of $\mathbf{Y}_t$.

Cointegration tests cover two situations:

- There is at most one cointegrating vector
  - Originally considered by Engle and Granger (1986), "Cointegration and Error Correction: Representation, Estimation and Testing," *Econometrica*. They developed a simple two-step residual-based testing procedure based on regression techniques.

- There are possibly $0 \leq r < n$ cointegrating vectors.
  - Originally considered by Johansen (1988), "Statistical Analysis of Cointegration Vectors," *Journal of Economics Dynamics and Control*. He developed a sophisticated sequential procedure for determining the existence of cointegration and for determining the number of cointegrating relationships based on maximum likelihood techniques.

Residual-Based Tests for Cointegration

Engle and Granger's two-step procedure for determining if the $(n \times 1)$ vector $\boldsymbol{\beta}$ is a cointegrating vector is as follows:

- Form the cointegrating residual $\boldsymbol{\beta}'\mathbf{Y}_t = u_t$
- Perform a unit root test on u_t to determine if it is $I(0)$.

The null hypothesis in the Engle-Granger procedure is no-cointegration and the alternative is cointegration.

There are two cases to consider.

1. The proposed cointegrating vector $\boldsymbol{\beta}$ is pre-specified (not estimated). For example, economic theory may imply specific values for the elements in $\boldsymbol{\beta}$ such as $\boldsymbol{\beta} = (1, -1)'$. The cointegrating residual is then readily constructed using the prespecified cointegrating vector.

2. The proposed cointegrating vector is estimated from the data and an estimate of the cointegrating residual $\hat{\boldsymbol{\beta}}'\mathbf{Y}_t = \hat{u}_t$ is formed.

Note: Tests for cointegration using a pre-specified cointegrating vector are generally much more powerful than tests employing an estimated vector.

Testing for Cointegration When the Cointegrating Vector Is Pre-specified

Let $\mathbf{Y}_t$ denote an $(n \times 1)$ vector of $I(1)$ time series, let $\boldsymbol{\beta}$ denote an $(n \times 1)$ prespecified cointegrating vector and let

$$u_t = \boldsymbol{\beta}'\mathbf{Y}_t = \text{cointegrating residual}$$

The hypotheses to be tested are

$$\begin{aligned} H_0 &: u_t = \boldsymbol{\beta}'\mathbf{Y}_t \sim I(1) \text{ (no cointegration)} \\ H_1 &: u_t = \boldsymbol{\beta}'\mathbf{Y}_t \sim I(0) \text{ (cointegration)} \end{aligned}$$

- Any unit root test statistic may be used to evaluate the above hypotheses. The most popular choices are the ADF and PP statistics, but one may also use the more powerful ERS and Ng-Perron tests.

- Cointegration is found if the unit root test rejects the no-cointegration null.

- It should be kept in mind, however, that the cointegrating residual may include deterministic terms (constant or trend) and the unit root tests should account for these terms accordingly.

Testing for Cointegration When the Cointegrating Vector Is Estimated

Let $\mathbf{Y}_t$ denote an $(n \times 1)$ vector of $I(1)$ time series and let $\boldsymbol{\beta}$ denote an $(n \times 1)$ unknown cointegrating vector. The hypotheses to be tested are

$$\begin{aligned} H_0 &: u_t = \boldsymbol{\beta}'\mathbf{Y}_t \sim I(1) \text{ (no cointegration)} \\ H_1 &: u_t = \boldsymbol{\beta}'\mathbf{Y}_t \sim I(0) \text{ (cointegration)} \end{aligned}$$

- Since $\boldsymbol{\beta}$ is unknown, to use the Engle-Granger procedure it must be first estimated from the data.

- Before $\boldsymbol{\beta}$ can be estimated some normalization assumption must be made to uniquely identify it.

- A common normalization is to specify $\mathbf{Y}_t = (y_{1t}, \mathbf{Y}_{2t}')'$ where $\mathbf{Y}_{2t} = (y_{2t}, \ldots, y_{nt})'$ is an $((n-1) \times 1)$ vector and the cointegrating vector is normalized as $\boldsymbol{\beta} = (1, -\boldsymbol{\beta}_2')'$.

Engle and Granger propose estimating the normalized cointegrating vector $\boldsymbol{\beta}_2$ by least squares from the regression

$$\begin{aligned} y_{1t} &= \boldsymbol{\gamma}'\mathbf{D}_t + \boldsymbol{\beta}_2'\mathbf{Y}_{2t} + u_t \\ \mathbf{D}_t &= \text{deterministic terms} \end{aligned}$$

and testing the no-cointegration hypothesis with a unit root test using the estimated cointegrating residual

$$\hat{u}_t = y_{1t} - \hat{\boldsymbol{\gamma}}'\mathbf{D}_t - \hat{\boldsymbol{\beta}}_2\mathbf{Y}_{2t}$$

The unit root test regression in this case is without deterministic terms (constant or constant and trend).

For example, if one uses the ADF test, the test regression is

$$\Delta\hat{u}_t = \pi\hat{u}_{t-1} + \sum_{j=1}^{p} \xi\Delta\hat{u}_{t-j} + error$$

Distribution Theory

- Phillips and Ouliaris (PO) (1990) show that ADF and PP unit root tests (t-tests and normalized bias) applied to the estimated cointegrating residual *do not* have the usual Dickey-Fuller distributions under the null hypothesis of no-cointegration.

- Due to the spurious regression phenomenon under the null hypothesis, the distribution of the ADF and PP unit root tests have asymptotic distributions that are functions of Wiener processes that

  – Depend on the deterministic terms in the regression used to estimate $\boldsymbol{\beta}_2$

  – Depend on the number of variables, $n-1$, in $\mathbf{Y}_{2t}$.

- These distributions are known as the *Phillips-Ouliaris* (PO) distributions, and are described in Phillips and Ouliaris (1990).

To further complicate matters, Hansen (1992) showed the appropriate PO distributions of the ADF and PP unit root tests applied to the residuals also depend on the trend behavior of y_{1t} and $\mathbf{Y}_{2t}$ as follows:

Case I: $\mathbf{Y}_{2t}$ and y_{1t} are both $I(1)$ without drift and $\mathbf{D}_t = 1$. The ADF and PP unit root test statistics follow the PO distributions, adjusted for a constant, with dimension parameter $n - 1$.

Case II: $\mathbf{Y}_{2t}$ is $I(1)$ with drift, y_{1t} may or may not be $I(1)$ with drift and $\mathbf{D}_t = 1$. The ADF and PP unit root test statistics follow the PO distributions, adjusted for a constant and trend, with dimension parameter $n - 2$. If $n - 2 = 0$ then the ADF and PP unit root test statistics follow the DF distributions adjusted for a constant and trend.

Case III: $\mathbf{Y}_{2t}$ is $I(1)$ without drift, y_{1t} is $I(1)$ with drift and $\mathbf{D}_t = (1, t)'$. The resulting ADF and PP unit root test statistics on the residuals follow the PO distributions, adjusted for a constant and trend, with dimension parameter $n - 1$.

Example: PO Critical Values

PO Critical Values: Case I		
n-1	1%	5%
1	-3.89	-3.36
2	-4.29	-3.74
3	-4.64	-4.09
4	-4.96	-4.41
5	-5.24	-4.71

Regression-Based Estimates of Cointegrating Vectors and Error Correction Models

Least Square Estimator

Least squares may be used to consistently estimate a normalized cointegrating vector. However, the asymptotic behavior of the least squares estimator is non-standard. The following results about the behavior of $\hat{\boldsymbol{\beta}}_2$ if $\mathbf{Y}_t$ is cointegrated are due to Stock (1987) and Phillips (1991):

- $T(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2)$ converges in distribution to a non-normal random variable not necessarily centered at zero.

- The least squares estimate $\hat{\boldsymbol{\beta}}_2$ is consistent for $\boldsymbol{\beta}_2$ and converges to $\boldsymbol{\beta}_2$ at rate T instead of the usual rate $T^{1/2}$. That is, $\hat{\boldsymbol{\beta}}_2$ is *super consistent*.

- $\hat{\boldsymbol{\beta}}_2$ is consistent even if $\mathbf{Y}_{2t}$ is correlated with u_t so that there is no asymptotic simultaneity bias.

- In general, the asymptotic distribution of $T(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2)$ is asymptotically biased and non-normal. The usual OLS formula for computing $\widehat{\text{avar}}(\hat{\boldsymbol{\beta}}_2)$ is incorrect and so the usual OLS standard errors are not correct.

- Even though the asymptotic bias goes to zero as T gets large $\hat{\boldsymbol{\beta}}_2$ may be substantially biased in small samples. The least squres estimator is also not efficient.

The above results indicate that the least squares estimator of the cointegrating vector $\boldsymbol{\beta}_2$ could be improved upon. A simple improvement is suggested by Stock and Watson (1993).

What causes the bias and non-normality in $\hat{\boldsymbol{\beta}}_2$?

Assume a triangular representation of the form

$$\begin{aligned} y_{1t} &= \mathbf{Y}_{2t}'\boldsymbol{\beta}_2 + u_{1t} \\ \mathbf{Y}_{2t} &= \mathbf{Y}_{2t-1} + \mathbf{u}_{2t} \end{aligned}$$

The bias and non-normality is a function of the time series structure of $\mathbf{u}_t = (u_{1t}, \mathbf{u}_{2t}')'$. Assume a Wold structure for $\mathbf{u}_t$

$$\begin{pmatrix} u_{1t} \\ \mathbf{u}_{2t} \end{pmatrix} = \begin{pmatrix} \psi_{11}(L) & \boldsymbol{\Psi}_{12}(L) \\ \boldsymbol{\Psi}_{21}(L) & \boldsymbol{\Psi}_{22}(L) \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \boldsymbol{\varepsilon}_{2t} \end{pmatrix}$$

$$\begin{pmatrix} \varepsilon_{1t} \\ \boldsymbol{\varepsilon}_{2t} \end{pmatrix} \sim \text{iid } N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \boldsymbol{\sigma}_{12} \\ \boldsymbol{\sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right)$$

Result: There is no asymptotic bias if $\mathbf{u}_t \sim WN(\mathbf{0}, \mathbf{I}_n)$

Stock and Watson's Efficient Lead/Lag Estimator

Stock and Watson (1993) provide a very simple method for obtaining an asymptotically efficient (equivalent to maximum likelihood) estimator for the normalized cointegrating vector $\boldsymbol{\beta}_2$ as well as a valid formula for computing its asymptotic variance. Let

$$\begin{aligned}\mathbf{Y}_t &= (y_{1t}, \mathbf{Y}_{2t}')' \\ \mathbf{Y}_{2t} &= (y_{2t}, \ldots, y_{nt})' \\ \boldsymbol{\beta} &= (1, -\boldsymbol{\beta}_2')'\end{aligned}$$

Stock and Watson's efficient estimation procedure is:

- Augment the cointegrating regression of y_{1t} on $\mathbf{Y}_{2t}$ with appropriate deterministic terms $\mathbf{D}_t$ with p leads and lags of $\Delta\mathbf{Y}_{2t}$

$$
\begin{aligned}
y_{1t} &= \boldsymbol{\gamma}'\mathbf{D}_t + \boldsymbol{\beta}_2'\mathbf{Y}_{2t} + \sum_{j=-p}^{p} \boldsymbol{\psi}_j'\Delta\mathbf{Y}_{2t-j} + u_t \\
&= \boldsymbol{\gamma}'\mathbf{D}_t + \boldsymbol{\beta}_2'\mathbf{Y}_{2t} + \boldsymbol{\psi}_0'\Delta\mathbf{Y}_{2t} \\
&\quad +\boldsymbol{\psi}_p'\Delta\mathbf{Y}_{2t+p} + \cdots + \boldsymbol{\psi}_1'\Delta\mathbf{Y}_{2t+1} \\
&\quad +\boldsymbol{\psi}_{-1}'\Delta\mathbf{Y}_{2t-1} + \cdots + \boldsymbol{\psi}_{-p}'\Delta\mathbf{Y}_{2t-p} + u_t
\end{aligned}
$$

- Estimate the augmented regression by least squares. The resulting estimator of $\boldsymbol{\beta}_2$ is called the *dynamic OLS* estimator and is denoted $\hat{\boldsymbol{\beta}}_{2,DOLS}$. It will be consistent, asymptotically normally distributed and efficient (equivalent to MLE) under certain assumptions (see Stock and Watson (1993))

- Asymptotically valid standard errors for the individual elements of $\hat{\boldsymbol{\beta}}_{2,DOLS}$ are given by the OLS standard errors multiplied by the ratio

$$\left(\frac{\hat{\sigma}_u^2}{\widehat{\mathsf{lrv}}(u_t)}\right)^{1/2}$$

  where $\hat{\sigma}_u^2$ is the OLS estimate of var(u_t) and $\widehat{\mathsf{lrv}}(u_t)$ is any consistent estimate of the long-run variance of u_t using the residuals $\hat{u}_t$ from. Alternatively, the Newey-West HAC standard errors may also be used.

- An alternative method to correct the standard errors utilizes a Cochrane-Orcutt GLS transformation of the error term. This is called the *dynamic GLS estimator*. Recently, Okagi and Ling (2005) have utilized this estimator to produce improved tests for cointegration based on Hausman-type tests.

Estimating Error Correction Models by Least Squares

Consider a bivariate $I(1)$ vector $\mathbf{Y}_t = (y_{1t}, y_{2t})'$ and assume that $\mathbf{Y}_t$ is cointegrated with cointegrating vector $\boldsymbol{\beta} = (1, -\beta_2)'$ so that $\boldsymbol{\beta}'\mathbf{Y}_t = y_{1t} - \beta_2 y_{2t}$ is $I(0)$. Suppose one has a consistent estimate $\hat{\beta}_2$ (by OLS or DOLS) of the cointegrating coefficient and is interested in estimating the corresponding error correction model for Δy_{1t} and Δy_{2t} using

$$\begin{aligned}
\Delta y_{1t} &= c_1 + \alpha_1(y_{1t-1} - \hat{\beta}_2 y_{2t-1}) \\
&\quad + \sum_j \psi_{11}^j \Delta y_{1t-j} + \sum_j \psi_{12}^j \Delta y_{2t-j} + \varepsilon_{1t} \\
\Delta y_{2t} &= c_2 + \alpha_2(y_{1t-1} - \hat{\beta}_2 y_{2t-1}) \\
&\quad + \sum_j \psi_{21}^j \Delta y_{1t-j} + \sum_j \psi_{22}^2 \Delta y_{2t-j} + \varepsilon_{2t}
\end{aligned}$$

- Because $\hat{\beta}_2$ is super consistent it may be treated as known in the ECM, so that the estimated disequilibrium error $y_{1t} - \hat{\beta}_2 y_{2t}$ may be treated like the known disequilibrium error $y_{1t} - \beta_2 y_{2t}$.

- Since all variables in the ECM are $I(0)$, the two regression equations may be consistently estimated using ordinary least squares (OLS).

- Alternatively, the ECM system may be estimated by seemingly unrelated regressions (SUR) to increase efficiency if the number of lags in the two equations are different.