Section B4: Unit Roots and Cointegration Analysis

B4.1. Cointegrated Time Series

Some pairs of economic time series (e.g., wholesale and retail prices of a single commodity) may be expected to follow similar patterns of change over time. Even when short-term conditions cause such series to diverge, economic and/or policy dynamics will eventually force them back into equilibrium. In the economics literature, such pairs of data series are called *cointegrated* series.

To analyze cointegrated time series, we conceptualize them as *stochastic processes*, i.e., processes subject to randomness, and define properties of these processes. The notation in this section differs from that of Sections 1 through 3.

B4.2. Definitions of Short and Long Memory Stochastic Processes

We begin by defining some properties of stochastic processes. The stochastic process $\{\varepsilon_t\}$ is *stationary* (or strictly stationary) if, for every collection of time indices $\{t_j\}_{j=1}^{n}$, the joint distribution of $\{\varepsilon_{t_j}\}_{j=1}^{n}$ is the same as the joint distribution of $\{\varepsilon_{t_j+h}\}_{j=1}^{n}$ for all integers $h \geq 1$. It follows that the ε_t are identically distributed.

Because strict stationarity is defined in terms of the joint distributions of subsets of the stochastic process, the strict stationary of an empirical data series is often difficult to establish. Many data series, however, are clearly *nonstationary,* e.g., series that clearly display trends or cycles.

Covariance stationarity is defined in terms of the moments of the stochastic process and is thus easier to establish for an empirical data series. The stochastic process $\{\varepsilon_t\}$ is covariance stationary if, for some constant c,

$$\begin{aligned} &E[\varepsilon_t] = c, \\ &Var[\varepsilon_t] = \sigma_\varepsilon^2 \text{ (constant variance), and} \\ &Cov[\varepsilon_{t,}\varepsilon_{t+h}] = f(h), \text{ for } h \geq 1. \end{aligned} \tag{B4.2.1}$$

If a strictly stationary process has a finite second moment, it is covariance stationary. A covariance stationary series, however, need not be stationary.

A covariance stationary process $\{\varepsilon_t\}$ is a white noise process if

$$\begin{aligned} &E[\varepsilon_t] = 0, \\ &E[\varepsilon_t \varepsilon_{t-h}] = 0 \text{ for } h \neq 0 \text{ (uncorrolated), and} \\ &Var[\varepsilon_t] = \sigma_\varepsilon^2 \text{ (constant variance).} \end{aligned} \tag{B4.2.2}$$

As $t \to \infty$, $\{\varepsilon_t\}$ will frequently cross the horizontal axis and frequently return to values near its mean 0. White noise is used to model unexplained short-term movements in a data series.

Let $\{\varepsilon_t\}$ be a white noise process, and let

$$x_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}. \qquad \text{(B4.2.3)}$$

If $a_j \to 0$ as $j \to \infty$, then $\{x_t\}$ is a *short memory* process. When $\{x_t\}$ is a short memory process with constant mean and finite variance, we say that $\{x_t\}$ is *integrated of order 0* and write $\{x_t\} \sim I(0)$. White noise, for example, is a short memory process. In a typical empirical time series generated by a short memory process, autocorrelation is present but limited. When $a_j \nrightarrow 0$ as $j \to \infty$, $\{x_t\}$ is a *long memory* process.

B4.2. Analyzing Short and Long Memory Stochastic Processes

If $\{y_t\}$ is a long memory process, an old shock to the system (ε_{t-h}, where h is large) still has an effect on y_t. If, however, $\{y_t\}$ is a short memory process, the effect of ε_{t-h} on y_t diminishes quickly as h increases, i.e., the process "forgets" old shocks. Although short memory processes are usually serially correlated, the covariance of the values between two time periods diminishes as the distance between the time periods grows ($Cov[y_{t-h}, y_t] \to 0$ as $h \to \infty$).

Differencing a Long Memory Process

In many cases, differencing a long memory process produces a short memory process. Let

$$y_t = y_{t-1} + \varepsilon_t, \qquad \text{(B4.2.4)}$$

where ε_t is a white noise process. Then, although $\{y_t\}$ is a long memory process, the differenced process

$$\Delta y_t = y_t - y_{t-1} = \varepsilon_t \qquad \text{(B4.2.5)}$$

is a short memory process. When $\{y_t\}$ is a long memory process and $\{\Delta y_t\}$ is a short memory process, we say that $\{y_t\}$ is *integrated of order 1* and write $\{y_t\} \sim I(1)$.

Cointegration

Before presenting a rigorous definition of cointegration, we state some facts about linear combinations of $I(0)$ and $I(1)$ processes. Let a and b be constants.

a) If $\{x_t\} \sim I(0)$, then $a + bx_t$ is $I(0)$.

b) If $\{x_t\} \sim I(1)$, then $a + bx_t$ is $I(1)$.

c) A linear combination of short memory processes is a short memory process, i.e., if $\{x_t\} \sim I(0)$, and $\{y_t\} \sim I(0)$, then $ax_t + by_t$ is $I(0)$.

d) A linear combination of a short memory process and a long memory process is a long memory process, i.e., if $\{x_t\} \sim I(0)$, and $\{y_t\} \sim I(1)$, then $ax_t + by_t$ is $I(1)$.

Definition of Cointegration

A linear combination of two long memory processes is a long memory process unless the two processes "share" the same long-term memory. Suppose $\{x_t\} \sim I(1)$ and $\{y_t\} \sim I(1)$. If there exist constants *m*, *a,* and *b* such that

1. $m + ax_t + by_t$ is $I(0)$ and
2. $E[m + ax_t + by_t] = 0$,

then $\{x_t\}$ and $\{y_t\}$ are *cointegrated.*

Intuitively, because cointegrated long memory series share the same long-term memory, their long-term memories cancel out in the linear combination, leaving a short memory series. In the econometrics literature, the term "cointegrated" is commonly used even when the relationship between the two series is not strictly linear (e.g., it may be log-linear).

Problem of Spurious Regression

In time series regression analysis, we assume that one time series can be expressed as a linear combination of other time series, modulo an error term. When the time series have long memories, we are assuming cointegration.

Traditional regression diagnostics can be deceptive in the presence of long memory series. In particular, we may see high values of R^2 and low standard errors, leading to inflated *t*-statistics. There is a high probability of regression diagnostics indicating a relationship between two independent, randomly generated $I(1)$ series (Newbold and Granger 1974). Differencing the series may eliminate this problem but allows only analysis of short-run changes. Use of additional statistical tests and diagnostics enables us to analyze long-run relationships between cointegrated series.

B4.3. Testing for Unit Roots

Definition of a Unit Root Process in the Linear Case

If

$$y_t = \alpha + \rho y_{t-1} + \varepsilon_t, \qquad \text{(B4.3.1)}$$

where $E(\varepsilon_t | y_{t-1}, y_{t-2}, \ldots, y_0) = 0$, then $\{y_t\}$ has a unit root if and only if $|\rho| = 1$.

When $\rho = 1$ and $\alpha = 0$, $\{y_t\}$ is a random walk *without drift.* When $\rho = 1$ and $\alpha \neq 0$, $\{y_t\}$ is a random walk *with drift* (Figure 5). Unit root processes are long memory processes.

Testing H_0: $|\rho| = 1$

Observe that

- When $\rho < 0$, $\{y_t\}$ is negatively autocorrelated, which is rare in economic time series.

- When $|\rho| > 1$, $\{y_t\}$ is blowing up (positive or negative).

So we are usually interested in testing

$$H_0: \rho = 1 \qquad \text{vs.} \qquad H_1: 0 < \rho < 1. \tag{B4.3.2}$$

Dickey-Fuller Test for a Unit Root
We set $\theta = \rho - 1$. Then

$$\begin{aligned} \Delta y_t &= \alpha + \rho y_{t-1} - y_{t-1} + \varepsilon_t \\ &= \alpha + \theta y_{t-1} + \varepsilon_t \end{aligned} \tag{B4.3.3}$$

Now we can test

$$H_0: \theta = 0 \qquad \text{vs.} \qquad H_1: \theta < 0. \tag{B4.3.4}$$

We can compute the usual *t*-statistic for the hypothesis test. However, under H_0, the *t*-statistic is not asymptotically normal. Dickey and Fuller (1979) tabulated the critical values for the asymptotic distribution, which is known as the Dickey-Fuller (DF) distribution.

Asymptotic Critical Values for the Dickey-Fuller (DF) Unit Root Test

Significance Level	1%	2.5%	5%	10%
Critical Value	-3.43	-3.12	-2.86	-2.57

We reject $H_0: \theta = 0$ against the alternative $H_1: \theta < 0$ if the *t*-statistic $t_{\hat{\theta}} < c$, where *c* is one of the critical values above.

If $t_{\hat{\theta}}$ is strongly negative, we have evidence for rejecting the hypothesis that $\{y_t\} \sim I(1)$ in favor of the hypothesis that $\{y_t\} \sim I(0)$. Otherwise, we suspect that $\{y_t\}$ has a unit root, and spurious regression is a concern.

Application of the asymptotic critical values for the DF test requires a large sample, i.e., a long time series. Critical values for small samples have also been tabulated and are available in statistical software packages such as SAS and R.

The *power* of a hypothesis test is the probability of *not* committing a Type II error. Type II error is failing to reject a false null hypothesis. Unit root tests have been criticized for their low power. There is a relatively high probability that these tests may indicate a unit root in a series with no unit roots. In addition to the DF test, other tests for autocorrelation and unit roots are discussed in the literature (e.g., Ljung-Box, Durbin-Watson). The DF test is popular because it is simple and robust.

The Cointegrating Parameter

Suppose we have determined that $\{y_t\}$ and $\{x_t\}$ both have unit roots.

- If $\{y_t\}$ and $\{x_t\}$ are *not* cointegrated, regression of one series on the other may be "spurious."

- If $\{y_t\}$ and $\{x_t\}$ are cointegrated, we are interested in the cointegrating parameter β such that $y_t - \beta x_t$ is an $I(0)$ process.

When we know or hypothesize the cointegrating parameter β, we can set $z_t = y_t - \beta x_t$ and perform the standard DF test on the series $\{z_t\}$. If we find evidence that $\{z_t\}$ has a unit root, we may conclude that $\{y_t\}$ and $\{x_t\}$ are not cointegrated. (In practice, we may also consider that our hypothesized value of β may be incorrect.) The null hypothesis in the DF test is that $\{z_t\}$ has a unit root, i.e., that $\{y_t\}$ and $\{x_t\}$ are *not* cointegrated.

When $\{y_t\}$ and $\{x_t\}$ are cointegrated, the OLS estimator $\hat{\beta}$ from the regression

$$y_t = \hat{\alpha} + \hat{\beta} x_t + \varepsilon_t \tag{B4.3.5}$$

is a consistent estimator of β. However, under the null hypotheses that $z_t = y_t - \beta x_t$ has a unit root, we must run a spurious regression to obtain $\hat{\beta}$.

To perform the DF test with an estimated value of β, we set

$$\hat{u}_t = y_t - \hat{\beta} x_t, \tag{B4.3.6}$$

where $\hat{\beta}$ is the OLS estimator of β. Davidson and McKinnon (1993) tabulated critical values for the DF test for the presence of a unit root in the series $\{\hat{u}_t\}$.

Asymptotic Critical Values for the Dickey-Fuller (DF) Unit Root Test with Estimated $\boldsymbol{\beta}$

Significance Level	1%	2.5%	5%	10%
Critical Value	-3.90	-3.59	-3.34	-3.04

To perform a cointegration test with an estimated value of β, we run the regression

$$\Delta\hat{u}_t = \tau + \theta\hat{u}_{t-1} + \varepsilon_t \tag{B4.3.7}$$

and test $H_0\colon \theta = 0$ against the alternative $H_1\colon \theta < 0$. If the *t*-statistic is below the critical value, we have evidence that $\{y_t\}$ and $\{x_t\}$ are cointegrated. When β is estimated, we must get a *t*-statistic more strongly negative to show cointegration. This is because OLS tends to produce residuals that look like an $I(0)$ sequence even when $\{y_t\}$ and $\{x_t\}$ are not cointegrated.

Cointegration Test for Series with Linear Time Trends

Suppose $\{y_t\}$ and $\{x_t\}$ display linear time trends. The trends need not be the same, e.g, $\{y_t\}$ may be increasing faster than $\{x_t\}$. Let $y_t = \gamma t + g_t$, and let $x_t = \zeta t + z_t$. The stochastic portions of the series, $\{g_t\}$ and $\{z_t\}$, may be cointegrated. We may remove the trends from x_t and y_t and then test for unit roots and the cointegration of $\{g_t\}$ with $\{z_t\}$. Alternatively, we may run the regression

$$y_t = \hat{\alpha} + \hat{\eta}t + \hat{\beta}x_t \tag{B4.3.8}$$

and apply the DF test to the residuals $\hat{e}_t$. In this case, we note the following:

- The asymptotic critical values differ from those of the standard DF test and are given in the table below.
- If we find that $\{g_t\}$ and $\{z_t\}$ are cointegrated with cointegration parameter β, this means that $y_t - \beta x_t$ is not an $I(1)$ process.
- It's possible that $y_t - \beta x_t$ will still have a linear time trend.

Asymptotic Critical Values for the Cointegration Test with Linear Time Trends

Significance Level	1%	2.5%	5%	10%
Critical Value	-4.32	-4.03	-3.78	-3.50

B4.4. Series with Complex Autocorrelation Structures

If we suspect that $\{y_t\} \sim I(1)$ and that $\{y_t\}$ has a complex autocorrelation structure, we may include lags of Δy_t in the autoregressive model on which we base the unit root test. In this case, we apply the **augmented Dickey-Fuller (ADF)** test.

We use the general model

$$\Delta y_t = \alpha + \theta y_{t-1} + \sum_{i=1}^{p} \gamma_i \Delta y_{t-i} + \varepsilon_t. \tag{B4.4.1}$$

To find the appropriate value of *p*, we may set *p* = 0 and then increase *p* until γ_p is statistically insignificant. We can use an ordinary *t*- test to test $H_0: \gamma_p = 0$ (or an F test for the joint significance of the γ_i). As in the DF test, we test

$$H_0: \theta = 0 \qquad \text{vs.} \qquad H_1: \theta < 0. \tag{B4.4.2}$$

The asymptotic critical values for the ADF test are the same as those for the DF test. Under H_0, $\{\Delta y_t\}$ follows an autoregressive model of order *p*; under H_1, $\{\Delta y_t\}$ follows an autoregressive model of order $p + 1$.

Vector Autoregressive (VAR) Models

If $\{x_t\}$ and $\{y_t\}$ are autoregressive series of order *p*, we may fit the autoregressive models

$$y_t = \delta_0 + \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{i=1}^{p} \beta_i x_{t-i} + \varepsilon_t,$$

and (B4.4.3)

$$x_t = \gamma_0 + \sum_{i=1}^{p} \tau_i y_{t-i} + \sum_{i=1}^{p} \varphi_i x_{t-i} + u_t.$$

The inclusion of the lagged values helps ensure that the residuals are not strongly autocorrelated. VAR models are often used for short-term forecasting.

Error Correction Models

Suppose we have established that $\{y_t\}$ and $\{x_t\}$ are cointegrated with

$$y_t = \beta x_t + \varepsilon_t. \tag{B4.4.4}$$

We can use an **error correction model** to study the short term dynamics of the relationship between $\{y_t\}$ and $\{x_t\}$. For example, consider the model

$$\Delta y_t = \alpha_0 + \gamma_0 \Delta x_t + \delta(y_{t-1} - \beta x_{t-1}) + \varepsilon_t, \tag{B4.4.5}$$

where $\delta < 0$. The term $\delta(y_{t-1} - \beta x_{t-1})$ is called the **error correction term**. The parameter δ indicates the speed at which $\{y_t\}$ and $\{x_t\}$ return to their equilibrium relationship after a "shock" creates a short-term disturbance in this relationship. Note that, if $y_{t-1} < \beta x_{t-1}$, then x_{t-1} is higher than its equilibrium value. In this case, the inclusion of the δ term will increase the model-based estimate of Δy_t. Conversely, if $y_{t-1} > \beta x_{t-1}$, the error correction term $\delta(y_{t-1} - \beta x_{t-1})$ will decrease the estimated value of Δy_t.

When the cointegrating parameter β is unknown, we may estimate it first and then fit the error correction model using the OLS estimator $\hat{\beta}$ in place of β. This is called the *Engle-Granger two-step procedure*.

B4.5. Multivariate Cointegration Analysis
In the multivariate case, we consider the k x T matrix

$$Y = \begin{bmatrix} y_{11} & \cdots & y_{1T} \\ \vdots & \ddots & \vdots \\ y_{k1} & \cdots & y_{kT} \end{bmatrix}.$$

Suppose each row $\boldsymbol{y_i} = [y_{i1} \ldots y_{iT}]$ of Y represents an $I(1)$ process, but certain linear combinations of the rows are $I(0)$. We are interested in identifying cointegrating relationships among the k time series.

The multivariate extension of the error correction model may be expressed as

$$\Delta \boldsymbol{y_t} = \boldsymbol{\pi_0} + \Pi \boldsymbol{y_{t-1}} + \sum_{i=1}^{p-1} \Phi_i \Delta \boldsymbol{y_{t-i}} + \boldsymbol{\varepsilon_t},$$

where $\Pi = \alpha\beta'$, α and β are k x r matrices, and Φ_i is a k x k matrix. The Johansen multivariate formulation assumes that the error terms are independent and identically distributed with $\boldsymbol{\varepsilon_t} \sim N_k(0, \Sigma)$. The r columns of β are called the cointegration vectors, and the rank r of β is called the cointegration rank. The parameters in α (which must be negative, like δ in the two-variable case) indicate corrections of short-run deviations from the long-run equilibrium relationships defined by the columns of β.

Johansen Lambda-Max and Trace Tests
Based on the normality assumption, the Johansen tests are applied sequentially to help identify the cointegration rank r. If we find that $r < k$, it *may* be appropriate to remove one or more time series from the system. (If one variable is not cointegrated with any of the others, all values of the β matrix corresponding to it will be near 0.)

- *Lambda-Max Test*
  Using a maximum generalized eigenvalue as a test statistic, we test null hypothesis that the cointegration rank r is equal $i \in \{0,1,\ldots,k-1\}$ against the alternative that $r = i + 1$. We apply the test sequentially starting with $i = 0$.

- *Trace Test*

Using the trace (sum of diagonal elements) of a diagonal matrix of generalized eigenvalues as a test statistic, we test null hypothesis that the cointegration rank r is equal $i \in \{0,1, \dots, k-1\}$ against the alternative that $r = i + 1$.

Applying Multivariate Cointegration Tests to Energy Data

Johansen's proof of the validity of the lambda-max and trace tests relies on maximum likelihood theory, assuming that the differenced data follow a Gaussian (or normal) distribution. Because this assumption is often problematic for energy data, especially energy price data, Johansen's technique is not used in many analyses performed within EIA. Seasonality also complicates the analysis of monthly data by the vector error correction model.