CHAPTER 6

COINTEGRATION AND ERROR-CORRECTION MODELS

Learning Objectives

1. Introduce the basic concept of cointegration and show that it applies in a variety of economic models.
2. Show that cointegration necessitates that the stochastic trends of nonstationary variables be linked.
3. Consider the dynamic paths of cointegrated variables. Since the trends of the variables are linked, the dynamic paths of such variables must respond to the current deviation from the equilibrium relationship.
4. Develop the Engle–Granger cointegration test. The econometric methods underlying the test procedures stem from the theory of simultaneous difference equations.
5. The Engle–Granger method is illustrated using simulated data.
6. Illustrate the Engle–Granger method using real exchange rate data.
7. Develop the Johansen full-information maximum likelihood cointegration test.
8. Show how to test restrictions on cointegrating vectors. Discuss inference in models with $I(1)$ and $I(2)$ variables.
9. Illustrate the Johansen test using simulated data.
10. Show how to estimate ADL models using nonstationary variables and develop the ADL cointegration test.
11. Compare the Engle–Granger, Johansen, and ADL cointegration tests using interest rate data.

This chapter explores an exciting development in econometrics: the estimation of a structural equation or a VAR containing nonstationary variables. In univariate models, we have seen that a stochastic trend can be removed by differencing. The resulting stationary series can be estimated using univariate Box–Jenkins techniques. At one time, the conventional wisdom was to generalize this idea and difference all nonstationary variables used in a regression analysis. However, the appropriate way to treat nonstationary variables is not so straightforward in a multivariate context. It is quite possible for there to be a linear combination of integrated variables that is stationary; such variables are said to be **cointegrated**. In the presence of cointegrated variables, it is possible to model the long-run model and the short-run dynamics simultaneously. Many economic models entail such cointegrating relationships.

1. LINEAR COMBINATIONS OF INTEGRATED VARIABLES

Since money demand studies stimulated much of the cointegration literature, we begin by considering a simple model of money demand. Theory suggests that individuals want to hold a real quantity of money balances, so that the demand for nominal money holdings should be proportional to the price level. Moreover, as real income and the associated number of transactions increase, individuals will want to hold increased money balances. Finally, since the interest rate is the opportunity cost of holding money, money demand should be negatively related to the interest rate. In logarithms, an econometric specification for such an equation can be written as

$$m_t = \beta_0 + \beta_1 p_t + \beta_2 y_t + \beta_3 r_t + e_t \tag{6.1}$$

where m_t = demand for money
p_t = price level
y_t = real income
r_t = interest rate
e_t = *stationary* disturbance term
β_i = parameters to be estimated

and all variables but the interest rate are expressed in logarithms

The hypothesis that the money market is in equilibrium allows the researcher to collect time-series data of the money supply (= money demand if the money market always clears), the price level, real income (possibly measured using real GDP), and an appropriate short-term interest rate. The behavioral assumptions require that $\beta_1 = 1$, $\beta_2 > 0$, and $\beta_3 < 0$; a researcher conducting such a study would certainly want to test these parameter restrictions. Be aware that the properties of the unexplained portion of the demand for money (i.e., the $\{e_t\}$ sequence) are an integral part of the theory. If the theory is to make any sense at all, any deviation in the demand for money must necessarily be temporary in nature. Clearly, if e_t has a stochastic trend, the errors in the model will be cumulative so that deviations from money market equilibrium will not be eliminated. Hence, a key assumption of the theory is that the $\{e_t\}$ sequence is stationary.

The problem confronting the researcher is that real GDP, the money supply, price level, and interest rate can all be characterized as nonstationary $I(1)$ variables. As such, each variable can meander without any tendency to return to a long-run level. However, the theory expressed in (6.1) asserts that there exists a linear combination of these nonstationary variables that is stationary! Solving for the error term, we can rewrite (6.1) as

$$e_t = m_t - \beta_0 - \beta_1 p_t - \beta_2 y_t - \beta_3 r_t \tag{6.2}$$

Since $\{e_t\}$ must be stationary, it follows that the linear combination of integrated variables given by the right-hand-side of (6.2) must also be stationary. Thus, the theory necessitates that the time paths of the four nonstationary variables $\{m_t\}$, $\{p_t\}$, $\{y_t\}$, and $\{r_t\}$ be linked. This example illustrates the crucial insight that has dominated much of the macroeconometric literature in recent years: *Equilibrium theories involving nonstationary variables require the existence of a combination of the variables that is stationary.*

The money demand function is just one example of a stationary combination of nonstationary variables. Within any equilibrium framework, the deviations from equilibrium must be temporary. Other important economic examples involving stationary combinations of nonstationary variables include the following:

1. *Consumption function theory.* A simple version of the permanent income hypothesis maintains that total consumption (c_t) is the sum of permanent consumption (c_t^p) and transitory consumption (c_t^t). Since permanent consumption is proportional to permanent income (y_t^p), we can let β be the constant of proportionality and write $c_t = \beta y_t^p + c_t^t$. Transitory consumption is necessarily a stationary variable, and both consumption and permanent income are reasonably characterized as $I(1)$ variables. As such, the permanent income hypothesis requires that the linear combination of two $I(1)$ variables given by $c_t - \beta y_t^p$ be stationary.
2. *Unbiased forward rate hypothesis.* One form of the efficient market hypothesis asserts that the forward (or futures) price of an asset should equal the expected value of that asset's spot price in the future. Foreign exchange market efficiency requires that the one-period forward exchange rate equal the expectation of the spot rate in the next period. Letting f_t denote the log of the one-period price of forward exchange in t and s_t the log of the spot price of foreign exchange in t, the theory asserts that $E_t s_{t+1} = f_t$. If this relationship fails, speculators can expect to make a pure profit on their trades in the foreign exchange market. If agents' expectations are rational, the forecast error for the spot rate in $t + 1$ will have a conditional mean equal to zero, so that $s_{t+1} - E_t s_{t+1} = \varepsilon_{t+1}$ where $E_t \varepsilon_{t+1} = 0$. Combining the two equations yields $s_{t+1} = f_t + \varepsilon_{t+1}$. Since $\{s_t\}$ and $\{f_t\}$ are $I(1)$ variables, the **unbiased forward rate hypothesis** necessitates that there be a linear combination of nonstationary spot and forward exchange rates that is stationary.
3. *Commodity market arbitrage and purchasing power parity.* Theories of spatial competition suggest that in the short run, prices of similar products in varied markets might differ. However, arbiters will prevent the various prices from moving too far apart even if the prices are nonstationary. Similarly, the prices of Apple computers and PCs have exhibited sustained declines. Economic theory suggests that these simultaneous declines are related to each other since a price discrepancy between these similar products cannot continually widen. Also, as we saw in Chapter 4, purchasing power parity places restrictions on the movements of nonstationary price levels and exchange rates. If e_t denotes the log of the price of foreign exchange and p_t and p_t^*

denote, respectively, the logs of domestic and foreign price levels, long-run PPP requires that the linear combination $e_t + p_t^* - p_t$ be stationary.

All of these examples illustrate the concept of **cointegration** as introduced by Engle and Granger (1987). Their formal analysis begins by considering a set of economic variables in long-run equilibrium when

$$\beta_1 x_{1t} + \beta_2 x_{2t} + \cdots + \beta_n x_{nt} = 0$$

Letting β and x_t denote the vectors $(\beta_1, \beta_2, \ldots, \beta_n)$ and $(x_{1t}, x_{2t}, \ldots, x_{nt})'$, the system is in long-run equilibrium when $\beta x_t = 0$. The deviation from long-run equilibrium—called the **equilibrium error**—is e_t, so that

$$e_t = \beta x_t$$

If the equilibrium is meaningful, it must be the case that the equilibrium error process is stationary. In a sense, the use of the term *equilibrium* is unfortunate because economic theorists and econometricians use the term in different ways. Economic theorists usually use the term to refer to an equality between desired and actual transactions. The econometric use of the term makes reference to any long-run relationship among nonstationary variables. Cointegration does not require that the long-run relationship be generated by market forces or by the behavioral rules of individuals. In Engle and Granger's use of the term, the equilibrium relationship may be causal, behavioral, or simply a reduced-form relationship among similarly trending variables. Engle and Granger (1987) provide the following definition of cointegration:

The components of the vector $x_t = (x_{1t}, x_{2t}, \ldots, x_{nt})'$ are said to be *cointegrated of order d, b*, denoted by $x_t \sim CI(d, b)$ if

1. All components of x_t are integrated of order d.
2. There exists a vector $\beta = (\beta_1, \beta_2, \ldots, \beta_n)$ such that the linear combination $\beta x_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots + \beta_n x_{nt}$ is integrated of order $(d - b)$ where $b > 0$. Note that the vector β is called the **cointegrating vector**.[1]

In terms of equation (6.1), if the money supply, price level, real income, and interest rate are all $I(1)$ and the linear combination $m_t - \beta_0 - \beta_1 p_t - \beta_2 y_t - \beta_3 r_t = e_t$ is stationary, then the variables are cointegrated of order (1, 1). The vector x_t is $(m_t, 1, p_t, y_t, r_t)'$ and the cointegrating vector β is $(1, -\beta_0, -\beta_1, -\beta_2, -\beta_3)$. The deviation from long-run money market equilibrium is e_t; since $\{e_t\}$ is stationary, this deviation is temporary in nature.

There are four important points to note about the definition:

1. Cointegration typically refers to a *linear* combination of nonstationary variables. Theoretically, it is quite possible that nonlinear long-run relationships exist among a set of integrated variables. However, as discussed in Chapter 7, the current state of econometric practice is just beginning to allow for tests of nonlinear cointegrating relationships. Also note that the cointegrating vector is not unique. If $(\beta_1, \beta_2, \ldots, \beta_n)$ is a cointegrating vector, then for any nonzero value of λ, $(\lambda\beta_1, \lambda\beta_2, \ldots, \lambda\beta_n)$ is also a cointegrating vector. Typically, one of the variables is used to *normalize* the cointegrating vector by fixing its

coefficient at unity. To normalize the cointegrating vector with respect to x_{1t}, simply select $\lambda = 1/\beta_1$.

2. From Engle and Granger's original definition, cointegration refers to variables that are integrated of the same order. Of course, this does not imply that all integrated variables are cointegrated; usually, a set of $I(d)$ variables is *not* cointegrated. Such a lack of cointegration implies no long-run equilibrium among the variables, so that they can wander arbitrarily far from each other. If two variables are integrated of different orders, they cannot be cointegrated. Suppose x_{1t} is $I(d_1)$ and x_{2t} is $I(d_2)$ where $d_2 > d_1$. Question 7 at the end of this chapter asks you to prove that any linear combination of x_{1t} and x_{2t} is $I(d_2)$.

Nevertheless, it is possible to find equilibrium relationships among groups of variables that are integrated of different orders. Suppose that x_{1t} and x_{2t} are $I(2)$ and that the other variables under consideration are $I(1)$. As such, there cannot be a cointegrating relationship between x_{1t} (or x_{2t}) and x_{3t} However, if x_{1t} and x_{2t} are $CI(2,1)$, there exists a linear combination of the form $\beta_1 x_{1t} + \beta_2 x_{2t}$ which is $I(1)$. It is possible that *this* combination of x_{1t} and x_{2t} is cointegrated with the $I(1)$ variables. Lee and Granger (1990) use the term **multicointegration** to refer to this type of circumstance.

3. There may be more than one independent cointegrating vectors for a set of $I(1)$ variables. The number of cointegrating vectors is called the **cointegrating rank** of x_t. For example, suppose that the monetary authorities followed a feedback rule such that they decreased the money supply when nominal GDP was high and increased the nominal money supply when nominal GDP was low. This feedback rule might be represented by

$$\begin{aligned} m_t &= \gamma_0 - \gamma_1(y_t + p_t) + e_{1t} \\ &= \gamma_0 - \gamma_1 y_t - \gamma_1 p_t + e_{1t} \end{aligned} \tag{6.3}$$

where $\{e_{1t}\}$ = a stationary error in the money supply feedback rule.

Given the money demand function in (6.1), there are two cointegrating vectors for the money supply, price level, real income, and the interest rate. Let β be the $(2 \cdot 5)$ matrix:

$$\beta = \begin{bmatrix} 1 & -\beta_0 & -\beta_1 & -\beta_2 & -\beta_3 \\ 1 & -\gamma_0 & \gamma_1 & \gamma_1 & 0 \end{bmatrix}$$

The two linear combinations given by βx_t are stationary. As such, the cointegrating rank of x_t is two. As a practical matter, if multiple cointegrating vectors are found, it may not be possible to identify the behavioral relationships from what may be reduced-form relationships. As shown below, if x_t has n nonstationary components, there may be as many as $n - 1$ linearly independent cointegrating vectors. Hence, if x_t contains only two variables, there can be *at most* one independent cointegrating vector.

4. Most of the cointegration literature focuses on the case in which each variable contains a single unit root. The reason is that traditional regression or time-series analysis applies when variables are $I(0)$ and few economic

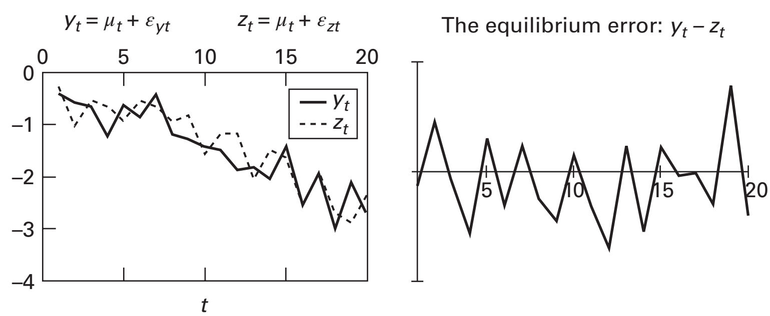
> variables are integrated of an order higher than unity. When it is unambiguous, many authors use the term *cointegration* to refer to the case in which variables are $CI(1, 1)$.

Worksheet 6.1 illustrates some of the important properties of cointegration relationships. In Case 1, both the $\{y_t\}$ and $\{z_t\}$ sequences were constructed so as to be random walk plus noise processes. Although the 20 realizations shown generally decline, extending the sample would eliminate this tendency. In any event, neither series shows any tendency to return to a long-run level, and formal Dickey–Fuller tests are not able

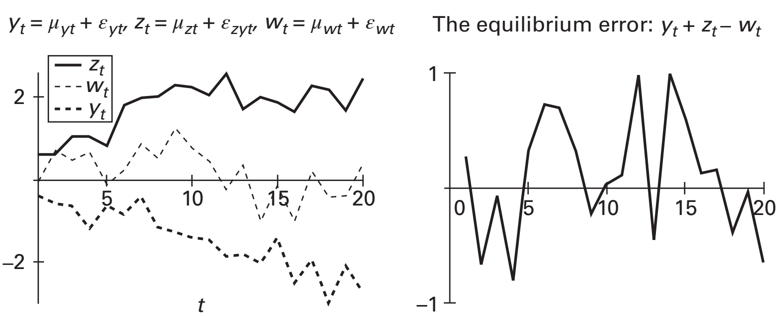
WORKSHEET 6.1

ILLUSTRATING COINTEGRATED SYSTEMS

CASE 1: The series $\{\mu_t\}$ is a random walk process and $\{\varepsilon_{yt}\}$ and $\{\varepsilon_{zt}\}$ are white noise. Hence, the $\{y_t\}$ and $\{z_t\}$ sequences are both random walk plus noise processes. Although each is nonstationary, the two sequences have the same stochastic trend; hence they are cointegrated such that the linear combination $(y_t - z_t)$ is stationary. The equilibrium error term $(\varepsilon_{yt} - \varepsilon_{zt})$ is an $I(0)$ process.



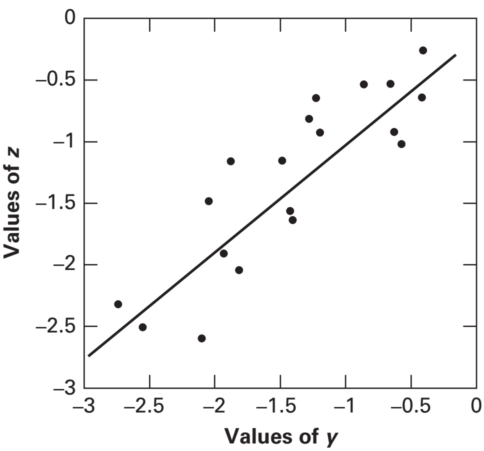
CASE 2: All three sequences are random walk plus noise processes. As constructed no two are cointegrated. However, the linear combination $(y_t + z_t - w_t)$ is stationary; hence, the three variables are cointegrated. The equilibrium error is an $I(0)$ process.



to reject the null hypothesis of a unit root in either series. Although each series is nonstationary, you can see that they do move together. In fact, the difference between the series $(y_t - z_t)$—shown in the second graph—is stationary; the *equilibrium error* term $e_t = (y_t - z_t)$ has a zero mean and a constant variance.

Case 2 illustrates cointegration among three random walk plus noise processes. As in Case 1, no series exhibits a tendency to return to a long-run level, and formal Dickey–Fuller tests are not able to reject the null hypothesis of a unit root in any of the three. In contrast to the previous case, no two of the series appear to be cointegrated; each series seems to "meander" away from the other two. However, as shown in the second graph, there exists a stationary linear combination of the three such that $e_t = y_t + z_t - w_t$. Thus, it follows that the dynamic behavior of *at least* one variable must be restricted by the values of the other variables in the system.

Figure 6.1 displays the information of Case 1 in a scatter plot of $\{y_t\}$ against the associated value of $\{z_t\}$; each of the 20 points represents the ordered pairs $(y_1, z_1), (y_2, z_2), \ldots, (y_{20}, z_{20})$. Comparing Worksheet 6.1 and Figure 6.1, you can see that low values in the $\{y_t\}$ sequence are associated with low values in the $\{z_t\}$ sequence and that values near zero in one series are associated with values near zero in the other. Since both series move together over time, there is a positive relationship between the two. The least-squares line in the scatter plot reveals this to be a strong positive association. In fact, this line is the "long-run" equilibrium relationship between the series, and the deviations from the line are the stationary deviations from long-run equilibrium.



The scatter plot was drawn using the $\{y_t\}$ and $\{z_t\}$ sequences from Case 1 of Worksheet 6.1. Since both series decline over time, there appears to be a positive relationship between the two. The equilibrium regression line is shown.

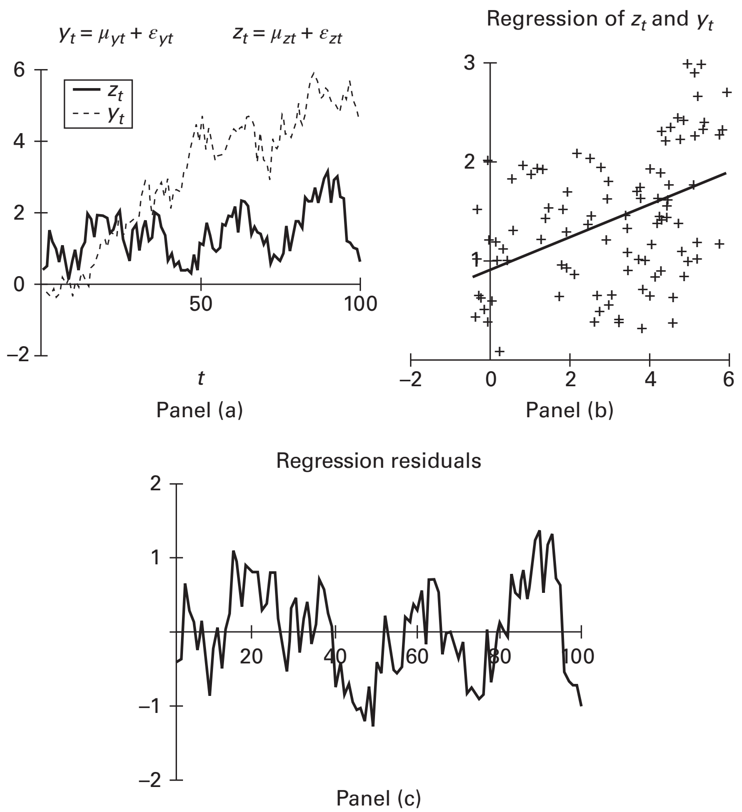
FIGURE 6.1 Scatter Plot of Cointegrated Variables

For comparison purposes, Panel (a) in Worksheet 6.2 shows the time paths of two random walk plus noise processes that are not cointegrated. Each seems to meander without any tendency to approach the other. The scatter plot shown in Panel (b) confirms the impression of no long-run relationship between the variables. The deviations from the straight line showing the regression of z_t on y_t are substantial. Plotting the regression residuals against time [see Panel (c)], suggests that the regression residuals are not stationary.

WORKSHEET 6.2

NONINTEGRATED VARIABLES

The $\{y_t\}$ and $\{z_t\}$ sequences are constructed to independent random walk plus noise processes. There is **no** cointegrating relationship between the two variables. As shown in graph (a), both seem to meander without any tendency to come together. Graph (b) shows the scatter plot of the two sequences and the regression line $z_t = \beta_0 + \beta_1 y_t$. However, this regression line is spurious. As shown in graph (c), the regression residuals are nonstationary.



2. COINTEGRATION AND COMMON TRENDS

Stock and Watson's (1988) observation that cointegrated variables share common stochastic trends provides a very useful way to understand cointegration relationships. For ease of exposition, return to the case in which the vector x_t contains only two variables so that $x_t = (y_t, z_t)'$. Ignoring cyclical and seasonal terms, we can write each variable as a random walk plus an irregular (but not necessarily a white noise) component:

$$y_t = \mu_{yt} + e_{yt} \tag{6.4}$$

$$z_t = \mu_{zt} + e_{zt} \tag{6.5}$$

where μ_{it} = a random walk process representing the stochastic trend in variable i

e_{it} = the stationary (irregular) component of variable i

If $\{y_t\}$ and $\{z_t\}$ are cointegrated of order (1,1), there must be nonzero values of β_1 and β_2 for which the linear combination $\beta_1 y_t + \beta_2 z_t$ is stationary. Consider the sum

$$\begin{aligned}\beta_1 y_t + \beta_2 z_t &= \beta_1(\mu_{yt} + e_{yt}) + \beta_2(\mu_{zt} + e_{zt}) \\ &= (\beta_1 \mu_{yt} + \beta_2 \mu_{zt}) + (\beta_1 e_{yt} + \beta_2 e_{zt})\end{aligned} \tag{6.6}$$

For $\beta_1 y_t + \beta_2 z_t$ to be stationary, the term $(\beta_1 \mu_{yt} + \beta_2 \mu_{zt})$ must vanish. After all, if either of the two trends appears in (6.6), the linear combination $\beta_1 y_t + \beta_2 z_t$ will also have a trend. Since the second term within parentheses is stationary, the necessary and sufficient condition for $\{y_t\}$ and $\{z_t\}$ to be $CI(1, 1)$ is

$$\beta_1 \mu_{yt} + \beta_2 \mu_{zt} = 0 \tag{6.7}$$

Clearly, μ_{yt} and μ_{zt} are variables whose realized values will be continually changing over time. Since we preclude both β_1 and β_2 from being equal to zero, it follows that (6.7) holds for all t if and only if

$$\mu_{yt} = -\beta_2 \mu_{zt} / \beta_1$$

For nonzero values of β_1 and β_2, the only way to ensure the equality is for the stochastic trends to be *identical* up to a scalar. Thus, up to the scalar $-\beta_2/\beta_1$, *two* $I(1)$ *stochastic processes* $\{y_t\}$ *and* $\{z_t\}$ *must have the same stochastic trend if they are cointegrated of order* (1, 1).

Return your attention to Worksheet 6.1. In Case 1, the $\{y_t\}$ and $\{z_t\}$ sequences were constructed so as to satisfy

$$y_t = \mu_t + \varepsilon_{yt}$$

$$z_t = \mu_t + \varepsilon_{zt}$$

$$\mu_t = \mu_{t-1} + \varepsilon_t$$

where ε_{yt}, ε_{zt}, and ε_t are independently distributed white-noise disturbances.

By construction, μ_t is a pure random walk process representing the same stochastic trend for both the $\{y_t\}$ and $\{z_t\}$ sequences. The value of μ_0 was initialized to zero, and three sets of 20 random numbers were drawn to represent the $\{\varepsilon_{yt}\}$, $\{\varepsilon_{zt}\}$, and $\{\varepsilon_t\}$ sequences. Using these realizations and the initial value of μ_0, the $\{y_t\}$, $\{z_t\}$, and $\{\mu_t\}$

sequences were constructed. As you can clearly determine, subtracting the realized value of z_t from y_t results in a stationary sequence:

$$y_t - z_t = (\mu_t + \varepsilon_{yt}) - (\mu_t + \varepsilon_{zt}) = \varepsilon_{yt} - \varepsilon_{zt}$$

To state the point using Engle and Granger's terminology, multiplying the cointegrating vector $\beta = (1, -1)$ by the vector by $x_t = (y_t, z_t)'$ yields the stationary sequence $\varepsilon_t = \varepsilon_{yt} - \varepsilon_{zt}$. Indeed, the equilibrium error term shown in the second graph of Worksheet 6.1 has all the hallmarks of a stationary process. The essential insight of Stock and Watson (1988) is that the parameters of the cointegrating vector must be such that they purge the trend from the linear combination. Any other linear combination of the two variables contains a trend so that the cointegrating vector is unique up to a normalizing scalar. Hence, $\beta_3 y_t + \beta_4 z_t$ cannot be stationary unless $\beta_3/\beta_4 = \beta_1/\beta_2$.

Recall that Case 2 of Worksheet 6.1 illustrates cointegration between three random walk plus noise processes. Each process is $I(1)$, and Dickey–Fuller unit root tests would not be able to reject the null hypothesis that each contains a unit root. As you can see in the lower portion of Worksheet 6.1, no pairwise combination of the series appears to be cointegrated. Each series seems to meander, and, as opposed to Case 1, no one single series appears to remain close to any other series. However, by construction, the trend in w_t is the simple summation of the trends in y_t and z_t:

$$\mu_{wt} = \mu_{yt} + \mu_{zt}$$

Here, the vector $x_t = (y_t, z_t, w_t)'$ has the cointegrating vector $(1, 1, -1)$, so that the linear combination $y_t + z_t - w_t$ is stationary. Consider

$$y_t + z_t - w_t = (\mu_{yt} + \varepsilon_{yt}) + (\mu_{zt} + \varepsilon_{zt}) - (\mu_{wt} + \varepsilon_{wt}) = \varepsilon_{yt} + \varepsilon_{zt} - \varepsilon_{wt}$$

This example illustrates the general point that cointegration will occur whenever the trend in one variable can be expressed as a linear combination of the trends in the other variable(s). In such circumstances it is always possible to find a vector β such that the linear combination $\beta_1 y_t + \beta_2 z_t + \beta_3 w_t$ does not contain a trend. The result easily generalizes to the case of n variables. Consider the vector representation:

$$x_t = \mu_t + e_t \tag{6.8}$$

where x_t = the vector $(x_{1t}, x_{2t}, \ldots, x_{nt})'$
μ_t = the vector of stochastic trends $(\mu_{1t}, \mu_{2t}, \ldots, \mu_{nt})'$
e_t = an $n \cdot 1$ vector of stationary components

If one trend can be expressed as a linear combination of the other trends in the system, it means that there exists a vector β such that

$$\beta_1 \mu_{1t} + \beta_2 \mu_{2t} + \cdots + \beta_n \mu_{nt} = 0$$

Premultiply (6.8) by this set of β_is to obtain

$$\beta x_t = \beta \mu_t + \beta e_t$$

Since $\beta\mu_t = 0$, it follows that $\beta x_t = \beta e_t$. Hence, the linear combination βx_t is stationary. As shown in Section 8, this argument easily generalizes to the case of multiple cointegrating vectors.

3. COINTEGRATION AND ERROR CORRECTION

A principal feature of cointegrated variables is that their time paths are influenced by the extent of any deviation from long-run equilibrium. After all, if the system is to return to long-run equilibrium, the movements of at least some of the variables must respond to the magnitude of the disequilibrium. Before proceeding further, be aware that we will be examining the time paths of multiple nonstationary time-series variables. To do so in a tractable way, we will need to draw relationship between the rank of a matrix and its characteristic roots. The required mathematics are provided in Appendix 6.1.

The relationship between long-term and short-term interest rates illustrates how variables might adjust to any discrepancies from the long-run equilibrium relationship. Clearly, the theory of the term structure of interest rates implies a long-run relationship between long- and short-term rates. If the gap between the long- and short-term rates is "large" relative to the long-run relationship, the short-term rate must ultimately rise relative to the long-term rate. Of course, the gap can be closed by (1) an increase in the short-term rate and/or a decrease in the long-term rate, (2) an increase in the long-term rate but a commensurately larger rise in the short-term rate, or (3) a fall in the long-term rate but a smaller fall in the short-term rate. Without a full dynamic specification of the model, it is not possible to determine which of the possibilities will occur. Nevertheless, the short-run dynamics must be influenced by the deviation from the long-run relationship.

The dynamic model implied by this discussion is one of **error correction**. In an error-correction model, the short-term dynamics of the variables in the system are influenced by the deviation from equilibrium. If we assume that both interest rates are $I(1)$, a simple error-correction model that could apply to the term structure of interest rates is

$$\Delta r_{St} = \alpha_S(r_{Lt-1} - \beta r_{St-1}) + \varepsilon_{St} \qquad \alpha_S > 0 \tag{6.9}$$

$$\Delta r_{Lt} = -\alpha_L(r_{Lt-1} - \beta r_{St-1}) + \varepsilon_{Lt} \qquad \alpha_L > 0 \tag{6.10}$$

where ε_{St} and ε_{Lt} are white-noise disturbance terms which may be correlated, r_{Lt} and r_{St} are the long- and short-term interest rates, and α_S, α_L, and β are parameters.

As specified, the short- and long-term interest rates change in response to stochastic shocks (represented by ε_{St} and ε_{Lt}) *and* in response to the previous period's deviation from long-run equilibrium. Everything else being equal, if this deviation happened to be positive (so that $r_{Lt-1} - \beta r_{St-1} > 0$), the short-term interest rate would rise and the long-term rate would fall. Long-run equilibrium is attained when $r_{Lt} = \beta r_{St}$ so that the expected change in each rate is zero.

Here you can see the relationship between error-correcting models and cointegrated variables. By assumption, Δr_{St} is stationary so that the left-hand side of (6.9) is $I(0)$. For (6.9) to be sensible, the right-hand side must be $I(0)$ as well. Given that ε_{St} is stationary, it follows that the linear combination $r_{Lt-1} - \beta r_{St-1}$ must also be stationary;

hence, the two interest rates must be cointegrated with the cointegrating vector $(1, -\beta)$. Of course, the identical argument applies to (6.10). The essential point to note is that the error-correction representation necessitates that the two variables be cointegrated of order *CI*(1, 1). This result is unaltered if we formulate a more general model by introducing the lagged changes of each rate into both equations:

$$\Delta r_{St} = a_{10} + \alpha_S(r_{Lt-1} - \beta r_{St-1}) + \Sigma a_{11}(i)\,\Delta r_{St-i} + \Sigma a_{12}(i)\,\Delta r_{Lt-i} + \varepsilon_{St} \quad (6.11)$$

$$\Delta r_{Lt} = a_{20} - a_L(r_{Lt-1} - \beta r_{St-1}) + \Sigma a_{21}(i)\,\Delta r_{St-i} + \Sigma a_{22}(i)\,\Delta r_{Lt-i} + \varepsilon_{Lt} \quad (6.12)$$

Again, ε_{St}, ε_{Lt}, and all terms involving Δr_{St-i} and Δr_{Lt-i} are stationary. Thus, the linear combination of interest rates $r_{Lt-1} - \beta r_{St-1}$ must also be stationary.

Inspection of (6.11) and (6.12) reveals a striking similarity to the VAR models of the previous chapter. This two-variable error-correction model is a bivariate VAR in first differences augmented by the error-correction terms $\alpha_S(r_{Lt-1} - \beta r_{St-1})$ and $-\alpha_L(r_{Lt-1} - \beta r_{St-1})$. Notice that α_S and α_L have the interpretation of **speed of adjustment** parameters. The larger α_S is, the greater the response of r_{St} to the previous period's deviation from long-run equilibrium. At the opposite extreme, very small values of α_S imply that the short-term interest rate is unresponsive to last period's equilibrium error. For the $\{\Delta r_{St}\}$ sequence to be unaffected by the long-term interest rate sequence, α_S *and* all the $a_{12}(i)$ coefficients must be equal to zero. Of course, at least one of the speed of adjustment terms in (6.11) and (6.12) must be nonzero. If both α_S and α_L are equal to zero, the long-run equilibrium relationship does not appear and the model is not one of error correction or cointegration.

The result can easily be generalized to the n-variable model. Formally, the $(n \cdot 1)$ vector of $I(1)$ variables $x_t = (x_{1t}, x_{2t}, \ldots, x_{nt})'$ has an error-correction representation if it can be expressed in the form:

$$\Delta x_t = \pi_0 + \pi x_{t-1} + \pi_1 \Delta x_{t-1} + \pi_2 \Delta x_{t-2} + \cdots + \pi_p \Delta x_{t-p} + \varepsilon_t \quad (6.13)$$

where π_0 = an $(n \cdot 1)$ vector of intercept terms with elements π_{i0}

$\pi_i = (n \cdot n)$ coefficient matrices with elements $\pi_{jk}(i)$

π = a matrix with elements π_{jk} such that one or more of the $\pi_{jk} \neq 0$

ε_t = an $(n \cdot 1)$ vector with elements ε_{it}

Note that the disturbance terms are such that ε_{it} may be correlated with ε_{jt}

Let all variables in x_t be $I(1)$. Now, if there is an error-correction representation of these variables as in (6.13), there is necessarily a linear combination of the $I(1)$ variables that is stationary. Solving (6.13) for πx_{t-1} yields

$$\pi x_{t-1} = \Delta x_t - \pi_0 - \Sigma \pi_i \Delta x_{t-i} - \varepsilon_t$$

Since each expression on the right-hand side is stationary, πx_{t-1} must also be stationary. Since π contains only constants, each row of π is a cointegrating vector of x_t. For example, the first row can be written as $(\pi_{11}x_{1t-1} + \pi_{12}x_{2t-1} + \cdots + \pi_{1n}x_{nt-1})$. Since each series is $I(1)$, $(\pi_{11}, \pi_{12}, \ldots, \pi_{1n})$ must be a cointegrating vector for x_t.

The key feature in (6.13) is the presence of the matrix π. There are two important points to note:

1. If all elements of π equal zero, (6.13) is a traditional VAR in first differences. In such circumstances there is no error-correction representation since Δx_t does not respond to the previous period's deviation from long-run equilibrium.
2. If one or more of the π_{jk} differs from zero, Δx_t responds to the previous period's deviation from long-run equilibrium. Hence, *estimating x_t as a VAR in first differences is inappropriate if x_t has an error-correction representation.* The omission of the expression πx_{t-1} entails a misspecification error if x_t has an error-correction representation as in (6.13).

A good way to examine the relationship between cointegration and error correction is to study the properties of the simple VAR model:

$$y_t = a_{11}y_{t-1} + a_{12}z_{t-1} + \varepsilon_{yt} \tag{6.14}$$

$$z_t = a_{21}y_{t-1} + a_{22}z_{t-1} + \varepsilon_{zt} \tag{6.15}$$

where ε_{yt} and ε_{zt} are white-noise disturbances that may be correlated with each other and, for simplicity, intercept terms have been ignored. Using lag operators, we can write (6.14) and (6.15) as

$$(1 - a_{11}L)y_t - a_{12}Lz_t = \varepsilon_{yt}$$

$$-a_{21}Ly_t + (1 - a_{22}L)z_t = \varepsilon_{zt}$$

The next step is to solve for y_t and z_t. Writing the system in matrix form, we obtain

$$\begin{bmatrix} (1 - a_{11}L) & -a_{12}L \\ -a_{21}L & (1 - a_{22}L) \end{bmatrix} \begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix}$$

Using Cramer's Rule or matrix inversion, we can obtain the solutions for y_t and z_t as

$$y_t = \frac{(1 - a_{22}L)\varepsilon_{yt} + a_{12}L\varepsilon_{zt}}{(1 - a_{11}L)(1 - a_{22}L) - a_{12}a_{21}L^2} \tag{6.16}$$

$$z_t = \frac{a_{21}L\varepsilon_{yt} + (1 - a_{11}L)\varepsilon_{zt}}{(1 - a_{11}L)(1 - a_{22}L) - a_{12}a_{21}L^2} \tag{6.17}$$

We have converted the two-variable first-order system represented by (6.14) and (6.15) into two univariate second-order difference equations of the type examined in Chapter 2. Note that both variables have the same inverse characteristic equation: $(1 - a_{11}L)(1 - a_{22}L) - a_{12}a_{21}L^2$. Setting $(1 - a_{11}L)(1 - a_{22}L) - a_{12}a_{21}L^2 = 0$ and solving for L yields the two roots of the inverse characteristic equation. In order to work with the characteristic roots (as opposed to the inverse characteristic roots), define $\lambda = 1/L$ and write the characteristic equation as

$$\lambda^2 - (a_{11} + a_{22})\lambda + (a_{11}a_{22} - a_{12}a_{21}) = 0 \tag{6.18}$$

Since the two variables have the same characteristic equation, the characteristic roots of (6.18) determine the time paths of both variables. The following remarks summarize the time paths of $\{y_t\}$ and $\{z_t\}$:

1. If both characteristic roots (λ_1, λ_2) lie inside the unit circle, (6.16) and (6.17) yield stable solutions for $\{y_t\}$ and $\{z_t\}$. If t is sufficiently large or if the initial conditions are such that the homogeneous solution is zero, the stability condition guarantees that the variables are stationary. The variables cannot be cointegrated of order (1, 1) since each is stationary.
2. If either root lies outside the unit circle, the solutions are explosive. Neither variable is difference stationary, so they cannot be $CI(1, 1)$. In the same way, if both characteristic roots are unity, the second difference of each variable will be stationary. Since each is $I(2)$, the variables cannot be $CI(1, 1)$.
3. As you can see from (6.14) and (6.15), if $a_{12} = a_{21} = 0$, the solution is trivial. For $\{y_t\}$ and $\{z_t\}$ to be unit root processes, it is necessary for $a_{11} = a_{22} = 1$. It follows that $\lambda_1 = \lambda_2 = 1$ and that the two variables evolve without any long-run equilibrium relationship; hence, the variables cannot be cointegrated.
4. For $\{y_t\}$ and $\{z_t\}$ to be $CI(1, 1)$, it is necessary for one characteristic root to be unity and the other to be less than unity in absolute value. In this instance, each variable will have the same stochastic trend and the first difference of each variable will be stationary. For example, if $\lambda_1 = 1$, (6.16) will have the form:

$$y_t = [(1 - a_{22}L)\varepsilon_{yt} + a_{12}L\varepsilon_{zt}]/[(1 - L)(1 - \lambda_2 L)]$$

or, multiplying by $(1 - L)$, we get

$$(1 - L)y_t = \Delta y_t = [(1 - a_{22}L)\varepsilon_{yt} + a_{12}L\varepsilon_{zt}]/(1 - \lambda_2 L)$$

which is stationary if $|\lambda_2| < 1$.

Thus, to ensure that the variables are $CI(1,1)$, we must set one of the characteristic roots equal to unity and the other to a value that is less than unity in absolute value. For the larger of the two roots to equal unity, the quadratic formula indicates that

$$0.5(a_{11} + a_{22}) + 0.5\sqrt{(a_{11}^2 + a_{22}^2) - 2a_{11}a_{22} + 4a_{12}a_{21}} = 1$$

so that after some simplification, the coefficients are seen to satisfy[2]

$$a_{11} = [(1 - a_{22}) - a_{12}a_{21}]/(1 - a_{22}) \tag{6.19}$$

Now consider the second characteristic root. Since a_{12} and/or a_{21} must differ from zero if the variables are cointegrated, the condition $|\lambda_2| < 1$ requires

$$a_{22} > -1 \tag{6.20}$$

and

$$a_{12}a_{21} + (a_{22})^2 < 1 \tag{6.21}$$

Equations (6.19), (6.20), and (6.21) are restrictions we must place on the coefficients of (6.14) and (6.15) if we want to ensure that the variables are cointegrated of order (1, 1). To see how these coefficient restrictions bear on the nature of the solution, write (6.14) and (6.15) as

$$\begin{bmatrix} \Delta y_t \\ \Delta z_t \end{bmatrix} = \begin{bmatrix} a_{11} - 1 & a_{12} \\ a_{21} & a_{22} - 1 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix} \tag{6.22}$$

Now, (6.19) implies that $a_{11} - 1 = -a_{12}a_{21}/(1 - a_{22})$ so that after a bit of manipulation, (6.22) can be written in the form

$$\Delta y_t = -[a_{12}a_{21}/(1 - a_{22})]y_{t-1} + a_{12}z_{t-1} + \varepsilon_{yt} \tag{6.23}$$

$$\Delta z_t = a_{21}y_{t-1} - (1 - a_{22})z_{t-1} + \varepsilon_{zt} \tag{6.24}$$

Equations (6.23) and (6.24) form an error-correction model. If both a_{12} and a_{21} differ from zero, we can normalize the cointegrating vector with respect to either variable. Normalizing with respect to y_t, we get

$$\Delta y_t = a_y(y_{t-1} - \beta z_{t-1}) + \varepsilon_{yt}$$
$$\Delta z_t = a_z(y_{t-1} - \beta z_{t-1}) + \varepsilon_{zt}$$

where $\alpha_y = -a_{12}a_{21}/(1 - a_{22})$

$\beta = (1 - a_{22})/a_{21}$

$\alpha_z = a_{21}$

You can see that y_t and z_t change in response to the previous period's deviation from the long-run equilibrium $y_{t-1} - \beta z_{t-1}$. If $y_{t-1} = \beta z_{t-1}$, y_t and z_t change only in response to ε_{yt} and ε_{zt} shocks. Moreover, if $\alpha_y < 0$ and $\alpha_z > 0$, y_t decreases and z_t increases in response to a positive deviation from long-run equilibrium. You should also be able to convince yourself that conditions (6.20) and (6.21) ensure that $\beta \neq 0$ and that at least one of the speed of adjustment parameters (i.e., α_y and α_z) is not equal to zero. Now, refer to (6.9) and (6.10); you can see this model is in exactly the same form as the interest rate example presented in the beginning of this section.

Although a_{12} and a_{21} cannot both equal zero, an interesting special case arises if one of these coefficients is zero. For example, if we set $a_{12} = 0$, the speed of adjustment coefficient α_y equals zero. In this case, y_t changes only in response to ε_{yt} as $\Delta y_t = \varepsilon_{yt}$.[3] The $\{z_t\}$ sequence does all of the correction to eliminate any deviation from long-run equilibrium. Since $\{y_t\}$ does not do any of the error-correcting, $\{y_t\}$ is said to be **weakly exogenous**.

To highlight some of the important implications of this simple model, we have shown the following:

1. *The restrictions necessary to ensure that the variables are CI*(1, 1) *guarantee that an error-correction model exists.* In our example, both $\{y_t\}$ and $\{z_t\}$ are unit root processes but the linear combination $y_t - \beta z_t$ is stationary; the normalized cointegrating vector is $[1, -(1 - a_{22})/a_{21}]$. The variables have an error-correction representation with speed of adjustment coefficients $\alpha_y = -a_{12}a_{21}/(1 - a_{22})$ and $\alpha_z = a_{21}$. It was also shown that an error-correction

model for $I(1)$ variables necessarily implies cointegration. This finding illustrates the **Granger representation theorem** stating that for any set of $I(1)$ variables, error correction and cointegration are equivalent representations.

2. *A cointegration necessitates coefficient restrictions in a VAR model.* It is important to realize that a cointegrated system can be viewed as a restricted form of a general VAR model. Let $x_t = (y_t, z_t)'$ and $\varepsilon_t = (\varepsilon_{yt}, \varepsilon_{zt})'$ so that we can write (6.22) in the form

$$\Delta x_t = \pi x_{t-1} + \varepsilon_t \tag{6.25}$$

Clearly, it is inappropriate to estimate a VAR of cointegrated variables using only first differences. Estimating (6.25) without the expression πx_{t-1} would eliminate the error-correction portion of the model. It is also important to note that the rows of π are *not* linearly independent if the variables are cointegrated. Multiplying each element in row 1 by $-(1 - a_{22})/a_{12}$ yields the corresponding element in row 2. Thus, the determinant of π is equal to zero, and y_t and z_t have the error-correction representation given by (6.23) and (6.24).

This two-variable example illustrates the very important insights of Johansen (1988) and Stock and Watson (1988) that *we can use the rank of π to determine whether or not two variables $\{y_t\}$ and $\{z_t\}$ are cointegrated.* Compare the determinant of π to the characteristic equation given by (6.18). If the largest characteristic root equals unity ($\lambda_1 = 1$), it follows that the determinant of π is zero and that π has a rank equal to unity. If π were to have a rank of zero, it would be necessary for $a_{11} = 1, a_{22} = 1$, and $a_{12} = a_{21} = 0$. The VAR represented by (6.14) and (6.15) would be nothing more than $\Delta y_t = \varepsilon_{yt}$ and $\Delta z_t = \varepsilon_{zt}$. In this case, both the $\{y_t\}$ and $\{z_t\}$ sequences are unit root processes without any cointegrating vector. Finally, if the rank of π is full, then neither characteristic root can be unity, so the $\{y_t\}$ and $\{z_t\}$ sequences are jointly stationary. After all, if there are two independent stationary relations for $\{y_t\}$ and $\{z_t\}$, both variables must be stationary.

3. In general, both variables in a cointegrated system will respond to a deviation from long-run equilibrium. However, it is possible that one (but not both) of the speed of adjustment parameters is zero. For example, if $\alpha_y = 0$, $\{y_t\}$ does not respond to the discrepancy from long-run equilibrium and $\{z_t\}$ does all of the adjustment. In this circumstance, $\{y_t\}$ is weakly exogenous because it does none of the error correction. As such, an econometric model for $\{z_t\}$ can be estimated and hypothesis testing can be conducted without reference to a specific model for $\{y_t\}$. Section 10 and Appendix 6.2 consider modeling in a cointegrated system when a variable is weakly exogenous.

Also, *it is necessary to reinterpret Granger causality in a cointegrated system.* In a cointegrated system, $\{y_t\}$ does not Granger cause $\{z_t\}$ if lagged values Δy_{t-i} do not enter the Δz_t equation *and* if z_t does not respond to the deviation from long-run equilibrium. Hence, $\{z_t\}$ must be weakly exogenous. If $a_{21} = 0$ in (6.24), $\{z_t\}$ is weakly exogenous and is not Granger caused by

$\{y_t\}$. Similarly, in the cointegrated system of (6.11) and (6.12), $\{r_{Lt}\}$ does not Granger cause $\{r_{St}\}$ if all $a_{12}(i) = 0$ and if $\alpha_S = 0$.

The *n*-Variable Case

Little is altered in the *n*-variable case. The relationship between cointegration, error correction, and the rank of the matrix π is invariant to adding additional variables to the system. The interesting feature introduced in the *n*-variable case is the possibility of multiple cointegrating vectors. Now consider a more general version of (6.25):

$$x_t = A_1 x_{t-1} + \varepsilon_t \tag{6.26}$$

where x_t = the $(n \cdot 1)$ vector $(x_{1t}, x_{2t}, \ldots, x_{nt})'$
ε_t = the $(n \cdot 1)$ vector $(\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{nt})'$
A_1 = an $(n \cdot n)$ matrix of parameters

Subtracting x_{t-1} from each side of (6.26) and letting I be an $(n \cdot n)$ identity matrix, we get

$$\begin{aligned}\Delta x_t &= -(I - A_1)x_{t-1} + \varepsilon_t \\ &= \pi x_{t-1} + \varepsilon_t\end{aligned} \tag{6.27}$$

where π is the $(n \cdot n)$ matrix $-(I - A_1)$ and π_{ij} denotes the element in row i and column j of π. As you can see, (6.27) is a special case of (6.13) such that all $\pi_i = 0$.

Again, the crucial issue for cointegration concerns the rank of the $(n \cdot n)$ matrix π. The only way for the rank of a matrix to be zero is for each of its elements to be zero. Hence, if the rank of π is zero, each element of π must equal zero so that there are no cointegrating vectors. In this instance, (6.27) is equivalent to an *n*-variable VAR in first differences:

$$\Delta x_t = \varepsilon_t$$

Here, each $\Delta x_{it} = \varepsilon_{it}$ so that all the $\{x_{it}\}$ sequences are unit root processes and there is no linear combination of the variables that is stationary.

At the other extreme, suppose that π is of full rank. The long-run solution to (6.27) is given by the *n*-independent equations:

$$\begin{aligned}&\pi_{11}x_{1t} + \pi_{12}x_{2t} + \pi_{13}x_{3t} + \cdots + \pi_{1n}x_{nt} = 0 \\ &\pi_{21}x_{1t} + \pi_{22}x_{2t} + \pi_{23}x_{3t} + \cdots + \pi_{2n}x_{nt} = 0 \\ &. \\ &. \\ &. \\ &\pi_{n1}x_{1t} + \pi_{n2}x_{2t} + \pi_{n3}x_{3t} + \cdots + \pi_{nn}x_{nt} = 0.\end{aligned} \tag{6.28}$$

Each of these *n* equations is an independent restriction on the long-run solution of the variables; the *n* variables in the system face *n* long-run constraints. In this case, each of the *n* variables contained in the vector x_t must be stationary with the long-run values given by the solution to 6.28. The variables cannot be *CI*(1, 1) since all are stationary.

In intermediate cases, in which the rank of π is equal to $r < n$, there are r cointegrating vectors. With r independent equations and n variables, there are $n - r$ stochastic trends in the system. If $r = 1$, there is a single cointegrating vector given by any row of the matrix π. Each $\{\Delta x_{it}\}$ sequence can be written in error-correction form. For example, we can write Δx_{1t} as

$$\Delta x_{1t} = \pi_{11} x_{1t-1} + \pi_{12} x_{2t-1} + \cdots + \pi_{1n}\, x_{nt-1} + \varepsilon_{1t}$$

or, normalizing with respect to x_{1t-1}, we can set $\alpha_1 = \pi_{11}$ and $\beta_{1j} = \pi_{1j}/\pi_{11}$ to obtain

$$\Delta x_{1t} = \alpha_1 (x_{1t-1} + \beta_{12} x_{2t-1} + \cdots + \beta_{1n} x_{nt-1}) + \varepsilon_{1t} \tag{6.29}$$

In the long run, the $\{x_{it}\}$ will satisfy the relationship

$$x_{1t} + \beta_{12} x_{2t} + \cdots + \beta_{1n} x_{nt} = 0$$

Hence, the normalized cointegrating vector is $(1, \beta_{12}, \beta_{13}, \ldots, \beta_{1n})$ and the speed of adjustment parameter is α_1. In the same way, with two cointegration vectors the long-run values of the variables will satisfy the two relationships

$$\pi_{11} x_{1t} + \pi_{12} x_{2t} + \cdots + \pi_{1n} x_{nt} = 0$$
$$\pi_{21} x_{1t} + \pi_{22} x_{2t} + \cdots + \pi_{2n} x_{nt} = 0$$

which can be appropriately normalized.

The main point here is that there are three important ways to test for cointegration. The Engle–Granger methodology seeks to determine whether the residuals of the equilibrium relationship are stationary. The Johansen (1988) methodology determines the rank of π and the error-correction method examines the speed of adjustment coefficients. The Engle–Granger approach is the subject of the next three sections. Sections 7 through 9 examine the Johansen (1988) methodology, and testing within the error-correction framework is examined in Section 10.

4. TESTING FOR COINTEGRATION: THE ENGLE–GRANGER METHODOLOGY

To explain the Engle–Granger testing procedure, let's begin with the type of problem likely to be encountered in applied studies. Suppose that two variables—say y_t and z_t—are believed to be integrated of order 1 and we want to determine whether there exists an equilibrium relationship between the two. Engle and Granger (1987) propose a four-step procedure to determine if two $I(1)$ variables are cointegrated of order $CI(1, 1)$.

STEP 1: Pretest the variables for their order of integration. By definition, cointegration necessitates that two variables be integrated of the same order. Thus, the first step in the analysis is to pretest each variable to determine its order of integration. The augmented Dickey–Fuller tests discussed in Chapter 4 can be used to infer the number of unit roots (if any) in each of the variables. If both variables are stationary, it is not necessary to proceed since standard time-series methods apply to stationary variables. If the variables are integrated of different orders, it is possible to conclude they are *not* cointegrated. However, as detailed in Section 5, if you have more than two variables such that some are $I(1)$ and some are $I(2)$, you may want to determine whether the variables are multicointegrated.

STEP 2: Estimate the long-run equilibrium relationship. If the results of Step 1 indicate that both $\{y_t\}$ and $\{z_t\}$ are $I(1)$, the next step is to estimate the long-run equilibrium relationship in the form

$$y_t = \beta_0 + \beta_1 z_t + e_t \tag{6.30}$$

If the variables are cointegrated, an OLS regression yields a "super-consistent" estimator of the cointegrating parameters β_0 and β_1. Stock (1987) proves that the OLS estimates of β_0 and β_1 converge faster than they do in OLS models using stationary variables. To explain, reexamine the scatter plot shown in Figure 6.1. You can see that the effect of the common trend dominates the effect of the stationary component; both variables seem to rise and fall in tandem. Hence, there is a strong linear relationship as shown by the regression line drawn in the figure.

In order to determine if the variables are actually cointegrated, denote the residual sequence from this equation by $\{\hat{e}_t\}$. Thus, the $\{\hat{e}_t\}$ series contains the estimated values of the deviations from the long-run relationship. If these deviations are found to be stationary, the $\{y_t\}$ and $\{z_t\}$ sequences are cointegrated of order (1, 1). It would be convenient if we could perform a Dickey–Fuller test on these residuals to determine their order of integration. Consider the autoregression of the residuals:

$$\Delta\hat{e}_t = a_1\hat{e}_{t-1} + \varepsilon_t \tag{6.31}$$

Since the $\{\hat{e}_t\}$ sequence is a residual from a regression equation (with a mean necessarily equal to zero), there is no need to include an intercept term; the parameter of interest in (6.31) is a_1. If we cannot reject the null hypothesis $a_1 = 0$, we can conclude that the residual series contains a unit root. Hence, we conclude that the $\{y_t\}$ and $\{z_t\}$ sequences are *not* cointegrated. The more precise wording is awkward because of a triple negative, but to be technically correct, *if it is not possible to reject the null hypothesis* $a_1 = 0$*, we cannot reject the hypothesis that the variables are not cointegrated.* Instead, the rejection of the null hypothesis implies that the residual sequence is stationary. Given that $\{y_t\}$ and $\{z_t\}$ were both found to be $I(1)$ and that the residuals are stationary, we can conclude that the series are cointegrated of order (1, 1).

In most applied studies it is not possible to use the Dickey–Fuller tables themselves. The problem is that the $\{\hat{e}_t\}$ sequence is generated from a regression equation; the researcher does not know the actual error e_t, only the estimate of the error $\hat{e}_t$. The methodology of fitting the regression in (6.30) selects values of β_0 and β_1 that minimize the sum of squared residuals. Since the residual variance is made as small as possible, the procedure is prejudiced toward finding a stationary error process in (6.31). Hence, the test statistic used to test the magnitude of a_1 must reflect this fact. Only if β_0 and β_1 were known in advance and used to construct the true $\{e_t\}$ sequence would an ordinary Dickey–Fuller table be appropriate. When you estimate the cointegrating vector, use the critical values provided in Table C in the *Supplementary Manual*. These critical values depend on sample size and the number of variables used in the analysis. For example, to test for cointegration between two variables using a sample size of 100, the critical value at the 5% significance level is –3.398.

If the residuals of (6.31) do not appear to be white noise, an augmented form of the test can be used instead of (6.31). Suppose that diagnostic checks indicate that the $\{\varepsilon_t\}$ sequence of (6.31) exhibits serial correlation. Instead of using the results from (6.31), estimate the autoregression:

$$\Delta\hat{e}_t = a_1\hat{e}_{t-1} + \sum_{i=1}^{n} a_{i+1}\Delta\hat{e}_{t-i} + \varepsilon_t \tag{6.32}$$

Again, if we reject the null hypothesis $a_1 = 0$, we can conclude that the residual sequence is stationary and that the variables are cointegrated.

STEP 3: Estimate the error-correction model. If the variables are cointegrated (i.e., if the null hypothesis of no cointegration is rejected), the residuals from the equilibrium regression can be used to estimate the error-correction model. If $\{y_t\}$ and $\{z_t\}$ are $CI(1, 1)$, the variables have the error-correction form:

$$\Delta y_t = \alpha_1 + \alpha_y[y_{t-1} - \beta_1 z_{t-1}] + \sum_{i=1} a_{11}(i)\Delta y_{t-i} + \sum_{i=1} a_{12}(i)\Delta z_{t-i} + \varepsilon_{yt} \tag{6.33}$$

$$\Delta z_t = \alpha_2 + \alpha_z[y_{t-1} - \beta_1 z_{t-1}] + \sum_{i=1} a_{21}(i)\Delta y_{t-i} + \sum_{i=1} a_{22}(i)\Delta z_{t-i} + \varepsilon_{zt} \tag{6.34}$$

where β_1 = the parameter of the cointegrating vector given by (6.30), ε_{yt}, and ε_{zt} = white-noise disturbances (which may be correlated with each other), and α_1, α_2, α_y, α_z, $\alpha_{11}(i)$, $\alpha_{12}(i)$, $\alpha_{21}(i)$, $\alpha_{22}(i)$ are all parameters.

Engle and Granger (1987) propose a clever way to circumvent the cross-equation restrictions involved in the direct estimation of (6.33) and (6.34). The magnitude of the residual $\hat{e}_{t-1}$ is the deviation from long-run equilibrium in period $(t - 1)$. Hence, it is possible to use the saved residuals $\{\hat{e}_{t-1}\}$ obtained in Step 2 as an estimate of the expression $y_{t-1} - \beta_1 z_{t-1}$ in (6.33) and (6.34). Thus, using the saved residuals from the estimation of the long-run equilibrium relationship, estimate the error-correcting model as

$$\Delta y_t = \alpha_1 + \alpha_y\hat{e}_{t-1} + \sum_{i=1}\alpha_{11}(i)\Delta y_{t-i} + \sum_{i=1}\alpha_{12}(i)\Delta z_{t-i} + \varepsilon_{yt} \tag{6.35}$$

$$\Delta z_t = \alpha_2 + \alpha_z\hat{e}_{t-1} + \sum_{i=1}\alpha_{21}(i)\Delta y_{t-i} + \sum_{i=1}\alpha_{22}(i)\Delta z_{t-i} + \varepsilon_{zt} \tag{6.36}$$

Other than the error-correction term $\hat{e}_{t-1}$, (6.35) and (6.36) constitute a VAR in first differences. This VAR can be estimated using the same methodology developed in Chapter 5. All of the procedures developed for a VAR apply to the system represented by the error-correction equations. Notably:

1. OLS is an efficient estimation strategy since each equation contains the same set of regressors.

2. Since all terms in (6.35) and (6.36) are stationary [i.e., Δy_t and its lags, Δz_t and its lags, and $\hat{e}_{t-1}$ are $I(0)$], the test statistics used in traditional VAR analysis are appropriate for (6.35) and (6.36). For example, lag lengths can be determined using a χ^2-test, and the restriction that all $\alpha_{jk}(i) = 0$ can be checked using an F-test. If there is a single cointegrating vector, restrictions concerning α_y or α_z can be conducted using a t-test.

STEP 4: Assess Model Adequacy. There are several procedures that can help determine whether the error-correction estimated model is appropriate.

1. You should be careful to assess the adequacy of the model by performing diagnostic checks to determine whether the residuals of the error-correction equations approximate white noise. If the residuals are serially correlated, lag lengths may be too short. Reestimate the model using lag lengths that yield serially uncorrelated errors. It may be that you need to allow longer lags of some variables than of others. If so, you can gain efficiency by estimating the near-VAR using the seemingly unrelated regressions (SUR) method. Out of sample forecasting exercises are also useful ways to select among alternative models.

2. The *speed of adjustment* coefficients α_y and α_z are of particular interest in that they have important implications for the dynamics of the system. As shown in Section 3, the values of α_y and α_z are directly related to the characteristic roots of the difference equation system. Direct convergence necessitates that α_y be negative and α_z be positive. If we focus on (6.36) it is clear that for any given value of $\hat{e}_{t-1}$, a large value of α_z is associated with a large value of Δz_t. If α_z is zero, the change in z_t does not at all respond to the deviation from long-run equilibrium in $(t-1)$. If α_z is zero and if all $\alpha_{21}(i) = 0$, then it can be said that $\{\Delta y_t\}$ does not Granger cause $\{\Delta z_t\}$. We know that α_y and/or α_z should be significantly different from zero if the variables are cointegrated. After all, if both α_y and α_z are zero, there is no error correction and (6.35) and (6.36) comprise nothing more than a VAR in first differences. Moreover, the absolute values of these speeds of adjustment coefficients must not be too large. The point estimates should imply that Δy_t and Δz_t converge to the long-run equilibrium relationship.

If all but one variable is weakly exogenous, you may want to estimate that variable using the error-correction technique described in Section 10.

3. As in a traditional VAR analysis, Lutkepohl and Reimers (1992) show that innovation accounting (i.e., impulse responses and variance decomposition analysis) can be used to obtain information concerning the interactions among the variables. As a practical matter, the two innovations ε_{yt} and ε_{zt} may be contemporaneously correlated if y_t has a contemporaneous effect on z_t and/or if z_t has a contemporaneous effect on y_t. In obtaining impulse response functions and variance decompositions, some method—such as a Choleski Decomposition—must be used to orthogonalize the innovations.

The shape of the impulse response functions and the results of the variance decompositions can indicate whether the dynamic responses of the variables conform to theory. Since all variables in (6.35) and (6.36) are $I(0)$, the impulse responses of Δy_t and Δz_t should converge to zero. You should reexamine your results from each step if you obtain a nondecaying or explosive impulse response function.

Before closing this section, a word of warning is in order. It is very tempting to use t-statistics to perform significance tests on the cointegrating vector. However, you must avoid this temptation since, in general, the coefficients do not have an asymptotic t-distribution. To explain, suppose you estimate (6.30) so that have a model in the form: $y_t = \beta_0 + \beta_1 z_t + e_t$. Even if the variables are cointegrated, the $\{e_t\}$ sequence is likely to be serially correlated. Moreover, since y_t and z_t are jointly determined variables, there is a simultaneity problem so that $\{z_t\}$ cannot be treated as an "independent" variable. There is one case in which the t-statistics are appropriate. Suppose that the cointegration relationship between $\{y_t\}$ and $\{z_t\}$ is such that

$$y_t = \beta_0 + \beta_1 z_t + \varepsilon_{1t}$$
$$\Delta z_t = \varepsilon_{2t}$$

where $E\varepsilon_{1t}\varepsilon_{2t} = 0$.

The notation is designed to illustrate the point that the residuals from both equations are uncorrelated white-noise disturbances. The set of assumptions is fairly restrictive in that the residuals from both equations must be serially uncorrelated and the cross-correlations must be zero. If these conditions hold, the OLS estimates of β_0 and β_1 can be tested using t-tests and F-tests. If the disturbances are not normally distributed, the asymptotic results are such that t-tests and F-tests are appropriate. Be aware that both conditions are necessary to perform such tests. If $E\varepsilon_{1t}\varepsilon_{2t} \neq 0$, $\{z_t\}$ is not exogenous since shocks to ε_{1t} affect z_t. Moreover, as in a standard regression, if the residuals of the cointegrating vector are serially correlated, inference concerning the coefficients is inappropriate. Phillips and Hansen (1990) develop a procedure that allows you to construct confidence intervals for the β_i in the presence of serial correlation and the lack of independence of the $\{z_t\}$ sequence. The details are discussed in Appendix 6.2 in the *Supplementary Manual*.

5. ILLUSTRATING THE ENGLE–GRANGER METHODOLOGY

Figure 6.2 shows three simulated variables that can be used to illustrate the Engle–Granger procedure. Inspection of the figure suggests that each is nonstationary, and there is no visual evidence that any pair is cointegrated. As detailed in Table 6.1,

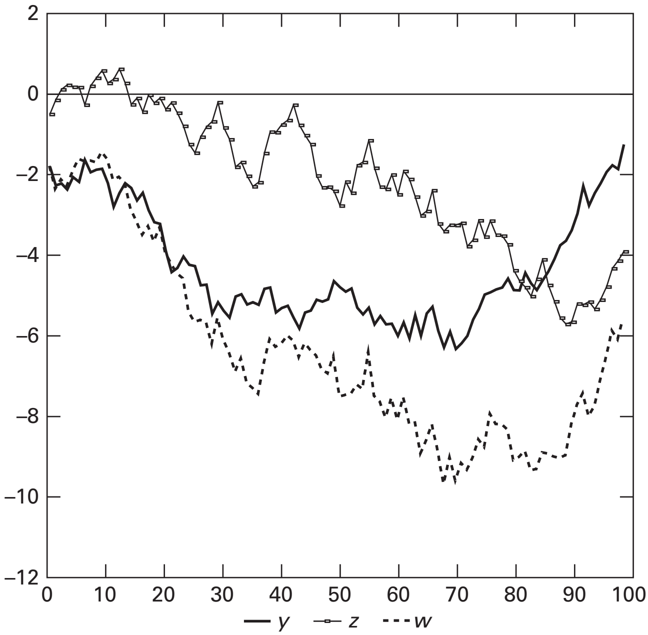


FIGURE 6.2 Three Cointegrated Series

Table 6.1 The Simulated Series

	$\{y_t\}$	$\{z_t\}$	$\{w_t\}$
Trend	$\mu_{yt} = \mu_{yt-1} + \varepsilon_{yt}$	$\mu_{zt} = \mu_{zt-1} + \varepsilon_{zt}$	$\mu_{wt} = \mu_{yt} + \mu_{zt}$
Pure Irregular	$\delta_{yt} = 0.5\delta_{yt-1} + \eta_{yt}$	$\delta_{zt} = 0.5\delta_{zt-1} + \eta_{zt}$	$\delta_{wt} = 0.5\delta_{wt-1} + \eta_{wt}$
Series	$y_t = \mu_{yt} + \delta_{yt}$	$z_t = \mu_{zt} + \delta_{zt} + 0.5\delta_{yt}$	$w_t = \mu_{wt} + \delta_{wt} + 0.5\delta_{yt} + 0.5\delta_{zt}$

each series is constructed as the sum of a stochastic trend component plus an autoregressive irregular component.

The first column of the table contains the formulas used to construct the $\{y_t\}$ sequence. First, 150 realizations of a white-noise process were drawn to represent the $\{\varepsilon_{yt}\}$ sequence. Initializing $\mu_{y0} = 0$, 150 values of the random walk process $\{\mu_{yt}\}$ were constructed using the formula $\mu_{yt} = \mu_{yt-1} + \varepsilon_{yt}$ (see the first cell of the table). Another 150 realizations of a white-noise process were drawn to represent the $\{\eta_{yt}\}$ sequence; given the initial condition $\delta_{y0} = 0$, these realizations were used to construct $\{\delta_{yt}\}$ as $\delta_{yt} = 0.5\delta_{yt-1} + \eta_{yt}$ (see the next lower cell). Adding the two constructed series yields 150 realizations for $\{y_t\}$. To help ensure randomness, only the last 100 observations are used in the simulated study. Hence, $\{y_t\}$ is the sum of a stochastic trend and a stationary (i.e., irregular) component.

The $\{z_t\}$ sequence was constructed in a similar fashion; the $\{\varepsilon_{zt}\}$ and $\{\eta_{zt}\}$ sequences are each represented by two different sets of 150 random numbers. The trend $\{\mu_{zt}\}$ and the autoregressive irregular term $\{\delta_{zt}\}$ were constructed as shown in the second column of the table. The $\{\delta_{zt}\}$ sequence can be thought of as a pure irregular component in the $\{z_t\}$ sequence. In order to introduce correlation between the $\{y_t\}$ and $\{z_t\}$ sequences, the irregular component in $\{z_t\}$ was constructed as the sum: $\delta_{zt} + 0.5\delta_{yt}$. In the third column you can see that the trend in $\{w_t\}$ is the simple summation of the trends in the other two series. As such, the three series have the cointegrating vector $(1, 1, -1)$. The irregular component in $\{w_t\}$ is the sum of pure innovation δ_{wt} and 50% of the innovations δ_{yt} and δ_{zt}.

Now pretend that we do not know the data-generating process. The issue is whether the Engle–Granger methodology can uncover the essential details of the process. Use the data on the file COINT6.XLS to follow along. The first step is to pretest the variables in order to determine their order of integration. Consider the augmented Dickey–Fuller regression equation for $\{y_t\}$:

$$\Delta y_t = \alpha_0 + \alpha_1 y_{t-1} + \sum_{i=1}^{n} \alpha_{i+1} \Delta y_{t-i} + \varepsilon_t$$

If the data happened to be quarterly, it would be natural to perform the augmented Dickey–Fuller tests using lag lengths that are multiples of 4 (i.e., $n = 4, 8, \ldots$). For each series, the results of the Dickey–Fuller test and the augmented test using 4 lags are reported in Table 6.2.

With 100 observations and a constant, the 5% critical value for the Dickey–Fuller test is −2.89. Since the absolute values of all t-statistics are well below this critical value, we cannot reject the null hypothesis of a unit root in any of the series. Of course, if there were any serious doubt about the presence of a unit root, we could use the procedures in Chapter 4 to test for the presence of the drift term. If various lag lengths yield different results, we would want to test for the most appropriate lag length.

The luxury of using simulated data is that we can avoid these potentially sticky problems and move on to Step 2. Since all three variables are presumed to be jointly determined, the long-run equilibrium regression can be estimated using either y_t, z_t or w_t as the "left-hand-side" variable. The three estimates of the long-run relationship (with t-values in parentheses) are

$$\begin{aligned}
y_t &= -0.048 \;-\; 0.927z_t \;+\; 0.977w_t + e_{yt} \\
&\quad\;(-0.58) \quad (-38.10) \quad (53.461) \\
z_t &= 0.0590 \;-\; 1.011y_t \;+ 1.026w_t + e_{zt} \\
&\quad\;(-0.67) \quad (-38.10) \quad (65.32) \\
w_t &= -0.085 \;+\; 0.990y_t + 0.953z_t + e_{wt} \\
&\quad\;(-1.01) \quad (53.46) \quad\;\; (65.32)
\end{aligned}$$

where e_{yt}, e_{zt}, and e_{wt} = the residuals from the three equilibrium regressions.

The essence of the test is to determine whether the residuals from the equilibrium regression are stationary. Again, in performing the test, there is no presumption that any one of the three residual series is preferable to any of the others. If we use each of the three series to estimate an equation in the form of (6.31) [or (6.32)], the estimated values of a_1 are given in Table 6.3.

Table 6.2 Estimated α_1 and the Associated t-statistic

	No Lags	4 Lags
Δy_t	−0.020 (−0.742)	−0.027 (−1.047)
Δz_t	−0.021 (−0.992)	−0.258 (−1.144)
Δw_t	−0.035 (−1.908)	−0.037 (−1.936)

Table 6.3 Estimated a_1 and the Associated t-statistic

	No lags	4 Lags
Δe_{yt}	−0.443 (−5.175)	−0.595 (−4.074)
Δe_{zt}	−0.452 (−5.379)	−0.593 (−4.226)
Δe_{wt}	−0.455 (−5.390)	−0.607 (−4.225)

From Table C, you can see that the critical values of the t-statistic as −3.828. Hence, using any one of the three equilibrium regressions, we can conclude that the series are cointegrated of order (1, 1). Fortunately, all three equilibrium regressions yield this same conclusion. We should be very wary of a result indicating that the variables are cointegrated using one variable for the normalization but are not cointegrated using another variable for the normalization. In such circumstances, it is possible that only a few of the variables are cointegrated. Suppose that x_{1t}, x_{2t}, and x_{3t} are three $I(1)$ variables and that x_{1t} and x_{2t} are cointegrated such that $x_{1t} - \beta_2 x_{2t}$ is stationary. A regression of x_{1t} on the other two variables should yield the stationary relationship $x_{1t} = \beta_2 x_{2t} + 0x_{3t}$. Similarly, a regression of x_{2t} on the other variables should yield the stationary relationship $x_{2t} = (1/\beta_2)x_{1t} + 0x_{3t}$. However, a regression of x_{3t} on x_{1t} and x_{2t} cannot reveal the cointegrating relationship. Nevertheless, the possibility of a contradictory result is a weakness of the test.

You must be careful in conducting significance tests on the estimated equilibrium regressions. As mentioned above, the coefficients *do not* have an asymptotic t-distribution unless the right-hand-side variables are actually independent and the error terms are serially uncorrelated. From Table 6.1, it should be clear that these assumptions are violated by the data generating process.

Step 3 entails estimating the error-correction model. Consider the first-order system shown with t-statistics in parentheses:

$$\underset{}{\Delta y_t} = \underset{(0.19)}{0.006} + \underset{(2.79)}{0.418}e_{wt-1} + \underset{(1.08)}{0.178}\Delta y_{t-1} + \underset{(1.94)}{0.313}\Delta z_{t-1} - \underset{(-2.27)}{0.368}\Delta w_{t-1} + \varepsilon_{yt} \quad (6.37)$$

$$\Delta z_t = \underset{(-1.12)}{-0.042} + \underset{(0.42)}{0.074e_{wt-1}} + \underset{(0.75)}{0.145\Delta y_{t-1}} + \underset{(1.38)}{0.262\Delta z_{t-1}} - \underset{(-1.63)}{0.313\Delta w_{t-1}} + \varepsilon_{zt} \quad (6.38)$$

$$\Delta w_t = \underset{(-0.90)}{-0.040} - \underset{(-0.33)}{0.069e_{wt-1}} + \underset{(0.68)}{0.156\Delta y_{t-1}} + \underset{(1.35)}{0.301\Delta z_{t-1}} - \underset{(-1.87)}{0.420\Delta w_{t-1}} + \varepsilon_{wt} \quad (6.39)$$

where $e_{wt-1} = w_{t-1} + 0.0852 - 0.9901y_{t-1} - 0.9535z_{t-1}$ so that e_{wt-1} is the lagged value of the residual from the equilibrium relationship using w_t as the dependent variable.

Equations (6.37) through (6.39) comprise a first-order VAR augmented with the single error-correction term e_{wt-1}. Again, there is an area of ambiguity since the residuals from any of the "equilibrium" relationships could have been used in the estimation. The signs of the speed of adjustment coefficients are in accord with convergence toward the long-run equilibrium. In response to a positive discrepancy in e_{wt-1}, both y_t and z_t tend to increase while w_t tends to decrease. The error-correction term, however, is significant only in (6.37).

Finally, the diagnostic methods discussed in the last section should be applied to (6.37) through (6.39) in order to assess the model's adequacy. Using actual data, lag-length tests and the properties of the residuals need to be considered. Moreover, innovation accounting could help determine whether the model is adequate. Question 2 at the end of the chapter asks you to perform some of these diagnostics.

The Engle–Granger Procedure with $I(2)$ Variables

Multicointegration refers to a situation in which a linear combination of $I(2)$ and $I(1)$ variables is integrated of order zero. For example, suppose that x_{1t} and x_{2t} are $I(2)$ and that z_t is $I(1)$. It is possible that a linear combination of x_{1t} and x_{2t} is $I(1)$ and that this combination is cointegrated with z_t. Hence, it is possible to have a long-run equilibrium relationship of the form

$$x_{1t} = \beta_2 x_{2t} + \alpha_1 z_t$$

However, a richer set of possibilities is given by the stationary relationship

$$x_{1t} = \beta_2 x_{2t} + \gamma_1 \Delta x_{2t} + \alpha_1 z_t$$

This specification allows for the possibility that the linear combination $x_{1t} - \beta_2 x_{2t}$ is $I(1)$ and cointegrated with the other $I(1)$ independent variables in the system: Δx_{2t} and z_t. To make sure you understand the issue, ask yourself if it is possible for β_2 to be zero. The answer is a resounding no. If $\beta_2 = 0$, the $I(2)$ variable x_{1t} cannot, by itself, be cointegrated with the $I(1)$ variables.

In principle, it is possible to check for multicointegration using a two-step procedure. First, search for a cointegrating relationship among the $I(2)$ variables and then use this relationship to check for a possible cointegrating relationship with the remaining $I(1)$ variables. Engsted, Gonzalo and Haldrup (1997) show that this procedure is effective only if the cointegrating vector for the first step is known. Otherwise, the second step is contaminated with the errors generated in the first step. In the most general form of their one-step procedure, you estimate an equation in the form

$$x_{1t} = a_0 + a_1 t + a_2 t^2 + \beta_2 x_{2t} + \beta_3 x_{3t} + \gamma_1 \Delta x_{2t} + \gamma_2 \Delta x_{3t} + \alpha_1 z_t + e_t \tag{6.40}$$

where x_{1t}, x_{2t}, and x_{3t} are $I(2)$ variables, z_t is a vector of $I(1)$ variables, and the deterministic regressors can include a quadratic time trend.

Hence, the test allows you to include up to two $I(2)$ variables and an unrestricted number of $I(1)$ variables as regressors. You might want to include the quadratic time trend if $\Delta^2 x_{1t}$ contains a drift. Since the key issue is the stationarity of the $\{e_t\}$ series, estimate a regression of the form

$$\Delta \hat{e}_t = \rho \hat{e}_{t-1} + \sum_{i=1}^{p} \rho_i \Delta \hat{e}_{t-i} + v_t$$

where $\{\hat{e}_t\}$ are the regression residuals from (6.40).

If it is possible to reject the null hypothesis $\rho = 0$, it is possible to conclude that there is multicointegration. In addition to sample size, the critical values of the t-statistic for the null hypothesis $\rho = 0$ depend on the number of $I(2)$ regressors ($m_2 = 1$ or 2), the number of $I(1)$ regressors ($m_1 = 0$ to 4), and the form of the deterministic regressors. The critical values are shown in Table D in the *Supplementary Manual*. Consider the U.K. money demand equations for the sample period 1963Q1 to 1989Q2 estimated by Haldrup (1994):

$$m_t = a_0 + 0.68 p_t + 1.57 y_t - 2.67 r_t - 2.55 \Delta p_t \tag{6.41}$$

and

$$m_t = a_0 + a_1 t + 0.89 p_t + 2.39 y_t - 2.69 r_t - 3.25 \Delta p_t \tag{6.42}$$

Pretesting the variables indicated that m_t (as measured by the log $M1$) and p_t (the log of the implicit price deflator) were $I(2)$ and that y_t (the log of total final expenditure) and r_t (a measure of the interest rate differential) were $I(1)$. The only variable needing explanation is the presence of Δp_t in the money demand function. The idea is to allow for the demand for money to depend on the inflation rate (i.e., change in the log of the price level) since high inflation should reduce the desire to hold money balances. Since there is a total of 105 observations, one $I(2)$ regressor (so that $m_2 = 1$), and three $I(1)$ regressors, the 5% critical values for models without and with the linear trend are −4.56 and −4.91, respectively. Using the residuals from the money demand equations given by (6.41) and (6.42), Haldrup found that the t-statistics for the null hypothesis $\rho = 0$ were −2.35 and −2.66, respectively. Hence, it is possible to conclude that the two regressions are spurious (i.e., it is not possible to reject the null hypothesis of no multicointegration).

Even though multicointegration fails, Haldrup goes on to experiment with various estimates of the error-correction mechanism. One interesting model (with standard errors in parentheses) is

$$\Delta^2 m_t = \underset{(0.02)}{-0.04\hat{e}_{t-1}} + \text{stationary regressors}$$

where the stationary regressors can include lagged values of $\Delta^2 m_t$ as well as current and lagged values of $\Delta^2 p_t$, Δy_t, Δp_t, and Δr_t. The point estimate is such that $\Delta^2 m_t$ is expected to decline in response to a positive discrepancy from the long-run relationship. The t-statistic of $-0.04/0.02 = 2$ suggests that the effect is just significant at the 5% level.

6. COINTEGRATION AND PURCHASING POWER PARITY

To illustrate the Engle–Granger methodology using actual data, reconsider the theory of purchasing power parity (PPP). Respectively, if e_t, p_t^*, and p_t denote the logarithms of the price of foreign exchange, the foreign price level, and the domestic price level, long-run PPP requires that $e_t + p_t^* - p_t$ be stationary. The unit root tests reported in Chapter 4 indicate that real exchange rates (defined as $r_t = e_t + p_t^* - p_t$) appear to be nonstationary. Cointegration offers an alternative method to check the theory; if PPP holds, the sequence formed by the sum $\{e_t + p_t^*\}$ should be cointegrated with the $\{p_t\}$ sequence. Call the constructed dollar value of the foreign price level f_t; that is, $f_t = e_t + p_t^*$. Long-run PPP asserts that there exists a linear combination of the form $f_t = \beta_0 + \beta_1 p_t + \mu_t$ such that $\{\mu_t\}$ is stationary *and* the cointegrating vector is such that $\beta_1 = 1$.

As reported in Chapter 4, in Enders (1988), I used price and exchange rate data for Germany, Japan, Canada, and the United States for both the Bretton Woods (1960–1971) and post-Bretton Woods (1973–1988) periods. Each series was converted into an index number such that each series was equal to unity at the beginning of its respective period (either 1960 or 1973). In the fixed exchange rate period, all values of $\{e_t\}$ were set equal to unity. Pretesting the data indicated that for each period, the U.S. price level $\{p_t\}$ and the dollar values of the foreign price levels $\{e_t + p_t^*\}$ both contained a single unit root. With differing orders of integration, it would have been possible to immediately conclude that long-run PPP had failed.

The next step was to estimate the long-run equilibrium relation by regressing each $f_t = e_t + p_{ft}$ on p_t such that

$$f_t = \beta_0 + \beta_1 p_t + \mu_t \tag{6.43}$$

Absolute PPP asserts $f_t = p_t$, so this version of the theory requires $\beta_0 = 0$ and $\beta_1 = 1$. The intercept β_0 is consistent with the relative version of PPP, requiring only that domestic and foreign price levels are proportional to each other. Unless there are compelling reasons to omit the constant, the recommended practice is to include an intercept term in the equilibrium regression. In fact, Engle and Granger's (1987) original Monte Carlo simulations all include intercept terms.

The estimated values of β_1 and their associated standard errors are reported in Table 6.4. Note that five of the six values are estimated to be quite a bit below unity. Be especially careful not to make too much of these findings. It is *not* appropriate to conclude that each value of β_1 is significantly different from unity simply because the values of $(1 - \beta_1)$ exceed two or three standard deviations. It is hard to overstate the

Table 6.4 The Equilibrium Regressions

	Germany	Japan	Canada
1973–1986			
Estimated β_1	0.5374	0.8938	0.7749
Standard Error	(0.0415)	(0.0316)	(0.0077)
1960–1971			
Estimated β_1	0.6660	0.7361	1.0809
Standard Error	(0.0262)	(0.0154)	(0.0200)

point that the assumptions underlying this type of t-test are not applicable because there is no presumption that p_t is the exogenous variable while f_t is the dependent variable, or that $\{\mu_t\}$ is white noise.

The residuals from each regression equation, called $\{\hat{\mu}_t\}$, were checked for unit roots. The unit root tests are straightforward because the residuals from a regression equation have a zero mean and do not have a time trend. The following two equations were estimated using the residuals from each long-run equilibrium relationship:

$$\Delta\hat{\mu}_t = a_1\hat{\mu}_{t-1} + \varepsilon_t \tag{6.44}$$

and

$$\Delta\hat{\mu}_t = a_1\hat{\mu}_{t-1} + \sum_{i=1}^{p} a_{i+1}\Delta\hat{\mu}_{t-i} + \varepsilon_t \tag{6.45}$$

Table 6.5 reports the estimated values of a_1 from (6.44) and from (6.45) using a lag length of four. It bears repeating that failure to reject the null hypothesis $a_1 = 0$ means we cannot reject the null of no cointegration. Alternatively, if $-2 < a_1 < 0$, it is possible to conclude that the $\{\hat{\mu}_t\}$ sequence does not have a unit root and that the $\{f_t\}$ and $\{p_t\}$ sequences are cointegrated. Also note that it is not appropriate to use the confidence intervals reported in Dickey and Fuller. The Dickey–Fuller statistics are inappropriate because the residuals used in (6.44) and (6.45) are not the actual error terms. Rather, these residuals are estimated error terms that are obtained from the estimate of the equilibrium regression. If we knew the magnitudes of the actual errors in each period, we could use the Dickey–Fuller tables.

Under the null hypothesis $a_1 = 0$, the critical values for the t-statistic depend on sample size. Comparing the results reported in Table 6.5 with the critical values provided by Table C indicates that only for Japan during the fixed exchange rate period it is possible to reject the null hypothesis of no cointegration. At the 5% significance level, the critical value of t is -3.398 for two variables and $T = 100$. Hence, at the 5% significance level we can reject the null of no cointegration (i.e., we accept the alternative that the variables are cointegrated) and find in favor of PPP. For the other countries in each time period, we cannot reject the null hypothesis of no cointegration and must conclude that PPP generally failed.

The third step in the methodology entails estimation of the error-correction model. Only the Japan/U.S. model needs estimation since it is the sole case for which

Table 6.5 Dickey–Fuller Tests of the Residuals

	Germany	Japan	Canada
1973–1986			
No lags			
Estimated a_1	−0.0225	−0.0151	−0.1001
Standard Error	(0.0169)	(0.0236)	(0.0360)
t-statistic for $a_1 = 0$	−1.331	−0.640	−2.781
4 lags			
Estimated a_1	−0.0316	−0.0522	−0.0983
Standard Error	(0.0170)	(0.0236)	(0.0388)
t-statistic for $a_1 = 0$	−1.859	−2.212	−2.533
1960–1971			
No lags			
Estimated a_1	−0.0189	−0.1137	−0.0528
Standard Error	(0.0196)	(0.0449)	(0.0286)
t-statistic for $a_1 = 0$	−0.966	−2.535	−1.846
4 lags			
Estimated a_1	−0.0294	−0.1821	−0.0509
Standard Error	(0.0198)	(0.0530)	(0.0306)
t-statistic for $a_1 = 0$	−1.468	−3.437	−1.663

cointegration holds. The final error-correction models for Japanese and U.S. price levels during the 1960–1971 period were estimated to be

$$\Delta f_t = \underset{(0.00044)}{0.00119} - \underset{(0.04184)}{0.10548}\hat{\mu}_{t-1} \tag{6.46}$$

$$\Delta p_t = \underset{(0.00033)}{0.00156} + \underset{(0.03175)}{0.01114}\hat{\mu}_{t-1} \tag{6.47}$$

where $\hat{\mu}_{t-1}$ is the lagged residual from the long-run equilibrium regression. Note that $\hat{\mu}_{t-1}$ is the estimated value of $f_{t-1} - \beta_0 - \beta_1 p_{t-1}$ and that standard errors are in parentheses.

Lag length tests (see the discussion of χ^2 and F-tests for lag length in Chapter 5) indicated that lagged values of Δf_{t-i} and Δp_{t-i} did not need to be included in the error-correction equations. Note that the point estimates in (6.46) and (6.47) indicate a direct convergence to long-run equilibrium. For example, in the presence of a one-unit deviation from long-run PPP in period $t - 1$, the Japanese price level converted into dollars falls by 0.10548 units and the U.S. price level rises by 0.01114 units. Both of these price changes in period t act to eliminate the positive discrepancy from long-run PPP present in period $t - 1$.

Notice the discrepancy between the magnitudes of the two speed of adjustment coefficients; in absolute value, the Japanese coefficient is approximately ten times that of the U.S. coefficient. As compared to the Japanese price level, the U.S. price level responded only slightly to a deviation from PPP. Moreover, the error-correction term is about 1/3 of a standard deviation from zero for the U.S. (0.01114/0.03175 = 0.3509) and approximately 2.5 standard deviations from zero for Japan (0.10548/0.4184 = 2.5210). Hence, at the 5% significance level, we can conclude that the speed of adjustment term is insignificantly different from zero for the United States but not for Japan. This result is consistent with the idea that the United States was a large country relative to Japan—movements in U.S. prices evolved independently of events in Japan, but movements in exchange rate adjusted Japanese prices responded to events in the United States.

You can update the study using the data contained on the file COINT_PPP.XLS. The file contains quarterly values of the U.K., Japanese, and Canadian wholesale prices and bilateral exchange rates with the United States. Germany is not included because the pre-unification data for Germany is not compatible with the more recent data. The file also contains the U.S. wholesale price level. Question 9 at the end of the chapter guides you through the process. The data starts in January 1973 and asks you to test for PPP by determining whether the three variables p_t, e_t and p_t^* are cointegrated.

7. CHARACTERISTIC ROOTS, RANK, AND COINTEGRATION

Although the Engle and Granger (1987) procedure is easily implemented, it does have several important defects. The estimation of the long-run equilibrium regression requires that the researcher place one variable on the left-hand side and use the others as regressors. For example, in the case of two variables, it is possible to run the Engle–Granger test for cointegration by using the residuals from either of the following two "equilibrium" regressions:

$$y_t = \beta_{10} + \beta_{11} z_t + e_{1t} \tag{6.48}$$

or

$$z_t = \beta_{20} + \beta_{21} y_t + e_{2t} \tag{6.49}$$

As the sample size grows infinitely large, asymptotic theory indicates that the test for a unit root in the $\{e_{1t}\}$ sequence becomes equivalent to the test for a unit root in the $\{e_{2t}\}$ sequence. Unfortunately, the large sample properties on which this result is derived may not be applicable to the sample sizes usually available to economists. In practice, it is possible to find that one regression indicates that the variables are cointegrated, whereas reversing the order indicates no cointegration. This is a very undesirable feature of the procedure because the test for cointegration should be invariant to the choice of the variable selected for normalization. The problem is obviously compounded using three or more variables since any of the variables can be selected as the left-hand-side variable. Moreover, in tests using three or more variables, we know that there may be more than one cointegrating vector. The method has no systematic procedure for the separate estimation of the multiple cointegrating vectors.

Another defect of the Engle–Granger procedure is that it relies on a *two-step* estimator. The first step is to generate the residual series $\{\hat{e}_t\}$, and the second step uses these generated errors to estimate a regression of the form $\Delta\hat{e}_t = a_1\hat{e}_{t-1} + \cdots$. Thus, the coefficient a_1 is obtained by estimating a regression using the residuals from another regression. Hence, any error introduced by the researcher in Step 1 is carried into Step 2. Fortunately, several methods have been developed that avoid these problems.

The Johansen (1988) and the Stock and Watson (1988) maximum likelihood estimators circumvent the use of two-step estimators *and* can estimate and test for the presence of multiple cointegrating vectors. Moreover, these tests allow the researcher to test restricted versions of the cointegrating vector(s) and the speed of adjustment parameters. Often, we want to determine whether it is possible to verify a theory by testing restrictions on the magnitudes of the estimated coefficients.

The Johansen (1988) procedure relies heavily on the relationship between the rank of a matrix and its characteristic roots. Appendix 6.1 reviews the essentials of these concepts; those of you wanting more details should review this material. For those wanting an intuitive explanation, notice that the Johansen procedure is nothing more than a multivariate generalization of the Dickey–Fuller test. In the univariate case, it is possible to view the stationarity of $\{y_t\}$ as being dependent on the magnitude of a_1; that is,

$$y_t = a_1 y_{t-1} + \varepsilon_t$$

or

$$\Delta y_t = (a_1 - 1)y_{t-1} + \varepsilon_t$$

If $(a_1 - 1) = 0$, the $\{y_t\}$ process has a unit root. Ruling out the case in which $\{y_t\}$ is explosive, if $(a_1 - 1) \neq 0$ we can conclude that the $\{y_t\}$ sequence is stationary. The Dickey–Fuller tables provide the appropriate statistics to formally test the null hypothesis $(a_1 - 1) = 0$. Now consider the simple generalization to n variables; as in (6.26), let

$$x_t = A_1 x_{t-1} + \varepsilon_t$$

so that

$$\begin{aligned} \Delta x_t &= A_1 x_{t-1} - x_{t-1} + \varepsilon_t \\ &= (A_1 - I)x_{t-1} + \varepsilon_t \\ &= \pi x_{t-1} + \varepsilon_t \end{aligned} \tag{6.50}$$

where x_t and $\varepsilon_t = (n \cdot 1)$ vectors

$A_1 =$ an $(n \cdot n)$ matrix of parameters

$I =$ an $(n \cdot n)$ identity matrix

π is defined to be $(A_1 - I)$

As indicated in the discussion surrounding (6.27), the rank of $(A_1 - I)$ equals the number of cointegrating vectors. By analogy to the univariate case, if $(A_1 - I)$ consists of all zeroes—so that $\text{rank}(\pi) = 0$—all of the $\{x_{it}\}$ sequences are unit root processes. Since there is no linear combination of the $\{x_{it}\}$ processes that is stationary, the variables are not cointegrated. If we rule out characteristic roots that are greater than unity and if $\text{rank}(\pi) = n$, (6.50) represents a convergent system of difference equations, so that all variables are stationary.

There are several ways to generalize (6.50). The equation is easily modified to allow for the presence of a drift term; simply let

$$\Delta x_t = A_0 + \pi x_{t-1} + \varepsilon_t \tag{6.51}$$

where $A_0 =$ the $(n \cdot 1)$ vector of constants $(a_{10}, a_{20}, \dots, a_{n0})'$

The effect of including the various a_{i0} is to allow for the possibility of a linear time trend in the data-generating process. You would want to include the drift term if the variables exhibited a decided tendency to increase or decrease. Here, the rank of π can be viewed as the number of cointegrating relationships existing in the "detrended" data. In the long run, $\pi x_{t-1} = 0$ so that each $\{\Delta x_{it}\}$ sequence has an expected value of a_{i0}. Aggregating all such changes over t yields the deterministic expression $a_{i0}t$.

Figure 6.3 illustrates the effects of including a drift in the data-generating process. Two random sequences with 100 observations each were generated; denote these sequences as $\{\varepsilon_{yt}\}$ and $\{\varepsilon_{zt}\}$. Initializing $y_0 = z_0 = 0$, we constructed the next 100 values of the $\{y_t\}$ and $\{z_t\}$ sequences as

$$\begin{bmatrix} \Delta y_t \\ \Delta z_t \end{bmatrix} = \begin{bmatrix} -0.2 & 0.2 \\ 0.2 & -0.2 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix}$$

so that the cointegrating relationship is

$$-0.2y_{t-1} + 0.2z_{t-1} = 0$$

or

$$y_t = z_t$$

In the top graph (a) of Figure 6.3, you can see that each sequence resembles a random walk process and that neither wanders too far from the other. The next graph (b) adds drift coefficients such that $a_{10} = a_{20} = 0.1$; now each series tends to increase by 0.1 units in each period. In addition to the fact that each sequence shares the same

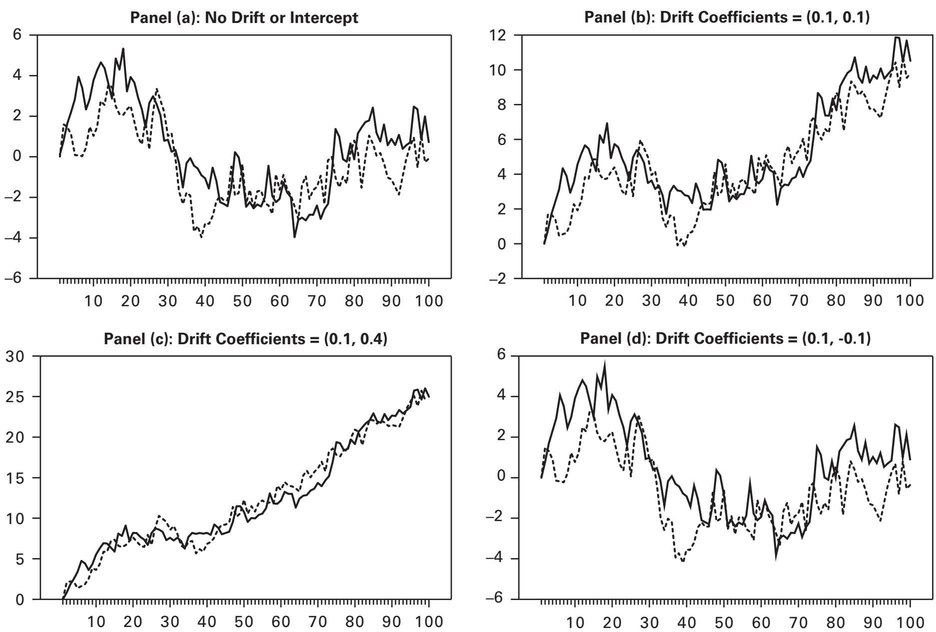


FIGURE 6.3 Drifts and Intercepts in Cointegrating Relationships

stochastic trend, note that each also has the same deterministic time trend. The fact that each has the same deterministic trend is *not* a result of the equivalence between a_{10} and a_{20}; since y_t and z_t are cointegrated, the general solution to (6.51) necessitates that each have the same linear trend. For verification, Panel (c) sets $a_{10} = 0.1$ and $a_{20} = 0.4$. Again, the sequences have the same stochastic and deterministic trends. As an aside, note that increasing a_{20} and decreasing a_{10} would have an ambiguous effect on the slope of the deterministic trend. This point will be important in a moment; by appropriately manipulating the elements of A_0 it is possible to include a constant in the cointegrating vector(s) without imparting a deterministic time trend to the system.

One way to include a constant in the cointegrating relationships is to restrict the values of the various a_{i0}. For example, if rank$(\pi) = 1$, the rows of π can differ by only a scalar, so that it is possible to write each $\{\Delta x_{it}\}$ sequence in (6.51) as

$$\begin{aligned}
\Delta x_{1t} &= \pi_{11}x_{1t-1} + \pi_{12}x_{2t-1} + \cdots + \pi_{1n}x_{nt-1} + a_{10} + \varepsilon_{1t} \\
\Delta x_{2t} &= s_2(\pi_{11}x_{1t-1} + \pi_{12}x_{2t-1} + \cdots + \pi_{1n}x_{nt-1}) + a_{20} + \varepsilon_{2t} \\
&\;\;\vdots \\
\Delta x_{nt} &= s_n(\pi_{11}x_{1t-1} + \pi_{12}x_{2t-1} + \cdots + \pi_{1n}x_{nt-1}) + a_{n0} + \varepsilon_{nt}
\end{aligned}$$

where s_i = scalars such that $s_i\pi_{1j} = \pi_{ij}$.

If the a_{i0} can be restricted such that $a_{i0} = s_i a_{10}$, it follows that all of the $\{\Delta x_{it}\}$ sequences can be written with the constant included in the cointegrating vector:

$$\begin{aligned}
\Delta x_{1t} &= (\pi_{11}x_{1t-1} + \pi_{12}x_{2t-1} + \cdots + \pi_{1n}x_{nt-1} + a_{10}) + \varepsilon_{1t} \\
\Delta x_{2t} &= s_2(\pi_{11}x_{1t-1} + \pi_{12}x_{2t-1} + \cdots + \pi_{1n}x_{nt-1} + a_{10}) + \varepsilon_{2t} \\
&\;\;\vdots \\
\Delta x_{nt} &= s_n(\pi_{11}x_{1t-1} + \pi_{12}x_{2t-1} + \cdots + \pi_{1n}x_{nt-1} + a_{10}) + \varepsilon_{nt}
\end{aligned}$$

or in compact form,

$$\Delta x_t = \pi^* x^*_{t-1} + \varepsilon_t \tag{6.52}$$

where

$$\begin{aligned}
x_t &= (x_t, x_{2t}, \ldots, x_{nt})' \\
x^*_{t-1} &= (x_{1t-1}, x_{2t-1}, \ldots, x_{nt-1}, 1)' \\
\pi^* &= \begin{bmatrix} \pi_{11} & \pi_{12} & \cdots & \pi_{1n} & a_{10} \\ \pi_{21} & \pi_{22} & \cdots & \pi_{2n} & a_{20} \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ \pi_{n1} & \pi_{n2} & \cdots & \pi_{nn} & a_{n0} \end{bmatrix}
\end{aligned}$$

The interesting feature of (6.52) is that the linear trend is purged from the system. In essence, the various a_{i0} have been altered in such a way that the general solution for each $\{x_{it}\}$ does not contain a time trend. The solution to the set of difference equations represented by (6.52) is such that all Δx_{it} are expected to equal zero when $\pi_{11}x_{1t-1} + \pi_{12}x_{2t-1} + \cdots + \pi_{1n}x_{nt-1} + a_{10} = 0$.

To highlight the difference between (6.51) and (6.52), the last graph (d) of Figure 6.3 illustrates the consequences of setting $a_{10} = 0.1$ and $a_{20} = -0.1$. You can see that neither sequence contains a deterministic trend. In fact, for the data shown in the figure, the trend will vanish so long as we select values of the drift terms maintaining the relationship $a_{10} = -a_{20.}$ (Question 1 at the end of this chapter will help you to demonstrate this result).

Some econometricians prefer to include an intercept term in the cointegrating vector along with a drift term. This makes sense if the variables contain a drift and if economic theory suggests that the cointegrating vector contains an intercept. However, it should be clear that the intercept in the cointegrating vector is not identified in the presence of a drift term. After all, some portion of the unrestricted drift can always be included in the cointegration vector. In terms of the example above, the system can always be written as

$$\begin{aligned}\Delta x_{1t} &= (\pi_{11}x_{1t-1} + \pi_{12}x_{2t-1} + \cdots + \pi_{1n}x_{nt-1} + b_{10}) + b_{11} + \varepsilon_{1t}\\ &\ \ \vdots\\ \Delta x_{nt} &= s_n(\pi_{11}x_{1t-1} + \pi_{12}x_{2t-1} + \cdots + \pi_{1n}x_{nt-1} + b_{10}) + b_{n1} + \varepsilon_{nt}\end{aligned}$$

where b_{i1} is defined to the value that satisfies $s_i b_{10} + b_{i1} = a_{10.}$

All that was done is to divide a_{10} into two parts and to place one part inside the cointegrating relationship. As such, some identification strategy is necessary since the proportion of the drift to include in the cointegrating vector is arbitrary. The popular software package EViews, for example, identifies the portion belonging in the cointegrating vector as the amount necessary to force the error-correction term to have a sample mean of zero. Nevertheless, as you can see from Figure 6.3, a drift term outside of the cointegrating relationship is necessary to capture the effects of a sustained tendency for the variables to increase (or decrease). Most researchers include drift terms if the data match Panels (b) or (c) of Figure 6.3. Otherwise, they include intercepts in the cointegrating vector or exclude the deterministic regressors altogether. If you are unsure, you can use the methods described in the next section to test whether the drifts can be appropriately restricted. Some software packages allow you to include a deterministic time trend in the model. However, it is best to avoid the use of a trend as an explanatory variable unless you have a good reason to include it in the model. Johansen (1994) discusses the role of the deterministic regressors in a cointegrating relationship.

As with the augmented Dickey–Fuller test, the multivariate model can also be generalized to allow for a higher-order autoregressive process. Consider

$$x_t = A_1 x_{t-1} + A_2 x_{t-2} + \cdots + A_p x_{t-p} + \varepsilon_t \tag{6.53}$$

where

x_t = the $(n \cdot 1)$ vector $(x_{1t}, x_{2t}, \ldots, x_{nt})'$

ε_t = an independently and identically distributed n-dimensional vector with zero mean and variance matrix Σ_ε.

Equation (6.53) can be put in a more usable form by adding and subtracting $A_p x_{t-p+1}$ to the right-hand side to obtain

$$x_t = A_1 x_{t-1} + A_2 x_{t-2} + A_3 x_{t-3} + \cdots + A_{p-2} x_{t-p+2} + (A_{p-1} + A_P) x_{t-p+1} - A_p \Delta x_{t-p+1} + \varepsilon_t$$

Next, add and subtract $(A_{p-1} + A_p)x_{t-p+2}$ to obtain

$$x_t = A_1 x_{t-1} + A_2 x_{t-2} + A_3 x_{t-3} + \cdots - (A_{p-1} + A_p)\Delta x_{t-p+2} - A_p \Delta x_{t-p+1} + \varepsilon_t$$

Just as in the augmented Dickey–Fuller test developed in Chapter 4, we can continue in this fashion to obtain

$$\Delta x_t = \pi x_{t-1} + \sum_{i=1}^{p-1} \pi_i \Delta x_{t-i} + \varepsilon_t \tag{6.54}$$

where $\pi = -\left(I - \sum_{i=1}^{p} A_i\right)$ and $\pi_i = -\sum_{j=i+1}^{p} A_j$

Again, the key feature to note in (6.54) is rank of the matrix π; the rank of π is equal to the number of independent cointegrating vectors. Clearly, if rank(π) = 0, the matrix is null and (6.54) is the usual VAR model in first differences. Instead, if π is of rank n, the vector process is stationary. In intermediate cases, if rank(π) = 1, there is a single cointegrating vector and the expression πx_{t-1} is the error-correction term. For other cases in which $1 < \text{rank}(\pi) < n$, there are multiple cointegrating vectors.

As detailed in Appendix 6.1, the number of distinct cointegrating vectors can be obtained by checking the significance of the characteristic roots of π. We know that the rank of a matrix is equal to the number of its characteristic roots that differ from zero. Suppose we obtained the matrix π and ordered the n characteristic roots such that $\lambda_1 > \lambda_2 > \cdots > \lambda_n$. If the variables in x_t are *not* cointegrated, the rank of π is zero and all of these characteristic roots will equal zero. Since ln(1) = 0, each of the expressions $\ln(1 - \lambda_i)$ will equal zero if the variables are not cointegrated. Similarly, if the rank of π is unity, $0 < \lambda_1 < 1$ so the first expression $\ln(1 - \lambda_1)$ will be negative and all the other $\lambda_i = 0$ so that $\ln(1 - \lambda_2) = \ln(1 - \lambda_3) = \cdots = \ln(1 - \lambda_n) = 0$.

In practice, we can obtain only estimates of π and its characteristic roots. The test for the number of characteristic roots that are insignificantly different from unity can be conducted using the following two test statistics:

$$\lambda_{\text{trace}}(r) = -T \sum_{i=r+1}^{n} \ln(1 - \hat{\lambda}_i) \tag{6.55}$$

$$\lambda_{\max}(r, r+1) = -T \ln(1 - \hat{\lambda}_{r+1}) \tag{6.56}$$

where $\hat{\lambda}_i$ = the estimated values of the characteristic roots (also called eigenvalues) obtained from the estimated π matrix

T = the number of usable observations

When the appropriate values of r are clear, these statistics are simply referred to as λ_{trace} and $\lambda_{\max}$.

The first statistic tests the null hypothesis that the number of distinct cointegrating vectors is less than or equal to r against a general alternative. From the previous discussion, it should be clear that λ_{trace} equals zero when all $\lambda_i = 0$. The further the estimated characteristic roots are from zero, the more negative is $\ln(1 - \hat{\lambda}_i)$ and the larger is the λ_{trace} statistic. The second statistic tests the null that the number of cointegrating vectors is r against the alternative of $r + 1$ cointegrating vectors. Again, if the estimated value of the characteristic root is close to zero, $\lambda_{\max}$ will be small.

Critical values of the λ_{trace} and the $\lambda_{\max}$ statistics are obtained using the Monte Carlo approach. The critical values are reproduced in Table E in the *Supplementary Manual*. The distribution of these statistics depends on two things:

1. The number of nonstationary components under the null hypothesis (i.e., $n - r$).
2. The form of the vector A_0. Use the top portion of Table E if you do not include either a constant in the cointegrating vector or a drift term. Use the middle portion of the table if you include a drift term A_0. Use the bottom portion of the table if you include a constant in the cointegrating vector.

Using quarterly data for Denmark over the sample period 1974:1 to 1987:3, Johansen and Juselius (1990) let the x_t vector be represented by

$$x_t = (m2_t, y_t, i_t^d, i_t^b)'$$

where $m2$ = log of the real money supply as measured by $M2$ deflated by a price index

y = log of real income

i^d = deposit rate on money representing a direct return on money holding

i^b = bond rate representing the opportunity cost of holding money

Including a constant in the cointegrating relationship (i.e., augmenting x_{t-1} with a constant), they report that the residuals from (6.54) appear to be serially uncorrelated. If we round off to two decimal places, the four characteristic roots of the estimated π matrix are given in the first column below:

	$\lambda_{\max}$ $-T\ \ln(1 - \hat{\lambda}_{r+1})$	λ_{trace} $-T\,\Sigma \ln(1 - \hat{\lambda}_i)$
$\hat{\lambda}_1 = 0.4332$	30.09	49.14
$\hat{\lambda}_2 = 0.1776$	10.36	19.05
$\hat{\lambda}_3 = 0.1128$	6.34	8.69
$\hat{\lambda}_4 = 0.0434$	2.35	2.35

The second column reports the various $\lambda_{\max}$ statistics as the number of usable observations ($T = 53$) multiplied by $\ln(1 - \hat{\lambda}_{r+1})$. For example, $-53\ \ln(1 - 0.0434) = 2.35$ and $-53\ \ln(1 - 0.1128) = 6.34$. The last column reports the λ_{trace} statistics as the summation of the $\lambda_{\max}$ statistics. Simple arithmetic reveals that $8.69 = 2.35 + 6.34$ and $19.05 = 2.35 + 6.34 + 10.36$.

To test the null hypothesis $r = 0$ against the general alternative $r = 1, 2, 3$, or 4, use the λ_{trace} statistic. Since the null hypothesis is $r = 0$ and there are four variables (i.e., $n = 4$), the summation in (6.55) runs from 1 to 4. If we sum over the four values, the calculated value of λ_{trace} is 49.14. Since Johansen and Juselius (1990) include the constant in the cointegrating vector, this calculated value of 49.14 is compared to the critical values reported in the bottom portion of Table E. For $n - r = 4$, the critical values of λ_{trace} are 49.65, 53.12, and 60.16 at the 10, 5, and 1% significance levels, respectively. Thus, at the 10% level, the restriction is *not* binding, so that the variables are *not* cointegrated using this test.

To make a point and to give you practice in using the table, suppose you want to test the null hypothesis $r \leq 1$ against the alternative $r = 2, 3$, or 4. Under this null hypothesis, the summation in (6.55) runs from 2 to 4 so that the calculated value of λ_{trace} is 19.05. For $n - r = 3$, the critical values of λ_{trace} are 32.00, 34.91, and 41.07 at the 10, 5, and 1% significance levels, respectively. The restriction $r = 0$ or $r = 1$ is not binding.

In contrast to the λ_{trace} statistic, the λ_{max} statistic has a specific alternative hypothesis. To test the null hypothesis $r = 0$ against the specific alternative $r = 1$, use equation (6.56). The calculated value of the $\lambda_{max}(0, 1)$ statistic is $-53\ \ln(1 - 0.4332) = 30.09$. For $n - r = 4$, the critical values of λ_{max} are 25.56, 28.14, 30.32, and 33.24 at the 10, 5, 2.5, and 1% significance levels, respectively. Hence, it is possible to reject the null hypothesis $r = 0$ at the 5% significance level (but not the 2.5% level) and conclude that there is only one cointegrating vector (i.e., $r = 1$). Before reading on, you should take a moment to examine the data and convince yourself that the null hypothesis $r = 1$ against the alternative $r = 2$ cannot be rejected at conventional levels. You should find that the calculated value of the λ_{max} statistic for $r = 1$ is 10.36 and that the critical value at the 10% level is 19.77. Hence, there is no significant evidence of more than one cointegrating vector.

The example illustrates the important point that the results of the λ_{trace} and λ_{max} tests can conflict. The λ_{max} test has the sharper alternative hypothesis. It is usually preferred for trying to pin down the number of cointegrating vectors.

8. HYPOTHESIS TESTING

In the Dickey–Fuller tests discussed in Chapter 4, it was important to correctly ascertain the form of the deterministic regressors. A similar situation applies in the Johansen procedure. As you can see in Table E, the critical values of the λ_{trace} and λ_{max} statistics are smallest without any deterministic regressors and largest with an intercept term included in the cointegrating vector. Instead of cavalierly positing the form of A_0, it is possible to test restricted forms of the vector.

One of the most interesting aspects of the Johansen procedure is that it allows for testing restricted forms of the cointegrating vector(s). In a money demand study, you might want to test restrictions concerning the long-run proportionality between money and prices, or the sizes of the income and interest rate elasticities of demand for money. In terms of equation (6.1) (i.e., $m_t = \beta_0 + \beta_1 p_t + \beta_2 y_t + \beta_3 r_t + e_t$), the restrictions of interest are $\beta_1 = 1, \beta_2 > 0$, and $\beta_3 < 0$.

The key insight to all such hypothesis tests is that *if there are r cointegrating vectors, only these r linear combinations of the variables are stationary*. All other linear combinations are nonstationary. Thus, suppose you reestimate the model restricting the parameters of π. If the restrictions are not binding, you should find that the number of cointegrating vectors has *not* diminished.

To test for the presence of an intercept in the cointegrating vector as opposed to the unrestricted drift A_0, estimate the two forms of the model. Denote the ordered characteristic roots of the unrestricted π matrix by $\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_n$ and the characteristic roots of the model with the intercept(s) in the cointegrating vector(s) by $\hat{\lambda}_1^*, \hat{\lambda}_2^*, \ldots, \hat{\lambda}_n^*$. Suppose that the unrestricted form of the model has r nonzero characteristic roots. Asymptotically, the statistic

$$-T \sum_{i=r+1}^{n} [\ln(1-\hat{\lambda}_i^*) - \ln(1-\hat{\lambda}_i)] \tag{6.57}$$

has a χ^2 distribution with $(n-r)$ degrees of freedom.

The intuition behind the test is that all values of $\ln(1-\hat{\lambda}_i^*)$ and $\ln(1-\hat{\lambda}_i)$ should be equivalent if the restriction is not binding. Hence, small values for the test statistic imply that it is permissible to include the intercept in the cointegrating vector. However, the likelihood of finding a stationary linear combination of the n variables is greater with the intercept in the cointegrating vector than if the intercept is absent from the cointegrating vector. Thus, a large value of $\hat{\lambda}_{r+1}^*$ [and a corresponding large value of $-T\ \ln(1-\hat{\lambda}_{r+1}^*)$], implies that the restriction artificially inflates the number of cointegrating vectors. Thus, as proven by Johansen (1991), if the test statistic is sufficiently large, it is possible to reject the null hypothesis of an intercept in the cointegrating vector(s) and conclude that there is a linear trend in the variables. This is precisely the case represented by the middle portion of Figure 6.3.

Johansen and Juselius (1990) test the restriction that their estimated Danish money demand function does not have a drift. Since they found only one cointegrating vector among $m2$, y, i^d, and i^b, set $n = 4$ and $r = 1$. The calculated value of the χ^2 statistic in (6.57) is 1.99. With three degrees of freedom, this is insignificant at conventional levels; they conclude that the variables do not have a linear time trend and find it appropriate to include the constant in the cointegrating vector.

In order to test other restrictions on the cointegrating vector, Johansen defines the two matrices α and β, both of dimension $(n \cdot r)$ where r is the rank of π. The properties of α and β are such that

$$\pi = \alpha\beta'$$

Note that β is the matrix of cointegrating parameters and α is the matrix of weights with which each cointegrating vector enters the n equations of the VAR. In a sense, α can be viewed as the matrix of the speed of adjustment parameters. Due to the cross-equation restrictions, it is not possible to estimate α and β using OLS.[4] However, using maximum likelihood estimation, it is possible to (1) estimate (6.54) as an error-correction model, (2) determine the rank of π, (3) use the r most significant cointegrating vectors to form β', and (4) select α such that $\pi = \alpha\beta'$. Question 5 at the end of this chapter asks you to find several such α and β' matrices.

It is easy to understand the process in the case of a single cointegrating vector. Given that rank(π) = 1, the rows of π are all linear multiples of each other. Hence, the equations in (6.54) have the form

$$\begin{aligned}
\Delta x_{1t} &= \pi_{11}x_{1t-1} + \pi_{12}x_{2t-1} + \cdots + \pi_{1n}x_{nt-1} + \cdots + \varepsilon_{1t} \\
\Delta x_{2t} &= s_2(\pi_{11}x_{1t-1} + \pi_{12}x_{2t-1} + \cdots + \pi_{1n}x_{nt-1}) + \cdots + \varepsilon_{2t} \\
&\vdots \\
\Delta x_{nt} &= s_n(\pi_{11}x_{1t-1} + \pi_{12}x_{2t-1} + \cdots + \pi_{1n}x_{nt-1}) + \cdots + \varepsilon_{nt}
\end{aligned}$$

where the s_i are scalars and, for notational simplicity, the matrices $\pi_i\Delta x_{t-i}$ have not been written out.

Now define $\alpha_i = s_i\pi_{11}$ and $\beta_i = \pi_{1i}/\pi_{11}$ so that each equation can be written as

$$\Delta x_{it} = \alpha_i(x_{1t-1} + \beta_2 x_{2t-1} + \cdots + \beta_n x_{nt-1}) + \cdots + \varepsilon_{it} \quad (i = 1, \ldots, n)$$

or in matrix form,

$$\Delta x_t = \sum_{i=1}^{p-1} \pi_i \Delta x_{t-i} + \alpha\beta' x_{t-1} + \varepsilon_t \tag{6.58}$$

where the single cointegrating vector is $\beta = (1, \beta_2, \beta_3, \ldots, \beta_n)'$ and the speed of adjustment parameters are given by $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)'$.

Once α and β' are determined, testing various restrictions on α and β' is straightforward if you remember the fundamental point that if there are r cointegrating vectors, only these r linear combinations of the variables are stationary. Thus, the test statistics involve comparing the number of cointegrating vectors under the null and alternative hypotheses. Again, let $\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_n$ and $\hat{\lambda}_1^*, \hat{\lambda}_2^*, \ldots, \hat{\lambda}_n^*$ denote the ordered characteristic roots of the unrestricted and restricted models, respectively. To test restrictions on β, form the test statistic

$$T\sum_{i=1}^{r}[\ln(1 - \hat{\lambda}_i^*) - \ln(1 - \hat{\lambda}_i)] \tag{6.59}$$

Asymptotically, this statistic has a χ^2 distribution with degrees of freedom equal to the number of restrictions placed on β. Small values of $\hat{\lambda}_i^*$ relative to $\hat{\lambda}_i$ (for $i \leq r$) imply a reduced number of cointegrating vectors. Hence, the restriction embedded in the null hypothesis is binding if the calculated value of the test statistic exceeds that in a χ^2 table. For example, Johansen and Juselius test the restriction that money and income move proportionally. Their estimated long-run equilibrium relationship is

$$m2_t = 1.03y_t - 5.21i_t^b + 4.22i_t^d + 6.06$$

They restrict the coefficient of income to be unity and find the restricted values of the $\hat{\lambda}_i^*$ to be such that

	$\hat{\lambda}_i^*$	$T\ln(1 - \hat{\lambda}_i^*)$
$i = 1$	0.433	−30.04
$i = 2$	0.172	−10.01
$i = 3$	0.044	−2.36
$i = 4$	0.006	−0.32

Given that the unrestricted model has $r = 1$ and $-T\ln(1 - \hat{\lambda}_1) = 30.09$, (6.59) becomes $-30.04 + 30.09 = 0.05$. Since there is only 1 restriction imposed on β, the test statistic has a χ^2 distribution with 1 degree of freedom. A χ^2 table indicates that 0.05 is not significant; hence, they conclude that the restriction is not binding.

Restrictions on α can be tested in the same way. The procedure is to restrict α and compare the r most significant characteristic roots for the restricted and unrestricted models using (6.59). If the calculated value of (6.59) exceeds that from a χ^2 table, with degrees of freedom equal to the number of restrictions placed on α, the restrictions can be rejected. For example, Johansen and Juselius (1990) test the restriction that only money demand (i.e., $m2_t$) responds to the deviation from long-run equilibrium. Formally, they test the restriction that $\alpha_2 = \alpha_3 = \alpha_4 = 0$. Restricting the three values of α_i to equal zero, they find the largest characteristic root in the restricted model is such that $T\ln(1 - \hat{\lambda}_1^*) = -23.42$. Since the unrestricted model is such that $T\ln(1 - \hat{\lambda}_1) = -30.09$, equation (6.59) becomes $-23.42 - (-30.09) = 7.67$. The χ^2 statistic with three degrees of freedom is 7.81 at the 5% significance level. Hence, they find mild support for the hypothesis that the restriction is not binding.

If there is a single cointegrating vector, the Engle–Granger and Johansen methods have the same asymptotic distribution. If it can be determined that only one cointegrating vector exists, it is also common to rely on the estimated error-correction model to test restrictions on α. If $r = 1$, and a single value of α is being tested, the usual t-statistic is asymptotically equivalent to the Johansen test.

Lag Length and Causality Tests

The simplest way to understand lag length tests is to consider the system in the form of (6.54)

$$\Delta x_t = \pi x_{t-1} + \sum_{i=1}^{p-1} \pi_i \Delta x_{t-i} + \varepsilon_t$$

Regardless of the rank of π, all of the Δx_{t-i} are stationary variables. Hence, we can use Rule 1 of Sims, Stock, and Watson (1990). Recall that the rule implies that the coefficients of interest on zero-mean stationary variables can be tested using a normal distribution. Since lag length depends solely on the values of the various π_i, a χ^2 distribution is appropriate to test any restriction concerning lag length. As in the case of any VAR, let Σ_u and Σ_r be the variance/covariance matrices of the unrestricted and restricted systems, respectively. As in Chapter 5, let c denote the maximum number of regressors contained in the longest equation. The test statistic

$$(T - c)(\log|\Sigma_r| - \log|\Sigma_u|)$$

can be compared to a χ^2 distribution with degrees of freedom equal to the number of restrictions in the system. Alternatively, you can use the multivariate AIC or SBC to determine the lag length. If you want to test the lag lengths for a single equation, an F-test is appropriate.

The rule also means that you cannot perform Granger causality tests in a cointegrated system using a standard F-test. First, suppose that rank $(\pi) = 0$ so that

$$\Delta x_t = \sum_{i=1}^{p-1} \pi_i \Delta x_{t-i} + \varepsilon_t$$

As such, Granger causality involves only stationary variables. Yet, this was precisely the case discussed in Chapter 5 when the variables in a VAR are not cointegrated. Hence, Granger causality tests can be conducted using a standard F distribution. However, if the variables are cointegrated, a Granger causality test involves the coefficients of π. Since these coefficients multiply nonstationary variables, it is not appropriate to use an F-statistic to test for Granger causality. After all, if rank$(\pi) \neq 0$, it is impossible to write the restrictions of the test as restrictions on a set of $I(0)$ variables. Block exogeneity tests are also ruled out too. If w_t is cointegrated with y_t or z_t, you cannot use a standard χ^2 test to determine whether w_t belongs in the equations for y_t and z_t.

To Difference or Not to Difference

We have reached a point where it is possible to address the issue of differencing the nonstationary variables in an unrestricted VAR. There is no question that differencing leads to a misspecification error if the variables are cointegrated. Suppose that the actual data-generating process is given by the cointegrated system of (6.54) but you estimate the following VAR in first differences:

$$\Delta x_t = \sum_{i=1}^{p-1} \pi_i \Delta x_{t-i} + \varepsilon_t$$

The system is misspecified since it excludes the long-run equilibrium relationships among the variables that are contained in πx_{t-1}. Given the misspecification error, all of the coefficient estimates, t-tests, F-tests, tests of cross-equation restrictions, impulse responses and variance decompositions are not representative of the true process. Hence, there is a substantial penalty to pay if you estimate a VAR in first differences when the data are actually cointegrated; differencing "throws away" information contained in the cointegrating relationship(s).

Why not simply estimate all VARs in levels? The answer is that it is preferable to use the first differences if the $I(1)$ variables are not cointegrated. There are three consequences if the $I(1)$ variables are not cointegrated and you estimate the VAR in levels:

1. Tests lose power because you estimate n^2 more parameters (one extra lag of each variable in each equation).
2. For a VAR in levels, tests for Granger causality conducted on the $I(1)$ variables do not have a standard F distribution. If you use first differences, you can use the standard F distribution to test for Granger causality.
3. When the VAR has $I(1)$ variables, the impulse responses at long forecast horizons are inconsistent estimates of the true responses. Since the impulse responses need not decay, any imprecision in the coefficient estimates will have a permanent effect on the impulse responses. If the VAR is estimated

in first differences, the impulse responses decay to zero and so the estimated responses are consistent.

The suggestion is that it is important to properly determine whether the $I(1)$ variables are cointegrated. You can perform lag length tests regardless of whether the variables are cointegrated. As such, the suggested methodology is to estimate an unrestricted VAR. Most researchers would begin with a lag length of approximately $T^{1/3}$. You may want to alter the number of lags to correspond to the seasonal frequency of the data. For example, with 100 observations of two variables using quarterly data, you might want to begin with 8 lags even though $T^{1/3}$ is approximately five. Select the most appropriate lag length and then perform a cointegration test. If the variables are not cointegrated, estimate the system in first differences. If the variables are cointegrated, you can work with the error-correction model. Since the error-correction term and all values of Δx_{t-i} are stationary, you can conduct inference on any variable (except those appearing within the cointegrating vectors) using the usual test statistics. Impulse responses and variance decompositions will yield consistent estimates of the actual values.

Tests on Multiple Cointegrating Vectors

If the rank of π exceeds one, it is not straightforward to interpret the cointegrating vectors. When there are multiple cointegrating vectors, any linear combination of these vectors is also a cointegrating vector. Fortunately, it is often possible to identify separate behavioral relationships by appropriately restricting the individual cointegrating vectors. The only complication is that you need to be clear about the number of restrictions you impose on the system. It is important to note that *if there are r cointegration relationships in an n-variable system, there exists a cointegrating vector for each subset of* $(n - r + 1)$ *variables.* For example, if there are two cointegrating vectors in a three-variable system, there is a cointegrating vector for each bilateral pair of the variables $(2 = n - r + 1)$. To demonstrate the point, let $x_t = (x_{1t}, x_{2t}, x_{3t}, x_{4t})'$ and suppose there are two cointegrating vectors for these four variables. If we normalize each vector with respect to x_{1t}, we can write the two independent relationship in $\beta' x_t = 0$ as

$$\begin{bmatrix} 1 & -\beta_{12} & -\beta_{13} & -\beta_{14} \\ 1 & -\beta_{22} & -\beta_{23} & -\beta_{24} \end{bmatrix} \begin{bmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \\ x_{4t} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Consider the $2 \cdot n$ matrix β' consisting of the cointegrating parameters. Subtract row 1 from row 2 to obtain

$$\begin{bmatrix} 1 & -\beta_{12} & -\beta_{13} & -\beta_{14} \\ 0 & -\beta_{22} + \beta_{12} & -\beta_{23} + \beta_{13} & -\beta_{24} + \beta_{14} \end{bmatrix}$$

Now, renormalize row 2 by dividing each of its elements by $(\beta_{12} - \beta_{22})$ to obtain

$$\begin{bmatrix} 1 & -\beta_{12} & -\beta_{13} & -\beta_{14} \\ 0 & 1 & -\beta_{23}^{*} & -\beta_{24}^{*} \end{bmatrix}$$

where $-\beta_{23}^* = (\beta_{13} - \beta_{23})/(\beta_{12} - \beta_{22})$ and $-\beta_{24}^* = (\beta_{14} - \beta_{24})/(\beta_{12} - \beta_{22})$. Hence, x_{2t}, x_{3t}, and x_{4t} are cointegrated such that $x_{2t} = \beta_{23}^* x_{3t} + \beta_{24}^* x_{4t}$. Similarly, add β_{12} times row 2 to row 1 to obtain

$$\begin{bmatrix} 1 & 0 & -\beta_{13}^* & -\beta_{14}^* \\ 0 & 1 & -\beta_{23}^* & -\beta_{24}^* \end{bmatrix}$$

where $\beta_{1j}^* = \beta_{1j} + \beta_{12}\beta_{2j}^*$

Thus, x_1, x_3, and x_4 are cointegrated such that $x_{1t} = \beta_{13}^* x_{3t} + \beta_{14}^* x_{4t}$. Since the labeling of the variables is irrelevant, it follows that there exists a cointegrating vector for each subset of three variables. More generally, β' will be an $r \cdot n$ matrix of cointegrating parameters, and each subset of $n - r + 1$ variables will be cointegrated. From the preceding discussion, it should be clear that standard row and column operations on β' *do not* entail restrictions on the cointegrating vectors. Such operations merely result in additional cointegrating vectors that are linear combinations of the original vectors.

EXAMPLE 1: VARIABLE EXCLUSION WITHIN AN EQUATION With multiple cointegrating vectors, you cannot test whether any one particular $\beta_{ij} = 0$ since this assumption does not restrict the cointegrating space. In the general case where β' is an $r \cdot n$ matrix, a testable exclusion restriction entails the exclusion of r or more variables from a cointegrating vector. Hence, excluding r variables from a cointegrating vector entails only one restriction. If the sample value of the χ^2 statistic with one degree of freedom (since there is only one restriction involved) exceeds a critical value, reject the null hypothesis that this set of variables contains a cointegrating relationship.

EXAMPLE 2: VARIABLE EXCLUSION ACROSS EQUATIONS Next, suppose that you want to test whether x_{4t} can be excluded from the set of cointegrating relationships. The restriction $\beta_{14} = \beta_{24} = 0$ entails only one restriction on the cointegrating space. In the general case where β' is an $r \cdot n$ matrix, the test $\beta_{1j} = \beta_{2j} = \cdots = \beta_{rj} = 0$ still involves only one restriction. This follows since x_{it} can be eliminated from $r - 1$ equations using simple row and column operations.

EXAMPLE 3: CONDITIONAL RESTRICTIONS It is also possible to restrict one cointegrating vector conditional on the values of all other cointegrating vectors. For example, you might want to determine if $(1, 0, \beta_{23}, \beta_{24})'$ is a cointegrating vector for the given normalized values of β_{12}, β_{13}, and β_{14}. Thus, you fix the values of β_{12}, β_{13}, and β_{14} and determine whether you can exclude x_{2t} from the second vector. Cutler, Davis, and Smith (1999) consider the identification issue in considerable detail. They examine the following four behavioral relationships in a seven variable system:

$$\begin{aligned} m_t &= d_0 + d_1 y_t + d_2 r_t + d_3 p_t + e_{1t} \\ c_t &= a_0 + a_1 y_t + a_2 r_t + e_{2t} \\ i_t &= b_0 + b_1 y_t + b_2 r_t + e_{3t} \\ im_t &= g_0 + g_1 y_t + g_2 r_t + e_{4t} \end{aligned}$$

where m_t = log of nominal money holdings
y_t = log of real income
r_t = real interest rate
c_t = log of real consumption
i_t = log of real investment
p_t = log of the price level
im_t = log of real imports
e_{1t}, e_{2t}, e_{3t}, and e_{4t} = stationary error terms

The first equation is the money demand equation. The next three equations are a simple consumption function, an investment function, and an import demand function, respectively. Consumption, investment, and imports are each assumed to be functions of only income and the interest rate. The issue is to determine whether it is possible to identify these four equations from a seven-variable system. Toward this end, they obtained estimates of a 7×7 π matrix over a number of sample periods. There were at least four cointegrating vectors in every case considered. Over the entire sample, 1960*Q*2 to 1990*Q*4, Cutler, Davis, and Smith (1999) found that they could not reject the restrictions at conventional significance levels (the *prob*-value was 16%).

The Test in the Presence of *I*(2) Variables

It is also possible to test for multicointegration using Johansen's methodology. Consider the VAR system:

$$\Delta^2 x_t = \pi x_{t-1} + \Gamma \Delta x_{t-1} + \sum_{i=1}^{p-2} \pi_i \Delta^2 x_{t-i} + \varepsilon_t \qquad (6.60)$$

The issue of multicointegration concerns the ranks of both π and Γ. In principle, it is possible to consider all possible orders of cointegration for the variables in the system. However, to illustrate the procedure, it is useful to begin with a three-variable system consisting of the three $I(2)$ variables x_{1t}, x_{2t}, and x_{3t} that are multicointegrated such that

$$\pi_{11}x_{1t} + \pi_{12}x_{2t} + \pi_{13}x_{3t} + \Gamma_{11}\Delta x_{1t} + \Gamma_{12}\Delta x_{2t} + \Gamma_{13}\Delta x_{3t} = 0$$

Let r denote the rank of π and r_1 denote the rank of Γ so that (6.60) is such that $r = r_1 = 1$. Clearly, if $r = 0$, multicointegration fails since there is no linear combination of the three $I(2)$ variables that forms an equilibrium relationship. If $r = 1$ and $r_1 = 0$, the equilibrium relationship has the form $\pi_{11}x_{1t} + \pi_{12}x_{2t} + \pi_{13}x_{3t} = 0$. As such, $\Delta^2 x_t = \pi x_{t-1} + I(0)$ variables so that $\pi_{11}x_{1t} + \pi_{12}x_{2t} + \pi_{13}x_{3t}$ is necessarily a stationary relationship—the variables are $CI(2, 2)$. All of this may seem straightforward, but there is a complicating factor when the ranks of π and Γ have to be estimated. To illustrate the point, suppose that the $I(2)$ variables are cointegrated such that

$$\pi_{11}x_{1t} + \pi_{12}x_{2t} + \pi_{13}x_{3t} \sim I(1)$$

where $\sim I(d)$ indicates the order of integration.

If you take the first difference, it follows that $\pi_{11}\Delta x_{1t} + \pi_{12}\Delta x_{2t} + \pi_{13}\Delta x_{3t}$ is $I(0)$. You should be able to figure out the problem. For any cointegrating vector in π, it is possible to estimate an identical cointegration vector for the first differences of the variables. Yet a linear combination of the two relationships is not stationary. Consider the result obtained by subtracting the $I(0)$ relationship from the $I(1)$ relationship:

$$\pi_{11}x_{1t} + \pi_{12}x_{2t} + \pi_{13}x_{3t} - \pi_{11}\Delta x_{1t} - \pi_{12}\Delta x_{2t} - \pi_{13}\Delta x_{3t}$$
$$= \pi_{11}x_{1t-1} + \pi_{12}x_{2t-1} + \pi_{13}x_{3t-1}$$

Since $\pi_{11}x_{1t-1} + \pi_{12}x_{2t-1} + \pi_{13}x_{3t-1}$ is $I(1)$, all that has been done is to change the time subscript for the variables in the cointegrating relationship. The point is that it is necessary to find cointegrating vectors in Γ that are not linear combinations of those in π.

If we take the more general case considered by Johansen (1995), let rank(π) = r and let s denote the number of cointegrating vectors in Γ that are orthogonal to those in π. In an n-variable system such that some of the variables are $I(2)$, you should be able to verify that:

1. If $r = 0$, there is no relationship among the variables that is stationary.
2. In a system with n variables, if $r + s = n - 1$, there is a unique multicointegrating vector. The number of $I(2)$ stochastic trends in an n-variable system is given by $n - r - s$.
3. The value of s must be such that $s < n - r$. For the analysis of $I(2)$ variables to be appropriate, the values of r and s must be such that $s + r < n$. If $s = n - r$, then x_t contains no $I(2)$ variables.

Johansen's cointegration test with $I(2)$ variables is actually a two-step procedure. In the first step, you estimate a model as in (6.60) to determine the rank of π. Determine the value of r using the λ_{trace} and λ_{max} statistics in the usual way. In the second step, you determine the value of s conditional on the value of r.[5] Let the null hypothesis be $s = s_0$ and consider

$$Q^*_{r,s} = -T\sum_{i=s_0+1}^{n}\ln(1 - \hat{\lambda}_i) \tag{6.61}$$

Hence, $Q^*_{r,s}$ is constructed in the same fashion as a λ_{trace} statistic. The principal differences are that you test the rank of Γ conditional on the value of r and that you obtain the number of cointegrating vectors orthogonal to those in π. As such, the critical values needed to determine the value of s have to be modified. Given the value of r, if the sample value of $Q^*_{r,s}$ exceeds the critical value calculated by Johansen, reject the null hypothesis $s = s_0$ in favor of the alternative $s > s_0$. For $r = 1$, the critical values at the 10, 5, and 1% significance levels are

Critical Values for $Q^*_{1,s}$

	$s = 0$	$s = 1$
10%	31.88	17.79
5%	34.80	19.99
1%	40.84	24.74

For example, let $r = 1$ and suppose that the sample value of $Q^*_{1,s}$ is found to be 35.00. As such, the null hypothesis $s = 0$ can be rejected at the 5% significance level.

9. ILLUSTRATING THE JOHANSEN METHODOLOGY

An interesting way to illustrate the Johansen methodology is to use exactly the same data shown in Figure 6.2. Recall that the data is contained in the file COINT6.XLS. Although the Engle–Granger technique did find that the simulated data were cointegrated, a comparison of the two procedures is useful. Use the following four steps when implementing the Johansen procedure.

STEP 1: It is good practice to pretest all variables to assess their order of integration. Plot the data to see if a linear time trend is likely to be present in the data-generating process. In most instances you will have variables that are integrated of the same order. In other cases, you can check for multicointegration.

The results of the test can be quite sensitive to the lag length, so it is important to be careful. The most common procedure is to estimate a vector autoregression using the *undifferenced* data. Then use the same lag-length tests as in a traditional VAR. Begin with the longest lag length deemed reasonable and test whether it can be shortened. For example, if we want to test whether lags 2 through 4 are important, we can estimate the following two VARs:

$$x_t = A_0 + A_1 x_{t-1} + A_2 x_{t-2} + A_3 x_{t-3} + A_4 x_{t-4} + e_{1t}$$
$$x_t = A_0 + A_1 x_{t-1} + e_{2t}$$

where x_t = the $(n \cdot 1)$ vector of variables
$A_0 = (n \cdot 1)$ matrix of intercept terms
$A_i = (n \cdot n)$ matrices of coefficients
e_{1t} and $e_{2t} = (n \cdot 1)$ vectors of error terms.

Estimate the first system with four lags of each variable in each equation and call the variance/covariance matrix of residuals Σ_4. Now estimate the second equation using only one lag of each variable in each equation and call the variance/covariance matrix of residuals Σ_1. Even though we are working with nonstationary variables, we can perform lag length tests using the likelihood ratio test statistic recommended by Sims (1980):

$$(T - c)(\log|\Sigma_1| - \log|\Sigma_4|)$$

where T = number of observations
c = number of parameters in the unrestricted system
$\log|\Sigma_i|$ = natural logarithm of the determinant of Σ_i

Following Sims, use the χ^2 distribution with degrees of freedom equal to the number of coefficient restrictions. Since each A_i has n^2 coefficients, constraining $A_2 = A_3 = A_4 = 0$ entails $3n^2$ restrictions. Alternatively, you can select lag length p using the multivariate generalizations of the AIC or SBC. In the model at hand, you should find that the general-to-specific method and the AIC select a lag length of 2 whereas the SBC selects a lag length of 1.

STEP 2: Estimate the model and determine the rank of π. Many time-series statistical software packages contain a routine to estimate the model. Here, it suffices to say that OLS is not appropriate because it is necessary to impose cross-equation restrictions on the π matrix. In most circumstances, you may choose to estimate the model in three forms: (1) with all elements of A_0 set equal to zero, (2) with a drift, or (3) with a constant term in the cointegrating vector.

For example, we can use the simulated data shown in Figure 6.2 so that $x_t = (y_t, z_t, w_t)'$. If we pretend that we do not know the form of the data-generating process, we might want to include an intercept term in the cointegrating vector(s). As we saw in the last section, it is possible to test for the presence of the intercept. If we follow the general-to-specific method and use a lag length of 2, the estimated model has the form

$$\Delta x_t = A_0 + \pi x_{t-1} + \pi_1 \Delta x_{t-1} + \varepsilon_t \tag{6.62}$$

where A_0 was constrained so as to force the intercept to appear in the cointegrating vector.

As always, carefully analyze the properties of the residuals of the estimated model. Any evidence that the errors are not white noise usually means that lag lengths are too short. Figure 6.4 shows deviations of y_t from the long-run relationship ($\mu_t = -0.01331 - y_t - 1.0350z_t + 1.0162w_t$) and one of the error sequences (i.e., the $\{\varepsilon_{yt}\}$ sequence that equals the residuals from the y_t equation in (6.62)). Both sequences conform to their theoretical properties in that the residuals from the long-run equilibrium appear to be stationary and the estimated values of the $\{\varepsilon_{yt}\}$ series approximate a white-noise process.

The estimated values of the characteristic roots of the π matrix in (6.62) are

$$\lambda_1 = 0.32600;\ \lambda_2 = 0.14032;\ \text{and}\ \lambda_3 = 0.033168$$

Since $T = 98$ (100 observations less the two lost as a result of using 2 lags), the calculated values of λ_{max} and λ_{trace} for the various possible values of r are reported in the center column of Table 6.6.

Consider the hypothesis that the variables are not cointegrated (so that the rank $\pi = 0$). Depending on the alternative hypothesis, we have a choice of two possible test statistics. If we are interested in the hypothesis that the variables are not cointegrated ($r = 0$) against the alternative of one or more

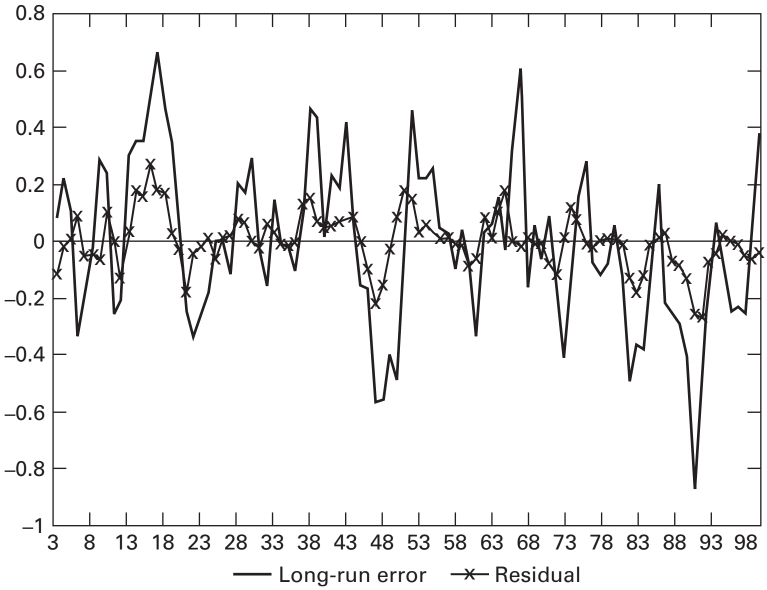


FIGURE 6.4 Long-Run and Short-Run Errors

Table 6.6 The $\lambda_{\max}$ and λ_{trace} Tests

Null Hypothesis	Alternative Hypothesis		95% Critical Value	90% Critical Value
λ_{trace} *tests*		λ_{trace} *value*		
$r = 0$	$r > 0$	56.786	34.91	32.00
$r \leq 1$	$r > 1$	18.123	19.96	17.85
$r \leq 2$	$r > 2$	3.306	9.24	7.52
$\lambda_{\max}$ *tests*		$\lambda_{\max}$ *value*		
$r = 0$	$r = 1$	38.663	22.00	19.77
$r = 1$	$r = 2$	14.817	15.67	13.75
$r = 2$	$r = 3$	3.306	9.24	7.52

cointegrating vectors ($r > 0$), we can calculate the $\lambda_{\text{trace}}(0)$ statistic:

$$\begin{aligned}\lambda_{\text{trace}}(0) &= -T\,[\ln(1-\lambda_1) + \ln(1-\lambda_2) + \ln(1-\lambda_3)] \\ &= -98\,[\ln(1-0.326) + \ln(1-0.14032) + \ln(1-0.033168)] \\ &= 56.786\end{aligned}$$

Since 56.786 exceeds the 5% critical value of the λ_{trace} statistic (in the bottom panel of Table E, the critical value is 34.91), it is possible to reject the null hypothesis of no cointegrating vectors and accept the alternative of one or more cointegrating vectors. Next, we can use the $\lambda_{\text{trace}}(1)$ statistic to test the null of $r \leq 1$ against the alternative of two or three cointegrating vectors.

In this case, the $\lambda_{trace}(1)$ statistic is

$$\begin{aligned}\lambda_{trace}(1) &= -T\,[\ln(1-\lambda_2)+\ln(1-\lambda_3)]\\ &= -98\,[\ln(1-0.14032)+\ln(1-0.033168)]\\ &= 18.123\end{aligned}$$

Since 18.123 is less than the 5% critical value of 19.96, we cannot reject the null hypothesis at this significance level. However, 18.123 does exceed the 10% critical value of 17.85; some researchers might reject the null and accept the alternative of two or three cointegrating vectors. The $\lambda_{trace}(2)$ statistic indicates no more than two cointegrating vectors at the 10% significance level.

The λ_{max} statistic does not help to clarify the issue. The null hypothesis of no cointegrating vectors ($r = 0$) against the specific alternative $r = 1$ is clearly rejected. The calculated value $\lambda_{max}(0, 1) = -98\ \ln(1 - 0.326) =$ 38.663 exceeds the 5% critical value of 22.00. Note that the test of the null hypothesis $r = 1$ against the specific alternative $r = 2$ cannot be rejected at the 5%, but can be rejected at the 10%, significance level. The calculated value of $\lambda_{max}(1, 2)$ is $-98\ \ln(1\ - 0.14032) = 14.817$, whereas the critical values at the 5 and 10% significance levels are 15.67 and 13.75, respectively. Even though the actual data-generating process contains only one cointegrating vector, the realizations are such that researchers willing to use the 10% significance level would incorrectly conclude that there are two cointegrating vectors. Failing to reject an incorrect null hypothesis is always an inherent danger of using wide confidence intervals.

STEP 3: Analyze the normalized cointegrating vector(s) and speed of adjustment coefficients. If we select $r = 1$, the estimated cointegrating vector $(\beta_0, \beta_1, \beta_2, \beta_3)$ is

$$\beta = (0.00553, 0.41532, 0.42988, -0.42207)$$

If we normalize with respect to β_1, the normalized cointegrating vector and the speed of adjustment parameters are

$$\begin{aligned}\beta &= (-0.01331, -1.0000, -1.0350, 1.0162)\\ \alpha_y &= 0.54627\\ \alpha_z &= 0.16578\\ \alpha_w &= 0.21895\end{aligned}$$

Recall that the data were constructed imposing the long-run relationship: $w_t = y_t + z_t$; hence, the estimated coefficients of the normalized β vector are close to their theoretical values of $(0, -1, -1, 1)$. Consider the following tests:

1. The test that $\beta_0 = 0$ entails one restriction on one cointegrating vector; hence, the likelihood ratio test has a χ^2 distribution with one degree of freedom. The calculated value of $\chi^2 = 0.011234$ is not significant at conventional levels. Hence, we cannot reject the null hypothesis that $\beta_0 = 0$.

Thus, it is possible to use the form of the model in which there is neither a drift nor an intercept in the cointegrating vector. Thus, to clarify the issue concerning the number of cointegrating vectors, it would be wise to reestimate the model excluding the constant from the cointegrating vector.

2. To restrict the normalized cointegrating vector such that $\beta_2 = -1$ and $\beta_3 = 1$ entails two restrictions on one cointegrating vector; hence, the likelihood ratio test has a χ^2 distribution with two degrees of freedom. The calculated value of $\chi^2 = 0.55350$ is not significant at conventional levels. Hence, we cannot reject the null hypothesis that $\beta_2 = -1$ and $\beta_3 = 1$.
3. To test the joint restriction $\beta = (0, -1, -1, 1)$ entails the three restrictions $\beta_0 = 0$, $\beta_2 = -1$, and $\beta_3 = 1$. The calculated value of χ^2 with three degrees of freedom is 1.8128 so that the significance level is 0.612. Hence, we cannot reject the null hypothesis that the cointegrating vector is $(0, -1, -1, 1)$.

STEP 4: Finally, innovation accounting and causality tests on the error-correction model of (6.62) could help to identify a structural model and determine whether the estimated model appears to be reasonable. Since the simulated data have no economic meaning, innovation accounting is not performed here.

10. ERROR-CORRECTION AND ADL TESTS

In the Engle–Granger method, it is possible to estimate the long-run equilibrium relationship from a regression of z_t on y_t or from a regression of y_t on z_t. In the Johansen method, all variables are treated symmetrically. Hence, either method can be used in circumstances when you do not want to explicitly specify a "dependent" variable and a set of "independent" variables. This can be especially advantageous if the variables are jointly determined and you are not sure how to disentangle the interdependence among them. For example, in a test for purchasing power parity, it is likely that the exchange rate and the two price levels all have strong effects on each other. In other circumstances, the selection of a dependent variable and the set of independent variables might be clear. As discussed in this section, there are potential benefits to be had by incorporating such information into a cointegration model. The starting point is to be precise about the econometric meaning of the term "exogenous." To begin with the simplest case, suppose that y_t and z_t are cointegrated of order (1, 1) and that the error-correcting model (ECM) is represented by

$$\Delta y_t = \alpha_1(y_{t-1} - \beta z_{t-1}) + e_{1t} \tag{6.63}$$

$$\Delta z_t = \alpha_2(y_{t-1} - \beta z_{t-1}) + e_{2t} \tag{6.64}$$

Notice that (6.63) and (6.64) are in reduced form and not in structural form. In order to allow for the possibility that the error terms are correlated, we can let the relationship between the error terms and the structural shocks be given by

$$\begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix}$$

where ε_{yt} and ε_{zt} are the structural innovations in Δy_t and Δz_t, and the c_{ij} are coefficients. As in the discussion of structural VARs in Section 10 of Chapter 5, the structural shocks are uncorrelated in that $E\varepsilon_{yt}\varepsilon_{zt} = 0$. Even though $E\varepsilon_{yt}\varepsilon_{zt} = 0$, e_{1t} and e_{2t} will generally be correlated if c_{12} and/or c_{21} differ from zero.

For now, suppose that the values of the c_{ij} are unknown. Nevertheless, it is always possible to construct an orthogonalization between the two errors such that

$$e_{1t} = \rho e_{2t} + v_t \tag{6.65}$$

where ρ is the regression coefficient of e_{1t} on e_{2t} and v_t is the innovation in e_{1t} that is not correlated with e_{2t}. If we substitute (6.64) and (6.65) into (6.63), we obtain

$$\begin{aligned}\Delta y_t &= \alpha_1(y_{t-1} - \beta z_{t-1}) + \rho e_{2t} + v_t \\ &= \alpha_1(y_{t-1} - \beta z_{t-1}) + \rho[\Delta z_t - \alpha_2(y_{t-1} - \beta z_{t-1})] + v_t \\ &= (\alpha_1 - \rho\alpha_2)\,(y_{t-1} - \beta z_{t-1}) + \rho\Delta z_t + v_t\end{aligned}$$

Now, if we let $\alpha = \alpha_1 - \rho\alpha_2$, we can write

$$\Delta y_t = \alpha(y_{t-1} - \beta z_{t-1}) + \rho\Delta z_t + v_t \tag{6.66}$$

In general, it is not appropriate to estimate (6.66) directly since it contains the jointly determined variables Δy_t and Δz_t. The general problem is that Δz_t will be correlated with the error term v_t so that there is a simultaneity problem. As such, OLS cannot be used to recover meaningful estimates of the parameters of the model. Even if the simultaneity problem is rectified, there is an identification problem since α_1 and α_2 cannot be separately identified from the OLS estimate of α. However, it is possible to specify conditions such that the simultaneity and identification problems disappear and that OLS is an efficient estimation and testing strategy. As will be shown below, the two conditions are $\alpha_2 = 0$ (so that z_t does not respond to the discrepancy from the long-run equilibrium relationship) and $c_{21} = 0$ (so that z_t does not respond to ε_{yt}). Thus, the two required assumptions are that z_t is weakly exogenous and causally prior to y_t.

Cointegration with Weak Exogeneity

Following Engle, Hendry, and Richard (1983), a variable x_{it} is weakly exogenous for the parameter set P if the marginal distribution of x_{it} contains no useful information for conducting inference on P. In a cointegrated system, if a variable does not respond to the discrepancy from the long-run equilibrium relationship, it is **weakly exogenous**. Hence, if the speed of adjustment parameter α_i is zero, the variable in question is weakly exogenous. In the example used by Johansen and Juselius (1990), it might be possible to argue that real income should be weakly exogenous. After all, in a full-employment environment, discrepancies between long-run money demand and supply would not be expected to change real income. For our purposes, the practical importance is that a weakly exogeneous variable does not experience the type of feedback that necessitates the use of a VAR.

To explain, suppose that you try to estimate an equation like (6.66) using OLS. You could use a two-step method, such as that employed in the Engle–Granger procedure, and regress y_t on z_t to obtain an estimate of β and then form the variable $y_{t-1} - \beta z_{t-1}$. However, at this point in time, the preference in the literature is to estimate the unrestricted equation

$$\Delta y_t = \beta_1 y_{t-1} + \beta_2 z_{t-1} + \beta_3 \Delta z_t + v_t \tag{6.67}$$

where from (6.66) the estimated coefficients are such that $\beta_1 = \alpha_1 - \rho\alpha_2, \beta_2 = (\alpha_1 - \rho\alpha_2)\beta$ and $\beta_3 = \rho$.

Since the coefficients of (6.67) are unrestricted, this form of the model is often called an **autoregressive distributed lag** to distinguish it from an ECM in the form of (6.66). Notice that the value of α_2 appears in the estimates for β_1 and β_2. However, if z_t is weakly exogenous (i.e., if $\alpha_2 = 0$), your coefficient estimates should be such that $\beta_1 = \alpha_1$, $\beta_2 = \alpha_1\beta$ and $\beta_3 = \rho$. Thus, you can identify α_1, β, and ρ from β_1, β_2, and β_3 since the OLS estimation of (6.67) is equivalent to estimating the equation

$$\Delta y_t = \alpha_1 y_{t-1} - \alpha_1 \beta z_{t-1} + \rho \Delta z_t + v_t \tag{6.68}$$

Although weak exogeneity allows the model to be identified, there is still the issue of properly testing (6.68) for cointegration. Since $\{y_t\}$ and $\{z_t\}$ are $I(1)$, the test statistics of the null hypothesis $\beta_1 = 0$ and $\beta_2 = 0$ in (6.67) are nonstandard and need to be tabulated. The usual way to test for cointegration is to use the t-statistic for the null hypothesis $\beta_1 = 0$ in (6.67).[6] After all, if $\beta_1 = 0$, there is no error-correction so that y_t is not cointegrated with z_t. Table F in the *Supplementary Manual*, uses the work of Ericsson and MacKinnon (2002) to calculate the appropriate critical values necessary to determine whether $\beta_1 < 0$. The critical values depend on the number of $I(1)$ regressors in the model (denoted by k), the adjusted sample size T^a, and the form of the deterministic regressors. For example, if you have an adjusted sample size with 100 observations and estimate a model with an intercept ($d = 1$) and two weakly exogenous variables ($k = 3$), Table F indicates that the appropriate critical values to test the null hypothesis $\beta_1 = 0$ are −4.181, −3.538, and −3.205 at the 1, 5, and 10% significance levels, respectively.

If you compare (6.67) with (6.63), you can see the benefit of employing weak exogeneity. Since the two representations are equivalent, e_{1t} is composed of Δz_t and v_t. Since (6.67) will have a smaller variance than the error term in (6.63), the coefficients of (6.67) can be estimated with more precision than the coefficient of (6.63). A second benefit ascribed to estimating such a model is that the coefficients of y_{t-1} and z_{t-1} are unrestricted. As such, the short-run dynamics for Δy_t are not dictated by long-run equilibrium relationship $y_{t-1} = \beta z_{t-1}$. In the Engle–Granger and Johansen approaches, the so-called **Common Factor Restriction** forces the short-run changes in Δy_t to be a constant proportion of the previous period's deviation from long-run equilibrium.

Inference on the Cointegrating Vector

Suppose you assume that weak exogeneity holds and conclude that the variables are cointegrated (so that $\alpha_1 < 0$ and $\alpha_2 = 0$). As such, it is possible to write (6.64) and (6.67) as

$$\Delta y_t = \alpha_1 (y_{t-1} - \beta z_{t-1}) + \rho \Delta z_t + v_t \tag{6.69}$$

and

$$\Delta z_t = e_{2t} \tag{6.70}$$

Now the question becomes: Can you conduct inference on α_1 and β in (6.69) using standard t-tests and F-tests? The answer, quite possibly, is yes! Since all variables in (6.69) are stationary, we are really operating within a standard OLS regression framework. A simultaneity problem exists if the regressors appearing in (6.69) depend on the error term v_t. Clearly, the $I(0)$ variable $y_{t-1} - \beta z_{t-1}$ is pre-determined so that there is no need to worry about the influence of v_t on the error-correction term. Hence, the key issue concerns the contemporaneous relationship between Δy_t and Δz_t. If Δz_t is unaffected by innovations in Δy_t, it is appropriate to conduct inference on (6.69) using a standard t-tests and F-tests.

Recall that the particular orthogonalization used in (6.65) is such that $e_{1t} = \rho e_{2t} + v_t$ where e_{2t} and v_t are uncorrelated. This is actually a Choleski decomposition in that Δz_t does not respond to innovations in Δy_t but Δy_t responds to innovations in Δz_t. It should be clear that actual error structure has this Choleski form only if $c_{21} = 0$. In other words, if $c_{21} = 0$, (6.65) is equivalent to $e_{1t} = \rho e_{2t} + \varepsilon_{yt}$ and $e_{2t} = \varepsilon_{zt}$. Given that $\Delta z_t = e_{2t}$ does not depend on ε_{yt}, there is no feedback from Δy_t to Δz_t so that it is possible to use standard inference on (6.68) or (6.69).

Thus, testing restrictions on α_1 is straightforward since it is the coefficient on the $I(0)$ variable $(y_{t-1} - \beta z_{t-1})$. As such, given that $\alpha_1 \neq 0$, it is appropriate to form confidence intervals on α_1 using a standard t-distribution. Similarly, given that $\beta \neq 0$, β can be written as the coefficient on the $I(0)$ variable $(y_{t-1}/\beta - \alpha_1 z_{t-1})$. Inference on β can also be conducted using a t-distribution. Finally, note that ρ is the coefficient on the stationary variable Δz_t. Hence, it is appropriate to construct confidence intervals for ρ using a t-distribution.

It is straightforward to generalize these results. Since z_t can actually be a vector of $I(1)$ variables, you can estimate (6.67) for y_t and a set of weakly exogenous variables z_t. For example, with two weakly exogenous variables, z_{1t} and z_{2t}, the error-correction model generalizes to

$$\Delta y_t = \alpha_1(y_{t-1} - \beta_1 z_{1t-1} - \beta_2 z_{2t-1}) + \beta_3 \Delta z_{1t} + \beta_4 \Delta z_{2t} + v_t$$

so that you estimate a model of the form

$$\Delta y_t = \alpha_1 y_{t-1} + b_1 z_{1t-1} + b_2 z_{2t-1} + \beta_3 \Delta z_{1t} + \beta_4 \Delta z_{2t} + v_t.$$

where $b_1 = -\beta_1/\alpha_1$ and $b_2 = -\beta_2/\alpha_1$.

To test for cointegration use the t-statistic for the null hypothesis $\alpha_1 = 0$. Since you have three $I(1)$ variables in the model, obtain the critical values from Table F such that $k = 3$. Of course, if we start from a higher order process, additional lags of Δy_{t-i}, Δz_{1t-i}, and Δz_{2t-i} should be added to the equation. As in the two-variable case, you need to assume that Δy_t has no contemporaneous effects on any values of the Δz_{it}.

11. COMPARING THE THREE METHODS

In this section, we compare the Engle–Granger, Johansen and ADL tests for cointegration using the three-month Treasury bill and 10-year interest rates using the data in QUARTERLY.XLS. Although we know that the spread acts as a stationary variable, the point of this section is to illustrate the use of the three testing methodologies. Since we have already verified that the individual rates act as $I(1)$ process we can skip the preliminary step of pretesting for unit roots. To keep the discussion on point, reasonable leg lengths for each test are simply reported. You can verify them in the exercises at the end of the chapter.

The Engle–Granger Methodology

Given that each rate acts as a unit-root process, we can begin by estimating the long-run equilibrium relationship

$$\begin{aligned} r_{Lt} = \underset{(13.23)}{1.642} + \underset{(43.15)}{0.915} r_{St} \end{aligned} \tag{6.71}$$

Next, we test the residuals from (6.71) for stationarity by estimating an equation in the form (6.32). If you experiment with various lag lengths, you will find that various lag lengths tests suggest a three lag model or a one lag model. If we adopt the SBC and use one lagged change, we obtain

$$\Delta\hat{e}_t = \underset{(-4.45)}{-0.155}\hat{e}_{t-1} + \underset{(2.96)}{0.201}\Delta\hat{e}_{t-1}$$

In a model with 2 variables with 208 usable observations, the 5% critical value shown in Table C is −3.368 and the 1% value is −3.95. As such, we can reject the null hypothesis of no cointegration. Since we are making no assumption concerning weak-exogeneity, it is clearly possible to carry out the analysis using r_{St} as the left-hand side variable. Reversing the variables in (6.71) yields

$$r_{St} = \underset{(-7.04)}{-1.103} + \underset{(43.15)}{0.982} r_{Lt}$$

In this form, the Engle–Granger test also supports the finding of cointegration since the regression of the residuals yields

$$\Delta\hat{e}_t = \underset{(-4.78)}{-0.172}\hat{e}_{t-1} + \underset{(3.24)}{0.219}\Delta\hat{e}_{t-1}$$

Notice that the two estimates of the long-run equilibrium relationship are somewhat different from each other. However, it is not possible to conduct inference on either of these cointegrating vectors unless you use the methods discussed in Appendix 6.2 in the *Supplementary Manual*. As an exercise, you can repeat the cointegration test using a three lag specification.

The Johansen Methodology

Let x_t denote the vector $[r_{Lt}, r_{St}]'$. If you estimate the unrestricted VAR in the form of (6.53) (i.e., if you estimate the VAR $x_t = A_0 + \sum A_i x_{t-i}$) you should find that the SBC selects a lag length of one whereas the AIC and general-to-specific methodology selects a lag length of eight. Again, for expositional purposes it is simplest to report the results of the one lag model. Given this lag length, it is possible to estimate the model in the form of (6.54). Since the interest rates do not continually increase or decrease over time, it seems reasonable to constrain the drift terms so that a constant appears in the cointegrating relationship. The estimated value of the π^* matrix is such that

$$\pi^* x^*_{t-1} = \begin{bmatrix} -1.048 & 1.102 & 0.956 \\ -0.446 & 0.100 & 2.133 \end{bmatrix} \begin{bmatrix} r_{Lt} \\ r_{St} \\ 1 \end{bmatrix}$$

The characteristic roots are such that $\lambda_1 = 0.1295$ and $\lambda_2 = 0.0136$ so that $-T\ln(1-\lambda_1) = 29.13$ and $-T\ \ln(1-\lambda_2) = 2.87$. To test the null hypothesis of no cointegration against the general alternative of 1 or 2 cointegrating vectors compare the sum $29.13 + 2.87 = 32.00$ to the 5% critical value of the λ_{trace} statistic shown in Table E. Since 32.00 exceeds the critical value of 19.96, reject the null and conclude that there is at least one cointegrating vector. To test the null of one cointegrating vector against the alternative of two cointegrating vectors, compare the sample value of 2.87 to the 5% critical value of 9.24. As such, we can conclude that there is only one cointegrating vector.

Normalizing the cointegrating vector with respect to r_{Lt} yields

$$\begin{aligned} r_{Lt} = {} & 0.912 + 1.051 r_{St} \\ & (2.65) \quad (17.88) \end{aligned}$$

A key difference between this estimate of the long-run equilibrium relationship and those from the Engle–Granger test is that standard inference can be performed on the coefficients of the cointegrating vector. For example, the likelihood ratio test for the null hypothesis that the coefficients on the long-term and short-term rates both equal unity is only 0.643 with a *prob*-value of 0.422. As such, we can conclude that the restriction is not binding. Hence, in the long-run, the 10-year rate tends to move 1:1 with the short-term rate. If you re-estimate the model imposing the restriction, you should find

$$\begin{aligned} \Delta r_{Lt} = {} & -0.098\ (r_{Lt-1} - 1.17 - r_{St-1}) + 0.185\Delta r_{Lt-1} + 0.002\Delta r_{St-1} \\ & (-2.32) \qquad (-7.10) \qquad\qquad (1.88) \qquad\qquad (0.03) \end{aligned} \tag{6.72}$$

$$\begin{aligned} \Delta r_{St} = {} & 0.084\ (r_{Lt-1} - 1.17 - r_{St-1}) + 0.053\Delta r_{Lt-1} + 0.229\Delta r_{St-1} \\ & (1.51) \qquad (-7.10) \qquad\qquad (0.41) \qquad\qquad (2.23) \end{aligned} \tag{6.73}$$

If you test for the presence of the intercept, you will find that the constant term in the cointegrating vector is highly significant. The important point is that the t-statistics on the error-correcting terms imply that the long-term rate adjusts to the discrepancy from the long-run equilibrium relationship, but the short-rate does not. In other words, r_{St} is weakly exogenous. Consider the dynamic adjustment mechanism if there is a positive 1-unit discrepancy from the long-run equilibrium relationship. The estimates imply that the long-term rate falls by −0.098 units and that the short-term rate does none of the adjusting. As such, the deviations from the long-run relationship are quite long lived.

The Error-Correction Methodology

In contrast to the Engle–Granger and Johansen methodologies, to use the error-correction test it is necessary to assume that one of the variables is weakly exogenous. Suppose that we were certain that the short-term interest rate did none of the adjustment necessary to restore the long-run equilibrium relationship. Given that the short-term rate is weakly exogenous, we can estimate an equation in the form

$$\Delta r_{Lt} = \beta_0 + \beta_1 r_{Lt-1} + \beta_2 r_{St-1} + \beta_3 \Delta r_{St} + A_1(L)\Delta r_{Lt-1} + A_2(L)\Delta r_{St-1} + v_t \quad (6.74)$$

Equation (6.74) looks very much like (6.72) except that elements of the cointegrating vector are unrestricted and the contemporaneous value of Δr_{St} is included. Since we are not treating all variables symmetrically, there is no need to constrain the lag length represented by the polynomial $A_1(L)$ to be the same as that from $A_2(L)$. However, for this case, it turns out that a lag length of six seems appropriate for each variable. Consider the estimated equation

$$\Delta r_{Lt} = \underset{(1.52)}{0.113} - \underset{(-4.45)}{0.171} r_{Lt-1} + \underset{(4.80)}{0.187} r_{St-1} + \underset{(15.92)}{0.612}\Delta r_{St} + A_1(L)\Delta r_{Lt-1} + A_2(L)\Delta r_{St-1} + v_t \quad (6.75)$$

The key point to note is that the t-statistic for the null hypothesis $\beta_1 = 0$ is −4.45. Given the presence of an intercept ($d = 1$), two $I(1)$ variables ($k = 2$), and that the estimation begins in $1961Q4$ ($T = 205$), the adjusted sample size is $T^a = 205 - (2^*2 - 1) - 1 = 201$. From Table F, the critical values at the 1, 5, and 10% significance levels are approximately −3.834, −3.231, and −2.916, respectively. Hence, we can reject the null hypothesis of no cointegration and conclude that the variables are cointegrated.

We can reparameterize (6.75) such that

$$\Delta r_{Lt} = -0.171(r_{Lt-1} - 1.09 r_{St-1} - 0.661) + 0.612\Delta r_{St} + A_1(L)\Delta r_{Lt-1} + A_2(L)\Delta r_{St-1} + v_t.$$

In this particular example, all three approaches find that the variables are cointegrated. The Engle–Granger approach indicates that the speed of adjustment parameter is −0.155 (or −0.172), but does not indicate which variable (or variables) does the adjustment. In response to a one-unit deviation from the long-run equilibrium, the Johansen approach indicates that the long-term rate adjusts by −0.098 units while the ECM approach indicates that it adjusts by −0.171 units. The Engle–Granger approach does not allow us to readily perform inference of the cointegrating vector, but the Johansen approach allows us to conclude that two rates move 1:1 in the long run.

So long as we are willing to assume $\beta_2 \neq 0$, it is possible to perform inference on the coefficient on r_{St-1} in the long-run equilibrium relationship. Clearly, it would have been possible to reparameterize (6.75) such that

$$\Delta r_{Lt} = -0.187(0.914 r_{Lt-1} - r_{St-1} - 0.604) + 0.612\Delta r_{St} + A_1(L)\Delta r_{Lt-1} + A_2(L)\Delta r_{St-1} + v_t \quad (6.76)$$

Hence, β_2 (= 0.187) is the coefficient on a stationary variable so that it has a standard t-distribution. Given that the standard error of β_2 is 0.038, a $\pm$1.96 standard deviation band runs from 0.111 to 0.263. Alternatively, we could have performed an F-test for the null hypothesis $\beta_1 = \beta_2$ in (6.72). A traditional F-test is appropriate since each coefficient has a t-distribution. With 1 degree of freedom in the numerator and 189 in the denominator, the sample value of $F = 2.86$ is significant at the 0.093 level. If you re-estimate the model such that $\beta_1 = \beta_2$, you should find

$$\Delta r_{Lt} = -0.175(r_{Lt-1} - r_{St-1} - 2.01) + 0.604\Delta r_{St} + A_1(L)\Delta r_{Lt-1} + A_2(L)\Delta r_{St-1} + v_t$$

If you are willing to abstract from the stationary dynamics, it is clear how to trace out the effects of a one-unit shock in Δr_{St}. All else equal, if $\Delta r_{St} = 1$, it follows that $\Delta r_{Lt} = 0.604$. In period $t + 1$, it follows that the discrepancy from the long-run equilibrium is −0.396 (= 0.604 − 1) and the change in the long-rate is (−0.396)(−0.175) = 0.069. In subsequent periods, the long rate keeps rising by 17.5% of the discrepancy from the long-run equilibrium. At this point, you could go on to perform the innovation accounting by estimating an equation of the form $\Delta r_{St} = A_3(L)\Delta r_{Lt} + A_4(L)\Delta r_{St} + e_{2t}$. Note that the equation is in first-differences since the Δr_{St} equation does not contain an error-correction term. Also note that the assumption that Δr_{St} is weakly exogenous implies a causal ordering of the innovations in that a v_t shock has no contemporaneous effect on Δr_{St} but an e_{2t} shock directly affects Δr_{Lt}.

12. SUMMARY AND CONCLUSIONS

Many economic theories imply that a linear combination of certain nonstationary variables must be stationary. For example, if the variables $\{x_{1t}\}$, $\{x_{2t}\}$, and $\{x_{3t}\}$ are $I(1)$ and the linear combination $e_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t}$ is stationary, the variables are said to be cointegrated of order (1, 1). The vector $(\beta_0, \beta_1, \beta_2, \beta_3)$ is called the cointegrating vector. Cointegrated variables share the same stochastic trends and so cannot drift too far apart. Cointegrated variables have an error-correction representation such that each responds to the deviation from "long-run equilibrium."

One way to check for cointegration is to examine the residuals from the long-run equilibrium relationship. If these residuals have a unit root, the variables cannot be cointegrated of order (1, 1). Another way to check for cointegration among $I(1)$ variables is to estimate a VAR in first differences and include the lagged level of the variables. The Johansen methodology uses the λ_{trace} and $\lambda_{\max}$ test statistics to determine if the variables are cointegrated and the number of cointegrating vectors. These tests are sensitive to the presence of the deterministic regressors included in the cointegrating vector(s). Restrictions on the cointegrating vector(s) and/or the speed of adjustment parameters can be tested using χ^2 statistics. You should be aware of the role of the deterministic regressors in a cointegration framework. Johansen (1994) shows how to test to determine whether there is a deterministic trend, drift terms outside of the cointegrating vector, or constants that all appear in the cointegrating vector. A third way to test for cointegration is to estimate the error-correction model. If only one variable adjusts to the discrepancy from the long-run equilibrium relationship, it can be preferable to estimate an autoregressive distributed lag model. It is straightforward to estimate the model using OLS and to perform hypothesis tests on the coefficients of the cointegrating vector. For more complicated situations, Appendix 6.2 discusses the Phillips-Hansen (1990) method of modeling in a single equation framework.

QUESTIONS AND EXERCISES

1. Let equations (6.14) and (6.15) contain intercept terms such that

$$y_t = a_{10} + a_{11}y_{t-1} + a_{12}z_{t-1} + \varepsilon_{yt} \quad \text{and} \quad z_t = a_{20} + a_{21}y_{t-1} + a_{22}z_{t-1} + \varepsilon_{zt}$$

a. Show that the solution for y_t can be written as

$$y_t = [(1 - a_{22}L)\varepsilon_{yt} + (1 - a_{22})a_{10} + a_{12}L\varepsilon_{zt} + a_{12}a_{20}]/[(1 - a_{11}L)(1 - a_{22}L) - a_{12}a_{21}L^2]$$

b. Find the solution for z_t.

c. Suppose that y_t and z_t are $CI(1, 1)$. Use the conditions in (6.19), (6.20), and (6.21) to write the error-correcting model. Compare your answer to (6.22) and (6.23). Show that the error-correction model contains an intercept term.

d. Show that $\{y_t\}$ and $\{z_t\}$ have the same deterministic time trend (i.e., show that the slope coefficients of the time trends are identical).

e. What is the condition such that the slope of the trend is zero? Show that this condition is such that the constant can be included in the cointegrating vector.

f. Modify (6.26) so that each equation has an explicit intercept. Specifically, let $x_t = A_0 + A_1x_{t-1} + \varepsilon_t$ where A_0 is an $(n \cdot 1)$ vector with elements a_{0i}. Suppose that the rank of π is 1. How are the solutions to 6.28 affected the presence of the intercepts? How is the error-correction representation in (6.29) affected?

2. The data file COINT6.XLS contains the three simulated series used in Sections 5 and 9.

a. Use the data to reproduce the results in Section 5.

b. Obtain the impulse responses and variance decompositions using the ordering such that $y_t \rightarrow z_t \rightarrow w_t$. Do these seem reasonable, given the way in which the variables were constructed?

c. Use the data to reproduce the results in Section 9.

d. Examine Table 6.1. Show that y_t and z_t, but not w_t, are weakly exogenous.

e. Use the data to compare the ECM test to the Engle–Granger and Johansen tests treating y_t and z_t as weakly exogenous.

3. In Question 9 of Chapter 4 you were asked to use the data on QUARTERLY.XLS to estimate the regression equation

$$\underset{}{INDPROD_t} = \underset{(29.90)}{30.48} + \underset{(36.58)}{0.04M1NSA_t}$$

a. Use the Engle–Granger test to show that the regression is spurious.

b. Examine the scatter plot of $INDPRO_t$ against $M1NSA_t$. How do you explain the fact that R^2 is close to unity and that the t-statistic on the money supply is 36.58?

c. Use the data on the file labeled REAL.XLS. Denote the natural logs of real GDP and consumption by ly_t and lc_t, respectively. Estimate the regression

$$lc_t = \underset{(-51.78)}{-0.962} + \underset{(494.19)}{1.06ly_t} \qquad R^2 = 0.999$$

If you perform the Engle–Granger test using four lags, you should find

$$\Delta\hat{e}_t = -0.092\hat{e}_{t-1} + \sum_{i=1}^{4}\beta_i\Delta\hat{e}_{t-i}$$

The t-statistic for $\hat{e}_{t-1}$ is −3.48. How do you interpret the consumption–income relationship?

4. The file labeled QUARTERLY.XLS contains the interest rates paid on U.S. 3-month, 5-year, and 10-year U.S. government securities. The data run from 1960Q1 to 2012Q4. The variables are labeled TBILL, R5 and R10, respectively.

   a. Pretest the variables to show that the rates all act as unit root processes. Specifically, perform augmented Dickey–Fuller tests using the lag length selected by deleting lags until the t-statistic on the last lag is significant at the 5% level. If you include an intercept (but no time trend) you should obtain:

Series	Lags	Estimated a_1	t-statistic
TBILL	7	−0.028	−1.61
R5	5	−0.013	−1.03
R10	7	−0.011	−0.78

   b. Estimate the cointegrating relationships using the Engle–Granger procedure. Perform augmented Dickey–Fuller tests on the residuals. Using TBILL as the "dependent" variable, you should find

$$\underset{}{\text{TBILL}_t} = \underset{(2.31)}{0.367} + \underset{(-13.44)}{2.7\text{R5}_t} - \underset{(20.78)}{1.91\text{R10}_t}$$

   where t-statistics are within parentheses.
   Perform the Engle–Granger test on the residuals from the equation above. Why is it appropriate to use eight lags in the augmented form of the test? If you use eight lags, you should find that the coefficient on the lagged residual (i.e., e_{t-1}) is −0.276 with a t-statistic of −4.08. The 5% critical value is about −3.76. Based on this data, do you conclude that the variables are cointegrated?

   c. Repeat part b using R10 as the dependent variable. If you use 6 lags in the augmented form of the Engle–Granger test (i.e., estimate $\Delta e_t = \alpha_1 e_{t-1} + \ldots$) you should find $a_1 =$ −0.105 and the t-statistic is −2.34. Using R10 as the dependent variable, are the three interest rates cointegrated?

   d. Estimate the model using the Johansen procedure. Use 7 lags and include a constant in the cointegrating vector. You should find the following:

Trace Tests				Maximum Eigenvalue Tests			
Null	Alternative	λ_{trace}	5% Value	Null	Alternative	λ_{max}	5% Value
$r = 0$	$r \geq 1$	45.50	34.91	$r = 0$	$r = 1$	37.83	22.00
$r \leq 1$	$r \geq 2$	7.67	19.96	$r = 1$	$r = 2$	6.89	15.67
$r \leq 2$	$r = 3$	0.78	9.24	$r = 2$	$r = 3$	0.78	9.24

   i. Explain why the λ_{trace} test strongly suggests there is exactly one cointegrating vector.
   ii. To what extent is this result reinforced by the λ_{max} test?
   Verify that the cointegrating vector is

$$1.99\text{TBILL}_t + 0.879\text{R5}_t - 1.67\text{R10}_t + 0.820 = 0$$

   Compare this result to your answer in part **b**.

e. Check to determine whether the individual interest rate pairs are cointegrated. In particular, is $R5_t$ with cointegrated $R10_t$?

f. Why might you be wary about testing for cointegration using the ADL test developed in Section 10?

5. In Question 4, the Engle–Granger methodology found that the long-run equilibrium relationship for the three interest rates was

$$TBILL_t = 0.367 - 1.91R5_t + 2.74R10_t$$

a. Estimate an error-correcting model using 2 lagged changes of each variable. Use the residuals from this long-run equilibrium relationship as the error-correction term and do not include intercepts. You should find that the error-corrections are such that

$$\Delta TBILL_t = 0.062e_{t-1} + \cdots \; t\text{-statistic for the error-correction term: } 0.73$$
$$\Delta R5_t = -0.161e_{t-1} + \cdots \; t\text{-statistic for the error-correction term: } -2.94$$
$$\Delta R10_t = -0.162e_{t-1} + \cdots \; t\text{-statistic for the error-correction term: } -2.52$$

where e_{t-1} is the lagged residual from your estimate in part the equilibrium relationship.

i. Verify that the multivariate AIC selects a model with 2 lagged changes of each variable. Perform the appropriate diagnostic tests on the system. In particular, determine whether the three series of residuals appear to be white noise. Are the lags lengths unnecessarily short?

ii. Discuss the nature of the adjustment. Are any of the rates weakly exogenous? In response to a deviation from the long-run relationship, how are the three rates predicted to change?

b. Use a Choleski decomposition such that the T-bill rate is causally prior to $R5_t$ and $R5_t$ is causally prior to $R10_t$.

c. Obtain the variance decompositions using the same ordering as you used in part **b**. Show that the preponderance of the forecast error variance of each rate is primarily due to the T-bill rate.

6. Suppose you estimate π to be

$$\pi = \begin{bmatrix} 0.6 & -0.5 & 0.2 \\ 0.3 & -0.25 & 0.1 \\ 1.2 & -1.0 & 0.4 \end{bmatrix}$$

a. Show that the determinant of π is zero.

b. Show that two of the characteristic roots are zero and that the third is 0.75.

c. Let $\beta' = (3 - 2.51)$ be the single cointegrating vector normalized with respect to x_{3t}. Find the $(3 \cdot 1)$ vector α such that $\pi = \alpha\beta'$. How would α change if you normalized β with respect to x_{1t}?

d. Describe how you could test the restriction $\beta_1 + \beta_2 = 0$.

Now suppose you estimate π to be

$$\pi = \begin{bmatrix} 0.8 & 0.4 & 0.0 \\ 0.1 & 0.1 & 0.0 \\ 0.75 & 0.25 & 0.5 \end{bmatrix}$$

e. Show that the three characteristic roots are 0.0, 0.5, and 0.9.

f. Select β such that

$$\beta = \begin{bmatrix} 0.8 & 0.75 \\ 0.4 & 0.25 \\ 0.0 & 0.5 \end{bmatrix}$$

Find the $(3 \cdot 2)$ matrix α such that $\pi = \alpha\beta'$.

7. Suppose that x_{1t} and x_{2t} are integrated of orders 1 and 2, respectively. You are to sketch the proof that any linear combination of x_{1t} and x_{2t} is integrated of order 2. Toward this end:

a. Allow x_{1t} and x_{2t} to be the random walk processes: $x_{1t} = x_{1t-1} + \varepsilon_{1t}$ and $x_{2t} = x_{2t-1} + \varepsilon_{2t}$.

i. Given the initial conditions x_{10} and x_{20}, show that the solutions for x_{1t} and x_{2t} have the form $x_{1t} = x_{10} + \Sigma\varepsilon_{1t-i}$ and $x_{2t} = x_{20} + \Sigma\varepsilon_{2t-i}$.

ii. Show that the linear combination $\beta_1 x_{1t} + \beta_2 x_{2t}$ will generally contain a stochastic trend.

iii. What assumption is necessary to ensure that x_{1t} and x_{2t} are $CI(1, 1)$?

b. Now let x_{2t} be integrated of order 2. Specifically, let $\Delta x_{2t} = \Delta x_{2t-1} + \varepsilon_{2t}$. Given initial condition for x_{20} and x_{21}, find the solution for x_{2t}. (You may allow ε_{1t} and ε_{2t} to be perfectly correlated.)

i. Is there any linear combination of x_{1t} and x_{2t} that contains only a stochastic trend?

ii. Is there any linear combination of x_{1t} and x_{2t} that does not contain a stochastic trend?

c. Provide an intuitive explanation for the statement: If x_{1t} and x_{2t} are integrated of orders d_1 and d_2 where $d_2 > d_1$, any linear combination of x_{1t} and x_{2t} is integrated of order d_2.

8. Chapter 6 of the *Programming Manual* uses the variables *Tbill* and *Tb1yr* on the file QUARTERLY.XLS to illustrate both the Johansen and Engle–Granger cointegration tests.

a. Verify that the t-statistics of the Dickey–Fuller tests using 7 lags are −1.61304 and −1.39320 for the *Tbill* (r_{St}) and *Tb1yr* (r_{Lt}), respectively.

b. Estimate the long-run relationship alternatively using *Tbill* and *Tb1yr* as the "independent" variable. For r_{St} as the left-hand-side variable, you should find $r_{St} = -0.187 + 0.936r_{Lt}$.

c. Estimate an equation in the form of (6.32) using 6 lags. The estimate of a_1 should be −0.372 with a t-statistic of −4.78. Use Table C to determine whether the variables are cointegrated. What happens if you use r_{Lt} as the left-hand-side variable in the long-run relationship?

d. Estimate the error-correction model and obtain the impulse response functions. Your results should look like those in Section 6.1 of the *Programming Manual*.

e. If you perform the Johansen test using seven lags you should find that the eigenvalues are 0.1523 and 0.0078. Calculate the λ_{max} and the λ_{trace} statistics as in (6.55) and (6.56). Use your results to the number of cointegrating vectors.

9. The file COINT_PPP.XLS contains monthly values of the Japanese, Canadian, and Swiss consumer price levels and the bilateral exchange rates with the United States. The file also contains the U.S. consumer price level. The names on the individual series should be self-evident. For example, *JAPANCPI* is the Japanese price level and *JAPANEX* is the bilateral Japanese/U.S. exchange rate. The starting date for all variables is January 1974 while the availability of the variables is such that most end near the end of 2013. The price indices have been normalized to equal 100 in January 1973 and only the U.S. price index is seasonally adjusted.

a. Form the log of each variable and pretest each for a unit root. Can the null hypothesis of a unit root be rejected for any of the series? How might you proceed if you found that the U.S. CPI was trend stationary?

b. Form the log of each variable. Estimate the long-run relationship between Japan and the U.S. as

$$\log(japanex) = 9.97 - 0.104 \log(japancpi) - 0.768 \log(uscpi)$$
$$(27.25) \quad (-0.98) \qquad\qquad (-17.05)$$

i. Do the point estimates of the slope coefficients seem to be consistent with long-run PPP?

ii. From the t-statistics, can you conclude that the Japanese CPI is not significant at the 5% level?

c. Let u_t denote the residuals from the long-run relationship. Use these residuals to perform the Engle–Granger test for cointegration. If you use eleven lagged changes, you should find

$$\Delta u_t = -0.025u_{t-1} + \sum_{i=1}^{11} a_i \Delta u_{t-1} + \varepsilon_t$$

The t-statistic on the coefficient for u_{t-1} is −3.44. From Table C, with three variables and 457 usable observations, the 5% and 10% critical values are about −3.760 and −3.464, respectively. Do you conclude that long-run PPP fails?

d. Repeat parts **i** and **ii** using Canada and Switzerland. If you use the residuals from the long-run equilibrium relationships you should find

$$\text{Canada (10 lags)} \quad \Delta u_t = -0.012u_{t-1} + \sum a_i \Delta u_{t-i} + \varepsilon_t; \quad t\text{-stat.} = -1.89$$
$$\text{Switzerland (10 lags)} \quad \Delta u_t = -0.027u_{t-1} + \sum a_i \Delta u_{t-i} + \varepsilon_t; \quad t\text{-stat.} = -3.02.$$

e. Although (at conventional significance levels) we reject the null hypothesis of long-run PPP between Japan and the United States, estimate the error-correction model for $ljapanex_t$. If you use 11 lagged changes of each variable, you should find

$$\Delta ljapanex_t = -0.005 - 0.030\hat{e}_{t-1} + \Sigma\Delta\beta_1 ljapanex_{t-i} + \Sigma\Delta\beta_2 ljapancpi_{t-i} + \Sigma\Delta\beta_3 luscpi_{t-i}$$

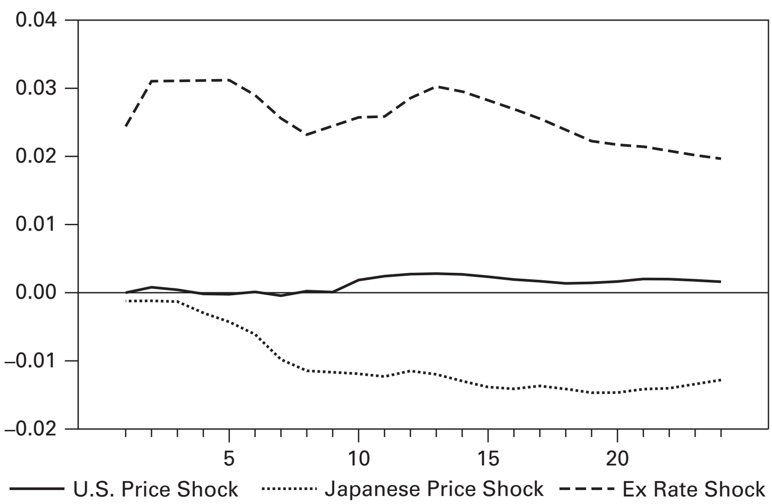


FIGURE 6.5 Responses of the Japanese Exchange Rate

where $\hat{e}_{t-1}$ is the residual from the equilibrium relationship above and eleven lagged changes are used for each variable. The t-statistic on the error correction term is -3.54. Which of the variables(s) can be said to be weakly exogenous?

f. Obtain the impulse functions using the ordering $luscpi_t \rightarrow ljapancpi_t \rightarrow ljapanex_t$. As in Figure 6.5, you should find that the U.S. price shock has little effect on the exchange rate but that a shock to the Japanese price level causes the yen to depreciate. The response of the exchange rate to its own shock is immediate and permanent.

g. Are the results of the cointegration test sensitive to the normalization (i.e., which of the variables is used as the 'dependent' variable) used in the equilibrium regression?

10. In Question 9d, you were asked to use the Engle–Granger procedure test for PPP among the variables log(*canex*), log(*cancpi*), and log(*uscpi*).

a. Now use the Johansen methodology and constrain the constant to the cointegrating vector to obtain:

Rank	λ_i	$\lambda_{\max}$	λ_{trace}
1	0.0535	25.647	35.987
2	0.0138	6.460	10.339
3	0.0083	3.879	3.879

Use Table E to show that there is a single cointegrating vector.

b. Consider the estimated cointegrating vector:

$$-0.949\ \log(canex) - 6.484\ \log(cancpi) + 1.600\ \log(uscpi) + 31.653 = 0$$

Normalize with respect to the exchange rate. Does the long-run relationship seem to be consistent with PPP?