Lecture 5
Co-integration, Error Correction Models

Chapter Contents

12.4

Cointegration

- As a general rule, nonstationary time-series variables should not be used in regression models to avoid the problem of spurious regression
  - There is an exception to this rule

- There is an important case when $e_t = y_t - \beta_1 - \beta_2 x_t$ is a stationary I(0) process
  - In this case y_t and x_t are said to be **cointegrated**
    - Cointegration implies that y_t and x_t share similar stochastic trends, and, since the difference e_t is stationary, they never diverge too far from each other

- The test for stationarity of the residuals is based on the test equation:

Eq. 12.7

$$\Delta\hat{e}_t = \gamma\hat{e}_{t-1} + v_t$$

  - The regression has no constant term because the mean of the regression residuals is zero.
  - We are basing this test upon estimated values of the residuals

Table 12.4 Critical Values for the Cointegration Test

Regression model	1%	5%	10%
(1) $y_t = \beta x_t + e_t$	−3.39	−2.76	−2.45
(2) $y_t = \beta_1 + \beta_2 x_t + e_t$	−3.96	−3.37	−3.07
(3) $y_t = \beta_1 + \delta t + \beta_2 x_t + e_t$	−3.98	−3.42	−3.13

Note: These critical values are taken from J. Hamilton (1994), *Time Series Analysis*, Princeton University Press, p. 766.

- There are three sets of critical values
  - Which set we use depends on whether the residuals are derived from:

Eq. 12.8a

$$Equation\ 1: \quad \hat{e}_t = y_t - bx_t$$

Eq. 12.8b

$$Equation\ 2: \quad \hat{e}_t = y_t - b_2 x_t - b_1$$

Eq. 12.8c

$$Equation\ 3: \quad \hat{e}_t = y_t - b_2 x_t - b_1 - \hat{\delta} t$$

- Consider the estimated model:

Eq. 12.9

$$\hat{B}_t = 1.140 + 0.914F_t, \quad R^2 = 0.881$$
$$(t) \quad (6.548) \quad (29.421)$$

  - The unit root test for stationarity in the estimated residuals is:

$$\Delta\hat{e}_t = -0.225\hat{e}_{t-1} + 0.254\Delta\hat{e}_{t-1}$$
$$(tau) \quad (-4.196)$$

- The null and alternative hypotheses in the test for cointegration are:

$$H_0: \text{the series are not cointegrated} \Leftrightarrow \text{residuals are nonstationary}$$
$$H_1: \text{the series are cointegrated} \Leftrightarrow \text{residuals are stationary}$$

  – Similar to the one-tail unit root tests, we reject the null hypothesis of no cointegration if $\tau \leq \tau_c$, and we do not reject the null hypothesis that the series are not cointegrated if $\tau > \tau_c$

- Consider a general model that contains lags of *y* and *x*
  - Namely, the autoregressive distributed lag (ARDL) model, except the variables are nonstationary:

$$y_t = \delta + \theta_1 y_{t-1} + \delta_0 x_t + \delta_1 x_{t-1} + v_t$$

- If y and x are cointegrated, it means that there is a long-run relationship between them
  - To derive this exact relationship, we set $y_t = y_{t-1} = y$, $x_t = x_{t-1} = x$ and $v_t = 0$
  - Imposing this concept in the ARDL, we obtain:

$$y(1 - \theta_1) = \delta + (\delta_0 + \delta_1)x$$

    - This can be rewritten in the form:

$$y = \beta_1 + \beta_2 x$$

- Add the term $-y_{t-1}$ to both sides of the equation:

$$y_t - y_{t-1} = \delta + (\theta_1 - 1)y_{t-1} + \delta_0 x_t + \delta_1 x_{t-1} + v_t$$

  - Add the term $-\delta_0 x_{t-1} + \delta_0 x_{t-1}$:

$$\Delta y_t = \delta + (\theta_1 - 1)y_{t-1} + \delta_0(x_t - x_{t-1}) + (\delta_0 + \delta_1)x_{t-1} + v_t$$

  - Manipulating this we get:

$$\Delta y_t = (\theta_1 - 1)\left(\frac{\delta}{(\theta_1 - 1)} + y_{t-1} + \frac{(\delta_0 + \delta_1)}{(\theta_1 - 1)} x_{t-1}\right) + \delta_0 \Delta x_t + v_t$$

Eq. 12.10

- Or:

$$\Delta y_t = -\alpha(y_{t-1} - \beta_1 - \beta_2 x_{t-1}) + \delta_0 \Delta x_t + v_t$$

- – This is called an error correction equation
- – This is a very popular model because:
  - It allows for an underlying or fundamental link between variables (the long-run relationship)
  - It allows for short-run adjustments (i.e. changes) between variables, including adjustments to achieve the cointegrating relationship

- For the bond and federal funds rates example, we have:

$$\Delta\hat{B}_t = -0.142(B_{t-1} - 1.429 - 0.777F_{t-1}) + 0.842\Delta F_t - 0.327\Delta F_{t-1}$$
$$(t) \quad (2.857) \qquad\qquad\qquad\qquad\qquad (9.387) \qquad\quad (3.855)$$

 – The estimated residuals are

$$\hat{e}_{t-1} = (B_{t-1} - 1.429 - 0.777F_{t-1})$$

- The result from applying the ADF test for stationarity is:

$$\Delta\hat{e}_t = -0.169\hat{e}_{t-1} + 0.180\Delta\hat{e}_{t-1}$$
$$(t) \quad (-3.929)$$

  – Comparing the calculated value (-3.929) with the critical value, we reject the null hypothesis and conclude that (*B*, *F*) are cointegrated

12.5
Regression with No-Cointegration

- How we convert nonstationary series to stationary series, and the kind of model we estimate, depend on whether the variables are **difference stationary** or **trend stationary**
  - In the former case, we convert the nonstationary series to its stationary counterpart by taking first differences
  - In the latter case, we convert the nonstationary series to its stationary counterpart by de-trending

- Consider the random walk model:

$$y_t = y_{t-1} + v_t$$

  - This can be rendered stationary by taking the first difference:

$$\Delta y_t = y_t - y_{t-1} = v_t$$

    - The variable y_t is said to be a **first difference stationary** series

12.5.1
First Difference Stationary

- A suitable regression involving only stationary variables is:

Eq. 12.11a

$$\Delta y_t = \theta \Delta y_{t-1} + \beta_0 \Delta x_t + \beta_1 \Delta x_{t-1} + e_t$$

- – Now consider a series y_t that behaves like a random walk with drift:

$$y_t = \alpha + y_{t-1} + v_t$$

with first difference:

$$\Delta y_t = \alpha + v_t$$

- The variable y_t is also said to be a **first difference stationary** series, even though it is stationary around a constant term

- Suppose that *y* and *x* are I(1) and not cointegrated
  - An example of a suitable regression equation is:

Eq. 12.11b

$$\Delta y_t = \alpha + \theta \Delta y_{t-1} + \beta_0 \Delta x_t + \beta_1 \Delta x_{t-1} + e_t$$

- Consider a model with a constant term, a trend term, and a stationary error term:

$$y_t = \alpha + \delta t + v_t$$

  – The variable y_t is said to be **trend stationary** because it can be made stationary by removing the effect of the deterministic (constant and trend) components:

$$y_t - \alpha - \delta t = v_t$$

- If y and x are two trend-stationary variables, a possible autoregressive distributed lag model is:

Eq. 12.12

$$y_t^* = \theta y_{t-1}^* + \beta_0 x_t^* + \beta_1 x_{t-1}^* + e_t$$

- As an alternative to using the de-trended data for estimation, a constant term and a trend term can be included directly in the equation:

$$y_t = \alpha + \delta t + \theta y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + e$$

where:

$$\alpha = \alpha_1(1 - \theta_1) - \alpha_2(\beta_0 + \beta_1) + \theta_1 \delta_1 + \beta_1 \delta_2$$

$$\delta = \delta_1(1 - \theta_1) - \delta_2(\beta_0 + \beta_1)$$

- If variables are stationary, or I(1) and cointegrated, we can estimate a regression relationship between the levels of those variables without fear of encountering a spurious regression
- If the variables are I(1) and not cointegrated, we need to estimate a relationship in first differences, with or without the constant term
- If they are trend stationary, we can either de-trend the series first and then perform regression analysis with the stationary (de-trended) variables or, alternatively, estimate a regression relationship that includes a trend variable

FIGURE 12.4 Regression with time-series data: nonstationary variables