Chapter 20

Unit Roots and Cointegration

20.1 Introduction

As we saw in the last chapter, the usual asymptotic results cannot be expected to apply if any of the variables in a regression model is generated by a nonstationary process. For example, in the case of the linear regression model $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u}$, the usual results depend on the assumption that the matrix $n^{-1}\boldsymbol{X}^{\top}\boldsymbol{X}$ tends to a finite, positive definite matrix as the sample size n tends to infinity. When this assumption is violated, some very strange things can happen, as we saw when we discussed "spurious" regressions between totally unrelated variables in Section 19.2. This is a serious practical problem, because a great many economic time series trend upward over time and therefore seem to violate this assumption.

Two obvious ways to keep standard assumptions from being violated when using such series are to detrend or difference them prior to use. But detrending and differencing are very different operations; if the former is appropriate, the latter will not be, and vice versa. Detrending a time series y_t will be appropriate if it is trend-stationary, which means that the DGP for y_t can be written as

$$y_t = \gamma_0 + \gamma_1 t + u_t, \tag{20.01}$$

where t is a time trend and u_t follows a stationary ARMA process. On the other hand, differencing will be appropriate if the DGP for y_t can be written as

$$y_t = \gamma_1 + y_{t-1} + u_t, \tag{20.02}$$

where again u_t follows a stationary ARMA process. If the u_t's were serially independent, (20.02) would be a random walk with drift, the drift parameter being γ_1. They will generally not be serially independent, however. As we will see shortly, it is no accident that the same parameter γ_1 appears in both (20.01) and (20.02).

The choice between detrending and differencing comes down to a choice between (20.01) and (20.02). The main techniques for choosing between them are various tests for what are called **unit roots**. The terminology comes from the literature on time-series processes. Recall from Section 10.5 that for an AR process $A(L)u_t = \varepsilon_t$, where $A(L)$ denotes a polynomial in the lag operator, the stationarity of the process depends on the roots of the polynomial equation $A(z) = 0$. If all roots are outside the unit circle, the process is stationary. If any root is equal to or less than 1 in absolute value, the process is not stationary. A root that is equal to 1 in absolute value is called a **unit root**. When a process has a unit root, as (20.02) does, it is said to be **integrated of order one** or $\boldsymbol{I(1)}$. A series that is $I(1)$ must be differenced once in order to make it stationary.

The obvious way to choose between (20.01) and (20.02) is to nest them both within a more general model. There is more than one way to do so. The most plausible model that includes both (20.01) and (20.02) as special cases is arguably

$$\begin{aligned} y_t &= \gamma_0 + \gamma_1 t + v_t; \;\; v_t = \alpha v_{t-1} + u_t \\ &= \gamma_0 + \gamma_1 t + \alpha\big(y_{t-1} - \gamma_0 - \gamma_1(t-1)\big) + u_t, \end{aligned} \tag{20.03}$$

where u_t follows a stationary process. This model was advocated by Bhargava (1986). When $|\alpha| < 1$, (20.03) is equivalent to the trend-stationary model (20.01); when $\alpha = 1$, it reduces to (20.02).

Because (20.03) is nonlinear in the parameters, it is convenient to reparametrize it as

$$y_t = \beta_0 + \beta_1 t + \alpha y_{t-1} + u_t, \tag{20.04}$$

where

$$\beta_0 \equiv \gamma_0(1-\alpha) + \gamma_1\alpha \quad \text{and} \quad \beta_1 \equiv \gamma_1(1-\alpha).$$

It is easy to verify that the estimates of α from least squares estimation of (20.03) and (20.04) will be identical, as will the estimated standard errors of those estimates if, in the case of (20.03), the latter are based on the Gauss-Newton regression. The only problem with the reparametrization (20.04) is that it hides the important fact that $\beta_1 = 0$ when $\alpha = 1$.

If y_{t-1} is subtracted from both sides, equation (20.04) becomes

$$\Delta y_t = \beta_0 + \beta_1 t + (\alpha - 1)y_{t-1} + u_t, \tag{20.05}$$

where Δ is the first-difference operator. If $\alpha < 1$, (20.05) is equivalent to the model (20.01), whereas, if $\alpha = 1$, it is equivalent to (20.02). Thus it is conventional to test the null hypothesis that $\alpha = 1$ against the one-sided alternative that $\alpha < 1$. Since this is a test of the null hypothesis that there is a unit root in the stochastic process which generates y_t, such tests are commonly called **unit root tests**.

At first glance, it might appear that a unit root test could be accomplished simply by using the ordinary t statistic for $\alpha - 1 = 0$ in (20.05), but this is not so. When $\alpha = 1$, the process generating y_t is integrated of order one. This means that y_{t-1} will not satisfy the standard assumptions needed