Chapter 7

Asymptotic Least Squares Theory: Part II

In the preceding chapter the asymptotic properties of the OLS estimator were derived under "standard" regularity conditions that require data to obey suitable LLN and CLT. Some important consequences of these conditions include root-T consistency and asymptotic normality of the OLS estimator. There are, however, various data that do not obey an LLN nor a CLT. For instance, the simple average of the random variables that follow a random walk diverges, as shown in Example 5.31. Moreover, the data that behave similarly to random walks are also not governed by LLN or CLT. Such data are said to be *integrated of order one*, denoted as $I(1)$; a precise definition of $I(1)$ series will be given in Section 7.1. $I(1)$ data are practically relevant because, since the seminal work of Nelson and Plosser (1982), it has been well documented in the literature that many economic and financial time series are better characterized as $I(1)$ variables.

This chapter is mainly concerned with estimation and hypothesis testing in linear regressions that involve $I(1)$ variables. When data are $I(1)$, the results of Section 6.2 are not applicable, and the asymptotic properties of the OLS estimator must be analyzed differently. As will be shown in subsequent sections, the OLS estimator remains consistent but with a faster convergence rate. Moreover, the normalized OLS estimator has a non-standard distribution in the limit which may be quite different from the normal distribution. This suggests that one should be careful in drawing statistical inferences from regressions with $I(1)$ variables. As far as economic interpretation is concerned, regressions with $I(1)$ variables are closely related economic equilibrium relations and hence play an important role in empirical studies. A more detailed analysis of $I(1)$ variables can be found in, e.g., Hamilton (1994).

7.1 $I(1)$ Variables

A time series $\{y_t\}$ is said to be $I(1)$ if it can be expressed as $y_t = y_{t-1} + \epsilon_t$ with ϵ_t satisfying the following condition.

[C1] $\{\epsilon_t\}$ is a weakly stationary process with mean zero and variance σ_ϵ^2 and obeys an FCLT:

$$\frac{1}{\sigma_* \sqrt{T}} \sum_{t=1}^{[Tr]} \epsilon_t = \frac{1}{\sigma_* \sqrt{T}} y_{[Tr]} \Rightarrow w(r), \qquad 0 \le r \le 1,$$

where w is the standard Wiener process, and

$$\sigma_*^2 = \lim_{T \to \infty} \operatorname{var}\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \epsilon_t \right),$$

which is known as the *long-run variance* of ϵ_t.

This definition is not the most general but is convenient for subsequent analysis. In view of (6.5), we can write

$$\operatorname{var}\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \epsilon_t \right) = \operatorname{var}(\epsilon_t) + 2 \sum_{j=1}^{T-1} \operatorname{cov}(\epsilon_t, \epsilon_{t-j}).$$

The existence of σ_*^2 implies that the variance of $\sum_{t=1}^{T} \epsilon_t$ is $O(T)$. Moreover, $\operatorname{cov}(\epsilon_t, \epsilon_{t-j})$ must be summable so that $\operatorname{cov}(\epsilon_t, \epsilon_{t-j})$ vanishes when j tends to infinity. For simplicity, a weakly stationary process will also be referred to as an $I(0)$ series. Under our definition, partial sums of an $I(0)$ series (e.g., $\sum_{i=1}^{t} \epsilon_i$) form an $I(1)$ series, while taking first difference of an $I(1)$ series (e.g., $y_t - y_{t-1}$) yields an $I(0)$ series. Note that A random walk is $I(1)$ with i.i.d. ϵ_t and $\sigma_*^2 = \sigma_\epsilon^2$. When $\{\epsilon_t\}$ is a stationary ARMA(p, q) process, $\{y_t\}$ is also known as an ARIMA$(p, 1, q)$ process. For example, many empirical evidences showing that stock prices and GDP are $I(1)$, and so are their log transformations. Yet, stock returns and GDP growth rates are found to be $I(0)$.

Analogous to a random walk, an $I(1)$ series y_t has mean zero and variance increasing linearly with t, and its autocovariances $\operatorname{cov}(y_t, y_s)$ do not decrease when $|t-s|$ increases, cf. Example 5.31. In contrast, an $I(0)$ series ϵ_t has a bounded variance, and $\operatorname{cov}(\epsilon_t, \epsilon_s)$ decays to zero when $|t-s|$ becomes large. Thus, an $I(1)$ series has increasingly large variations and smooth sample paths, yet an $I(0)$ series is not as smooth as $I(1)$ series and has smaller variations. To illustrate, we plot in Figure 7.1 two sample paths of an ARIMA process: $y_t = y_{t-1} + \epsilon_t$, where $\epsilon_t = 0.3\epsilon_{t-1} + u_t - 0.4u_{t-1}$ with u_t i.i.d. $\mathcal{N}(0, 1)$.

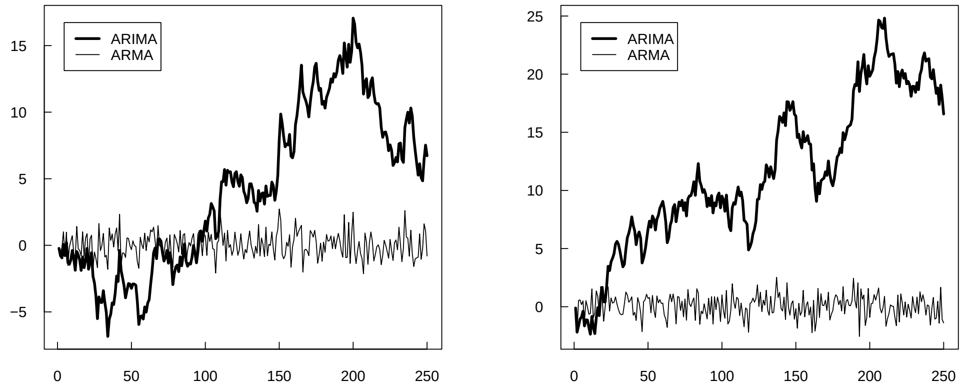


Figure 7.1: Sample paths of ARIMA and ARMA series.

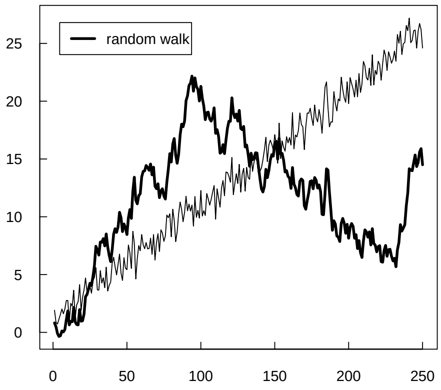
For comparison, we also include the sample paths of ϵ_t in the figure. It can be seen that ARIMA paths (thick lines) wander away from the mean level and exhibit large swings over time, whereas the ARMA paths (thin lines) are jagged and fluctuate around the mean level.

During a given time period, an $I(1)$ series may look like a series that follows a deterministic time trend: $y_t = a_o + b_o t + \epsilon_t$, where ϵ_t are $I(0)$. Such a series, also known as a trend stationary series, has finite variance and becomes stationary when the trend function is removed. It should be noted that an $I(1)$ series is fundamentally different from a trend stationary series, in that the former has unbounded variance and its behavior can not be captured by a deterministic trend function. Figure 7.2 illustrates the difference between the sample paths of a random walk $y_t = y_{t-1} + \epsilon_t$ and a trend stationary series $y_t = 1+0.1t+\epsilon_t$, where ϵ_t are i.i.d. $\mathcal{N}(0,1)$. It is clear that the variation of random walk paths is much larger than that of trend stationary paths.

7.2 Autoregression of an $I(1)$ Variable

Given the specification $y_t = \boldsymbol{x}_t'\boldsymbol{\beta}+e_t$ such that [B2] holds: $y_t = \boldsymbol{x}_t'\boldsymbol{\beta}_o+\epsilon_t$ with $\mathbb{E}(\boldsymbol{x}_t\epsilon_t) = \mathbf{0}$, the OLS estimator of $\boldsymbol{\beta}$ can be expressed as

$$\hat{\boldsymbol{\beta}}_T = \boldsymbol{\beta}_o + \left(\sum_{t=1}^{T} \boldsymbol{x}_t\boldsymbol{x}_t'\right)^{-1} \left(\sum_{t=1}^{T} \boldsymbol{x}_t\epsilon_t\right).$$



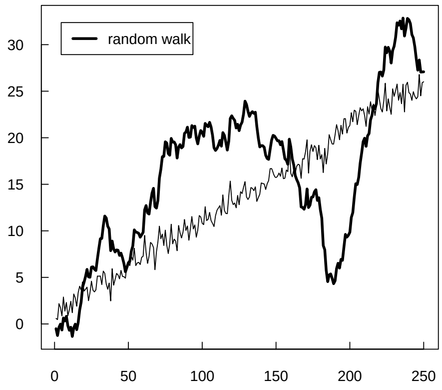


Figure 7.2: Sample paths of random walk and trend stationary series.

A generic approach to establishing OLS consistency is to show that, under suitable conditions, the second term on the right-hand side converges to zero in some probabilistic sense. Theorem 6.1 ensures this by imposing [B1] on data. With [B1] and [B2], we have $\sum_{t=1}^{T} \boldsymbol{x}_t \boldsymbol{x}_t' = O_{\mathbb{P}}(T)$ and $\sum_{t=1}^{T} \boldsymbol{x}_t \epsilon_t = o_{\mathbb{P}}(T)$, so that

$$\hat{\boldsymbol{\beta}}_T = \boldsymbol{\beta}_o + o_{\mathbb{P}}(1),$$

showing weak consistency of $\hat{\boldsymbol{\beta}}_T$. When [B3] also holds, the CLT effect yields a more precise order of $\sum_{t=1}^{T} \boldsymbol{x}_t \epsilon_t$, namely, $O_{\mathbb{P}}(T^{1/2})$. It follows that

$$\hat{\boldsymbol{\beta}}_T = \boldsymbol{\beta}_o + O_{\mathbb{P}}(T^{-1/2}),$$

which establishes root-T consistency of $\hat{\boldsymbol{\beta}}_T$. On the other hand, the asymptotic properties of the OLS estimator must be derived without resorting to LLN and CLT when y_t and $\boldsymbol{x}_t$ are $I(1)$.

7.2.1 Asymptotic Properties of the OLS Estimator

To illustrate, we first consider the simplest AR(1) specification:

$$y_t = \alpha y_{t-1} + e_t. \tag{7.1}$$

Suppose that $\{y_t\}$ is a random walk such that $y_t = \alpha_o y_{t-1} + \epsilon_t$ with $\alpha_o = 1$ and ϵ_t i.i.d. random variables with mean zero and variance σ_ϵ^2. From Examples 5.31 we know that $\{y_t\}$ does not obey a LLN. Moreover, $\sum_{t=2}^T y_{t-1}\epsilon_t = O_{\mathbb{P}}(T)$ by Example 5.32, and $\sum_{t=2}^T y_{t-1}^2 = O_{\mathbb{P}}(T^2)$ by Example 5.43. The OLS estimator of α is thus

$$\hat{\alpha}_T = 1 + \frac{\sum_{t=2}^T y_{t-1}\epsilon_t}{\sum_{t=2}^T y_{t-1}^2} = 1 + O_{\mathbb{P}}(T^{-1}). \tag{7.2}$$

This shows that $\hat{\alpha}_T$ converges to $\alpha_o = 1$ at the rate T^{-1}, which is in sharp contrast with the convergence rate of the OLS estimator discussed in Chapter 6. Thus, the estimator $\hat{\alpha}_T$ is a T-consistent estimator and also known as a *super consistent* estimator.

When y_t is an $I(1)$ series, it is straightforward to derive the following asymptotic results; all proofs are deferred to the Appendix.

Lemma 7.1 *Let $y_t = y_{t-1} + \epsilon_t$ be an $I(1)$ series with ϵ_t satisfying [C1]. Then,*

(i) $T^{-3/2}\sum_{t=1}^T y_{t-1} \Rightarrow \sigma_* \int_0^1 w(r)\,\mathrm{d}r$;

(ii) $T^{-2}\sum_{t=1}^T y_{t-1}^2 \Rightarrow \sigma_*^2 \int_0^1 w(r)^2\,\mathrm{d}r$;

(iii) $T^{-1}\sum_{t=1}^T y_{t-1}\epsilon_t \Rightarrow \frac{1}{2}[\sigma_*^2 w(1)^2 - \sigma_\epsilon^2] = \sigma_*^2 \int_0^1 w(r)\,\mathrm{d}w(r) + \frac{1}{2}(\sigma_*^2 - \sigma_\epsilon^2)$.

As Lemma 7.1 holds for general $I(1)$ processes, the assertions (i) and (ii) are generalizations of the results in Example 5.43. The assertion (iii) is also a more general result than Example 5.32 and gives a precise limit of $\sum_{t=2}^T y_{t-1}\epsilon_t/T$. The weak limit of the normalized OLS estimator of α,

$$T(\hat{\alpha}_T - 1) = \frac{\sum_{t=2}^T y_{t-1}\epsilon_t/T}{\sum_{t=2}^T y_{t-1}^2/T^2},$$

now can be derived from Lemma 7.1.

Theorem 7.2 *Let $y_t = y_{t-1} + \epsilon_t$ be an $I(1)$ series with ϵ_t satisfying [C1]. Given the specification $y_t = \alpha y_{t-1} + e_t$, the normalized OLS estimator of α is such that*

$$T(\hat{\alpha}_T - 1) \Rightarrow \frac{\frac{1}{2}\left[w(1)^2 - \sigma_\epsilon^2/\sigma_*^2\right]}{\int_0^1 w(r)^2\,\mathrm{d}r}.$$

where w is the standard Wiener process. In particular, when y_t is a random walk,

$$T(\hat{\alpha}_T - 1) \Rightarrow \frac{\frac{1}{2}\left[w(1)^2 - 1\right]}{\int_0^1 w(r)^2\,\mathrm{d}r},$$

which does not depend on σ_ϵ^2 and σ_*^2.

Theorem 7.2 shows that for an autoregression of $I(1)$ variables, the OLS estimator is T-consistent and has a non-standard distribution in the limit. It is worth mentioning that $T(\hat{\alpha}_T - 1)$ has a weak limit depending on the (nuisance) parameters σ_ϵ^2 and σ_*^2 in general and hence is not asymptotically pivotal, unless ϵ_t are i.i.d. It is also interesting to observe from Theorem 7.2 that OLS consistency is not affected even when y_{t-1} and correlated ϵ_t are both present, in contrast with Example 6.5. This is the case because $\sum_{t=2}^T y_{t-1}^2$ grows much too fast (at the rate T^2) and hence is able to wipe out the effect of $\sum_{t=2}^T y_{t-1}\epsilon_t$ (which grows at the rate T) when T becomes large.

Consider now the specification with a constant term:

$$y_t = c + \alpha y_{t-1} + e_t, \tag{7.3}$$

and the OLS estimators $\hat{c}_T$ and $\hat{\alpha}_T$. Define $\bar{y}_{-1} = \sum_{t=1}^{T-1} y_t/(T-1)$. The lemma below is analogous to Lemma 7.1.

Lemma 7.3 *Let $y_t = y_{t-1} + \epsilon_t$ be an $I(1)$ series with ϵ_t satisfying [C1]. Then,*

(i) $T^{-2}\sum_{t=1}^T (y_{t-1} - \bar{y}_{-1})^2 \Rightarrow \sigma_*^2 \int_0^1 w^*(r)^2 \,\mathrm{d}r;$

(ii) $T^{-1}\sum_{t=1}^T (y_{t-1} - \bar{y}_{-1})\epsilon_t \Rightarrow \sigma_*^2 \int_0^1 w^*(r)\,\mathrm{d}w(r) + \frac{1}{2}(\sigma_*^2 - \sigma_\epsilon^2),$

where w is the standard Wiener process and $w^*(t) = w(t) - \int_0^1 w(r)\, dr$.

Lemma 7.3 is concerned with "de-meaned" y_t, i.e., $y_t - \bar{y}$. In analogy with this term, the process w^* is also known as the "de-meaned" Wiener process. It can be seen that the sum of squares of de-meaned y_t also grows at the rate T^2 and that the sum of the products of de-meaned y_t and ϵ_t grows at the rate T. These rates are the same as those based on y_t, as shown in Lemma 7.1. The consistency of the OLS estimator now can be easily established, as in Theorem 7.2.

Theorem 7.4 *Let $y_t = y_{t-1} + \epsilon_t$ be an $I(1)$ series with ϵ_t satisfying [C1]. Given the specification $y_t = c + \alpha y_{t-1} + e_t$, the normalized OLS estimators of α and c are such that*

$$T(\hat{\alpha}_T - 1) \Rightarrow \frac{\int_0^1 w^*(r)\,\mathrm{d}w(r) + \frac{1}{2}(1 - \sigma_\epsilon^2/\sigma_*^2)}{\int_0^1 w^*(r)^2\,\mathrm{d}r} =: A,$$

$$\sqrt{T}\hat{c}_T \Rightarrow A\left(\sigma_* \int_0^1 w(r)\,\mathrm{d}r\right) + \sigma_* w(1).$$

where w is the standard Wiener process and $w^*(t) = w(t) - \int_0^1 w(r)\, dr$. In particular, when y_t is a random walk,

$$T(\hat{\alpha}_T - 1) \Rightarrow \frac{\int_0^1 w^*(r)\,\mathrm{d}w(r)}{\int_0^1 w^*(r)^2\,\mathrm{d}r}.$$

Theorem 7.4 shows again that, for the autoregression of an $I(1)$ variable that contains a constant term, the normalized OLS estimators are not asymptotically pivotal unless ϵ_t are i.i.d. It should be emphasized that the OLS estimators of the intercept and slope coefficients have different rates of convergence. The estimator of the latter is T-consistent, whereas the estimator of the former remains root-T consistent but does not have a limiting normal distribution.

Remarks:

1. While the asymptotic normality of the OLS estimator obtained under standard conditions is invariant with respect to model specifications, Theorems 7.4 and 7.2 indicate that the limiting results for autoregressions with an $I(1)$ variable are not. This is a great disadvantage because the asymptotic analysis need to be carried out for different specifications when the data are $I(1)$ series.

2. All the results in this sub-section are based on the data $y_t = y_{t-1} + \epsilon_t$ which does not involve an intercept. These results would break down if $y_t = c_o + y_{t-1} + \epsilon_t$ with a non-zero c_o; such series are said to be $I(1)$ with *drift*. It is easily seen that, when a drift is present,

$$y_t = c_o\, t + \sum_{i=1}^{t} \epsilon_i,$$

which contains a deterministic trend and an $I(1)$ series without drift. Such series exhibits large swings around the trend function and has much larger variation than $I(1)$ series without drift. See Exercise 7.2 for some properties of this series.

7.2.2 Tests of Unit Root

What we have learnt from the preceding subsections are: (1) The behavior of an $I(1)$ series is quite different from that of an $I(0)$ series, and (2) the presence of an $I(1)$ variable in a regression renders standard asymptotic results invalid. It is thus practically important to determine whether the data are in fact $I(1)$. Given the specifications (7.1) and (7.3), the hypothesis of interest is $\alpha_o = 1$; tests of this hypothesis are usually referred to as tests of *unit root.*

A leading unit-root test is the t statistic of $\alpha_o = 1$ in the specification (7.1) or (7.3). For the former, the t statistic is

$$\tau_0 = \frac{\left(\sum_{t=2}^T y_{t-1}^2\right)^{1/2}(\hat{\alpha}_T - 1)}{\hat{\sigma}_T},$$

where $\hat{\sigma}_T^2 = \sum_{t=2}^T (y_t - \hat{\alpha}_T y_{t-1})^2/(T-2)$ is the standard OLS variance estimator; for the latter, the t statistic is

$$\tau_c = \frac{\left[\sum_{t=2}^T (y_{t-1} - \bar{y}_{-1})^2\right]^{1/2}(\hat{\alpha}_T - 1)}{\hat{\sigma}_T},$$

where $\hat{\sigma}_T^2 = \sum_{t=2}^T (y_t - \hat{c}_T - \hat{\alpha}_T y_{t-1})^2/(T-3)$. In view of Theorem 7.2 and Theorem 7.4, it is easy to derive the weak limits of these statistics under the null hypothesis of $\alpha_o = 1$.

Theorem 7.5 *Let y_t be a random walk. Given the specifications (7.1) and (7.3), we have, respectively,*

$$\begin{aligned} \tau_0 &\Rightarrow \frac{\frac{1}{2}[w(1)^2 - 1]}{\left[\int_0^1 w(r)^2 \, \mathrm{d}r\right]^{1/2}}, \\ \tau_c &\Rightarrow \frac{\int_0^1 w^*(r) \, \mathrm{d}w(r)}{\left[\int_0^1 w^*(r)^2 \, \mathrm{d}r\right]^{1/2}}, \end{aligned} \tag{7.4}$$

where w is the standard Wiener process and $w^*(t) = w(t) - \int_0^1 w(r) \, dr$.

These statistics were first analyzed by Dickey and Fuller (1979) and their weak limits were derived in Phillips (1987). In addition to the specifications (7.1) and (7.3), Dickey and Fuller (1979) also considered the specification with the intercept and a time trend:

$$y_t = c + \alpha y_{t-1} + \beta\left(t - \frac{T}{2}\right) + e_t; \tag{7.5}$$

the t-statistic of $\alpha_o = 1$ is denoted as τ_t. The weak limit of τ_t is different from those in Theorem 7.5 but can be derived similarly; we omit the detail. The t-statistics τ_0, τ_c, and τ_t and the F tests considered in Dickey and Fuller (1981) are now known as the *Dickey-Fuller tests.*

The limiting distributions of the Dickey-Fuller tests are all non-standard but can be easily simulated; see e.g., Fuller (1996, p. 642) and Davidson and MacKinnon (1993,

Table 7.1: Some percentiles of the Dickey-Fuller distributions.

Test	1%	2.5%	5%	10%	50%	90%	95%	97.5%	99%
τ_0	−2.58	−2.23	−1.95	−1.62	−0.51	0.89	1.28	1.62	2.01
τ_c	−3.42	−3.12	−2.86	−2.57	−1.57	−0.44	−0.08	0.23	0.60
τ_t	−3.96	−3.67	−3.41	−3.13	−2.18	−1.25	−0.94	−0.66	−0.32

p. 708). These distributions will be referred to as the Dickey-Fuller distributions; some of their percentiles reported in Fuller (1996) are summarized in Table 7.1. We can see that these distributions are not symmetric about zero and are all skewed to the left. In particular, τ_c assumes negatives values about 95% of times, and τ_t is virtually a non-positive random variable.

To implement the Dickey-Fuller tests, we may, corresponding to the specifications (7.1), (7.3) and (7.5), estimate one of the following specifications:

$$\begin{aligned} \Delta y_t &= \theta y_{t-1} + e_t, \\ \Delta y_t &= c + \theta y_{t-1} + e_t, \\ \Delta y_t &= c + \theta y_{t-1} + \beta\Big(t - \frac{T}{2}\Big) + e_t, \end{aligned} \tag{7.6}$$

where $\Delta y_t = y_t - y_{t-1}$. Clearly, the hypothesis $\theta_o = 0$ for these specifications is equivalent to $\alpha_o = 0$ for (7.1), (7.3) and (7.5). It is also easy to verify that the weak limits of the normalized estimators in (7.6), $T\hat{\theta}_T$, are the same as the respective limits of $T(\hat{\alpha}_T - 1)$ under the null hypothesis. Consequently, the t-ratios of $\theta_o = 0$ also have the Dickey-Fuller distributions with the critical values given in Table 7.1. Using the t-ratios of (7.6) as unit-root tests is convenient in practice because they are routinely reported by econometrics packages.

A major drawback of the Dickey-Fuller tests is that they can only check if the data series is a random walk. When $\{y_t\}$ is a general $I(1)$ process, the dependence of ϵ_t renders the limits of Theorem 7.5 invalid, as shown in the result below.

Theorem 7.6 *Let $y_t = y_{t-1} + \epsilon_t$ be an $I(1)$ series with ϵ_t satisfying [C1]. Then,*

$$\begin{aligned} \tau_0 &\Rightarrow \frac{\sigma_*}{\sigma_\epsilon}\left(\frac{\frac{1}{2}[w(1)^2 - \sigma_\epsilon^2/\sigma_*^2]}{\left[\int_0^1 w(r)^2\,\mathrm{d}r\right]^{1/2}}\right), \\ \tau_c &\Rightarrow \frac{\sigma_*}{\sigma_\epsilon}\left(\frac{\int_0^1 w^*(r)\,\mathrm{d}w(r) + \frac{1}{2}(1 - \sigma_\epsilon^2/\sigma_*^2)}{\left[\int_0^1 w^*(r)^2\,\mathrm{d}r\right]^{1/2}}\right), \end{aligned} \tag{7.7}$$

where w is the standard Wiener process and $w^*(t) = w(t) - \int_0^1 w(r)\ dr$.

Theorem 7.6 includes Theorem 7.5 as a special case because the limits in (7.7) would reduce to those in (7.4) when ϵ_t are i.i.d. (so that $\sigma_*^2 = \sigma_\epsilon^2$). These results also suggest that the nuisance parameters σ_ϵ^2 and σ_*^2 may be eliminated by proper corrections of the statistics τ_0 and τ_c.

Let $\hat{e}_t$ denote the OLS residuals of the specification (7.1) or (7.3) and s_{Tn}^2 denote a Newey-West type estimator of σ_*^2 based on $\hat{e}_t$:

$$s_{Tn}^2 = \frac{1}{T-1}\sum_{t=2}^{T}\hat{e}_t^2 + \frac{2}{T-1}\sum_{s=1}^{T-2}\kappa\left(\frac{s}{n}\right)\sum_{t=s+2}^{T}\hat{e}_t\hat{e}_{t-s},$$

with κ a kernel function and $n = n(T)$ its bandwidth; see Section 6.3.2. Phillips (1987) proposed the following modified τ_0 and τ_c statistics:

$$\begin{aligned} Z(\tau_0) &= \frac{\hat{\sigma}_T}{s_{Tn}}\,\tau_0 - \frac{\frac{1}{2}(s_{Tn}^2 - \hat{\sigma}_T^2)}{s_{Tn}\left(\sum_{t=2}^{T} y_{t-1}^2/T^2\right)^{1/2}}, \\ Z(\tau_c) &= \frac{\hat{\sigma}_T}{s_{Tn}}\,\tau_c - \frac{\frac{1}{2}(s_T^2 - \hat{\sigma}_T^2)}{s_{Tn}\left[\sum_{t=2}^{T}(y_{t-1} - \bar{y}_{-1})^2\right]^{1/2}}; \end{aligned}$$

see also Phillips and Perron (1988) for the modifications of other unit-root tests. The Z-type tests are now known as the *Phillips-Perron tests.* It is quite easy to verify that the limits of $Z(\tau_0)$ and $Z(\tau_c)$ are those in (7.4) and do not depend on the nuisance parameters. Thus, the Phillips-Perron tests are asymptotically pivotal and capable of testing whether $\{y_t\}$ is a general $I(1)$ series.

Corollary 7.7 *Let $y_t = y_{t-1} + \epsilon_t$ be an $I(1)$ series with ϵ_t satisfying [C1]. Then,*

$$\begin{aligned} Z(\tau_0) &\Rightarrow \frac{\frac{1}{2}\left[w(1)^2 - 1\right]}{\left[\int_0^1 w(r)^2\,\mathrm{d}r\right]^{1/2}}, \\ Z(\tau_c) &\Rightarrow \frac{\int_0^1 w^*(r)\,\mathrm{d}w(r)}{\left[\int_0^1 w^*(r)^2\,\mathrm{d}r\right]^{1/2}}. \end{aligned}$$

where w is the standard Wiener process and $w^*(t) = w(t) - \int_0^1 w(r)\ dr$.

Said and Dickey (1984) introduced a different approach to circumventing the nuisance parameters in the limit. Note that the correlations in a weakly stationary process may be "filtered out" by a linear AR model with a proper order, say, k. For example, when $\{\epsilon_t\}$ is a weakly stationary ARMA process,

$$\epsilon_t - \gamma_1 \epsilon_{t-1} - \cdots - \gamma_k \epsilon_{t-k} = u_t$$

are serially uncorrelated for some k and some parameters $\gamma_1, \ldots, \gamma_k$. Basing on this idea Said and Dickey (1984) suggested the following "augmented" specifications:

$$\begin{aligned}
\Delta y_t &= \theta y_{t-1} + \sum_{j=1}^{k} \gamma_j \Delta y_{t-j} + e_t, \\
\Delta y_t &= c + \theta y_{t-1} + \sum_{j=1}^{k} \gamma_j \Delta y_{t-j} + e_t, \\
\Delta y_t &= c + \theta y_{t-1} + \beta\Big(t - \frac{T}{2}\Big) + \sum_{j=1}^{k} \gamma_j \Delta y_{t-j} + e_t,
\end{aligned} \tag{7.8}$$

where γ_j are unknown parameters. Compared with the specifications in (7.6), the augmented regressions in (7.8) contain k lagged differences Δy_{t-j}, $j = 1, \ldots, k$, which are ϵ_{t-j} under the null hypothesis. These differences are included to capture possible correlations among ϵ_t. After controlling these correlations, the resulting t-ratios of $\theta_o = 0$ turn out to have the Dickey-Fuller distributions, as the t-ratios for (7.6), and are known as the *augmented Dickey-Fuller tests*. Compared with the Phillips-Perron tests, these tests are capable of testing whether $\{y_t\}$ is a general $I(1)$ series without non-parametric kernel estimation for σ_*^2. Yet, one must choose a proper lag order k for the augmented specifications in (7.8).

7.3 Tests of Stationarity against $I(1)$

Instead of testing $I(1)$ series directly, Kwiatkowski, Phillips, Schmidt, and Shin (1992) proposed testing the property of stationarity against $I(1)$ seires. Their tests, obtained along the line in Nabeya and Tanaka (1988), are now known as the KPSS test.

Recall that the process $\{y_t\}$ is said to be *trend stationary* if

$$y_t = a_o + b_o t + \epsilon_t,$$

where ϵ_t satisfy [C1]. That is, y_t fluctuates around a deterministic trend function. When $b_o = 0$, the resulting process is a *level stationary* process, in the sense that it moves around the mean level a_o. A trend stationary process achieves stationarity by removing the deterministic trend, whereas a level stationary process is itself stationary and hence an $I(0)$ series. The KPSS test is of the following form:

$$\eta_T = \frac{1}{T^2 \, s^2_{Tn}} \sum_{t=1}^{T} \left(\sum_{i=1}^{t} \hat{e}_i \right)^2 ,$$

where s^2_{Tn} is, again, a Newey-West estimator of σ^2_*, computed using the model residuals $\hat{e}_t$. To test the null of trend stationarity, $\hat{e}_t = y_t - \hat{a}_T - \hat{b}_T \, t$ are the residuals of regressing y_t on the constant one and the time trend t. For the null of level stationarity, $\hat{e}_t = y_t - \bar{y}$ are the residuals of regressing y_t on the constant one.

To see the null limit of η_T, consider first the level stationary process $y_t = a_o + \epsilon_t$. The partial sums of $\hat{e}_t = y_t - \bar{y}$ are such that

$$\sum_{t=1}^{[Tr]} \hat{e}_t = \sum_{t=1}^{[Tr]} (\epsilon_t - \bar{\epsilon}) = \sum_{t=1}^{[Tr]} \epsilon_t - \frac{[Tr]}{T} \sum_{t=1}^{T} \epsilon_t, \quad r \in (0,1].$$

Then by a suitable FCLT,

$$\begin{aligned} \frac{1}{\sigma_* \sqrt{T}} \sum_{t=1}^{[Tr]} \hat{e}_t &= \frac{1}{\sigma_* \sqrt{T}} \sum_{t=1}^{[Tr]} \epsilon_t - \frac{[Tr]}{T} \left(\frac{1}{\sigma_* \sqrt{T}} \sum_{t=1}^{T} \epsilon_t \right) \\ &\Rightarrow w(r) - r w(1), \quad r \in (0,1]. \end{aligned}$$

That is, properly normalized partial sums of the residuals behave like a Brownian bridge with $w^0(r) = w(r) - r w(1)$. Similarly, given the trend stationary process $y_t = a_0 + b_0 \, t + \epsilon_t$, let $\hat{e}_t = y_t - \hat{a}_T - \hat{b}_T \, t$ denote the OLS residuals. Then, it can be shown that

$$\frac{1}{\sigma_* \sqrt{T}} \sum_{t=1}^{[Tr]} \hat{e}_t \Rightarrow w(r) + (2r - 3r^2) w(1) - (6r - 6r^2) \int_0^1 w(s) \, \mathrm{d}s, \quad r \in (0,1];$$

see Exercise 7.4. Note that the limit on the right is a functional of the standard Wiener process which, similar to a Brownian bridge, is a "tide-down" process in the sense that it is zero with probability one at $r = 1$.

The limits of η_T under the null hypothesis are summarized in the following theorem; some percentiles of their distributions are collected in Table 7.2.

Theorem 7.8 *let $y_t = a_o + \epsilon_t$ be a level stationary process with ϵ_t satisfying [C1]. Then, η_T computed from $\hat{e}_t = y_t - \bar{y}$ is such that*

$$\eta_T \Rightarrow \int_0^1 w^0(r)^2 \, \mathrm{d}r,$$

Table 7.2: Some percentiles of the distributions of the KPSS test.

Test	1%	2.5%	5%	10%
level stationarity	0.739	0.574	0.463	0.347
trend stationarity	0.216	0.176	0.146	0.119

where w^0 is the Brownian bridge. let $y_t = a_o + b_o t + \epsilon_t$ be a trend stationary process with ϵ_t satisfying [C1]. Then, η_T computed from from the OLS residuals $\hat{e}_t = y_t - \hat{a}_T - \hat{b}_T t$ is such that

$$\eta_T \Rightarrow \int_0^1 f(r)^2 \,\mathrm{d}r,$$

where $f(r) = w(r) + (2r - 3r^2)w(1) - (6r - 6r^2)\int_0^1 w(s)\,\mathrm{d}s$ and w is the standard Wiener process.

These tests have power against $I(1)$ series because η_T would diverge under $I(1)$ alternatives. This is the case because T^2 in η_T is not be a proper normalizing factor when the data are $I(1)$. It is worth mentioning that the KPSS tests also have power against other alternatives, such as stationarity with mean changes and trend stationarity with trend breaks. Thus, rejecting the null of stationarity does not imply that the series being tested must be $I(1)$.

7.4 Regressions of $I(1)$ Variables

From the preceding section we have seen that the asymptotic behavior of the OLS estimator in an auotregression changes dramatically when $\{y_t\}$ is an $I(1)$ series. It is then reasonable to expect that the OLS asymptotics would also be quite different from that in Chapter 6 if the dependent variable and regressors of a regression model are both $I(1)$ series.

7.4.1 Spurious Regressions

In a classical simulation study, Granger and Newbold (1974) found that, while two independent random walks should have no relationship whatsoever, regressing one random walk on the other typically yields a significant t-ratio. Thus, one would falsely reject the null hypothesis of no relationship between two independent random walks. This is known as the problem of *spurious regression*. Phillips (1986) provided analytic results showing why such a spurious inference may arise.

To illustrate, we consider a simple linear specification:

$$y_t = \alpha + \beta x_t + e_t.$$

Let $\hat{\alpha}_T$ and $\hat{\beta}_T$ denote the OLS estimators for α and β, respectively. Also denote their t-ratios as $t_\alpha = \hat{\alpha}_T/s_\alpha$ and $t_\beta = \hat{\beta}_T/s_\beta$, where s_α and s_β are the OLS standard errors for $\hat{\alpha}_T$ and $\hat{\beta}_T$. We are interested in the case that $\{y_t\}$ and $\{x_t\}$ are $I(1)$ series: $y_t = y_{t-1}+u_t$ and $x_t = x_{t-1} + v_t$, where $\{u_t\}$ and $\{v_t\}$ are mutually independent processes satisfying the following condition.

[C2] $\{u_t\}$ and $\{v_t\}$ are two weakly stationary processes have mean zero and respective variances σ_u^2 and σ_v^2 and both obey an FCLT with respective long-run variances:

$$\sigma_y^2 = \lim_{T\to\infty} \frac{1}{T} \operatorname{IE}\left(\sum_{t=1}^T u_t\right)^2, \qquad \sigma_x^2 = \lim_{T\to\infty} \frac{1}{T} \operatorname{IE}\left(\sum_{t=1}^T v_t\right)^2.$$

In the light of Lemma 7.1, the limits below are immediate:

$$\frac{1}{T^{3/2}} \sum_{t=1}^T y_t \Rightarrow \sigma_y \int_0^1 w_y(r)\,\mathrm{d}r,$$
$$\frac{1}{T^2} \sum_{t=1}^T y_t^2 \Rightarrow \sigma_y^2 \int_0^1 w_y(r)^2\,\mathrm{d}r,$$

where w_y is a standard Wiener processes. Similarly,

$$\frac{1}{T^{3/2}} \sum_{t=1}^T x_t \Rightarrow \sigma_x \int_0^1 w_x(r)\,\mathrm{d}r,$$
$$\frac{1}{T^2} \sum_{t=1}^T x_t^2 \Rightarrow \sigma_x^2 \int_0^1 w_x(r)^2\,\mathrm{d}r,$$

where w_x is also a standard Wiener process which, due to mutual independence between $\{u_t\}$ and $\{v_t\}$, is independent of w_y. As in Lemma 7.3, we also have

$$\frac{1}{T^2} \sum_{t=1}^T (y_t - \bar{y})^2 \Rightarrow \sigma_y^2 \int_0^1 w_y(r)^2\,\mathrm{d}r - \sigma_y^2 \left(\int_0^1 w_y(r)\,\mathrm{d}r\right)^2 =: \sigma_y^2 m_y,$$
$$\frac{1}{T^2} \sum_{t=1}^T (x_t - \bar{x})^2 \Rightarrow \sigma_x^2 \int_0^1 w_x(r)^2\,\mathrm{d}r - \sigma_x^2 \left(\int_0^1 w_x(r)\,\mathrm{d}r\right)^2 =: \sigma_x^2 m_x,$$

where $w_y^*(t) = w_y(t) - \int_0^1 w_y(r)\,\mathrm{d}r$ and $w_x^*(t) = w_x(t) - \int_0^1 w_x(r)\,\mathrm{d}r$ are two mutually independent, "de-meaned" Wiener processes. Analogous to the limits above, it is also easy to show that

$$\begin{aligned} &\frac{1}{T^2}\sum_{t=1}^{T}(y_t - \bar{y})(x_t - \bar{x}_t) \\ &\qquad \Rightarrow \sigma_y\sigma_x\left(\int_0^1 w_y(r)w_x(r)\,\mathrm{d}r - \int_0^1 w_y(r)\,\mathrm{d}r\int_0^1 w_x(r)\,\mathrm{d}r\right) \\ &\qquad =: \sigma_y\sigma_x m_{yx}. \end{aligned}$$

The following results on $\hat{\alpha}_T$, $\hat{\beta}_T$ and their t-ratios now can be easily derived from the limits above.

Theorem 7.9 *Let $y_t = y_{t-1} + u_t$ and $x_t = x_{t-1} + v_t$, where $\{u_t\}$ and $\{v_t\}$ are two mutually independent processes satisfying [C2]. Then for the specification $y_t = \alpha + \beta x_t + e_t$, we have:*

(i) $\hat{\beta}_T \Rightarrow \dfrac{\sigma_y\, m_{yx}}{\sigma_x\, m_x}$,

(ii) $T^{-1/2}\hat{\alpha}_T \Rightarrow \sigma_y\left(\displaystyle\int_0^1 w_y(r)\,\mathrm{d}r - \frac{m_{yx}}{m_x}\int_0^1 w_x(r)\,\mathrm{d}r\right)$,

(iii) $T^{-1/2}\,t_\beta \Rightarrow \dfrac{m_{yx}}{(m_y m_x - m_{yx}^2)^{1/2}}$,

(iv) $T^{-1/2}\,t_\alpha \Rightarrow \dfrac{m_x\int_0^1 w_y(r)\,\mathrm{d}r - m_{yx}\int_0^1 w_x(r)\,\mathrm{d}r}{\left[(m_y m_x - m_{yx}^2)\int_0^1 w_x(r)^2\,\mathrm{d}r\right]^{1/2}}$,

where w_x and w_y are two mutually independent, standard Wiener processes.

When y_t and x_t are mutually independent, the true parameters of this regression should be $\alpha_o = \beta_o = 0$. The first two assertions of Theorem 7.9 show, however, that the OLS estimators do not converge in probability to zero. Instead, $\hat{\beta}_T$ has a non-degenerate limiting distribution, whereas $\hat{\alpha}_T$ diverges at the rate $T^{1/2}$. Theorem 7.9(iii) and (iv) further indicate that t_α and t_β both diverge at the rate $T^{1/2}$. Thus, one would easily infer that these coefficients are significantly different from zero if the critical values of the t-ratio were taken from the standard normal distribution. These results together suggest that, when the variables are $I(1)$ series, one should be extremely careful in drawing statistical inferences from the t tests, for the t tests do not have the standard normal distribution in the limit.

Nelson and Kang (1984) also showed that, given the time trend specification:

$$y_t = a + b\,t + e_t,$$

it is likely to falsely infer that the time trend is significant in explaining the behavior of y_t when $\{y_t\}$ is a random walk. This is known as the problem of *spurious trend*. Phillips and Durlauf (1986) analyzed this problem as in Theorem 7.9 and demonstrated that the F test of $b_o = 0$ diverges at the rate T. The divergence of the F test (and hence the t-ratio) explains why an incorrect inference would result.

7.4.2 Cointegration

The results in the preceding sub-section indicate that the relation between $I(1)$ variables found using standard asymptotics may be a spurious one. They do *not* mean that there can be no relation between $I(1)$ variables. In this section, we formally characterize the relations between $I(1)$ variables.

Consider two variables y and x that obey an equilibrium relationship $ay - bx = 0$. With real data (y_t, x_t), $z_t := ay_t - bx_t$ are understood as equilibrium errors because they need not be zero all the time. When y_t and x_t are both $I(1)$, a linear combination of them is, in general, also an $I(1)$ series. Thus, $\{z_t\}$ would be an $I(1)$ series that wanders away from zero and has growing variances over time. If that is the case, $\{z_t\}$ rarely crosses zero (the horizontal axis), so that the equilibrium condition entails little empirical restriction on z_t.

On the other hand, when y_t and x_t are both $I(1)$ but involve the same random walk q_t such that $y_t = q_t + u_t$ and $x_t = cq_t + v_t$, where $\{u_t\}$ and $\{v_t\}$ are two $I(0)$ series. It is then easily seen that

$$z_t = cy_t - x_t = cu_t - v_t,$$

which is a linear combination of $I(0)$ series and hence is also $I(0)$. This example shows that when two $I(1)$ series share the same trending (random walk) component, it is possible to find a linear combination of these series that annihilates the common trend and becomes an $I(0)$ series. In this case, the equilibrium condition is empirically relevant because z_t is $I(0)$ and hence must cross zero often.

Formally, two $I(1)$ series are said to be *cointegrated* if a linear combination of them is $I(0)$. The concept of *cointegration* was originally proposed by Granger (1981) and Granger and Weiss (1983) and subsequently formalized in Engle and Granger (1987).

This concept is readily generalized to characterize the relationships among d $I(1)$ time series. Let $\boldsymbol{y}_t$ be a d-dimensional vector $I(1)$ series such that each element is an $I(1)$ series. These elements are said to be cointegrated if there exists a $d \times 1$ vector, $\boldsymbol{\alpha}$, such that $z_t = \boldsymbol{\alpha}'\boldsymbol{y}_t$ is $I(0)$. We say that the elements of $\boldsymbol{y}_t$ are CI(1,1) for simplicity, indicating that a linear combination of the elements of $\boldsymbol{y}_t$ is capable of reducing the integrated order by one. The vector $\boldsymbol{\alpha}$ is referred to as a *cointegrating vector*.

When $d > 2$, there may be more than one cointegrating vector. Clearly, if $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ are two cointegrating vectors, so are their linear combinations. Hence, we are primarily interested in the cointegrating vectors that are *not* linearly dependent. The space spanned by linearly independent cointegating vectors is the *cointegrating space*; the number of linearly independent cointegrating vectors is known as the *cointegrating rank* which is the dimension of the cointegrating space. If the cointegrating rank is r, we can put these r linearly independent cointegrating vectors together and form the $d \times r$ matrix $\boldsymbol{A}$ such that $\boldsymbol{z}_t = \boldsymbol{A}'\boldsymbol{y}$ is a vector $I(0)$ series. Note that for a d-dimensional vector $I(1)$ series $\boldsymbol{y}_t$, the cointegrating rank is at most $d-1$. For if the cointegrating rank is d, $\boldsymbol{A}$ would be a $d \times d$ nonsingular matrix, so that $\boldsymbol{A}^{-1}\boldsymbol{z}_t = \boldsymbol{y}_t$ must be a vector $I(0)$ series as well. This contradicts the assumption that $\boldsymbol{y}_t$ is a vector of $I(1)$ series.

A simple way to find a cointegration relationship among the elements of $\boldsymbol{y}_t$ is to regress one element, say, $y_{1,t}$ on all other elements, $\boldsymbol{y}_{2,t}$, as suggested by Engle and Granger (1987). A *cointegrating regression* is

$$y_{1,t} = \boldsymbol{\alpha}'\boldsymbol{y}_{2,t} + \boldsymbol{z}_t,$$

where the vector $(1\ \boldsymbol{\alpha}')'$ is the cointegrating vector with the first element normalized to one, and z_t are the regression errors and also the equilibrium errors. The estimated cointegrating regression is

$$y_{1,t} = \hat{\boldsymbol{\alpha}}_T'\boldsymbol{y}_{2,t} + \hat{\boldsymbol{z}}_t,$$

where $\hat{\boldsymbol{\alpha}}_T$ is the OLS estimate, and $\hat{z}_t$ are the OLS residuals which approximate the equilibrium errors. It should not be surprising to find that the estimator $\hat{\boldsymbol{\alpha}}_T$ is T-consistent, as in the case of autoregression with a unit root.

When the elements of $\boldsymbol{y}_t$ are cointegrated, they must be determined jointly. The equilibrium errors z_t are thus also correlated with $\boldsymbol{y}_{2,t}$. As far as the consistency of the OLS estimator is concerned, the correlations between z_t and $\boldsymbol{y}_{2,t}$ do not matter asymptotically, but they would result in finite-sample bias and efficiency loss. To correct these correlations and obtain more efficient estimates, Saikkonen (1991) proposed estimating a modified co-integrating regression that includes additional k leads and lags of $\Delta \boldsymbol{y}_{2,t} = \boldsymbol{y}_{2,t} - \boldsymbol{y}_{2,t-1}$:

$$y_{1,t} = \boldsymbol{\alpha}' \boldsymbol{y}_{2,t} + \sum_{j=-k}^{k} \Delta \boldsymbol{y}'_{2,t-j} \boldsymbol{b}_j + e_t.$$

It has been shown that the OLS estimator of $\boldsymbol{\alpha}$ is asymptotically efficient in the sense of Saikkonen (1991, Definition 2.2); we omit the details. Phillips and Hansen (1990) proposed a different way to compute efficient estimates.

When cointegration exists, the true equilibrium errors z_t should be an $I(0)$ series; otherwise, they should be $I(1)$. One can then verify the cointegration relationship by applying unit-root tests, such as the augmented Dickey-Fuller test and the Phillips-Perron test, to $\hat{z}_t$. The null hypothesis that a unit root is present is equivalent to the hypothesis of *no cointegration*. Failing to reject the null hypothesis of no cointegration suggests that the regression is in fact a spurious one, in the sense of Granger and Newbold (1974).

To implement a unit-root test on cointegration residuals $\hat{z}_T$, a difficulty is that $\hat{z}_T$ is not a raw series but a result of OLS fitting. Thus, even when z_t may be $I(1)$, the residuals $\hat{z}_t$ may not have much variation and hence behave like a stationary series. Consequently, the null hypothesis would be rejected too often if the original Dickey-Fuller critical values were used. Engle and Granger (1987), Engle and Yoo (1987), and Davidson and MacKinnon (1993) simulated proper critical values for the unit-root tests on cointegrating residuals. Similar to the unit-root tests discussed earlier, these critical values are all "model dependent." In particular, the critical values vary with d, the number of variables (dependent variables and regressors) in the cointegrating regression. Let τ_c denote the t-ratio of an auxiliary autoregression on $\hat{z}_t$ with a constant term. Table 7.3 summarizes some critical values of the τ_c test of no cointegration based on Davidson and MacKinnon (1993).

The cointegrating regression approach has some drawbacks in practice. First, the estimation result is somewhat arbitrary because it is determined by the dependent variable in the regression. As far as cointegration is conerned, any variable in the vector series could serve as a dependent variable. Although the choice of the dependent variable does not matter asymptotically, it does affect the estimated cointegration relationships in finite samples. Second, this approach is more suitable for finding only one cointegrating relationship, despite that Engle and Granger (1987) proposed estimating multiple cointegration relationships by a vector regression. It is now typical to adopt the maximum likelihood approach of Johansen (1988) to estimate the cointegrating space directly.

Table 7.3: Some percentiles of the distributions of the cointegration τ_c test.

d	1%	2.5%	5%	10%
2	−3.90	−3.59	−3.34	−3.04
3	−4.29	−4.00	−3.74	−3.45
4	−4.64	−4.35	−4.10	−3.81

Cointegration has an important implication. When the elements of $\boldsymbol{y}_t$ are cointegrated such that $\boldsymbol{A}'\boldsymbol{y}_t = \boldsymbol{z}_t$, then there must exist an error correction model (ECM) in the sense that

$$\Delta\boldsymbol{y}_t = \boldsymbol{B}\boldsymbol{z}_{t-1} + \boldsymbol{C}_1\Delta\boldsymbol{y}_{t-1} + \cdots + \boldsymbol{C}_k\Delta\boldsymbol{y}_{t-k} + \nu_t,$$

where $\boldsymbol{B}$ is $d \times r$ matrix of coefficients associated with the vector of equilibrium errors and $\boldsymbol{C}_j$, $j = 1, \ldots, k$, are the coefficient matrices associated with lagged differences. It must be emphasized that cointegration characterizes the long-run equilibrium relationship among the variables because it deals with the *levels* of $I(1)$ variables. On the other hand, the corresponding ECM describes short-run dynamics of these variables, in the sense that it is a dynamic vector regression on the *differences* of these variables. Thus, the long-run equilibrium relationships are useful in explaining the short-run adjustment when cointegration exists.

The result here also indicates that, when cointegration exists, a vector AR model of $\Delta\boldsymbol{y}_t$ is misspecified because it omits the important variable $\boldsymbol{z}_{t-1}$, the lagged equilibrium errors. Omitting this variable would render the estimates in the AR model of $\Delta\boldsymbol{y}_t$ inconsistent. Therefore, it is important to identify the cointegrating relationship before estimating an ECM. On the other hand, the Johansen approach mentioned above permits joint estimation of the cointegrating space and ECM. In practice, an ECM can be estimated by replacing $\boldsymbol{z}_{t-1}$ with the residuals of a cointegrating regression $\hat{\boldsymbol{z}}_{t-1}$ and then regressing $\Delta\boldsymbol{y}_t$ on $\hat{\boldsymbol{z}}_{t-1}$ and lagged $\Delta\boldsymbol{y}_t$. Note that standard asymptotic theory applies here because ECM involves only stationary variables when cointegration exists.

Appendix

Proof of Lemma 7.1: By invoking a suitable FCLT, the proofs of the assertions (i) and (ii) are the same as those in Example 5.43. To prove (iii), we apply the formula of summation by parts. For two sequences $\{a_t\}$ and $\{b_t\}$, put $A_n = \sum_{t=1}^{n} a_t$ for $n \geq 0$ and $A_{-1} = 0$. Summation by parts is such that, for $0 \leq p \leq q$,

$$\sum_{t=p}^{q} a_t b_t = \sum_{t=p}^{q-1} A_t(b_t - b_{t+1}) + A_q b_q - A_{p-1} b_p.$$

Now, setting $a_t = \epsilon_t$ and $b_t = y_{t-1}$ we get $A_t = y_t$ and

$$\sum_{t=1}^{T} y_{t-1}\epsilon_t = y_T y_{T-1} - \sum_{t=1}^{T-1} y_t \epsilon_t = y_T^2 - \sum_{t=1}^{T} \epsilon_t^2 - \sum_{t=1}^{T} y_{t-1}\epsilon_t.$$

Hence,

$$\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\epsilon_t = \frac{1}{2}\left(\frac{1}{T}y_T^2 - \frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2\right) \Rightarrow \frac{1}{2}[\sigma_*^2 w(1)^2 - \sigma_\epsilon^2].$$

Note that the stochastic integral $\int_0^1 w(r)\,\mathrm{d}w(r) = [w(1)^2 - 1]/2$; see e.g., Davidson (1994, p. 507). An alternative expression of the weak limit of $\sum_{t=1}^{T} y_{t-1}\epsilon_t / T$ is thus

$$\sigma_*^2 \int_0^1 w(r)\,\mathrm{d}w(r) + \frac{1}{2}(\sigma_*^2 - \sigma_\epsilon^2). \quad \square$$

Proof of Theorem 7.2: The first assertion follows directly from Lemma 7.1 and the continuous mapping theorem. The second assertion follows from the first by noting that $\sigma_*^2 = \sigma_\epsilon^2$ when ϵ_t are i.i.d. $\square$

Proof of Lemma 7.3: By Lemma 7.1(i) and (ii), we have

$$\begin{aligned}
&\frac{1}{T^2}\sum_{t=1}^{T}(y_{t-1} - \bar{y}_{-1})^2 \\
&\qquad = \frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}^2 - \left(\frac{1}{T^{3/2}}\sum_{t=1}^{T} y_{t-1}\right)^2 \\
&\qquad \Rightarrow \sigma_*^2 \int_0^1 w(r)^2\,\mathrm{d}r - \sigma_*^2\left(\int_0^1 w(r)\,\mathrm{d}r\right)^2.
\end{aligned}$$

It is easy to show that the limit on the right-hand side is just $\sigma_*^2 \int_0^1 w^*(r)^2\,\mathrm{d}r$, as asserted in (i). To prove (ii), note that

$$\frac{1}{T}\bar{y}_{-1}\sum_{t=1}^{T}\epsilon_t = \left(\frac{1}{T^{3/2}}\sum_{t=1}^{T} y_{t-1}\right)\left(\frac{1}{\sqrt{T}}y_T\right) \Rightarrow \sigma_*^2\left(\int_0^1 w(r)\,\mathrm{d}r\right)w(1).$$

It follows from Lemma 7.1(iii) that

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}(y_{t-1}-\bar{y}_{-1})\epsilon_t \\
&\qquad \Rightarrow \sigma_*^2\int_0^1 w(r)\,\mathrm{d}w(r)+\frac{1}{2}(\sigma_*^2-\sigma_\epsilon^2)-\sigma_*^2\left(\int_0^1 w(r)\,\mathrm{d}r\right)w(1) \\
&\qquad = \sigma_*^2\int_0^1 w^*(r)\,\mathrm{d}w(r)+\frac{1}{2}(\sigma_*^2-\sigma_\epsilon^2),
\end{aligned}$$

where the last equality is due to the fact that $\int_0^1 \mathrm{d}w(r)=w(1)$. □

Proof of Theorem 7.4: The assertions on $T(\hat{\alpha}_T-1)$ follow directly from Lemma 7.3. For $\hat{c}_T$, we have

$$\begin{aligned}
\sqrt{T}\hat{c}_T &= T(1-\hat{\alpha}_T)\frac{1}{\sqrt{T}}\bar{y}_{-1}+\frac{\sqrt{T}}{T-1}\sum_{t=2}^{T}\epsilon_t \\
&\Rightarrow A\left(\sigma_*\int_0^1 w(r)\,\mathrm{d}r\right)+\sigma_* w(1). \quad \square
\end{aligned}$$

Proof of Theorem 7.5: First note that

$$\hat{\sigma}_T^2=\frac{1}{T-2}\sum_{t=2}^{T}\epsilon_t^2-\frac{\hat{\alpha}_T-1}{T-2}\sum_{t=2}^{T}y_{t-1}\epsilon_t\xrightarrow{\text{a.s.}}\sigma_\epsilon^2,$$

by Kolmogorov's SLLN. When ϵ_t are i.i.d., σ_ϵ^2 is also the long-run variance σ_*^2. The t statistic τ_0 is thus

$$\tau_0=\frac{(\sum_{t=2}^{T}y_{t-1}^2)/T^2)^{1/2}\,T(\hat{\alpha}_T-1)}{\hat{\sigma}_T}\Rightarrow\frac{\frac{1}{2}\big[w(1)^2-1\big]}{\big[\int_0^1 w(r)^2\,\mathrm{d}r\big]^{1/2}},$$

by Lemma 7.1(ii) and the second assertion of Theorem 7.2. Similarly, the weak limit of τ_c follows from Lemma 7.3(i) and the second assertion of Theorem 7.4. □

Proof of Theorem 7.6: When $\{y_t\}$ is a general $I(1)$ process, it follows from Lemma 7.1(ii) and the first assertion of Theorem 7.2 that

$$\tau_0=\frac{(\sum_{t=2}^{T}y_{t-1}^2)/T^2)^{1/2}\,T(\hat{\alpha}_T-1)}{\hat{\sigma}_T}\Rightarrow\frac{\sigma_*}{\sigma_\epsilon}\left(\frac{\frac{1}{2}\big[w(1)^2-\sigma_\epsilon^2/\sigma_*^2\big]}{(\int_0^1 w(r)^2\,\mathrm{d}r)^{1/2}}\right).$$

The second assertion can be derived similarly. □

Proof of Corollary 7.7: The limits follow straightforwardly from Theorem 7.6. □

Proof of Theorem 7.8: For $\hat{e}_t = y_t - \bar{y}$, $T^{-1/2}\sum_{i=1}^{[Tr]} \hat{e}_i \Rightarrow \sigma_* w^0(r)$, as shown in the text. As s_{Tn} is consistent for σ_*, the first assertion follows from the continuous mapping theorem. The proof for the second assertion is left to Exercise 7.4. □

Proof of Theorem 7.9: The assertion (i) follows easily from the expression

$$\hat{\beta}_T = \frac{\sum_{t=1}^T (x_t - \bar{x})(y_t - \bar{y})/T^2}{\sum_{t=1}^T (x_t - \bar{x})^2/T^2}.$$

To prove (ii), we note that

$$\begin{aligned} T^{-1/2}\hat{\alpha}_T &= \frac{1}{T^{3/2}}\sum_{t=1}^T y_t - \frac{1}{T^{3/2}}\hat{\beta}_T \sum_{t=1}^T x_t \\ &\Rightarrow \sigma_y \int_0^1 w_y(r)\,\mathrm{d}r - \left(\frac{\sigma_y m_{yx}}{\sigma_x m_x}\right)\left(\sigma_x \int_0^1 w_x(r)\,\mathrm{d}r\right). \end{aligned}$$

In Exercise 7.5, it is shown that the OLS variance estimator $\hat{\sigma}_T^2$ diverges such that

$$\hat{\sigma}_T^2/T \Rightarrow \sigma_y^2\left(m_y - \frac{m_{yx}^2}{m_x}\right).$$

It follows that

$$\frac{1}{\sqrt{T}}\, t_\beta = \frac{\hat{\beta}_T\left[\sum_{t=1}^T (x_t - \bar{x})^2/T^2\right]^{1/2}}{\hat{\sigma}_T/\sqrt{T}} \Rightarrow \frac{m_{yx}/m_x^{1/2}}{(m_y - m_{yx}^2/m_x)^{1/2}}.$$

This proves (iii). For the assertion (iv), we have

$$\begin{aligned} \frac{1}{\sqrt{T}}\, t_\alpha &= \frac{\hat{\alpha}_T}{\hat{\sigma}_T\sqrt{T}}\left[\frac{T\sum_{t=1}^T (x_t - \bar{x})^2}{\sum_{t=1}^T x_t^2}\right]^{1/2} \\ &\Rightarrow \frac{m_x^{1/2}\left[\int_0^1 w_y(r)\,\mathrm{d}r - (m_{yx}/m_x)\int_0^1 w_x(r)\,\mathrm{d}r\right]}{\left[(m_y - m_{yx}^2/m_x)\int_0^1 w_x(r)^2\,\mathrm{d}r\right]^{1/2}}. \end{aligned}$$ □

Exercises

7.1 For the specification (7.3), derive the weak limit of the t-ratio for $c = 0$.

7.2 Suppose that $y_t = c_o + y_{t-1} + \epsilon_t$ with $c_o \neq 0$. Find the orders of $\sum_{t=1}^T y_t$ and $\sum_{t=1}^T y_t^2$ and compare these ordier with those in Lemma 7.1.

7.3 Given the specification $y_t = c + \alpha y_{t-1} + e_t$, suppose that $y_t = c_o + y_{t-1} + \epsilon_t$ with $c_o \neq 0$ and ϵ_t i.i.d. Find the weak limit of $T(\hat{\alpha}_T - 1)$ and compare with the result in Theorem 7.4.

7.4 Given $y_t = a_0 + b_0\, t + \epsilon_t$, let $\hat{e}_t = y_t - \hat{a}_T - \hat{b}_T\, t$ denote the OLS residuals. Show that

$$\frac{1}{\sigma_* \sqrt{T}} \sum_{t=1}^{[Tr]} \hat{e}_t \Rightarrow w(r) + (2r - 3r^2)w(1) - (6r - 6r^2) \int_0^1 w(s)\, \mathrm{d}s, \quad r \in (0, 1].$$

7.5 As in Section 7.4.1, consider the specification $y_t = \alpha + \beta x_t + e_t$, where y_t and x_t are two independent random walks. Let $\hat{\sigma}_T^2 = \sum_{t=1}^T (y_t - \hat{\alpha}_T - \hat{\beta}_T x_t)^2/(T-2)$ denote the standard OLS variance estimator. Show that

$$\hat{\sigma}_T^2/T \Rightarrow \sigma_y^2 \left(m_y - \frac{m_{yx}^2}{m_x} \right).$$

7.6 As in Section 7.4.1, consider the specification $y_t = \alpha + \beta x_t + e_t$, where y_t and x_t are two independent random walks. Let d denote the Durbin-Watson statistic. Granger and Newbold (1974) also observed that it is typical to have a small value of d. Prove that

$$T\, d \Rightarrow \frac{(\sigma_u^2/\sigma_y^2) + (m_{yx}/m_x)^2(\sigma_v^2/\sigma_x^2)}{m_y - m_{yx}^2/m_x},$$

and explain how this result is related to Granger and Newbold's observation.

References

Davidson, James (1994). *Stochastic Limit Theory*, New York, NY: Oxford University Press.

Davidson, R. and James G. MacKinnon (1993). *Estimation and Inference in Econometrics*, New York, NY: Oxford University Press.

Dickey, David A. and Wayne A. Fuller (1979). Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association*, **74**, 427–431.

Dickey, David A. and Wayne A. Fuller (1981). Likelihood ratio statistics for autoregressive time series with a unit root, *Econometrica*, **49**, 1057–1072.

Engle, Robert F. and Clive W. J. Granger (1987). Co-integration and error correction: Representation, estimation and testing, *Econometrica*, **55**, 251–276.

Engle, Robert F. and Byung S. Yoo (1987). Forecasting and testing co-integrated systems, *Journal of Econometrics*, **35**, 143–159.

Fuller, Wayne A. (1996). *Introduction to Statistical Time Series*, second edition, New York, NY: John Wiley and Sons.

Granger, Clive W. J. (1981). Some properties of time series data and their use in econometric model specification, *Journal of Econometrics*, **16**, 121–130.

Granger, Clive W. J. and Paul Newbold (1974). Spurious regressions in econometrics, *Journal of Econometrics*, **2**, 111–120.

Granger, Clive W. J. and Andrew A. Weiss (1983). Time series analysis of error-correction models, in Samuel Karlin, Takeshi Amemiya, & Leo A. Goodman ed., *Studies in Econometrics, Time Series, and Multivariate Statistics*, 255-278, New York: Academic Press.

Hamilton, James D. (1994). *Time Series Analysis*, Priceton, NJ: Princeton University Press.

Johansen, Søren (1988). Statistical analysis of cointegration vectors, *Journal of Economic Dynamics and Control*, **12**, 231–254.

Kwiawtkowski, Denis, Peter C. B. Phillips, Peter Schmidt, and Yongcheol Shin (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, **54**, 159–178.

Nabeya, Seiji and Katsuto Tanaka (1988). Asymptotic theory of a test for the constancy of regression coefficients against the random walk alternative, *Annals of Statistics*, **16**, 218–235.

Nelson, Charles R. and Heejoon Kang (1984). Pitfalls in the use of time as an explanatory variable in regression, *Journal of Business and Economic Statistics*, **2**, 73–82.

Nelson, Charles R. and Charles I. Plosser (1982). Trends and random walks in macroeconomic time series: Some evidence and implications, *Journal of Monetary Economics*, **10**, 139–162.

Phillips, Peter C. B. (1986). Understanding spurious regressions in econometrics, *Journal of Econometrics*, **33**, 311–340.

Phillips, Peter C. B. (1987). Time series regression with a unit root, *Econometrica*, **55**, 277–301.

Phillips, Peter C. B. and Steven N. Durlauf (1986). Multiple time series regression with integrated processes, *Review of Economic Studies*, **53**, 473–495.

Phillips, Peter C. B. and Bruce E. Hansen (1990). Statistical inference in instrumental variables regression with I(1) processes, *Review of Economic Studies*, **57**, 99–125.

Phillips, Peter C. B. and Pierre Perron (1988). Testing for a unit root in time series regression, *Biometrika*, **75**, 335–346.

Said, Said E. and David A. Dickey (1984). Testing for unit roots in autoregressive-moving average models of unknown order, *Biometrika*, **71**, 599–607.

Saikkonen, Pentti (1991). Asymptotically efficient estimation of cointegration regressions, *Econometric Theory*, **7**, 1–21.