Cointegration Analysis

16

16.1 Summary

This chapter is addressed to the analysis of cointegrated variables. Properties like superconsistency of the LS estimator and conditions for asymptotic normality are extensively discussed. Error-correction is the reverse of cointegration, which is why we provide an introduction to the analysis of error-correction models as well. In particular, we discuss cointegration testing. In 2003, Clive W.J. Granger was awarded the Nobel prize for introducing the concept of cointegration. Finally, we stress once more the effect of linear time trends underlying the series.

16.2 Error-Correction and Cointegration

Before cointegration had been launched, so-called **error-correction models** impressed due to their empirical performance, cf. e.g. Davidson, Hendry, Srba, and Yeo (1978). Today we know that these models are just the other side of the cointegration coin. By way of example, the key statement of Granger's representation theorem from Engle and Granger (1987) is illustrated, which demonstrates the fact that error-correction and cointegration are equivalent.

Autoregressive Distributed Lag Model

Let us consider a dynamic regression model in which $\{y_t\}$ has an autoregressive structure on the one hand and which is explained by (lagged) exogenous variables x_{t-j} on the other:

$$ y_t = a_1 \, y_{t-1} + \cdots + a_p \, y_{t-p} + c_0 \, x_t + c_1 \, x_{t-1} + \cdots + c_\ell \, x_{t-\ell} + \varepsilon_t . $$

Hence, this is an extension of the AR(p) process. Because of the additional exogenous explanatory variables, we sometimes speak of ARX(p, ℓ) models, although such processes are more often called **autoregressive distributed lag models**, ARDL(p, ℓ). We assume that $\{x_t\}$ is integrated of order one. In order to have cointegration with $\{y_t\}$, the ARDL model has to be stable. We adopt the stability condition from Proposition 3.4:

$$1 - a_1 z - \cdots - a_p z^p = 0 \quad \Rightarrow \quad |z| > 1 \,.$$

Example 16.1 (ARDL(2,2)) With $p = \ell = 2$ we consider

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + c_0 x_t + c_1 x_{t-1} + c_2 x_{t-2} + \varepsilon_t \,.$$

Due to the assumed stability, the parameter b can be defined:

$$b := \frac{c_0 + c_1 + c_2}{1 - a_1 - a_2} \,.$$

In fact, the denominator is not only different from zero but positive if stability is given (which we again know from Proposition 3.4): $1 - a_1 - a_2 > 0$. Thus, the following parameter γ is negative:

$$\gamma := -(1 - a_1 - a_2) < 0 \,.$$

Elementary manipulations lead to the reparameterization (cf. Problem 16.1)

$$\Delta y_t = \gamma \left[y_{t-1} - b \, x_{t-1} \right] - a_2 \, \Delta y_{t-1} + c_0 \, \Delta x_t - c_2 \, \Delta x_{t-1} + \varepsilon_t \,. \tag{16.1}$$

In this equation differences of y are related to their own lags and differences of x. In addition, they depend on a linear combination of lagged levels (in square brackets). This last aspect is the one constituting error-correction models. The cointegration relation $y = b \, x$ is understood as a long-run equilibrium relation and $v_{t-1} = y_{t-1} - b \, x_{t-1}$ as a deviation from it in $t - 1$. This deviation from the equilibrium again influences the increments of y_t. The involved linear combination of y_{t-1} and x_{t-1} needs to be stationary because $\{\Delta y_t\}$ is stationary by assumption. Hence, in the example it is obvious that such a relation between differences and levels implies cointegration. Indeed, it is the lagged deviation from the equilibrium v_{t-1} influencing the increments of Δy_t with a negative sign. If y_{t-1} is greater than the equilibrium value, $v_{t-1} > 0$, then this affects the change from y_{t-1} to y_t negatively, i.e. y is corrected towards the equilibrium, and vice versa for values below the equilibrium value. What economists know as deviation from the equilibrium is called "error" (in the sense of a deviation from a set target) in engineering, which explains the name error-correction model. ∎

The example can be generalized. Cointegrated ARDL models of arbitrary order can always be formulated as error-correction models. This is not surprising against the background of Granger's representation theorem.

Granger's Representation Theorem

Error-correction adjustment is the downside to cointegration. The relation between cointegration and error-correction is explained by the following proposition where we, however, will not spell out all the technical details. The result goes back to Granger, cf. Engle and Granger (1987) or Johansen (1995, Theorem 4.2).

Proposition 16.1 (Representation Theorem) *Let $\{y_t\}$ and $\{x_t\}$ be integrated of order one. They are cointegrated if and only if they have an error-correction representation,*

$$\Delta y_t = \gamma\, v_{t-1} + \sum_{j=1}^{p} a_j\, \Delta y_{t-j} + \sum_{j=1}^{\ell} \alpha_j\, \Delta x_{t-j} + \varepsilon_t\,, \tag{16.2}$$

$$\Delta x_t = \gamma_x\, v_{t-1} + \sum_{j=1}^{p_x} a_j^{(x)}\, \Delta y_{t-j} + \sum_{j=1}^{\ell_x} \alpha_j^{(x)}\, \Delta x_{t-j} + \varepsilon_{x,t}\,, \tag{16.3}$$

$$v_t = y_t - b\, x_t \sim I(0)\,, \quad b \neq 0\,, \tag{16.4}$$

where at least one of the so-called adjustment coefficients γ or γ_x is different from zero.

Of course, not all a_j ($a_j^{(x)}$) and α_j ($\alpha_j^{(x)}$) need to be different from zero. The error sequences $\{\varepsilon_t\}$ and $\{\varepsilon_{x,t}\}$ are white noise and may be contemporaneously correlated. Frequently in practice, additional contemporaneous differences of the respective other variable are hence incorporated on the right-hand side. For Eq. (16.2) this means e.g. the inclusion of $\alpha_0\, \Delta x_t$. Then, one sometimes also speaks of the conditional or structural error-correction equation.

Cointegration and the Long-Run Variance Matrix

In (14.6) the symmetric long-run variance matrix Ω of a stationary vector has been defined. We now consider an I(1) vector $z_t' = (z_{1,t}, z_{2,t})$ such that $\Delta z_t = (w_{1,t}, w_{2,t})'$ is integrated of order zero, which is why Ω cannot be equal to the zero matrix. Nevertheless, the matrix does not have to be invertible. It rather holds: If the vector $\{z_t\}$ is cointegrated, then Ω has the reduced rank one and is not invertible. Equivalently, this means: If Ω is invertible, then $\{z_{1,t}\}$ and $\{z_{2,t}\}$ are not cointegrated. To this, we consider the following example.

Example 16.2 (Ω under Cointegration) In this example, let $\{z_{2,t}\}$ be a random walk, the cointegration parameter be one, and the deviation from the equilibrium $v_t = \varepsilon_{1,t}$ be iid and independent of $\{\Delta z_{2,t}\}$:

$$z_{1,t} = z_{2,t} + \varepsilon_{1,t},$$

$$z_{2,t} = z_{2,t-1} + \varepsilon_{2,t}.$$

Hence, if $\{\varepsilon_{1,t}\}$ and $\{\varepsilon_{2,t}\}$ are independent with variances σ_1^2 and σ_2^2, then one shows with $w_{1,t} = \Delta z_{1,t} = \varepsilon_{2,t} + \varepsilon_{1,t} - \varepsilon_{1,t-1}$ and $w_{2,t} = \Delta z_{2,t} = \varepsilon_{2,t}$:

$$\Gamma_w(0) = \begin{pmatrix} \sigma_2^2 + 2\sigma_1^2 & \sigma_2^2 \\ \sigma_2^2 & \sigma_2^2 \end{pmatrix} \quad \text{and} \quad \Gamma_w(1) = \begin{pmatrix} -\sigma_1^2 & 0 \\ 0 & 0 \end{pmatrix}.$$

For $h > 1$, $\Gamma_w(h) = 0$. Thus, it holds:

$$\Omega_w = \sigma_2^2 \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},$$

i.e. the matrix is of rank one and not invertible. ∎

Conversely, it holds as well that full rank of Ω follows from the absence of cointegration. This is to be illustrated by the following example.

Example 16.3 (Ω without Cointegration) Now, let $\{z_{1,t}\}$ and $\{z_{2,t}\}$ be two random walks independent of each other:

$$z_{1,t} = z_{1,t-1} + \varepsilon_{1,t},$$

$$z_{2,t} = z_{2,t-1} + \varepsilon_{2,t}.$$

Since $\{\varepsilon_{1,t}\}$ and $\{\varepsilon_{2,t}\}$ are independent with variances σ_1^2 and σ_2^2, one shows with $w_{1,t} = \varepsilon_{1,t}$ and $w_{2,t} = \varepsilon_{2,t}$:

$$\Gamma_w(0) = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}.$$

For $h > 0$, $\Gamma_w(h) = 0$. Thus, it holds:

$$\Omega_w = \Gamma_w(0),$$

where this matrix has full rank 2 in the case of positive variances and is thus invertible. ∎

Hence, the presence of cointegration of the I(1) vector $\{z_t\}$ depends on the matrix Ω. The examples show that no cointegration of $\{z_t\}$ is equivalent to the full rank of Ω, cf. Phillips (1986).

Linearly Independent Cointegration Vectors

For more than two I(1) variables linearly independent cointegration vectors can exist. In such a situation there can be no talk of "the true" cointegration vector. Each and every linear combination of independent cointegration vectors is itself again a cointegration vector. Although this cannot occur under our assumption of a bivariate vector, we want to become aware of this problem by means of a three-dimensional example.

Example 16.4 (Three Interest Rates) Assume z_1, z_2 and z_3 to be interest rates for one-month, two-month and three-month loans integrated of order one. Then one expects (due to the expectations hypothesis of the term structure) the interest rate differentials $z_1 - z_2$ and $z_2 - z_3$ to provide stable relations:

$$z_{1,t} - z_{2,t} = v_{1,t} \sim I(0),$$
$$z_{2,t} - z_{3,t} = v_{2,t} \sim I(0).$$

Here, $b_1' = (1, -1, 0)$ and $b_2' = (0, 1, -1)$ are linearly independent, and both are cointegrating vectors for $z_t' = (z_{1,t}, z_{2,t}, z_{3,t})$ as $v_{1,t}$ and $v_{2,t}$ are both assumed to be I(0):

$$\begin{pmatrix} b_1' \\ b_2' \end{pmatrix} \begin{pmatrix} z_{1,t} \\ z_{2,t} \\ z_{3,t} \end{pmatrix} = \begin{pmatrix} v_{1,t} \\ v_{2,t} \end{pmatrix}.$$

Hence, the uniqueness of the cointegration vector is lost. It is rather that b_1 and b_2 form a basis for the cointegration space. Each vector contained in the plane they span, i.e. each linear combination of b_1 and b_2, is itself again a cointegration vector. For example for

$$\beta_\alpha = b_1 + (1 - \alpha)b_2 = \begin{pmatrix} 1 \\ -\alpha \\ \alpha - 1 \end{pmatrix}$$

one obtains a stationary relation comprising all three variables from z_t,

$$z_{1,t} = \alpha z_{2,t} + (1 - \alpha)z_{3,t} + v_{1,t} + (1 - \alpha)v_{2,t},$$

where $v_{1,t} + (1 - \alpha)v_{2,t}$ is I(0). In particular for $\alpha = 0$ one sees that $z_{1,t}$ and $z_{3,t}$ alone are also cointegrated with the cointegration vector $\beta_0' = b_1' + b_2' = (1, 0, -1)$. The cointegration vectors b_1, b_2 as well as β_0 provide economically reasonable and theoretically secured statements on the interest rate differentials. However, β_α for

an arbitrary α is just as good a cointegrating vector. Hence,

$$z_1 = \alpha z_2 + (1 - \alpha)z_3$$

for $\alpha \neq 1$ or $\alpha \neq 0$ is as well a "true" long-run equilibrium relation, even if, in contrast to the interest rate differentials, it is not readily amenable to an economic interpretation. These problems with the interpretation of more than one linearly independent cointegration vectors are of a fundamental nature and cannot be solved unless one makes a priori (economically plausible) assumptions on the form of the cointegrating vectors. The cointegration analysis is not the life belt saving us from the lack of identification: With purely statistical methods one generally cannot make economic statements. ∎

For only two I(1) variables $\{y_t\}$ and $\{x_t\}$, linearly independent cointegration vectors cannot exist. We show this by contradiction. Hence, let us assume with $b_1 \neq b_2$ from $\mathbb{R}$ that two linearly independent relations exist. We collect them row-wise in the matrix B:

$$B = \begin{pmatrix} 1 & -b_1 \\ 1 & -b_2 \end{pmatrix}.$$

Then, it holds by assumption that

$$B \begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} y_t - b_1 x_t \\ y_t - b_2 x_t \end{pmatrix} = \begin{pmatrix} v_{1,t} \\ v_{2,t} \end{pmatrix}$$

is a vector of I(0) variables $\{v_{1,t}\}$ and $\{v_{2,t}\}$. Due to the independence of the cointegration vectors, B is invertible:

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = B^{-1} \begin{pmatrix} v_{1,t} \\ v_{2,t} \end{pmatrix}.$$

This yields $\{y_t\}$ and $\{x_t\}$ as linear combinations of $\{v_{i,t}\}$, $i = 1, 2$, from which it follows that $\{y_t\}$ and $\{x_t\}$ themselves have to be I(0), which contradicts the assumption.

16.3   Cointegration Regressions

In the case of bivariate cointegration, the LS estimator of a static regression of y_t on x_t tends to the true value with the sample size (i.e. with rate n). For this fast rate of convergence the term **superconsistency** has been coined in the literature. At the same time limiting normality only arises under additional assumptions.

Superconsistent Estimation

We consider the LS estimator regressing y_t on x_t under the assumption (16.4) that cointegration is present. For the sake of simplicity, we again do not allow for an intercept (which is why we assume that the I(1) processes have the starting value zero):

$$y_t = \hat{b}\,x_t + \hat{v}_t, \quad t = 1, \ldots, n. \tag{16.5}$$

Then we write for the LS estimator:

$$\hat{b} - b = \frac{\sum\limits_{t=1}^{n} x_t v_t}{\sum\limits_{t=1}^{n} x_t^2}$$

or

$$n\left(\hat{b} - b\right) = \frac{n^{-1} \sum\limits_{t=1}^{n} x_t v_t}{n^{-2} \sum\limits_{t=1}^{n} x_t^2}.$$

To be able to apply the functional limit theory from Chap. 14, we now define

$$w_t = \begin{pmatrix} v_t \\ \Delta x_t \end{pmatrix}, \quad \text{i.e.} \quad z_t = \begin{pmatrix} \sum\limits_{i=1}^{t} v_i \\ x_t \end{pmatrix} \tag{16.6}$$

instead of $z_t' = (y_t, x_t)$ as in Sect. 15.4 without cointegration. Then it holds with the results from Proposition 14.4(b) and (c):

$$n(\hat{b} - b) \overset{d}{\longrightarrow} \frac{\int\limits_{0}^{1} B_2(s)dB_1(s) + \sum\limits_{h=0}^{\infty} \mathrm{E}(\Delta x_t v_{t+h})}{\int\limits_{0}^{1} B_2^2(s)ds}.$$

As the LS estimator tends to the true value with the sample size n instead of with only $n^{0.5}$ as for the stationary regression model, since Stock (1987) and Engle and Granger (1987) it has become common usage to speak of superconsistency of the static cointegration estimator from (16.5); however, this result has been known from Phillips and Durlauf (1986) already.

Note that the estimation of b is consistent despite possible correlation between error term v_t and regressor x_t (or Δx_t). Insofar the cointegration regression knocks out the simultaneity bias (or "Haavelmo bias"): Superconsistency is a strong

asymptotic argument for single equation regressions despite possibly existing dependencies through simultaneous relations between the individual equations, i.e. despite correlation between regressors and error term. According to this, the cointegration approach can be understood as a reaction to the simultaneous equation methodology of former decades. At the same time, it is not clear anymore which variable constitutes the endogenous left-hand side and which quantity identifies the exogenous regressor. Beside (16.4), it also holds as a "true relation" that

$$x_t = \frac{y_t}{b} - \frac{v_t}{b}.$$

Hence, if x_t is regressed on y_t, then one would obtain analogously a superconsistent estimator for b^{-1}. This vehemently contrasts the results of the stationary standard econometrics where the (asymptotic) validity of LS crucially depends on the correct specification of the single equation and exogeneity assumptions.

Further Asymptotic Properties

In Problems 16.2 and 16.3 we prove the further properties of the following proposition. Here, the standard errors of the t-statistic, the uncentered coefficient of determination and the Durbin-Watson statistic are defined as in Sect. 15.4.

Proposition 16.2 (Cointegration Regression) *For cointegrated I(1) processes $\{x_t\}$ and $\{y_t\}$ it holds with (16.4) and the notation introduced that*

$$(a)\ n(\hat{b} - b) \quad \xrightarrow{d} \quad \frac{\int_0^1 B_2(s)dB_1(s) + \Delta_{xv}}{\int_0^1 B_2^2(s)ds},$$

$$(b)\ n(1 - R_{uc}^2) \xrightarrow{d} \frac{\gamma_1(0)}{b^2 \int_0^1 B_2^2(s)ds},$$

$$(c)\ s^2 \quad \xrightarrow{p} \quad \gamma_1(0),$$

$$(d)\ t_b = \frac{\hat{b}-b}{s_b} \quad \xrightarrow{d} \quad \frac{\int_0^1 B_2(s)dB_1(s) + \Delta_{xv}}{\sqrt{\gamma_1(0) \int_0^1 B_2^2(s)ds}},$$

$$(e)\ dw \quad \xrightarrow{p} \quad 2\left(1 - \rho_v(1)\right),$$

where

$$\Delta_{xv} := \sum_{h=0}^{\infty} E(\Delta x_t v_{t+h}) \quad and \quad \gamma_1(0) := Var(v_t)$$

as $n \to \infty$.

Three remarks are to help with the interpretation. (i) Note again that supercon-sistency holds even if the regressors x_t correlate with the error terms v_t. This nice property, however, comes at a price: Unfortunately, the limiting distributions from Proposition 16.2(a) and (d) are (without further assumptions) generally not Gaussian. (ii) As a rule, for regressions involving trending (integrated) variables one empirically observes values of the coefficient of determination near one, which is explained by Proposition 16.2(b). Due to that, for trending (integrated) time series the coefficient of determination cannot be interpreted as usual: Since $\{y_t\}$ does not have a constant variance, the coefficient of determination does not give the percentage of the variance explained by the regression. (iii) At no point it was assumed that the error terms, $\{v_t\}$, are white noise. For the first order residual autocorrelation, it holds

$$\hat{\rho}_{\hat{v}}(1) = \frac{\sum_{t=1}^{n-1} \hat{v}_t \hat{v}_{t+1}}{\sum_{t=1}^{n} \hat{v}_t^2} \xrightarrow{p} \rho_v(1) = \frac{\mathrm{E}(v_t v_{t+1})}{\mathrm{Var}(v_t)} \, ,$$

which is just reflected by the behavior of the Durbin-Watson statistic.

Asymptotic Normality

The price for the superconsistency without exogeneity assumption is that the limiting distribution of the t-statistic is generally not Gaussian anymore. However, if v_t and Δx_s are stochastically independent for all t and s, then Krämer (1986) shows that asymptotic normality arises. This assumption, however, is stronger than necessary. It suffices to require, first, $\Delta_{xv} = 0$ and, second, that the Brownian motions B_1 and B_2 are independent. For independence of B_1 and B_2 we only need

$$\omega_{12} = \sum_{h=-\infty}^{\infty} \mathrm{E}(\Delta x_t v_{t+h}) = 0 \qquad (16.7)$$

as under this condition it holds that $B_i = \omega_i W_i$. Due to Proposition 10.4 the following corollary is obtained (also cf. Problem 16.4).

Corollary 16.1 (Asymptotic Normality) *If Δ_{xv} from Proposition 16.2 is zero, then it holds under (16.7) that*

$$t_b \xrightarrow{d} \mathcal{N}\left(0, \frac{\omega_1^2}{\gamma_1(0)}\right),$$

as $n \to \infty$.

If one has consistent estimators for the variance and the long-run variance, then the t-statistic can be modified as follows and can be applied with standard normal

distribution asymptotics under Corollary 16.1:

$$\tau_b := \sqrt{\frac{\hat{\gamma}_1(0)}{\hat{\omega}_1^2}} \, t_b \xrightarrow{d} \mathcal{N}(0, 1).$$

For the estimation of the long-run variance, we refer to the remarks in Sect. 15.2, in particular Example 15.2. As $\{v_t\}$ itself is not observable, $\hat{\gamma}_1(0)$ and $\hat{\omega}_1^2$ have to be calculated from $\hat{v}_t = y_t - \hat{b}x_t$.

Efficient Estimation

The assumptions $\Delta_{xv} = 0$ and $\omega_{12} = 0$ required for Corollary 16.1 are often not met in practice. We now consider modifications of LS resulting in limiting normality that get around such restrictions. To that end we have a closer look at the LS limit from Proposition 16.2 called now $\mathcal{L}(\hat{b})$:

$$n\,(\hat{b} - b) \xrightarrow{d} \mathcal{L}(\hat{b}).$$

We define the process $B_{1\cdot2}$,

$$B_{1\cdot2}(s) := B_1(s) - \omega_{12}\omega_2^{-2}B_2(s), \tag{16.8}$$

in such a way that it does not correlate with B_2:

$$E(B_{1\cdot2}(r)B_2(s)) = \min(r, s)\omega_{12} - \min(r, s)\omega_{12}\omega_2^{-2}\omega_2^2 = 0.$$

Because of normality the two processes are thus independent. The variance of the new process is

$$\text{Var}(B_{1\cdot2}(s)) = E(B_{1\cdot2}^2(s)) = s\omega_{1\cdot2}^2, \quad \omega_{1\cdot2}^2 := \omega_1^2 - \omega_{12}^2\omega_2^{-2}.$$

With

$$B_1(s) = B_{1\cdot2}(s) + B_2(s)\omega_2^{-2}\omega_{12}$$

we may now decompose the LS limit as follows:

$$\mathcal{L}(\hat{b}) = \left(\int_0^1 B_2^2(s)ds\right)^{-1}\left[\int_0^1 B_2(s)dB_{1\cdot2}(s) + \int_0^1 B_2(s)dB_2(s)\omega_2^{-2}\omega_{12} + \Delta_{xv}\right]$$
$$= \mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_3.$$

The first component, $\mathcal{L}_1$, is conditionally normal with mean zero, i.e.

$$\mathcal{L}_1 | B_2 \sim \mathcal{N}\left(0, \omega_{1\cdot 2}^2 \left(\int_0^1 B_2^2(s)ds\right)^{-1}\right),$$

which is true by Proposition 10.4. The second component defined as a multiple of $\int_0^1 B_2(s)dB_2(s) = (B_2^2(1) - 1)/2$ is stochastic and introduces skewness into $\mathcal{L}(\hat{b})$, while finally this distribution is shifted deterministically by Δ_{xv}. Consequently, for arbitrary $\varepsilon > 0$ one has

$$P(|\mathcal{L}_1| < \varepsilon) > P(|\mathcal{L}(\hat{b})| < \varepsilon),$$

see Saikkonen (1991, Theorem 3.1) Saikkonen. In that sense, $\mathcal{L}_1$ is more concentrated around zero than $\mathcal{L}(\hat{b})$ in general, such that intuitively the LS estimator is closest to zero for $\omega_{12} = \Delta_{xv} = 0$. This is the intuition for the following definition: Let $\hat{b}^+$ denote a cointegration estimator with

$$n(\hat{b}^+ - b) \xrightarrow{d} \mathcal{L}_1 \quad \text{where} \quad \mathcal{L}_1 | B_2 \sim \mathcal{N}\left(0, \omega^2 \left(\int_0^1 B_2^2(s)ds\right)^{-1}\right), \qquad (16.9)$$

for some positive constant ω; then $\hat{b}^+$ is said to be efficient (see Saikkonen 1991 for a more general discussion). To further justify this notion of efficiency we note that full information maximum likelihood estimation of a cointegrated system results in exactly this distribution, see Phillips (1991).

Efficient cointegration regressions are not only interesting because they achieve the lower bound for the standard error; more importantly, related t-type statistics are asymptotically normal, which allows for standard inference. Suppose we have an efficient estimator satisfying (16.9), and that $\hat{\omega}$ (typically computed from cointegration residuals) is consistent for ω. Then we define the t-type statistic

$$t^+ = \frac{\hat{b}^+ - b}{\hat{\omega}} \sqrt{\sum_{t=1}^n x_t^2}.$$

From the previous discussion it follows, see Phillips and Park (1988):

$$t^+ \xrightarrow{d} \mathcal{N}(0, 1).$$

Consequently, the caveat of superconsistent LS cointegration regression, lacking normality in general, is overcome by efficient estimators.

Let us repeat once more: LS cointegration estimation is efficient under the not very realistic assumption that $\omega_{12} = \Delta_{xv} = 0$. Several modifications of LS

achieving efficiency without this assumption have been proposed. First, the so-called dynamic LS estimator suggested independently by Saikkonen (1991) and Stock and Watson (1993) is settled in the time domain, see also Phillips and Loretan (1991); second, so-called frequency domain based modifications of LS have been suggested by Phillips and Hansen (1990) ("fully modified LS") or Park (1992) ("canonical cointegrating regression"); they all meet (16.9) and $t^+ \sim \mathcal{N}(0, 1)$, asymptotically.

Linear Time Trends

In Sect. 15.2 we considered linear time trends and I(1) processes, i.e. so-called integrated processes with drift:

$$x_t = \mu + x_{t-1} + e_t, \quad \mu \neq 0. \tag{16.10}$$

By repeated substitution one obtains

$$x_t = x_0 + \mu t + \sum_{j=1}^{t} e_j,$$

i.e. $\{x_t\}$ consists of a linear trend of the slope μ and an I(1) component; and of a starting value whose influence can be neglected such that we set $x_0 = 0$ w.l.o.g. In addition, let the cointegration relation (16.4) hold true. Consequently, $\{y_t\}$ as well exhibits a linear trend of the slope $b\,\mu$. The cointegration relation (16.4) simultaneously eliminates the deterministic linear time trend and the stochastic I(1) trend from both series. In this case the static LS regression from (16.5) yields an even faster rate of convergence ($n^{1.5}$ instead of n) and simultaneously, the limiting distribution is Gaussian. The following proposition is a special case of the more general results from West (1988). We prove it in Problem 16.5.

Proposition 16.3 (West) *We assume cointegrated I(1) processes $\{x_t\}$ and $\{y_t\}$ with drift (i.e. (16.10) with (16.4)). Then it holds for the regression (16.5) without intercept that*

$$n^{1.5}(\hat{b} - b) \xrightarrow{d} \mathcal{N}\left(0, \tfrac{3\omega_1^2}{\mu^2}\right)$$

as $n \to \infty$, where ω_1^2 is the long-run variance of $\{v_t\}$ from (16.4).

For the sake of completeness, note that the asymptotics from Proposition 16.3 is qualitatively retained if the regression is run *with* intercept. However, the variance of the Gaussian distribution is changed. It becomes $12\omega_1^2/\mu^2$. For practical inference, μ^2 has to be estimated from $\{x_t\}$ or $\{\Delta x_t\}$, while ω_1^2 can be estimated consistently

from the cointegration residuals $\hat{v}_t$. For $\mu = 1$ the limiting distribution from Proposition 16.3 equals that from (15.5), which is not coincidental, see also the proof in Problem 16.5: The scalar I(1) regressor with drift is dominated by the linear time trend; hence, the cointegration regression amounts to a trend stationary regression.

Note, however, that Proposition 16.3 holds only in our special case that x_t is a scalar I(1) variable. If one has a vector of I(1) regressors with drift instead of a scalar variable, then the asymptotic normality in general does not hold anymore, and the very fast convergence with rate $n^{1.5}$ is lost as well. Instead, Hansen (1992) proved results in line with Proposition 16.2 if there are several I(1) regressors of which at least one has a drift.

16.4 Cointegration Testing

If one regresses nonstationary (integrated) time series on each other, the interpretation of the regression outcome largely depends on whether the series are cointegrated or not. Hence, one has to test for the absence or presence of cointegration.

Residual-Based Dickey-Fuller Test

The idea of the following test for the null hypothesis of no cointegration dates back to Engle and Granger (1987), although a rigorous asymptotic treatment was provided later by Phillips and Ouliaris (1990). The idea is very simple. Without cointegration any linear combination of I(1) variables results in a series that too has a unit root. Hence, the Dickey-Fuller test is applied to LS residuals, which are computed from a regression with intercept:

$$\hat{u}_t = y_t - \hat{\alpha} - \hat{\beta} x_t .$$

Due to the included intercept, $\{\hat{u}_t\}$ are zero mean by construction. Hence, the DF regression in the second step may be run w.l.o.g. without intercept:

$$\hat{u}_t = \bar{a}\,\hat{u}_{t-1} + \bar{e}_t , \quad t = 1, \ldots, n .$$

The LS estimator $\bar{a}$ converges to 1 under the null hypothesis of a residual unit root with the rate known from Sect. 15.3. However, $\hat{\beta}$ does not converge to 0, but a limit characterized in Proposition 15.5. Consequently, the asymptotic distribution of $n(\bar{a} - 1)$ does not only depend on one WP but rather on two. Let $\bar{t}_a(2)$, involving 2 I(1) processes, denote the t-statistic related to $\bar{a} - 1$. Then the limit depends on two

standard Wiener processes W_1 and W_2:

$$\bar{t}_a(2) \xrightarrow{d} \overline{\mathcal{DF}}(W_1, W_2) \ . \tag{16.11}$$

Interestingly, this limit is free of nuisance parameters as long as $\{\Delta y_t\}$ and $\{\Delta x_t\}$ are white noise; in particular, it does not depend on the eventual correlation between $\{\Delta y_t\}$ and $\{\Delta x_t\}$; see also Problem 16.7. If $\{\Delta y_t\}$ and $\{\Delta x_t\}$ are not white noise processes, then $\bar{a}$ may be computed from a lag-augmented regression, or a modification to the test statistic in line with Phillips (1987) has to be applied. Critical values or p-values are most often taken from MacKinnon (1991, 1996). Here, we do not present a definition of the functional shape of the limit $\overline{\mathcal{DF}}(W_1, W_2)$; rather, to give at least an idea thereof, we consider now explicitly the less complicated case without constant. So, $\hat{u}_t$ and $\hat{\beta}$ are now from a regression without intercept,

$$\hat{u}_t = y_t - \hat{\beta} x_t \ .$$

We use a new notation to denote the subsequent DF regression

$$\hat{u}_t = \breve{a}\,\hat{u}_{t-1} + \breve{\varepsilon}_t \ , \quad t = 1, \ldots, n \ .$$

In Problem 16.7 the following result is given.

Proposition 16.4 (Phillips & Ouliaris) *Let $\{z_t\}$ with $z_t' = (y_t, x_t)$ be a random walk such that $\{\Delta y_t\}$ and $\{\Delta x_s\}$ are white noise processes not correlated for $t \neq s$, although we do allow for contemporaneous correlation. In case of no cointegration it holds with the notation introduced above that*

$$n(\breve{a} - 1) \xrightarrow{d} \frac{\int_0^1 U(t)\,dU(t)}{\int_0^1 U^2(t)\,dt} \ , \quad n \to \infty \ ,$$

with

$$U(t) := W_1(t) - \frac{\int_0^1 W_1(s)W_2(s)\,ds}{\int_0^1 W_2^2(s)\,ds} W_2(t) \ ,$$

where W_1 and W_2 are two independent Wiener processes.

The functional shape of this limit corresponds exactly to the one in (1.11); only that the WP W is replaced by U, which is, however, no longer a WP. Not surprisingly, the new process U is defined as residual from a projection of W_1 (corresponding to y) on W_2 (corresponding to x). Once more this shows the power and elegance of the functional limit theory approach introduced in Chap. 14.

Residual-Based KPSS Test

It comes in natural to apply also the KPSS test to regression residuals. As in Sect. 15.3 the hypotheses are now exchanged: We test for the null hypothesis of cointegration against the alternative of no cointegration. Working with LS cointegration residuals, $\hat{v}_t = y_t - \hat{b}x_t$, we have under the null hypothesis

$$\hat{v}_t = v_t - (\hat{b} - b)\, x_t \,,$$

where

$$n\,(\hat{b} - b) \xrightarrow{d} b_\infty$$

with the limiting distribution b_∞ characterized in Proposition 16.2. Interestingly, we observe by a FCLT (Proposition 14.1) that

$$n^{0.5}\,(\hat{b} - b)\, x_{\lfloor rn \rfloor} \;\Rightarrow\; b_\infty B_2(r)\,,$$

with B_2 being the Brownian motion behind $\{x_t\}$ from $z_t' = \left(\sum_{j=1}^t v_j,\, x_t \right)$. Therefore, it holds that $(\hat{b} - b)\, x_t$ converges to zero with growing sample size, and the empirical residuals are proxies of the unobserved cointegration deviation: $\hat{v}_t \approx v_t$. Hence, it is tempting to believe that a KPSS test applied to the sequence $\{\hat{v}_t\}$ behaves as if applied to $\{v_t\}$. This, however, is not correct, as we will demonstrate next, since the limit characterized in Proposition 15.4 is not recovered when working with cointegration residuals.

The KPSS test builds on the partial sum process $S_t = \sum_{j=1}^t \hat{v}_j$. Mimicking the proof of Proposition 14.2(a) in Problem 14.3, we obtain the following FCLT for the partial sum process:

$$n^{-0.5} S_{\lfloor rn \rfloor} \;=\; n^{-0.5} \sum_{j=1}^{\lfloor rn \rfloor} v_j - n\,(\hat{b} - b)\, n^{-1.5} \sum_{j=1}^{\lfloor rn \rfloor} x_j$$

$$\Rightarrow B_1(r) - b_\infty \int_0^r B_2(s)\, ds\,.$$

Notwithstanding that $\hat{v}_t \approx v_t$ we must thus not jump at the conclusion that the residual effect is negligible: The more careful analysis showed that the limit of the partial sum process depends on the distribution b_∞ arising from the cointegration regression.

What is more, we know that the LS limit b_∞ from Proposition 16.2 is plagued by the nuisance parameters Δ_{xv} and ω_{12}, except for the special case of Corollary 16.1. Therefore, Shin (1994) suggested to apply the KPSS test not with LS residuals but

with residuals from an efficient regression (now with intercept),

$$S_t^+ = \sum_{j=1}^{t} \hat{v}_j^+ \,, \quad \hat{v}_j^+ = y_t - \hat{a}^+ - \hat{b}^+ x_t \,,$$

where $\hat{b}^+$ is efficient in the sense of (16.9). Efficient cointegration regressions rely on removing Δ_{xv} and ω_{12} consistently. Hence, Shin (1994) showed that the limiting distribution of the KPSS test applied to efficient residuals is free of nuisance parameters, and he provided critical values. Let $\overline{\eta}^+(2)$ denote the residual-based KPSS statistic building on $v_j^+ = y_t - \hat{a}^+ - \hat{b}^+ x_t$, thus involving two I(1) variables. Under the null hypothesis of cointegration it holds asymptotically that (Shin, 1994, Thm. 2)

$$\overline{\eta}^+(2) \xrightarrow{d} \overline{\mathcal{CM}}(W_1, W_2) \,,$$

where

$$\overline{\mathcal{CM}}(W_1, W_2) = \int_0^1 \left[W_1(s) - sW_1(1) - \frac{\int_0^s \underline{W}_2(r)dr \int_0^1 \underline{W}_2(r)dW_1(r)}{\int_0^1 \underline{W}_2^2(r)dr} \right]^2 ds \,,$$
(16.12)

and $\underline{W}$ is again short for a demeaned Wiener process. For related work see Harris and Inder (1994) or Leybourne and McCabe (1994), although the latter paper considered only the case of LS residuals.

Error-Correction Test

The third test we look into is not residual-based. The analysis rather relies on the error-correction equation (16.2):

$$\Delta y_t = \gamma \, v_{t-1} + \text{differences} + \varepsilon_t \,.$$

The fact that we restrict the analysis to the error-correction equation of $\{y_t\}$ and that we ignore Eq. (16.3) has to be justified by the assumption

$$\gamma_x = 0 \,. \tag{16.13}$$

This assumptions implies that, when cointegration is present, only Δy_t reacts to the deviation from the equilibrium of the previous period. Because of the assumption (16.13), absence of cointegration means $\gamma = 0$, which is the null hypothesis. In order that there is an adjustment to the equilibrium in the case of

cointegration, $\gamma < 0$ under the alternative:

$$H_0 : \gamma = 0 \qquad \text{vs.} \qquad H_1 : \gamma < 0.$$

To provide further intuition for the test statistic, we rewrite the error-correction equation (16.2) by inserting the definition of v_{t-1}:

$$\Delta y_t = \gamma (y_{t-1} - b x_{t-1}) + \sum_{j=1}^{p} a_j \,\Delta y_{t-j} + \sum_{j=1}^{\ell} \alpha_j \,\Delta x_{t-j} + \varepsilon_t$$

$$= \gamma y_{t-1} + \theta x_{t-1} + \sum_{j=1}^{p} a_j \,\Delta y_{t-j} + \sum_{j=1}^{\ell} \alpha_j \,\Delta x_{t-j} + \varepsilon_t . \qquad (16.14)$$

Here, we defined

$$\theta = -\gamma b \,,$$

where the null hypothesis of course implies $\theta = 0$. Hence, one may test the null hypothesis by means of an F-type test statistic for $\gamma = \theta = 0$, which has been investigated by Boswijk (1994). Alternatively, one may employ a t-type test specifically for $\gamma = 0$ only as proposed by Banerjee, Dolado, and Mestre (1998). The following proposition characterizes the asymptotic behavior of the LS estimator $\hat{\gamma}$, cf. Banerjee, Dolado, and Mestre (1998, Proposition 1). The Wiener processes W_1 and W_2 are adopted from Proposition 14.4 with $z_t' = (y_t, x_t)$. In order to get a limiting distribution free of nuisance parameters, we assume that Δx_t and ε_s are uncorrelated at arbitrary points in time:

$$\mathrm{E}(\Delta x_t \varepsilon_s) = 0 \,; \qquad (16.15)$$

see also the proof in Problem 16.6.

Proposition 16.5 (BDM) *Let the I(1) processes $\{x_t\}$ and $\{y_t\}$ be not cointegrated, and let the exogeneity assumption (16.15) be fulfilled. Then, it holds for the LS estimator from the regression (16.14) that*

$$n \hat{\gamma} \overset{d}{\to} \frac{\int\limits_0^1 W_2^2 (s)\, ds \int\limits_0^1 W_1 (s)\, dW_1 (s) - \int\limits_0^1 W_1 (s)\, W_2 (s)\, ds \int\limits_0^1 W_2 (s)\, dW_1 (s)}{\int\limits_0^1 W_2^2(s)\, ds \int\limits_0^1 W_1^2(s)\, ds - \left(\int\limits_0^1 W_1(s) W_2(s)\, ds \right)^2}$$

as $n \to \infty$.

Obviously, this limiting distribution can be reshaped into the following form (which lends itself for a multivariate generalization with vectors $\{x_t\}$):

$$
n\,\hat{\gamma} \overset{d}{\to} \frac{\int\limits_0^1 W_1(s)dW_1(s) - \int\limits_0^1 W_1(s)W_2(s)ds \left(\int\limits_0^1 W_2^2(s)ds\right)^{-1} \int\limits_0^1 W_2(s)dW_1(s)}{\int\limits_0^1 W_1^2(s)ds - \int\limits_0^1 W_1(s)W_2(s)ds \left(\int\limits_0^1 W_2^2(s)ds\right)^{-1} \int\limits_0^1 W_1(s)W_2(s)ds}\,.
$$

In fact, Banerjee et al. (1998) suggest the computation of the t-statistic relating to $\hat{\gamma}$: $t_\gamma(2)$. As limiting distribution under H_0 one obtains:

$$
t_\gamma(2) \overset{d}{\to} \mathcal{BDM}(W_1, W_2) \tag{16.16}
$$

$$
= \frac{\int\limits_0^1 W_1(s)dW_1(s) - \int\limits_0^1 W_1(s)W_2(s)ds \left(\int\limits_0^1 W_2^2(s)ds\right)^{-1} \int\limits_0^1 W_2(s)dW_1(s)}{\sqrt{\int\limits_0^1 W_1^2(s)ds - \int\limits_0^1 W_1(s)W_2(s)ds \left(\int\limits_0^1 W_2^2(s)ds\right)^{-1} \int\limits_0^1 W_1(s)W_2(s)ds}}\,.
$$

Again, in most practical situations an intercept will be included in (16.14). If the corresponding test statistic is called $\bar{t}_\gamma(2)$, then it holds under the null hypothesis that

$$
\bar{t}_\gamma(2) \overset{d}{\to} \overline{\mathcal{BDM}}(W_1, W_2)\,. \tag{16.17}
$$

This limit has the same functional shape as $\mathcal{BDM}(W_1, W_2)$, only that W_i have to be replaced by the demeaned analogs $\underline{W}_i$, $i = 1, 2$.

Simulated critical values for conducting the tests can be found in Banerjee et al. (1998). One rejects for small values. From Ericsson and MacKinnon (2002) p-values are available, too.

Linear Time Trends

It has been mentioned in this and the previous chapter that in practice one would run regressions with an intercept to account for non-zero means of the series. But how should one proceed if the mean function follows a linear time trend, i.e. if the series are I(1) with drift? This is a quite realistic assumption for many economic and financial time series where positive growth rates are plausible. One might consider the analysis of detrended data, see Sect. 15.2. Note that the regression of detrended series is equivalent to including a linear time trend in the regression (see Frisch & Waugh, 1933):

$$
y_t = \hat{\alpha} + \hat{\delta}\,t + \hat{\beta}\,x_t + \hat{u}_t\,.
$$

This is why we call such regressions also **detrended regressions**. Similarly, one may augment the error-correction regression (16.14) by a linear time trend (and a constant). Many economists, however, do not run detrended regression or detrend the series even if the data display a linear time trend by eyeball inspection. Economically, it is often more meaningful to "explain" one trend by another instead of regressing deviations from linear trends on each other. Also statistically the regression of detrended data may not seem advisable since power losses are to be expected (see Hamilton, 1994, Sect. 19.2).

Running regressions with intercept only, i.e. without detrending, in the presence of linear time trends in the regressors has some subtle implications, however. Generally, the presence of a linear time trend in the data not accounted for in the regression will affect the limiting distributions. Just remember that a simple (or bivariate) cointegration regression in the presence of a linear time trend (Proposition 16.3) resembles more the detrending of a trend stationary process (Proposition 15.1) than a cointegration regression without linear trend (Proposition 16.2). More precisely, a linear time trend in $\{x_t\}$ will dominate the stochastic unit root in the following sense: If $\{x_t\}$ is I(1) with drift, $E(\Delta x_t) = \mu + e_t$, $\mu \neq 0$, or

$$x_t = x_0 + \mu\, t + \sum_{j=1}^{t} e_j, \quad t = 1, \ldots n,$$

then this process grows with rate n (and not $n^{0.5}$, see Proposition 14.1):

$$\frac{x_{\lfloor rn \rfloor}}{n} \;\Rightarrow\; 0 + \mu\, r + 0, \quad \mu \neq 0.$$

This provides an intuition for the following finding by Hansen (1992, Theorem 7): If $\{x_t\}$ is I(1) with drift and $\{y_t\}$ and $\{x_t\}$ are not cointegrated, and if a regression-based DF test for no cointegration is computed from a regression with intercept but without detrending, then the limiting distribution of the t-type DF statistic is not given by $\overline{\mathcal{DF}}(W_1, W_2)$ from (16.11), but rather by the detrended univariate distribution $\widetilde{\mathcal{DF}}$ given in (15.10). A corresponding result was established for the error-correction test by Hassler (2000a): If $\{x_t\}$ and $\{y_t\}$ are I(1) but not cointegrated, and if $\{x_t\}$ is integrated with drift and the error-correction test for no cointegration is computed from a regression with intercept but without detrending, then the limiting distribution of the t-statistic is not given by $\overline{\mathcal{BDM}}(W_1, W_2)$ from (16.17), but by the detrended Dickey-Fuller distribution $\widetilde{\mathcal{DF}}$. And similarly: If $\{x_t\}$ is I(1) with drift and cointegrated with $\{y_t\}$, and if a regression-based KPSS test for cointegration is computed from a regression with intercept only, then the limiting distribution of the KPSS statistic is not given by $\overline{\mathcal{CM}}(W_1, W_2)$ from (16.12), but rather by the detrended univariate distribution $\widetilde{\mathcal{CM}}$ given in (15.11); see Hassler (2000b). Hence, we have the following proposition.

Proposition 16.6 (Hansen & Hassler) *Consider the test statistics $\bar{t}_a(2)$, $\bar{t}_\gamma(2)$ or $\overline{\eta}^+(2)$ computed without detrending to test for the null hypothesis of (no)*

cointegration of the I(1) processes $\{x_t\}$ and $\{y_t\}$. With $\widetilde{\mathcal{DF}}$ from (15.10) and $\widetilde{\mathcal{CM}}$ from (15.11) it holds under the respective null hypotheses that

$$
\begin{aligned}
(a)\ \bar{t}_a(2) &\xrightarrow{d} \begin{cases} \overline{\mathcal{DF}}(W_1, W_2)\,, & \text{if } E(\Delta x_t) = 0 \\ \widetilde{\mathcal{DF}}\,, & \text{if } E(\Delta x_t) \neq 0 \end{cases}, \\[2mm]
(b)\ \bar{t}_\gamma(2) &\xrightarrow{d} \begin{cases} \overline{\mathcal{BDM}}(W_1, W_2)\,, & \text{if } E(\Delta x_t) = 0 \\ \widetilde{\mathcal{DF}}\,, & \text{if } E(\Delta x_t) \neq 0 \end{cases}, \\[2mm]
(c)\ \overline{\eta}^{+}(2) &\xrightarrow{d} \begin{cases} \overline{\mathcal{CM}}(W_1, W_2)\,, & \text{if } E(\Delta x_t) = 0 \\ \widetilde{\mathcal{CM}}\,, & \text{if } E(\Delta x_t) \neq 0 \end{cases},
\end{aligned}
$$

as $n \to \infty$.

Proposition 16.6 is not restricted to bivariate regressions, but carries over to the general multiple regression case as follows:

Consider single-equation regressions estimated by LS (or efficient variants thereof); regressions with intercept only on k I(1) regressors, of which at least one has a drift, result under the null hypothesis (of cointegration or no cointegration, respectively) in a limit as if one runs a detrended regression on $k-1$ I(1) regressors.

For $k = 1$ this reproduces Proposition 16.6. For a proof for the residual-based DF test with $k > 1$ see again Hansen (1992), and also the lucid discussion by Hamilton (1994, p. 596, 597); for a proof for the error-correction test see Hassler (2000a), and for the residual-based KPSS test see Hassler (2001) for $k > 1$.

In view of Proposition 16.6 one may identify two strategies when testing cointegration from regressions with intercept only; we restrict the discussion to the case of a scalar regressor x_t. First, one might ignore the possibility of linear trends and always work with critical values from $\overline{\mathcal{DF}}(W_1, W_2)$, $\overline{\mathcal{BDM}}(W_1, W_2)$ or $\overline{\mathcal{CM}}(W_1, W_2)$ provided for the case of regressions with intercept only under no drift; we call this strategy S_I (I for "ignoring"), and of course it is not correct if $\{x_t\}$ displays a linear trend in mean. Second, one may always account for the possibility of linear time trends and work with critical values from $\widetilde{\mathcal{DF}}$ or $\widetilde{\mathcal{CM}}$; let us call this strategy S_A (A for "account"), and note that it is only appropriate if $\{x_t\}$ is indeed dominated by a linear time trend. In a numerical example we discuss the consequences of S_I and S_A.

Example 16.5 (Testing under the Suspicion of Time Trends) We consider tests at a nominal significance level of 5 %. Let $\tilde{c}_1$ and $\bar{c}_2$ denote critical values from $\widetilde{\mathcal{DF}}$ or $\widetilde{\mathcal{CM}}$ and $\overline{\mathcal{DF}}(W_1, W_2)$, $\overline{\mathcal{BDM}}(W_1, W_2)$ or $\overline{\mathcal{CM}}(W_1, W_2)$, respectively. For the residual-based DF test by Phillips and Ouliaris (1990) we take asymptotic critical values from MacKinnon (1991): $\tilde{c}_1 = -3.41$ and $\bar{c}_2 = -3.34$. Coincidently, these critical values are not very distant. Strategy S_I results in a slightly too liberal test (rejecting more often than with 5 % probability) in the presence of a drift, while S_A is mildly conservative (rejecting less often than in 5 % of all cases) in the absence of

a linear trend in the regressor. So, for the residual-based DF test, Proposition 16.6 is not so relevant, since the distributions happen to differ not that much, and $\tilde{c}_1 \approx \bar{c}_2$. For the error-correction test by Banerjee et al. (1998), however, matters are not quite so harmless, since stronger size distortions are caused by a larger difference of the asymptotic critical values: $\tilde{c}_1 = -3.41$ and $\bar{c}_2 = -3.19$. With the residual-based KPSS test, things change qualitatively and quantitatively. Critical values from Kwiatkowski, Phillips, Schmidt, and Shin (1992) and Shin (1994) are $\tilde{c}_1 = 0.146$ and $\bar{c}_2 = 0.314$, and thus differ dramatically. Since one rejects for too large values, strategy S_I implies in the presence of a linear time trend a very conservative test; it will hardly reject the null hypothesis, which comes at a price of power of course. The other way round, without linear time trends strategy S_A will reject the true null hypothesis much too often resulting in an intolerably liberal test. ■

Clearly, none of the strategies S_I or S_A is generally acceptable when testing under the possibility of linear time trends. Fortunately, one often has strong a priori beliefs regarding the absence or presence of a linear time trend in the regressor. If one is convinced that a linear time trend is present in $\{x_t\}$, then one would apply e.g. $\bar{t}_a(2)$ or $\bar{t}_\gamma(2)$ with critical values from $\widetilde{\mathcal{DF}}$; if one believes that there is no linear time trend behind $\{x_t\}$, then critical values from $\overline{\mathcal{DF}}(W_1, W_2)$ or $\overline{\mathcal{BDM}}(W_1, W_2)$ must be recommended.

If one is not sure about the absence or presence of a linear time trend in the data, then there are (at least) two more strategies beyond S_I or S_A one may employ. As a third strategy, one may always test from detrended data. This clearly circumvents size distortions, but comes at a price of power losses as has been acknowledged for instance by Hansen (1992) or Hamilton (1994, p. 598). Fourth, one may rely on a pretest whether the regressor follows a linear trend or not. If a linear time trend in $\{x_t\}$ is significant, then one would apply e.g. $\bar{\eta}^+(2)$ with critical values from $\widetilde{\mathcal{CM}}$; if not, then critical values from $\overline{\mathcal{CM}}(W_1, W_2)$ should be applied. Such a pretesting strategy, however, will be troubled in small samples by the problem of controlling the significance level when carrying out a sequence of conditional tests (multiple testing). A recommendation whether the strategy of generally detrending or the strategy of pretesting is to be preferred, when the presence or absence of a linear time trend is debatable, will require future research.

16.5 Problems and Solutions

Problems

16.1 Show the equivalence of (16.1) and the ARDL(2,2) parameterization from Example 16.1.

16.2 Prove statements (b), (c) and (d) from Proposition 16.2.

16.3 Prove statement (e) from Proposition 16.2.

16.4 Prove Corollary 16.1.

16.5 Prove Proposition 16.3.

16.6 Prove Proposition 16.5 for the special case that no lagged differences are required to obtain white noise errors $\{\varepsilon_t\}$:

$$\Delta y_t = \gamma y_{t-1} + \theta x_{t-1} + \varepsilon_t .$$

16.7 Prove Proposition 16.4.

Solutions

16.1 Using γ and b, one obtains with $\Delta = 1 - L$ from (16.1):

$$y_t = y_{t-1} - (1 - a_1 - a_2)y_{t-1} + (c_0 + c_1 + c_2)x_{t-1}$$
$$- a_2(y_{t-1} - y_{t-2}) + c_0(x_t - x_{t-1}) - c_2(x_{t-1} - x_{t-2}) + \varepsilon_t$$
$$= a_1 y_{t-1} + a_2 y_{t-2} + c_0 x_t + c_1 x_{t-1} + c_2 x_{t-2} + \varepsilon_t .$$

Hence, the claim is already proved.

16.2 We proceed in the same way as in the proof of Proposition 15.5, only that we work under (16.6) when appealing to Proposition 14.4. We start with

$$n\left(1 - R_{uc}^2\right) = \frac{n^{-1} \sum\limits_{t=1}^{n} \hat{v}_t^2}{n^{-2} \sum\limits_{t=1}^{n} y_t^2} = \frac{s^2}{n^{-2} \sum\limits_{t=1}^{n} y_t^2} .$$

The numerator on the right-hand side is just

$$s^2 = n^{-1} \sum_{t=1}^{n} \left(y_t - \hat{b} x_t\right)^2$$

$$= n^{-1} \sum_{t=1}^{n} \left(b x_t - \hat{b} x_t + v_t\right)^2$$

$$= n^{-1} \sum_{t=1}^{n} \left[(b - \hat{b})^2 x_t^2 + 2(b - \hat{b})x_t v_t + v_t^2\right]$$

$$= n(b - \hat{b})^2 \frac{\sum\limits_{t=1}^{n} x_t^2}{n^2} + 2(b - \hat{b}) \frac{\sum\limits_{t=1}^{n} x_t v_t}{n} + \frac{\sum\limits_{t=1}^{n} v_t^2}{n} .$$

The first of the three remaining terms tends to zero as $\sum x_t^2$ is of order n^2 and $(b - \hat{b})$ is of order n^{-1}; correspondingly, the second expression tends to zero as $\sum x_t\, v_t$ grows with n; finally, the third term converges to $\mathrm{Var}(v_t)$ as a law of large numbers holds for $\{v_t^2\}$. This proves Proposition 16.2(c). By the same arguments, one establishes:

$$n^{-2} \sum_{t=1}^{n} y_t^2 = n^{-2} \sum_{t=1}^{n} \left(b^2 x_t^2 + 2\, b\, x_t\, v_t + v_t^2\right) \overset{d}{\to} b^2 \int_0^1 B_2^2(s)\, ds + 0 + 0 \,.$$

Hence, Proposition 16.2(b) is proved as well.

Finally, the behavior of the t-statistic with $s_b^2 = s^2/\sum x_t^2$ is clear again by Proposition 14.4:

$$t_b = \frac{\hat{b} - b}{s_b} = \frac{\frac{1}{n}\sum\limits_{t=1}^{n} x_t\, v_t}{s\sqrt{\frac{1}{n^2}\sum\limits_{t=1}^{n} x_t^2}} \overset{d}{\longrightarrow} \frac{\int\limits_0^1 B_2(s)dB_1(s) + \sum\limits_{h=0}^{\infty} \mathrm{E}\left(\varDelta x_t\, v_{t+h}\right)}{\sqrt{\gamma_1(0)\int\limits_0^1 B_2^2(s)ds}} \,.$$

16.3 In order to analyze the behavior of the Durbin-Watson statistic, we only need to study the numerator,

$$n^{-1} \sum_{t=2}^{n} (\hat{v}_t - \hat{v}_{t-1})^2 = n^{-1} \sum_{t=2}^{n} \left((b - \hat{b})\varDelta x_t + \varDelta v_t\right)^2$$

$$= (b - \hat{b})^2 \frac{\sum\limits_{t=2}^{n}(\varDelta x_t)^2}{n} + 2(b - \hat{b})\frac{\sum\limits_{t=2}^{n} \varDelta x_t\, \varDelta v_t}{n} + \frac{\sum\limits_{t=2}^{n}(\varDelta v_t)^2}{n} \,.$$

As $(b - \hat{b})$ tends to zero, there remains asymptotically

$$n^{-1} \sum_{t=2}^{n}(\varDelta \hat{v}_t)^2 \overset{p}{\longrightarrow} \mathrm{Var}(\varDelta v_t) = 2\,\mathrm{Var}(v_t) - 2\,\mathrm{Cov}(v_t, v_{t-1}) \,.$$

Hence, it holds as claimed:

$$dw = \frac{n^{-1} \sum\limits_{t=2}^{n}(\varDelta \hat{v}_t^2)}{s^2} \overset{p}{\longrightarrow} 2\,(1 - \rho_v(1)) \,,$$

as s^2 approaches $\mathrm{Var}(v_t)$ with n growing.

16.4 The result will follow from Proposition 16.2. By (16.7), we have $\omega_{12} = 0$. Due to the resulting diagonality of $\varOmega$, $\varOmega^{0.5}$ is diagonal as well, cf. $\varOmega_2$ from

Example 14.6. Hence it holds:

$$\begin{pmatrix} B_1(s) \\ B_2(s) \end{pmatrix} = \begin{pmatrix} \omega_1 W_1(s) \\ \omega_2 W_2(s) \end{pmatrix},$$

and ω_2 cancels from the limiting distribution of Proposition 16.2(d) (using the second assumption $\Delta_{xv} = 0$):

$$\frac{\omega_2 \int_0^1 W_2(s)dW_1(s)\omega_1}{\sqrt{\gamma_1(0)\omega_2^2 \int_0^1 W_2^2(s)ds}} = \frac{\omega_1}{\sqrt{\gamma_1(0)}} \frac{\int_0^1 W_2(s)dW_1(s)}{\sqrt{\int_0^1 W_2^2(s)ds}}.$$

According to Proposition 10.4, the stochastic quotient on the right-hand side follows a standard normal distribution. Hence, it holds that

$$\frac{\sqrt{\gamma_1(0)}}{\omega_1} t_b \overset{d}{\to} \mathcal{N}(0, 1),$$

which proves the corollary.

16.5 By assumption, it holds:

$$n^{-3} \sum_{t=1}^n x_t^2 = n^{-3} \left[\mu^2 \sum_{t=1}^n t^2 + 2\mu \sum_{t=1}^n \left(t \sum_{j=1}^t e_j \right) + \sum_{t=1}^n \left(\sum_{j=1}^t e_j \right)^2 \right].$$

We know from Proposition 14.2(c) and (e) that the second and the third expression in square brackets have to be divided by $n^{2.5}$ and n^2, respectively, such that they converge. However, in front of the square bracket there is n^{-3}, such that it holds with (15.2):

$$n^{-3} \sum_{t=1}^n x_t^2 \overset{d}{\to} \frac{\mu^2}{3}.$$

Hence we have the denominator of the LS estimator under control:

$$\hat{b} = b + \frac{\sum_{t=1}^n x_t v_t}{\sum_{t=1}^n x_t^2}.$$

In order to crack the numerator, we consider

$$\frac{x_{\lfloor sn \rfloor}}{n} = \frac{x_0}{n} + \frac{\mu \lfloor sn \rfloor}{n} + \frac{\sum_{j=1}^{\lfloor sn \rfloor} e_j}{n},$$

and due to Proposition 14.1 it holds

$$\frac{x_{\lfloor sn \rfloor}}{n} \approx \frac{\mu \lfloor sn \rfloor}{n} \xrightarrow{d} \mu s \, ,$$

and x_t is dominated by a linear trend, i.e. x_t behaves just as μt. Thus, as for the detrending in the trend stationary case, we obtain with a standard Wiener process V that (see Sect. 15.2):

$$n^{1.5}(\hat{b} - b) = \frac{n^{-1.5} \sum\limits_{t=1}^{n} x_t v_t}{n^{-3} \sum\limits_{t=1}^{n} x_t^2}$$

$$\xrightarrow{d} \frac{\mu \omega_1 \int\limits_0^1 s dV(s)}{\mu^2/3}$$

$$= \frac{3 \omega_1}{\mu} \int\limits_0^1 s dV(s)$$

$$\sim \mathcal{N}\left(0, \frac{3 \omega_1^2}{\mu^2}\right) \, ,$$

where $\omega_1 V(s)$ is the Brownian motion corresponding to $\sum_{j=1}^{t} v_j$, and ω_1^2 is the long-run variance of $\{v_t\}$. The normality of the integral follows from Example 9.2. Hence, the claim is proved.

16.6 Because of the simplifying assumption we consider the regression of (16.14) without differences. As LS estimator for the vector

$$\psi := \begin{pmatrix} \gamma \\ \theta \end{pmatrix}$$

one hence obtains for a sample $t = 1, \ldots, n$:

$$\hat{\psi} = D^{-1} \sum_{t=1}^{n} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} \Delta y_t \, , \tag{16.18}$$

where D has the form

$$D = \begin{pmatrix} \sum\limits_{t=1}^{n} y_{t-1}^2 & \sum\limits_{t=1}^{n} y_{t-1} x_{t-1} \\ \sum\limits_{t=1}^{n} x_{t-1} y_{t-1} & \sum\limits_{t=1}^{n} x_{t-1}^2 \end{pmatrix} \, .$$

Plugging in $\Delta y_t = \varepsilon_t$ under H_0, we obtain

$$\hat{\psi} = D^{-1} \sum_{t=1}^{n} \binom{y_{t-1}}{x_{t-1}} \varepsilon_t \,.$$

In the case of no cointegration, for using Proposition 14.4 we choose

$$z_t = \binom{y_t}{x_t} \,.$$

Then it holds for the matrix D:

$$n^{-2}D \xrightarrow{d} \begin{pmatrix} \int_0^1 B_1^2(s)\,ds & \int_0^1 B_1(s)\,B_2(s)\,ds \\ \int_0^1 B_1(s)\,B_2(s)\,ds & \int_0^1 B_2^2(s)\,ds \end{pmatrix} . \tag{16.19}$$

For the inverse, this implies

$$n^2 D^{-1} \xrightarrow{d} \frac{1}{\det} \begin{pmatrix} \int_0^1 B_2^2(s)\,ds & -\int_0^1 B_1(s)\,B_2(s)\,ds \\ -\int_0^1 B_1(s)\,B_2(s)\,ds & \int_0^1 B_1^2(s)\,ds \end{pmatrix} ,$$

where "det" stands for the determinant of the limiting matrix from (16.19). Here, the known inversion formula for (2×2)-matrices was applied:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{\det} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}, \quad \det = a\,d - b\,c \,. \tag{16.20}$$

Note that Proposition 14.4(b) was applied with z_{t-1} instead of z_t. This is unproblematic and can be justified by similar arguments like those leading to Corollary 14.1 in the univariate case.

Next, we analyze with Proposition 14.4(c)

$$n^{-1} \sum_{t=1}^{n} \binom{y_{t-1}}{x_{t-1}} \Delta y_t = n^{-1} \sum_{t=1}^{n} (z_t - w_t)\, w_{1,t}$$

$$\xrightarrow{d} \int_0^1 B(s)\, dB_1(s) + \sum_{h=0}^{\infty} E\,(w_t w_{1,t+h}) - E\,(w_t w_{1,t})$$

$$= \int_0^1 B(s)\, dB_1(s) + \sum_{h=1}^{\infty} \mathrm{E}\,(w_t w_{1,t+h})$$

$$= \int_0^1 B(s)\, dB_1(s) + 0\,,$$

where we used that $w_{1,t} = \Delta y_t = \varepsilon_t$ is free from serial correlation and is uncorrelated with Δx_s at each point in time, see (16.15):

$$\mathrm{E}\,(w_t w_{1,t+h}) = 0\,, \quad h > 0\,.$$

Thus, under the null hypothesis of no cointegration, we obtain that $\hat{\gamma}$ tends to zero. For this purpose we consider the first row of the limit of $n^2 D^{-1}$ multiplied by $\int B(s)\, dB_1(s)$:

$$n\hat{\gamma} \xrightarrow{d} \frac{\int_0^1 B_2^2(s)\, ds \int_0^1 B_1(s)\, dB_1(s) - \int_0^1 B_1(s) B_2(s)\, ds \int_0^1 B_2(s)\, dB_1(s)}{\det}.$$

This is almost the claim as "det" is defined as the determinant of the limit of $n^{-2}D$. Finally, note that $B_i = \omega_i W_i$ holds as $\Delta x_t = w_{2,t}$ and $\Delta y_s = \varepsilon_s = w_{1,s}$ are uncorrelated. Thus, the long-run variances cancel from the limiting distribution and one obtains the required result.

16.7 Under the null hypothesis of no cointegration we define $z_t' = (y_t, x_t)$ with long-run variance matrix of full rank. The corresponding vector Brownian motion $B' = (B_1, B_2)$ can be written in terms of independent WPs $W' = (W_1, W_2)$ as

$$B(t) = T W(t) = \begin{pmatrix} t_{11} W_1(t) + \frac{\omega_{12}}{\omega_2} W_2(t) \\ \omega_2 W_2(t) \end{pmatrix}.$$

Here, T is the triangular decomposition given in (14.7) with $TT' = \Omega$. The limit of $\hat{\beta}$ from Proposition 15.5(a) hence becomes

$$\omega_2 \beta_{\infty} = \frac{\int_0^1 B_1(s) W_2(s)\, ds}{\int_0^1 W_2^2(s)\, ds} = \frac{t_{11} \int_0^1 W_1(s) W_2(s)\, ds + \frac{\omega_{12}}{\omega_2} \int_0^1 W_2^2(s)\, ds}{\int_0^1 W_2^2(s)\, ds}. \qquad (16.21)$$

With this result one obtains a FCLT for the residuals $\hat{u}_t = y_t - \hat{\beta} x_t$:

$$n^{-0.5} \hat{u}_{\lfloor rn \rfloor} \Rightarrow B_1(r) - \beta_{\infty} B_2(r)$$
$$= (1, -\beta_{\infty})\, T\, W(r)$$

$$= t_{11} \left(W_1(r) - \frac{\int_0^1 W_1(s)W_2(s)\,ds}{\int_0^1 W_2^2(s)\,ds} W_2(r) \right)$$

$$= t_{11} U(r)\,.$$

Unfortunately, however, Proposition 14.2 does not apply directly since $U(r)$ is not a WP. Still, with the techniques we used to prove Proposition 14.2(e) and (f) in Problems 14.3 and 14.5, we can establish (omitting details)

$$n^{-2} \sum_{t=1}^n \hat{u}_{t-1}^2 \xrightarrow{d} t_{11}^2 \int_0^1 U^2(t)\,dt\,,$$

$$n^{-1} \sum_{t=1}^n \hat{u}_{t-1} \Delta \hat{u}_t \xrightarrow{d} t_{11}^2 \frac{U^2(1)}{2} - \frac{1}{2} \left(\omega_1^2 - 2\beta_\infty \omega_{12} + \beta_\infty^2 \omega_2^2 \right)\,,$$

where the last limit arises because $\{\Delta z_t\}$ is white noise such that ω_i^2 and ω_{12} coincide with the (co)variances. Further, note by $B = TW$ that

$$\omega_{12}\beta_\infty = \frac{t_{11}\omega_{12}}{\omega_2} \frac{\int_0^1 W_1(s)W_2(s)ds}{\int_0^1 W_2^2(s)ds} + \frac{\omega_{12}^2}{\omega_2^2}\,. \qquad (16.22)$$

Remember from (14.7) that

$$t_{11}^2 = \omega_1^2 - \frac{\omega_{12}^2}{\omega_2^2}\,.$$

Consequently, for

$$\check{a} = 1 + \frac{\sum_{t=1}^n \hat{u}_{t-1}\Delta \hat{u}_t}{\sum_{t=1}^n \hat{u}_{t-1}^2}$$

we get by (16.21) and (16.22) that

$$n(\check{a} - 1) \xrightarrow{d} \frac{\frac{U^2(1)}{2} - \frac{1}{2}\left(1 + \left(\frac{\int_0^1 W_1(s)W_2(s)ds}{\int_0^1 W_2^2(s)ds}\right)^2\right)}{\int_0^1 U^2(t)dt}\,.$$

The numerator of this limit may be condensed. Use the product rule from Example 11.5 to obtain

$$W_1(1)W_2(1) = \int_0^1 W_1(s)dW_2(s) + \int_0^1 W_2(s)dW_1(s)\,.$$

It follows that

$$\frac{U^2(1)}{2} = \frac{W_1^2(1)}{2} - \frac{\int_0^1 W_1(s)W_2(s)ds}{\int_0^1 W_2^2(s)ds} \left(\int_0^1 W_1(s)dW_2(s) + \int_0^1 W_2(s)dW_1(s) \right)$$

$$+ \left(\frac{\int_0^1 W_1(s)W_2(s)ds}{\int_0^1 W_2^2(s)ds} \right)^2 \frac{W_2^2(1)}{2}.$$

Once more by Ito's lemma $\frac{W_i^2(1)}{2} = \int_0^1 W_i(s)dW_i(s) + \frac{1}{2}$, such that

$$\frac{U^2(1)}{2} = \int_0^1 W_1(s)dW_1(s) + \frac{1}{2}$$

$$- \frac{\int_0^1 W_1(s)W_2(s)ds}{\int_0^1 W_2^2(s)ds} \left(\int_0^1 W_1(s)dW_2(s) + \int_0^1 W_2(s)dW_1(s) \right)$$

$$+ \left(\frac{\int_0^1 W_1(s)W_2(s)ds}{\int_0^1 W_2^2(s)ds} \right)^2 \left(\int_0^1 W_2(s)dW_2(s) + \frac{1}{2} \right)$$

$$= \int_0^1 U(t)\,dU(t) + \frac{1}{2} \left(1 + \left(\frac{\int_0^1 W_1(s)W_2(s)ds}{\int_0^1 W_2^2(s)ds} \right)^2 \right).$$

This provides the expression for the limiting distribution given in Proposition 16.4 as required.

References

Banerjee, A., Dolado, J. J., & Mestre R. (1998). Error-correction mechanism tests for cointegration in a single-equation framework. *Journal of Time Series Analysis, 19*, 267–283.

Boswijk, H. P. (1994). Testing for an unstable root in conditional and structural error correction models. *Journal of Econometrics, 63*, 37–60.

Davidson, J., Hendry, D. F., Srba, F., & Yeo S. (1978). Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom. *Economic Journal, 88*, 661–692.

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

Ericsson, N. R., & MacKinnon, J. G. (2002). Distributions of error correction tests for cointegration. *Econometrics Journal, 5*, 285–318.

Frisch, R., & Waugh, F. V. (1933). Partial time regressions as compared with individual trends. *Econometrica, 1*, 387–401.

Hamilton, J. (1994). *Time series analysis*. Princeton: Princeton University Press.

Hansen, B. E. (1992). Efficient estimation and testing of cointegrating vectors in the presence of deterministic trends. *Journal of Econometrics, 53*, 87–121.

Harris, D., & Inder, B. (1994). A test of the null hypothesis of cointegration. In C. P. Hargreaves (Ed.), *Nonstationary time series analysis and cointegration* (pp. 133–152). Oxford/New York: Oxford University Press.

Hassler, U. (2000a). Cointegration testing in single error-correction equations in the presence of linear time trends. *Oxford Bulletin of Economics and Statistics, 62*, 621–632.

Hassler, U. (2000b). The KPSS test for cointegration in case of bivariate regressions with linear trends. *Econometric Theory, 16*, 451–453.

Hassler, U. (2001). The effect of linear time trends on the KPSS test for cointegration. *Journal of Time Series Analysis, 22*, 283–292.

Johansen, S. (1995). *Likelihood-based inference in cointegrated vector autoregressive models*. Oxford/New York: Oxford University Press.

Krämer, W. (1986). Least squares regression when the independent variable follows an ARIMA process. *Journal of the American Statistical Association, 81*, 150–154.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics, 54*, 159–178.

Leybourne, S. J., & McCabe, B. P. M. (1994). A simple test for cointegration. *Oxford Bulletin of Economics and Statistics, 56*, 97–103.

MacKinnon, J. G. (1991). Critical values for co-integration tests. In R. F. Engle, & C. W. J. Granger (Eds.), *Long-run economic relationships* (pp. 267–276). Oxford/New York: Oxford University Press.

MacKinnon, J. G. (1996). Numerical distribution functions for unit root and cointegration tests. *Journal of Applied Econometrics, 11*, 601–618.

Park, J. Y. (1992). Canonical cointegrating regressions. *Econometrica, 60*, 119–143.

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

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

Phillips, P. C. B. (1991). Optimal inference in cointegrated systems. *Econometrica, 59*, 283–306.

Phillips, P. C. B., & Durlauf, S. N. (1986). Multiple time series regression with integrated processes. *Review of Economic Studies, LIII*, 473–495.

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

Phillips, P. C. B., & Loretan, M. (1991). Estimating long-run economic equilibria. *Review of Economic Studies, 58*, 407–436.

Phillips, P. C. B., & Ouliaris, S. (1990). Asymptotic properties of residual based tests for cointegration. *Econometrica, 58*, 165–193.

Phillips, P. C. B., & Park, J. Y. (1988). Asymptotic equivalence of ordinary least squares and generalized least squares in regressions with integrated regressors. *Journal of the American Statistical Association, 83*, 111–115.

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

Shin, Y. (1994). A residual-based test of the Null of cointegration against the alternative of no cointegration. *Econometric Theory, 10*, 91–115.

Stock, J. H. (1987). Asymptotic properties of least squares estimators of cointegrating vectors. *Econometrica, 55*, 1035–1056.

Stock, J. H., & Watson, M. W. (1993). A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica, 61*, 783–820.

West, K. D. (1988). Asymptotic normality, when regressors have a unit root. *Econometrica, 56*, 1397–1418.