

Review Session 6

Sangmyung Ha

Fall 2026

1 Cointegrating Regression

Let $z_t = (y_t, x_t')' \in \mathbb{R}^{m+1}$ be $I(1)$. This means that the first differenced series Δz_t will be stationary, and under very general assumptions, the partial sum process

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \Delta z_t = \frac{1}{\sqrt{n}} z_{[nr]}$$

will converge to a Brownian motion with variance given by the long-run variance of Δz_t

$$\Omega_z := \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} z_n z_n'$$

Suppose $\Omega_z > 0$. Then, for any nonzero $v \in \mathbb{R}^{m+1}$, we have

$$v' \Omega_z v > 0$$

This implies that $v' z_n$ cannot be $I(0)$ for any $v \in \mathbb{R}^{m+1}$. To see this, note that the variance of $v' z_n$ is given by

$$\mathbb{E}(v' z_n)^2 = v' \mathbb{E} z_n z_n' v = n v' \Omega_z v + o(n)$$

which grows at the rate n . The components of the vector z_t are said to be *co-integrated* if

1. All components of z_t are $I(1)$.
2. There exists a vector $\alpha \in \mathbb{R}^{m+1}$ so that $\alpha' z_t \sim I(0)$.

In this case, the vector $\alpha \in \mathbb{R}^{m+1}$ is called the co-integrating vector. If $\alpha \in \mathbb{R}^{m+1}$ is a co-integrating vector, we see that

$$\alpha' \Omega_z \alpha = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}(\alpha' z_n)^2$$

is zero so that Ω_z is necessarily singular. We see that a co-integrating vector $\alpha \in \mathbb{R}^{m+1}$ must be an eigenvector associated with the eigenvalue zero. It is possible to have multiple linearly independent co-integrating vectors, in which case the eigenspace of Ω_z associated with the eigenvalue zero will have multiple dimensions.

1.1 Estimation of co-integrating vector

Suppose there exists a unique co-integrating vector α such that $\alpha' z_t = u_t \sim I(0)$. Without loss of generality, we may scale α so that its first element is one. This means that we may write

$$u_t = \alpha' z_t = y_t - \beta' x_t$$

where $\alpha = (1, -\beta')'$. We may estimate β by running OLS in the regression¹

$$y_t = \beta' x_t + u_t$$

To derive the limit distribution of the OLS estimator $\hat{\beta}$, we consider the invariance principle applied to $w_t = \begin{pmatrix} u_t \\ \Delta x_t \end{pmatrix}$. Note that we have

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} w_t = \begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} u_t \\ \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \Delta x_t \end{pmatrix} = \begin{pmatrix} B_{1n}(r) \\ B_{2n}(r) \end{pmatrix} \rightarrow_d \begin{pmatrix} B_1(r) \\ B_2(r) \end{pmatrix} = B(r)$$

where $\text{Var}(B(1)) = \Omega = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{pmatrix}$. Here, $\Omega = \sum_{k=-\infty}^{\infty} \Gamma(k)$ is the long run variance of w_t . Likewise, we define $\Sigma = \Gamma(0)$ and $\Delta = \sum_{k=0}^{\infty} \Gamma(k)$ to be the variance and one-sided long-run variance of w_t with elements given by²

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}, \quad \Delta = \begin{pmatrix} \delta_{11} & \delta_{12} \\ \delta_{21} & \delta_{22} \end{pmatrix}$$

Then, we have

$$\begin{aligned} n(\hat{\beta} - \beta) &= \left(\sum_{t=1}^n \frac{x_t}{\sqrt{n}} \frac{x_t'}{\sqrt{n}} \frac{1}{n} \right)^{-1} \left(\sum_{t=1}^n \frac{x_t}{\sqrt{n}} \frac{u_t}{\sqrt{n}} \right) \\ &\rightarrow_d \left(\int_0^1 B_2(r) B_2(r)' dr \right)^{-1} \left(\int_0^1 B_2(r) dB_1(r) + \delta'_{12} \right) \end{aligned}$$

where δ'_{12} is asymptotic bias given by

$$\delta_{12} = \sum_{k=0}^{\infty} \mathbb{E}[u_t \Delta x'_{t-k}]$$

Although we have found the limit distribution, inference on $\hat{\beta}$ is cumbersome because of the nuisance parameter δ'_{12} and correlation between B_1 and B_2 . To fix this, we may use (1) Canonical Cointegrating Regression or (2) Fully modified OLS.

¹In the usual regression, β is identified by orthogonality of x_t and u_t , which motivates using the least squares estimator. In the cointegrating regression, x_t and u_t need not be orthogonal. However, OLS can still be consistent because $\frac{1}{n} \sum_{t=1}^n x_t x_t'$ is of higher order than $\frac{1}{n} \sum_{t=1}^n x_t u_t$.

²Note that the delta notation Δ for the one-sided long run variance is different from the notation for first difference Δ .

1.2 Canonical Cointegrating Regression

Consider a stationary transformation of y_t^* and x_t^* given by

$$\begin{aligned} y_t^* &= y_t - \beta' \Delta_2' \Sigma^{-1} w_t - \omega_{12} \Omega_{22}^{-1} \Delta x_t \\ x_t^* &= x_t - \Delta_2' \Sigma^{-1} w_t \end{aligned}$$

where $\Delta_2 = \begin{pmatrix} \delta_{12} \\ \delta_{22} \end{pmatrix} \in \mathbb{R}^{m+1 \times m}$. This implies that

$$\begin{aligned} y_t^* &= x_t' \beta + u_t - \beta' \Delta_2' \Sigma^{-1} w_t - \omega_{12} \Omega_{22}^{-1} \Delta x_t \\ &= (x_t - \Delta_2' \Sigma^{-1} w_t)' \beta + u_t - \omega_{12} \Omega_{22}^{-1} \Delta x_t \\ &= x_t^{*'} \beta + u_t^* \end{aligned}$$

where

$$u_t^* = u_t - \omega_{12} \Omega_{22}^{-1} \Delta x_t$$

The CCR estimator $\hat{\beta}^*$ is defined as the regression coefficient from running y_t^* on x_t^* .

$$\begin{aligned} \hat{\beta}^* &= \left(\sum_{t=1}^n x_t^* x_t^{*'} \right)^{-1} \sum_{t=1}^n x_t^* y_t^* \\ &= \beta + \left(\sum_{t=1}^n x_t^* x_t^{*'} \right)^{-1} \sum_{t=1}^n x_t^* u_t^* \end{aligned}$$

We apply the invariance principle to the vector of transformed variables

$$w_t^* = \begin{pmatrix} u_t^* \\ \Delta x_t^* \end{pmatrix} = \begin{pmatrix} 1 & -\omega_{12} \Omega_{22}^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} u_t \\ \Delta x_t \end{pmatrix} - \begin{pmatrix} 0 \\ \Delta_2' \Sigma^{-1} \Delta w_t \end{pmatrix}$$

Given that the IP holds for w_t , it also holds for w_t^* and the limit Brownian motion is given by

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} w_t^* &= \begin{pmatrix} 1 & -\omega_{12} \Omega_{22}^{-1} \\ 0 & I \end{pmatrix} \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} w_t - \begin{pmatrix} 0 \\ \Delta_2' \Sigma^{-1} \frac{1}{\sqrt{n}} (w_{[nr]} - w_0) \end{pmatrix} \\ &\rightarrow_d \begin{pmatrix} 1 & -\omega_{12} \Omega_{22}^{-1} \\ 0 & I \end{pmatrix} B(r) \\ &= \begin{pmatrix} B_1(r) - \omega_{12} \Omega_{22}^{-1} B_2(r) \\ B_2(r) \end{pmatrix} \\ &= \begin{pmatrix} B_{1,2}(r) \\ B_2(r) \end{pmatrix} \end{aligned}$$

since $\frac{1}{\sqrt{n}} (w_{[nr]} - w_0) = O_p\left(\frac{1}{\sqrt{n}}\right)$. Since B_1 and B_2 are jointly normal, $B_{1,2}$ defined by

$$B_{1,2}(r) = B_1(r) - \omega_{12} \Omega_{22}^{-1} B_2(r)$$

is the projection residual $B_1(r) - \mathbb{E}[B_1(r)|B_2(r)]$. Also, for all $r, s \in [0, 1]$, we have

$$\begin{aligned}\mathbb{E}[B_2(s)B_{1,2}(r)] &= \mathbb{E}[B_2(s)B_1(r)] - \mathbb{E}[B_2(s)B_2(r)']\Omega_{22}^{-1}\omega_{21} \\ &= \min(s, r)\{\omega_{21} - \Omega_{22}\Omega_{22}^{-1}\omega_{21}\} \\ &= 0.\end{aligned}$$

so that B_2 and $B_{1,2}$ are independent processes. Naturally, the long-run variance of w_t^* is block diagonal and is given by

$$\begin{aligned}\Omega^* &= \begin{pmatrix} 1 & -\omega_{12}\Omega_{22}^{-1} \\ 0 & I \end{pmatrix} \Omega \begin{pmatrix} 1 & -\omega_{12}\Omega_{22}^{-1} \\ 0 & I \end{pmatrix}' \\ &= \begin{pmatrix} 1 & -\omega_{12}\Omega_{22}^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \Omega_{22} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\Omega_{22}^{-1}\omega_{12}' & I \end{pmatrix} \\ &= \begin{pmatrix} \omega_{11} - \omega_{12}\Omega_{22}^{-1}\omega_{21} & 0 \\ 0 & \Omega_{22} \end{pmatrix}\end{aligned}$$

Using the fundamental lemmas, we have

$$\begin{aligned}n(\hat{\beta}^* - \beta) &= \left(\sum_{t=1}^n \frac{x_t^*}{\sqrt{n}} \frac{x_t^{*'}}{\sqrt{n}} \frac{1}{n} \right)^{-1} \sum_{t=1}^n \frac{x_t^*}{\sqrt{n}} \frac{u_t^*}{\sqrt{n}} \\ &\rightarrow_d \left(\int_0^1 B_2(r)B_2'(r)dr \right)^{-1} \left(\int_0^1 B_2(r)dB_{1,2}(r) + \delta_{12}^{*'} \right)\end{aligned}$$

where $\delta_{12}^{*'}$ is asymptotic bias given by

$$\delta_{12}^* = \sum_{k=0}^{\infty} \mathbb{E}[u_t^* \Delta x_{t-k}^{*'}]$$

Let $a = (1, -\omega_{12}\Omega_{22}^{-1})'$, so that $u_t^* = a'w_t$. Since $\Gamma(k) \rightarrow 0$, telescoping gives

$$\sum_{k=0}^{\infty} \mathbb{E}[\Delta w_{t-k} w_t'] = \lim_{K \rightarrow \infty} \sum_{k=0}^K \mathbb{E}[\Delta w_{t-k} w_t'] = \lim_{K \rightarrow \infty} \mathbb{E}[(w_t - w_{t-K-1})w_t'] = \lim_{K \rightarrow \infty} \{\Sigma - \Gamma(K+1)'\} = \Sigma.$$

Consequently,

$$\begin{aligned}\delta_{12}^{*'} &= \sum_{k=0}^{\infty} \mathbb{E}[\Delta x_{t-k}^* u_t^*] \\ &= \sum_{k=0}^{\infty} \mathbb{E}[(\Delta x_{t-k} - \Delta_2' \Sigma^{-1} \Delta w_{t-k})w_t']a \\ &= \left(\Delta_2' - \Delta_2' \Sigma^{-1} \sum_{k=0}^{\infty} \mathbb{E}[\Delta w_{t-k} w_t'] \right) a \\ &= (\Delta_2' - \Delta_2' \Sigma^{-1} \Sigma) a = 0.\end{aligned}$$

This implies that the limit distribution of CCR estimator is given by

$$n(\hat{\beta}^* - \beta) \rightarrow_d \left(\int_0^1 B_2(r)B_2'(r)dr \right)^{-1} \int_0^1 B_2(r)dB_{1,2}(r).$$

1.3 Fully modified OLS

The fully modified OLS procedure directly modifies the OLS estimator by

$$\hat{\beta}^+ = \left(\sum_{t=1}^n x_t x_t' \right)^{-1} \left(\sum_{t=1}^n x_t y_t^+ - n\delta^+ \right)$$

where

$$\begin{aligned} y_t^+ &= y_t - \omega_{12} \Omega_{22}^{-1} \Delta x_t \\ \delta^+ &= \delta'_{12} - \delta'_{22} \Omega_{22}^{-1} \omega_{21} \end{aligned}$$

Note that we have

$$\begin{aligned} \hat{\beta}^+ &= \left(\sum_{t=1}^n x_t x_t' \right)^{-1} \left(\sum_{t=1}^n x_t y_t^+ - n\delta^+ \right) \\ &= \left(\sum_{t=1}^n x_t x_t' \right)^{-1} \left(\sum_{t=1}^n x_t (y_t - \omega_{12} \Omega_{22}^{-1} \Delta x_t) - n\delta^+ \right) \\ &= \beta + \left(\sum_{t=1}^n x_t x_t' \right)^{-1} \left(\sum_{t=1}^n x_t (u_t - \omega_{12} \Omega_{22}^{-1} \Delta x_t) - n\delta^+ \right) \\ &= \beta + \left(\sum_{t=1}^n x_t x_t' \right)^{-1} \left(\sum_{t=1}^n x_t u_t - \sum_{t=1}^n x_t \omega_{12} \Omega_{22}^{-1} \Delta x_t - n(\delta'_{12} - \delta'_{22} \Omega_{22}^{-1} \omega_{21}) \right) \end{aligned}$$

Thus, we see that

$$\begin{aligned} n(\hat{\beta}^+ - \beta) &= \left(\sum_{t=1}^n \frac{x_t}{\sqrt{n}} \frac{x_t'}{\sqrt{n}} \frac{1}{n} \right)^{-1} \left(\sum_{t=1}^n \frac{x_t}{\sqrt{n}} \frac{u_t}{\sqrt{n}} - \left(\sum_{t=1}^n \frac{x_t}{\sqrt{n}} \frac{\Delta x_t'}{\sqrt{n}} \right) \Omega_{22}^{-1} \omega'_{12} - \delta'_{12} + \delta'_{22} \Omega_{22}^{-1} \omega_{21} \right) \\ &\rightarrow_d \left(\int_0^1 B_2(r) B_2(r)' dr \right)^{-1} \left(\int_0^1 B_2(r) dB_1(r) - \left(\int_0^1 B_2(r) dB_2(r)' \right) \Omega_{22}^{-1} \omega'_{12} \right) \end{aligned}$$

In particular because the asymptotic bias from $\int_0^1 B_2(r) dB_1(r)$ and $\int_0^1 B_2(r) dB_2(r)'$ are δ'_{12} and δ'_{22} respectively and $\omega'_{12} = \omega_{21}$. Since $B_{1,2}$ is defined by

$$B_{1,2}(r) = B_1(r) - \omega_{12} \Omega_{22}^{-1} B_2(r)$$

The asymptotic distribution of FM-OLS is identical to that of the CCR estimator.

2 Unit Root Tests

Suppose we want to test whether a time series y_t has a unit root. To this end, we run the regression

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

and test $\alpha = 1$. Note that the OLS estimator for $\hat{\alpha}$ is given by

$$\begin{aligned}\hat{\alpha} &= \left(\sum_{t=2}^n y_{t-1}^2 \right)^{-1} \left(\sum_{t=2}^n y_{t-1} y_t \right) \\ &= 1 + \left(\sum_{t=2}^n y_{t-1}^2 \right)^{-1} \left(\sum_{t=2}^n y_{t-1} u_t \right)\end{aligned}$$

under the null. Also under the null, $y_t = \sum_{i=1}^t u_i \sim I(1)$ and we rewrite our expression so that the fundamental lemmas may be applied.

$$\begin{aligned}n(\hat{\alpha} - 1) &= \left(\sum_{t=2}^n \frac{y_{t-1}}{\sqrt{n}} \frac{y_{t-1}}{\sqrt{n}} \frac{1}{n} \right)^{-1} \left(\sum_{t=2}^n \frac{y_{t-1}}{\sqrt{n}} \frac{u_t}{\sqrt{n}} \right) \\ &\rightarrow_d \left(\int_0^1 B(r)^2 dr \right)^{-1} \left(\int_0^1 B(r) dB(r) + \lambda \right)\end{aligned}$$

where we've used IP on $w_t = u_t$. $B(r)$ is a Brownian motion with variance $\omega = \sum_{k=-\infty}^{\infty} \mathbb{E}[u_t u_{t-k}]$ and $\lambda = \sum_{k=1}^{\infty} \mathbb{E}[u_t u_{t-k}]$. The limit distribution is not normal and depends on the nuisance parameters (ω, λ) . Furthermore, dividing by the usual OLS standard error does not get rid of the nuisance parameters. To see this, note that the t-statistic obtained under the classical OLS assumptions is given by

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

The limit distribution of $t(\hat{\alpha})$ is given by

$$\begin{aligned}t(\hat{\alpha}) &= \frac{\hat{\alpha} - 1}{\hat{\sigma} \left(\sum_{t=2}^n y_{t-1}^2 \right)^{-1/2}} \\ &= \hat{\sigma}^{-1} \left(\sum_{t=2}^n \frac{y_{t-1}}{\sqrt{n}} \frac{y_{t-1}}{\sqrt{n}} \frac{1}{n} \right)^{-1/2} \left(\sum_{t=2}^n \frac{y_{t-1}}{\sqrt{n}} \frac{u_t}{\sqrt{n}} \right) \\ &\rightarrow_d \sigma^{-1} \left(\int_0^1 B(r)^2 dr \right)^{-1/2} \left(\int_0^1 B(r) dB(r) + \lambda \right)\end{aligned}$$

where we may easily show $\hat{\sigma} \rightarrow_p \sigma$ using the expression

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{t=1}^n \hat{u}_t^2$$

with

$$\hat{u}_t = y_t - \hat{\alpha} y_{t-1} = u_t + (1 - \hat{\alpha}) y_{t-1}$$

Again, we see that the limit distribution of $t(\hat{\alpha})$ still depends on nuisance parameters (σ, λ) .

2.1 Dickey-Fuller Test

If we are willing to assume that u_t is orthogonal white noise, the long-run variance of $w_t = u_t$ is just the contemporaneous variance σ^2 . Also, we have $\lambda = 0$. This implies that

$$\begin{aligned} n(\hat{\alpha} - 1) &\rightarrow_d \left(\int_0^1 B(r)^2 dr \right)^{-1} \left(\int_0^1 B(r) dB(r) + \lambda \right) \\ &= \left(\int_0^1 (\sigma W(r))^2 dr \right)^{-1} \left(\int_0^1 \sigma W(r) d(\sigma W(r)) \right) \\ &= \left(\int_0^1 W(r)^2 dr \right)^{-1} \left(\int_0^1 W(r) dW(r) \right) \end{aligned}$$

since $B(r) = \sigma W(r)$ where $W(r)$ is a standard Brownian motion. Also, we have

$$\begin{aligned} t(\hat{\alpha}) &\rightarrow_d \sigma^{-1} \left(\int_0^1 B(r)^2 dr \right)^{-1/2} \left(\int_0^1 B(r) dB(r) + \lambda \right) \\ &= \sigma^{-1} \left(\int_0^1 (\sigma W(r))^2 dr \right)^{-1/2} \left(\int_0^1 \sigma W(r) d(\sigma W(r)) \right) \\ &= \left(\int_0^1 W(r)^2 dr \right)^{-1/2} \left(\int_0^1 W(r) dW(r) \right) \end{aligned}$$

Both limiting distributions are free of nuisance parameters.

2.2 Augmented Dickey-Fuller Test

In many cases, it is not possible to assume that u_t is an orthogonal white noise. In this case, we may assume that u_t is an invertible AR(p) process given by

$$u_t = \sum_{k=1}^p \alpha_k u_{t-k} + \varepsilon_t$$

where $\varepsilon_t \sim WN(0, \sigma^2)$. Then, we have

$$\begin{aligned} y_t &= \alpha y_{t-1} + u_t \\ &= \alpha y_{t-1} + \sum_{k=1}^p \alpha_k u_{t-k} + \varepsilon_t \end{aligned}$$

under the null of $\alpha = 1$, we have $\Delta y_t = u_t$. Thus,

$$y_t = \alpha y_{t-1} + \sum_{k=1}^p \alpha_k \Delta y_{t-k} + \varepsilon_t$$

To test $\alpha = 1$ in this regression, we need to obtain the limit distribution of the OLS estimator $\hat{\alpha}$ in the regression above. To this end, we introduce the following Corollary.

Corollary 1 (Corollary 4 (EM1 p.47)). Suppose $z_t = (y_t, x_t')' \sim I(1)$ and $v_t \sim I(0)$. We assume existence of $u_t \sim I(0)$ such that

$$y_t = v_t' \alpha + x_t' \beta + u_t \quad (1)$$

The limiting distribution of the OLS estimator $\hat{\beta}$ from the regression (1) can be obtained from

$$n(\hat{\beta} - \beta) = \left(\frac{1}{n^2} \sum_{t=1}^n \tilde{x}_t \tilde{x}_t' \right)^{-1} \left(\frac{1}{n} \sum_{t=1}^n \tilde{x}_t u_t \right)$$

where $\tilde{x}_t$ is the projection residual given by

$$\tilde{x}_t = x_t - \left(\sum_{t=1}^n x_t v_t' \right) \left(\sum_{t=1}^n v_t v_t' \right)^{-1} v_t$$

Suppose

$$\frac{1}{n} \sum_{t=1}^n v_t u_t \rightarrow_p 0$$

Then, we have

$$\begin{aligned} n(\hat{\beta} - \beta) &= \left(\frac{1}{n^2} \sum_{t=1}^n \tilde{x}_t \tilde{x}_t' \right)^{-1} \left(\frac{1}{n} \sum_{t=1}^n \tilde{x}_t u_t \right) \\ &= \left(\frac{1}{n^2} \sum_{t=1}^n x_t x_t' \right)^{-1} \left(\frac{1}{n} \sum_{t=1}^n x_t u_t \right) + o_p(1) \end{aligned}$$

Proof. Given in our lecture slides. □

Using Corollary 4, the limit distribution of the OLS estimator from the regression

$$y_t = \alpha y_{t-1} + \sum_{k=1}^p \alpha_k \Delta y_{t-k} + \varepsilon_t$$

can be obtained from the expression

$$n(\hat{\alpha} - 1) = \left(\frac{1}{n^2} \sum_{t=2}^n y_{t-1}^2 \right)^{-1} \left(\frac{1}{n} \sum_{t=2}^n y_{t-1} \varepsilon_t \right) + o_p(1)$$

under the null of $\alpha = 1$ because we have

$$\frac{1}{n} \sum_{t=1}^n \Delta y_{t-k} \varepsilon_t = \frac{1}{n} \sum_{t=1}^n u_{t-k} \varepsilon_t \rightarrow_p 0$$

for all $k = 1, \dots, p$ due to invertibility of the $u_t \sim AR(p)$. We use the invariance principle in $w_t = \begin{pmatrix} \varepsilon_t \\ \triangle y_t \end{pmatrix} = \begin{pmatrix} \varepsilon_t \\ u_t \end{pmatrix} = \begin{pmatrix} \varepsilon_t \\ \frac{1}{\alpha(L)} \varepsilon_t \end{pmatrix}$ with $\alpha(L) = I - \alpha_1 L - \dots - \alpha_p L^p$. Using Beveridge-Nelson decomposition, we have

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} u_t = \frac{1}{\alpha(1)} \left(\frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \varepsilon_t \right) + o_p(1)$$

so that the limit Brownian motion given by applying IP to w_t is

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} w_t \rightarrow_d \begin{pmatrix} B(r) \\ \frac{1}{\alpha(1)} B(r) \end{pmatrix}$$

with $B(r) = \sigma W(r)$. Thus,

$$\begin{aligned} n(\hat{\alpha} - 1) &= \left(\frac{1}{n^2} \sum_{t=2}^n y_{t-1}^2 \right)^{-1} \left(\frac{1}{n} \sum_{t=2}^n y_{t-1} \varepsilon_t \right) + o_p(1) \\ &\rightarrow_d \left(\int_0^1 \left(\frac{B(r)}{\alpha(1)} \right)^2 dr \right)^{-1} \left(\int_0^1 \left(\frac{B(r)}{\alpha(1)} \right) dB(r) \right) \end{aligned}$$

because $\lambda = \sum_{k=1}^{\infty} \mathbb{E}[\varepsilon_t u_{t-k}] = 0$ due to invertibility of u_t . Thus, we have

$$n(\hat{\alpha} - 1) \rightarrow_d \alpha(1) \left(\int_0^1 W(r)^2 dr \right)^{-1} \left(\int_0^1 W(r) dW(r) \right)$$

This means that as long as we have consistent estimates for $\hat{\alpha}_k, k = 1, \dots$, we have

$$\frac{n(\hat{\alpha} - 1)}{\widehat{\alpha(1)}} \rightarrow_d \left(\int_0^1 W(r)^2 dr \right)^{-1} \left(\int_0^1 W(r) dW(r) \right)$$

Finally, we have

$$t(\hat{\alpha}) = \frac{\hat{\alpha} - 1}{\hat{\sigma} (\sum_{t=2}^n \tilde{y}_{t-1}^2)^{-1/2}} = \frac{\hat{\alpha} - 1}{\hat{\sigma} (\sum_{t=2}^n y_{t-1}^2)^{-1/2}} \sqrt{\frac{\sum_{t=2}^n \tilde{y}_{t-1}^2}{\sum_{t=2}^n y_{t-1}^2}}.$$

Here, we may show

$$\frac{\hat{\alpha} - 1}{\hat{\sigma} (\sum_{t=2}^n y_{t-1}^2)^{-1/2}} = O_p(1), \quad \frac{\sum_{t=2}^n \tilde{y}_{t-1}^2}{\sum_{t=2}^n y_{t-1}^2} = 1 + o_p(1)$$

so that

$$\begin{aligned}
t(\hat{\alpha}) &= \frac{\hat{\alpha} - 1}{\hat{\sigma} \left(\sum_{t=2}^n y_{t-1}^2 \right)^{-1/2}} + o_p(1) \\
&= \hat{\sigma}^{-1} \left(\sum_{t=2}^n \frac{y_{t-1}}{\sqrt{n}} \frac{y_{t-1}}{\sqrt{n}} \frac{1}{n} \right)^{-1/2} \left(\sum_{t=2}^n \frac{y_{t-1}}{\sqrt{n}} \frac{\varepsilon_t}{\sqrt{n}} \right) + o_p(1) \\
&\rightarrow_d \sigma^{-1} \left(\int_0^1 \left(\frac{B(r)}{\alpha(1)} \right)^2 dr \right)^{-1/2} \left(\int_0^1 \left(\frac{B(r)}{\alpha(1)} \right) dB(r) \right) \\
&= \left(\int_0^1 W(r)^2 dr \right)^{-1/2} \left(\int_0^1 W(r) dW(r) \right)
\end{aligned}$$

as long as we use $\hat{\sigma} \rightarrow_p \sigma$.

2.3 Unit root test with intercept and/or time trends

Suppose we want to test whether a time series y_t has a unit root after demeaning and/or accounting for time trend. This can be done by testing $\alpha = 1$ in the regression

$$y_t = \pi' c_t + \alpha y_{t-1} + u_t \quad (2)$$

where c_t may include the intercept term. We may estimate α in (2) in two ways.

1. Jointly regress y_t on (c_t, y_{t-1}) and collect the coefficient in front of y_{t-1} .
2. Regress y_t on c_t in the first step. Obtain the residual

$$\hat{y}_t^0 = y_t - \left(\sum_{t=1}^n y_t c_t' \right) \left(\sum_{t=1}^n c_t c_t' \right)^{-1} c_t$$

Run $AR(1)$ on $\hat{y}_t^0$ using the regression

$$\hat{y}_t^0 = \alpha \hat{y}_{t-1}^0 + e_t$$

It can be shown that the two estimators have the same limit distribution under the null.

3 Test for Cointegrating Relation

Suppose we want to test whether a time series $z_t = (y_t, x_t')'$ has a cointegrating vector. This is equivalent to testing for the existence of β such that

$$e_t = y_t - x_t' \beta$$

is $I(0)$. We may do this in two steps

1. Regress y_t on x_t .

2. Test for unit root on $\hat{e}_t = y_t - x_t' \hat{\beta}$ using the Dickey-Fuller test or the Augmented Dickey-Fuller test. In other words, test for $\alpha = 1$ in the regressions

$$\hat{e}_t = \alpha \hat{e}_{t-1} + u_t$$

or

$$\hat{e}_t = \alpha \hat{e}_{t-1} + \sum_{k=1}^p \alpha_k \Delta \hat{e}_{t-k} + \varepsilon_t$$

Let

$$\frac{1}{\sqrt{n}} \begin{pmatrix} y_{[nr]} \\ x_{[nr]} \end{pmatrix} \rightarrow_d \begin{pmatrix} B_1(r) \\ B_2(r) \end{pmatrix}$$

with $\text{Var}(B(1)) = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \Omega_{22} \end{pmatrix}$. Then, $\begin{pmatrix} B_1 \\ B_2 \end{pmatrix}$ can be written as

$$\begin{pmatrix} B_1 \\ B_2 \end{pmatrix} = \begin{pmatrix} \omega_{11}^{1/2} \sqrt{1-\rho^2} & \omega_{12} \Omega_{22}^{-1/2} \\ 0 & \Omega_{22}^{1/2} \end{pmatrix} \begin{pmatrix} W_1 \\ W_2 \end{pmatrix} \quad (3)$$

where $\rho^2 = \omega_{12} \Omega_{22}^{-1} \omega_{21} / \omega_{11}$ and (W_1, W_2) are two independent standard Brownian motions. We see that

$$\text{Var}(B(1)) = \begin{pmatrix} \omega_{11}^{1/2} \sqrt{1-\rho^2} & \omega_{12} \Omega_{22}^{-1/2} \\ 0 & \Omega_{22}^{1/2} \end{pmatrix} \begin{pmatrix} \omega_{11}^{1/2} \sqrt{1-\rho^2} & 0 \\ \Omega_{22}^{-1/2} \omega'_{12} & \Omega_{22}^{1/2} \end{pmatrix} = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \Omega_{22} \end{pmatrix}$$

as desired. Let

$$\begin{aligned} \hat{e}_t &= y_t - \hat{\beta}' x_t \\ &= y_t - \left(\sum_{t=1}^n y_t x_t' \right) \left(\sum_{t=1}^n x_t x_t' \right)^{-1} x_t \end{aligned}$$

Then, we have

$$\begin{aligned} \frac{1}{\sqrt{n}} \hat{e}_{[nr]} &= \frac{1}{\sqrt{n}} y_{[nr]} - \left(\sum_{t=1}^n \frac{y_t}{\sqrt{n}} \frac{x_t'}{\sqrt{n}} \frac{1}{n} \right) \left(\sum_{t=1}^n \frac{x_t}{\sqrt{n}} \frac{x_t'}{\sqrt{n}} \frac{1}{n} \right)^{-1} \frac{1}{\sqrt{n}} x_{[nr]} \\ &\rightarrow_d B_1(r) - \left(\int_0^1 B_1(r) B_2(r)' dr \right) \left(\int_0^1 B_2(r) B_2(r)' dr \right)^{-1} B_2(r) \end{aligned}$$

Using the relation (3), we may show

$$\begin{aligned} &B_1(r) - \left(\int_0^1 B_1(r) B_2(r)' dr \right) \left(\int_0^1 B_2(r) B_2(r)' dr \right)^{-1} B_2(r) \\ &= \left(\omega_{11}^{1/2} \sqrt{1-\rho^2} \right) \left(W_1(r) - \left(\int_0^1 W_1(r) W_2(r)' dr \right) \left(\int_0^1 W_2(r) W_2(r)' dr \right)^{-1} W_2(r) \right) \\ &:= \left(\omega_{11}^{1/2} \sqrt{1-\rho^2} \right) Q(r) \end{aligned}$$

where $Q(r) := W_1(r) - \left(\int_0^1 W_1(r) W_2(r)' dr \right) \left(\int_0^1 W_2(r) W_2(r)' dr \right)^{-1} W_2(r)$. Suppose we perform a Dickey-Fuller test on $\hat{e}_t$ under the assumption that u_t in the regression

$$\hat{e}_t = \alpha \hat{e}_{t-1} + u_t$$

is given by an orthogonal white noise process. The estimator for α is given by

$$\hat{\alpha} = \left(\sum_{t=2}^n \hat{e}_{t-1}^2 \right)^{-1} \left(\sum_{t=2}^n \hat{e}_{t-1} \hat{e}_t \right)$$

Under the null of $\alpha = 1$, we have

$$\begin{aligned} n(\hat{\alpha} - 1) &= \left(\sum_{t=2}^n \frac{\hat{e}_{t-1}}{\sqrt{n}} \frac{\hat{e}_{t-1}}{\sqrt{n}} \frac{1}{n} \right)^{-1} \left(\sum_{t=2}^n \frac{\hat{e}_{t-1}}{\sqrt{n}} \frac{u_t}{\sqrt{n}} \right) \\ &\rightarrow_d \left(\int_0^1 Q(r)^2 dr \right)^{-1} \left(\int_0^1 Q(r) dQ(r) \right) \end{aligned}$$

The limiting distribution differs from the usual Dickey-Fuller distribution because $Q(r)$ is a residual from projecting Brownian paths rather than a standard Brownian motion. Consequently, the residual-based DF and ADF tests require Engle-Granger cointegration critical values, which depend on the number m of $I(1)$ regressors and the deterministic terms included in the first-stage regression.