Cointegración

El concepto de cointegración se refiere a la relación de largo plazo que debe existir entre los niveles de algunas variables, según lo predice alguna teoría, ya sea económica o de alguna otra índole. Para entender el concepto y su utilidad, conviene considerar en principio el siguiente modelo

$$Z_{1t} = \beta_0 + \beta_1 Z_{2t} + \varepsilon_t \qquad (1)$$

en donde β_0 y β_1 son parámetros y $\{\varepsilon_t\}$ constituye una sucesión de errores aleatorios, mientras que $\{Z_{1t}\}$ y $\{Z_{2t}\}$ son dos sucesiones de variables aleatorias, la primera de las cuales se considera explicada por la segunda.

Los supuestos que permiten la estimación de los parámetros por Mínimos Cuadrados Ordinarios requieren que los errores sean no-correlacionados entre sí, con media cero y varianza constante. Dichos supuestos no se satisfacen necesariamente si las series de tiempo $\{Z_{1t}\}$ y $\{Z_{2t}\}$ son no-estacionarias, como se verá a continuación.

Considérese la siguiente situación, planteada originalmente en el estudio de Granger y Newbold (1974)[1], en la cual se tiene que $\{Z_{1t}\}$ y $\{Z_{2t}\}$ son no-estacionarios, denominados integrados de orden 1, I(1), porque tienen una tendencia subyacente de tipo polinomial de orden 1, o sea

$$Z_{1t} = Z_{1,t-1} + \varepsilon_{1t} \quad \text{y} \quad Z_{2t} = Z_{2,t-1} + \varepsilon_{2t} \qquad (2)$$

con $\{\varepsilon_{1t}\}$ y $\{\varepsilon_{2t}\}$ procesos de ruido blanco (o sea, que no contienen información) independientes, con media cero y varianzas σ_1^2 y σ_2^2 respectivamente, de manera tal que $\{Z_{1t}\}$ y $\{Z_{2t}\}$ son dos caminatas aleatorias independientes entre sí.

En esta situación, la ecuación (1) carece de interpretación y cualquier asociación que pudiera encontrarse entre Z_1 y Z_2 es de carácter espurio, particularmente la asociación lineal. Sin embargo, Granger y Newbold efectuaron un estudio de simulación en el cual generaron muestras de acuerdo con los procesos (2) y para cada muestra estimaron la regresión (1). Como resultado obtuvieron que, al nivel de significación del 5%, se rechazó la hipótesis nula $H_0 : \beta_1 = 0$ en favor de la alternativa $H_A : \beta_1 \neq 0$, en aproximadamente el 75% de los casos.

Adicionalmente, se observó que las regresiones tendían a mostrar un coeficiente de determinación $\left(R^2\right)$ sensiblemente alto, mientras que los residuos mostraban autocorrelación. La explicación teórica de los resultados encontrados por Granger y Newbold fue expuesta por Phillips (1986)[2], quien demostró que los estimadores de MCO para β_0 y β_1 son inconsistentes, y en particular el valor estimado de β_1 se aleja cada vez más de cero (el verdadero valor de β_1) conforme el tamaño de muestra aumenta.

Además, Phillips también demostró que los residuos de la regresión estimada se comportan como un proceso con raíz unitaria, de manera tal que el estimador de MCO de la varianza del error (suma de cuadrados residual entre grados de libertad) diverge conforme el tamaño de muestra aumenta. Este hecho invalida las pruebas t y F, puesto que al aumentar el tamaño de muestra, más probable será que se rechacen las hipótesis nulas correspondientes a dichas pruebas, lo cual conduce a concluir erróneamente que sí hay asociación entre Z_{1t} y Z_{2t}.

---

[1] Granger, C.W.J. y Newbold, P. (1974) *Spurious Regressions in Econometrics*. Journal of Econometrics 2, 111 – 120.
[2] Phillips, P.C.B. (1986) *Understanding Spurious Regressions in Econometrics*. Journal of Econometrics 33, 311-340.

Para poder apreciar mejor las consecuencias de trabajar con un modelo del tipo (1), cuando $\{Z_{1t}\}$ y $\{Z_{2t}\}$ siguen procesos individuales de la forma (2), considérese el comportamiento de los errores

$$\begin{aligned}\varepsilon_t &= Z_{1t} - \beta_0 - \beta_1 Z_{2t} \\ &= Z_{1,t-1} + \varepsilon_{1t} - \beta_0 - \beta_1 Z_{2,t-1} - \beta_1 \varepsilon_{2t} \\ &\ldots \\ &= Z_{1,0} - \beta_0 - \beta_1 Z_{2,0} + \sum_{i=1}^{t} \varepsilon_{1,i} - \beta_1 \sum_{i=1}^{t} \varepsilon_{2,i} \end{aligned} \tag{3}$$

con $Z_{1,0}$ y $Z_{2,0}$ valores fijos, tales que $Z_{1,0} = \beta_0 + \beta_1 Z_{2,0}$. Es claro entonces que

$$E(\varepsilon_t) = 0,\ \mathrm{Var}(\varepsilon_t) = t(\sigma_1^2 + \beta_1^2 \sigma_2^2) \ \text{ y } \ \mathrm{Cov}(\varepsilon_t, \varepsilon_{t+k}) = \mathrm{Var}(\varepsilon_t) \text{ para toda t y } k \geq 1 \tag{4}$$

por lo cual se sigue que la correlación entre ε_t y ε_{t+1} es igual a $\sqrt{t/(t+1)}$, ya que

$$t(\sigma_1^2 + \beta_1^2\sigma_2^2) / \surd t(\sigma_1^2 + \beta_1^2\sigma_2^2)\ (t+1)(\sigma_1^2 + \beta_1^2\sigma_2^2) = t / \surd t(t+1)$$

la cual tiende a 1 conforme t crece. Enders (1995, Cap. 4)[3] considera cuatro casos distintos en relación con el modelo (1), estos son como sigue.

Caso 1. Tanto $\{Z_{1t}\}$ como $\{Z_{2t}\}$ son estacionarios, o sea I(0). Este es el caso que corresponde al modelo de regresión lineal simple **clásico**, en donde no se presenta ninguna anomalía. En lugar de los modelos (2) se suponen procesos del tipo

$$Z_{1t} = \rho_1 Z_{1,t-1} + \varepsilon_{1,t} \ \text{ y } \ Z_{2t} = \rho_2 Z_{2,t-1} + \varepsilon_{2,t} \ \text{ con } |\rho_1|<1 \text{ y } |\rho_2|<1 \tag{5}$$

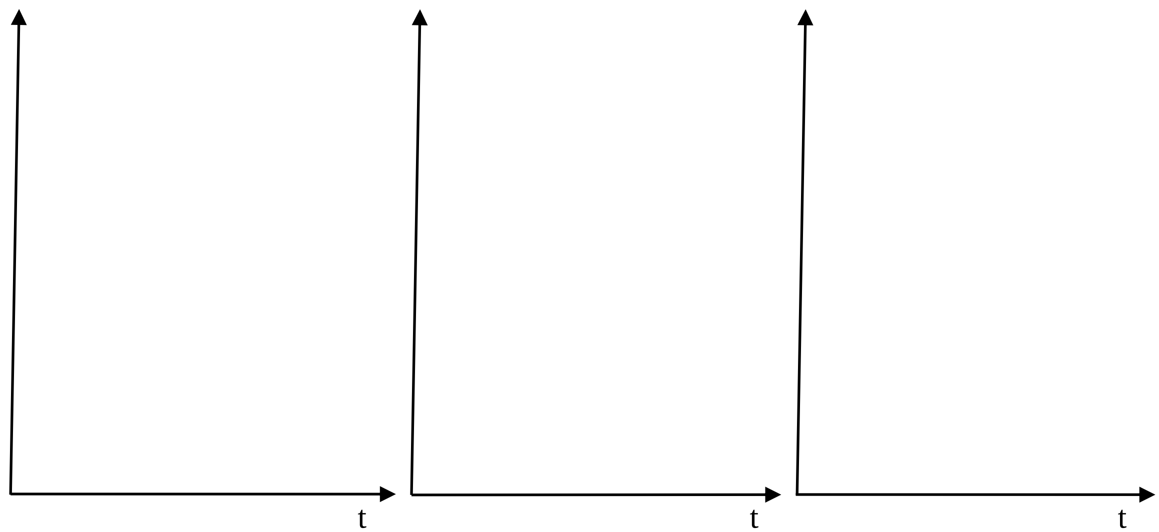
los cuales dan origen al error de la regresión

[3] Enders, W. (1995) *Applied Econometric Time Series*. New York: John Wiley & Sons.

$$\varepsilon_t = Z_{1,0} - \beta_0 - \beta_1 Z_{2,0} + \sum_{i=1}^{t} \rho_1^i \varepsilon_{1,i} - \beta_1 \sum_{i=1}^{t} \rho_2^i \varepsilon_{2,i} \qquad (6)$$

que también resulta ser estacionario.

Gráficas del Caso 1

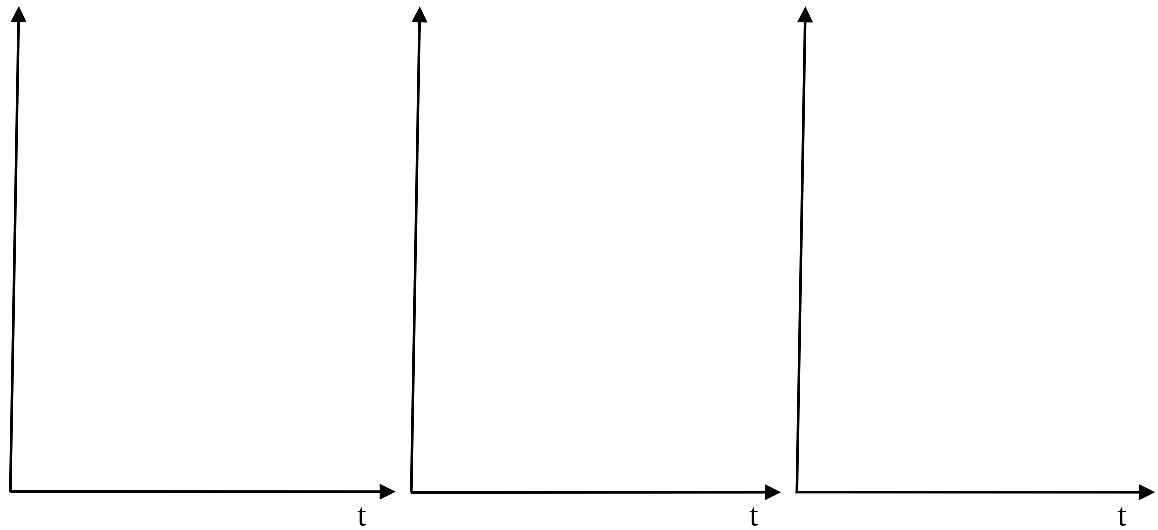


Caso 2. Uno de los dos procesos $\{Z_{1t}\}$ o $\{Z_{2t}\}$ es I(0) y el otro es I(1). En este caso, el modelo de regresión carece de sentido y sus resultados son inválidos. Para ver esto, supóngase que la relación (5) se cumple, pero con $\rho_1 = 1$, de forma tal que (6) se convierte en

$$\varepsilon_t = Z_{1,0} - \beta_0 - \beta_1 Z_{2,0} + \sum_{i=1}^{t} \varepsilon_{1,i} - \beta_1 \sum_{i=1}^{t} \rho_2^i \varepsilon_{2,i} \, . \qquad (7)$$

Aquí se puede apreciar que $\{\varepsilon_t\}$ no es estacionario, puesto que $\sum_{i=1}^{t} \varepsilon_{1,i}$ implica la presencia de una tendencia estocástica.

Gráficas del Caso 2



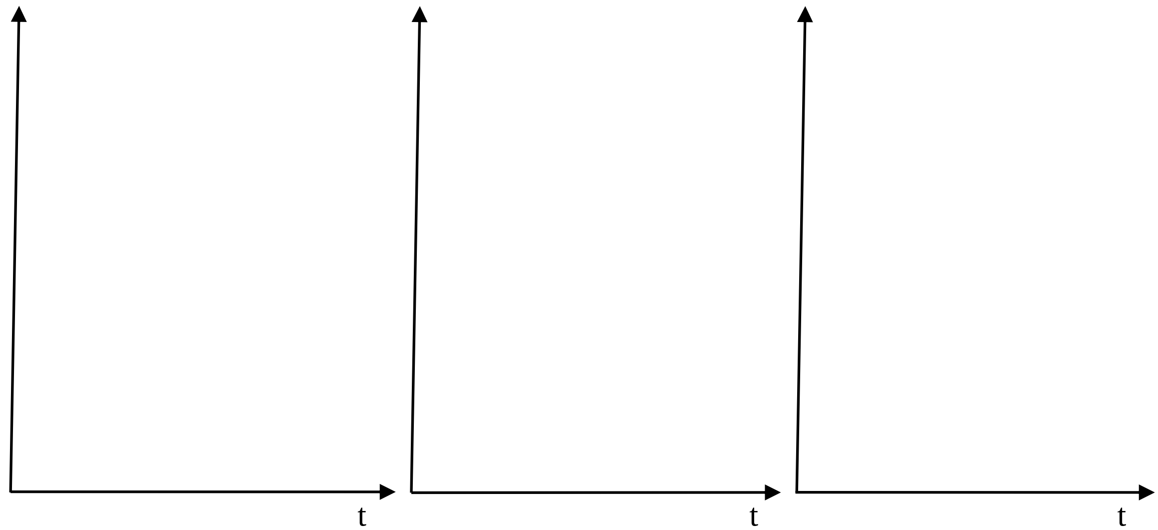
En este caso, una posible solución consiste en transformar la variable Z_1, para trabajar con sus diferencias $Z_{1,t} = Z_{1,t} - Z_{1,t-1}$ y regresar entonces al caso anterior, pero se debe cuidar la interpretación de los resultados que se obtengan.

Caso 3. $\{Z_{1t}\}$ y $\{Z_{2t}\}$ son I(1), al igual que los errores de la regresión $\{\varepsilon_t\}$. A la regresión que surge en este caso se le conoce como **espuria** y sus resultados nuevamente son inválidos, pues los errores asociados a la regresión tienen efectos permanentes, como se puede ver en la expresión (3). La manera más sencilla de corregir este problema consiste en replantear el modelo (1) para escribirlo en diferencias, o sea,

$$\nabla Z_{1t} = \beta_0 + \beta_1 \nabla Z_{2t} + \delta_t \qquad (8)$$

en donde $\{\nabla Z_{1t}\}, \{\nabla Z_{2t}\}$ y $\{\delta_t\}$ son estacionarios y corresponden al Caso 1 visto previamente, pero debe notarse que ahora se pierde la información referente a las series en niveles.

Gráficas del Caso 3



Caso 4. Los dos procesos $\{Z_{1t}\}$ y $\{Z_{2t}\}$ son I(1), mientras que $\{\varepsilon_t\}$ ~I(0). En estas circunstancias se dice que Z_1 y Z_2 están **cointegradas**. Un ejemplo de esta situación ocurre cuando

$$Z_{1t} = \mu_t + \varepsilon_{1t} \quad \text{y} \quad Z_{2t} = \mu_t + \varepsilon_{2t} \quad \text{con} \quad \mu_t = \mu_{t-1} + \varepsilon_{0t} \tag{9}$$

donde $\{\varepsilon_{1t}\}$, $\{\varepsilon_{2t}\}$ y $\{\varepsilon_{0t}\}$ son procesos de ruido blanco. Aquí se tiene que $\{Z_{1t}\}$ y $\{Z_{2t}\}$ son I(1) pero $Z_{1t} - Z_{2t} = \varepsilon_{1t} - \varepsilon_{2t}$ es una nueva variable I(0), obtenida como combinación lineal de Z_1 y Z_2.

Para entender mejor la idea de cointegración, considérese el caso en que, tanto $\{Z_{1t}\}$ como $\{Z_{2t}\}$ son I(1). La relación que liga a estos procesos es la regresión lineal simple

$$Z_{1t} = \beta_0 + \beta_1 Z_{2t} + \varepsilon_t \tag{10}$$

y también se cumple que

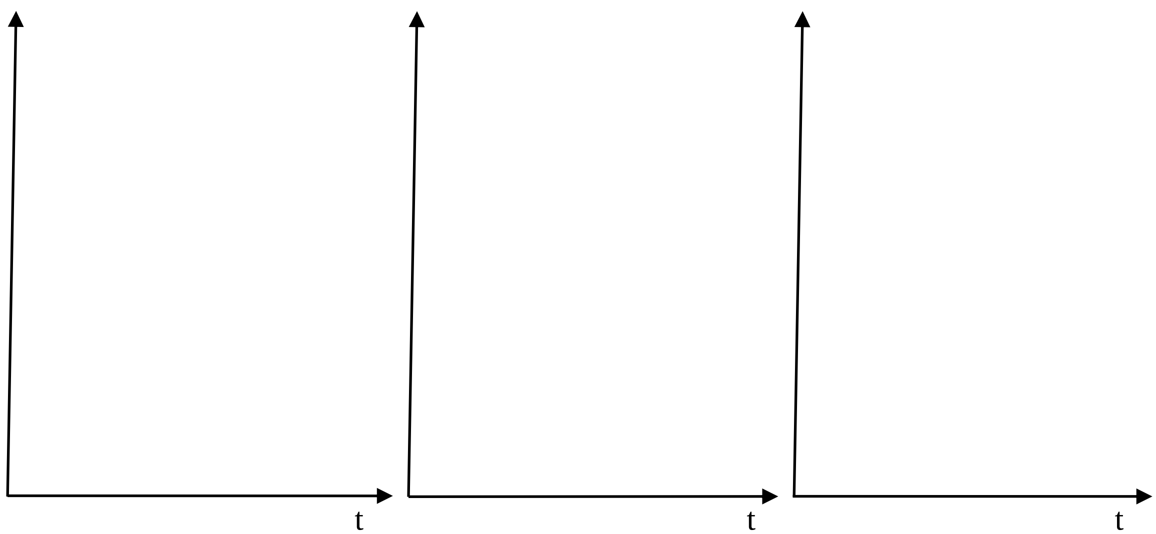
$$Z_{1t} + \alpha Z_{2t} = \delta_t \tag{11}$$

donde ε_t sigue una caminata aleatoria y δ_t se comporta como un proceso AR(1), o sea,

$$\varepsilon_t = \varepsilon_{t-1} + \varepsilon_{1t} \text{ y } \delta_t = \rho\delta_{t-1} + \varepsilon_{2t} \text{ con } |\rho| < 1. \tag{12}$$

Además, se supone que $\{\varepsilon_{1t}\}$ y $\{\varepsilon_{2t}\}$ son ruidos blancos con media cero, varianzas σ_1^2 y σ_2^2, así como covarianza contemporánea dada por $\mathrm{Cov}(\varepsilon_{1t}, \varepsilon_{2t}) = \sigma_{12}$.

Gráficas del Caso 4



Al resolver el sistema (10) – (11), con $\alpha \neq -\beta_1$, se obtiene

$$Z_{1t} = (\beta_1 + \alpha)^{-1}\beta_0\alpha + \alpha(\beta_1 + \alpha)^{-1}\varepsilon_t + \beta_1(\beta_1 + \alpha)^{-1}\delta_t$$

$$Z_{2t} = -(\beta_1 + \alpha)^{-1}\beta_0 - (\beta_1 + \alpha)^{-1}\varepsilon_t + (\beta_1 + \alpha)^{-1}\delta_t \tag{13}$$

de tal manera que la no-estacionariedad de $\{Z_{1t}\}$ y $\{Z_{2t}\}$ se debe a que estos procesos dependen linealmente de $\{\varepsilon_t\}$, que sigue una caminata aleatoria.

Por otro lado, $\{Z_{1t} + \alpha Z_{2t}\}$ es I(0) puesto que $\{\delta_t\}$ lo es, por lo cual $Z_{1t} + \alpha Z_{2t} = 0$ es una relación de equilibrio (hacia el cual tienden los procesos $\{Z_{1t}\}$ y $\{Z_{2t}\}$ en el largo plazo). Dicha relación surge al aplicar el vector, denominado de cointegración, $(1, \alpha)'$ al proceso bivariado $\{\mathbf{Z}_t\}$ formado por $\mathbf{Z}_t = (Z_{1t}, Z_{2t})'$. Desde luego, si el coeficiente ρ de la expresión (12) es igual a uno, lo que se tiene es una relación espuria entre Z_1 y Z_2.