Clase 3 – Curso series de tiempo en R - CEPAL

Cointegración y modelo de corrección de errores

NACIONES UNIDAS

CEPAL

Contenidos de la clase

- **Estacionariedad**
- **Cointegración**
- **Modelo de corrección de errores**
- **Ejercitación en R con datos de comercio exterior y de empleo**

¿Por qué interesa saber si una serie es estacionaria?

Si se quiere entender la relación entre dos o más variables que utilizan el análisis de regresión **se necesita dar por sentada algún tipo de estabilidad en el tiempo**.

Si se permite que la relación entre dos variables cambie de forma arbitraria en cada período, no se puede esperar aprender mucho acerca de cómo un cambio en una variable afecta a la otra si sólo se tiene acceso a una sola realización de serie de tiempo.

Al establecer un modelo de regresión múltiple para los datos de series de tiempo, **se está suponiendo cierta forma de estacionariedad en la que los coeficientes no cambian en el tiempo**.

Concepto de estacionariedad

Un proceso estocástico que tiene media y varianza finita es estacionario en covarianza si para cada t y $k \geq 1$.

1. La **media** es constante

$$E(y) = \mu$$

2. La **varianza** es constante

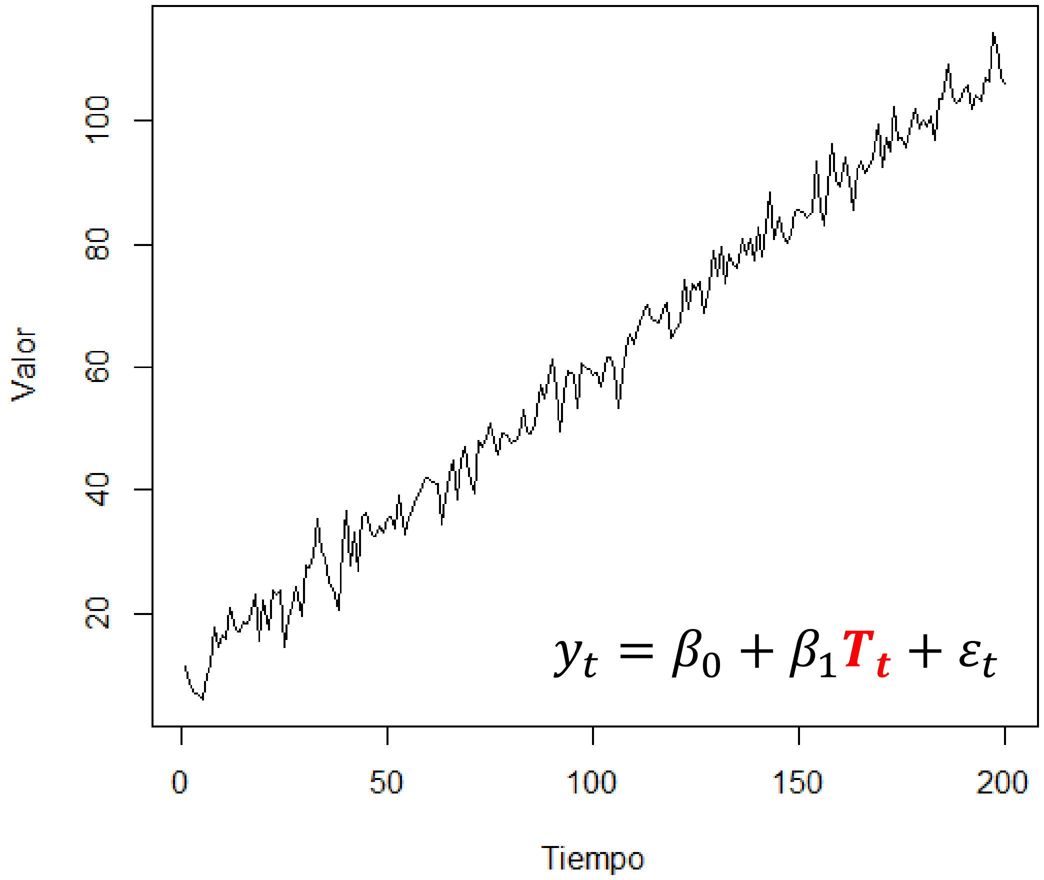
$$Var(y_t) = \sigma_y^2$$

3. La **covarianza** entre y_t y y_{t+k} depende sólo de la distancia entre los dos términos (k), y no de la ubicación del periodo inicial (t)

$$Cov(y_t, y_{t+k}) = \gamma_s$$

¿Cuáles son los motivos por los que una serie puede ser no estacionaria?

Serie de tiempo con tendencia determinística



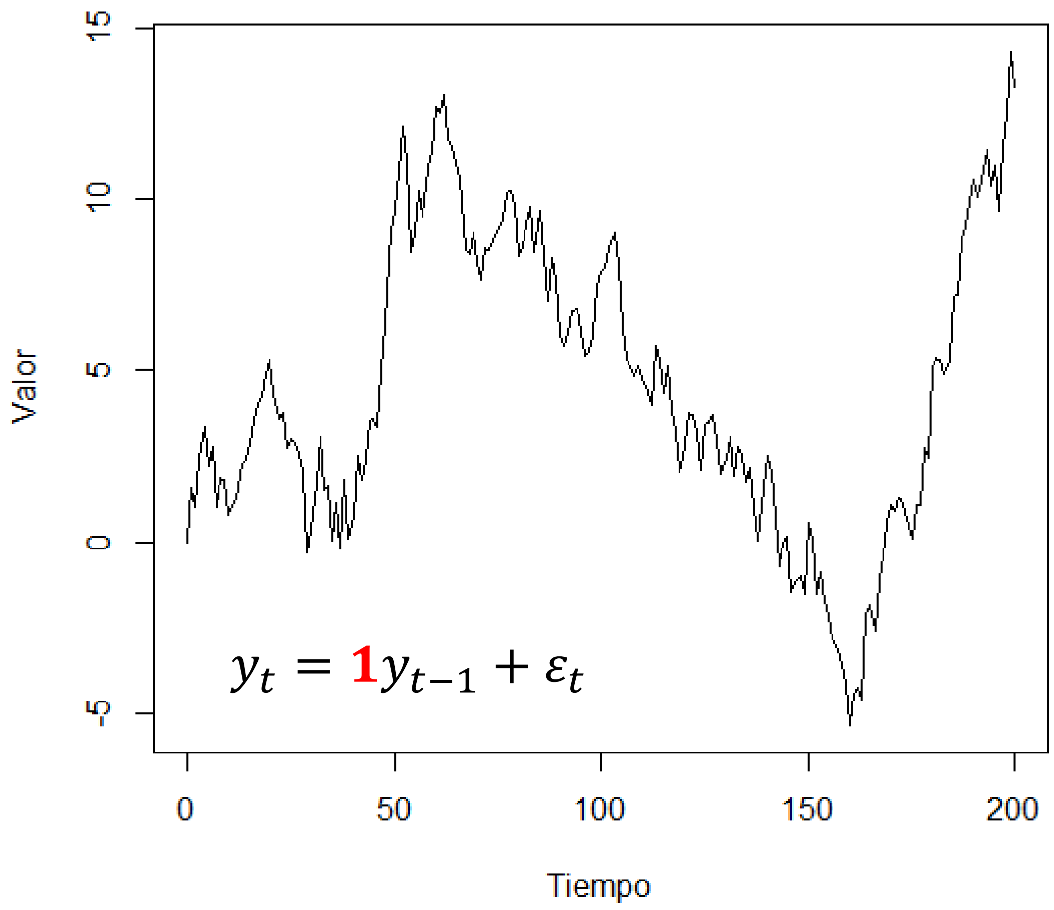
1. Tendencia determinística

El proceso generador de la serie tiene un componente que es una función no aleatoria del tiempo.

Un ejemplo de serie con una tendencia determinística muy clara es la población.

¿Cuáles son los motivos por los que una serie puede ser no estacionaria?

Serie de tiempo con tendencia estocástica



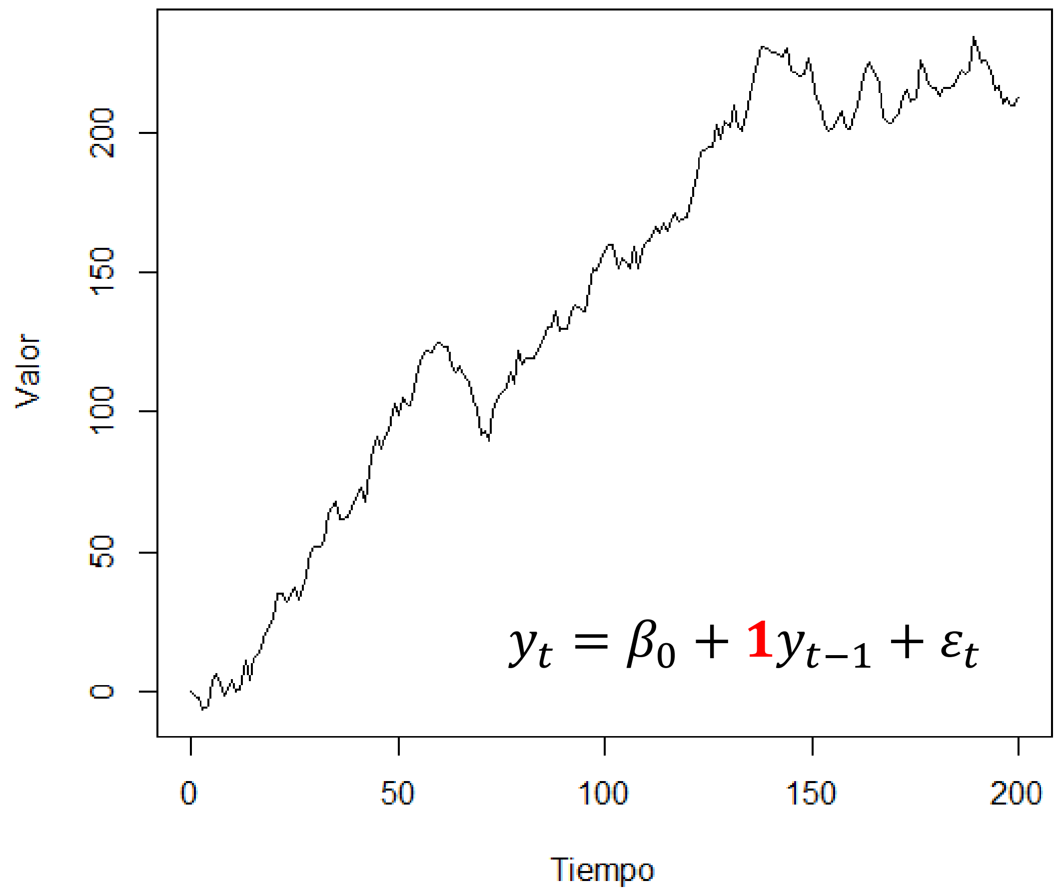
2. Tendencia estocástica

El proceso generador de la serie tiene un componente que es una función aleatoria del tiempo.

El modelo más conocido que genera una serie con tendencia estocástica es el “**random walk**” **(o paseo aleatorio)**. Este es un modelo autorregresivo de orden 1 (AR-1), cuyo coeficiente es igual a 1. El hecho de que el coeficiente del componente autorregresivo sea igual a 1 hace que sea una serie con elevada persistencia.

¿Cuáles son los motivos por los que una serie puede ser no estacionaria?

Serie de tiempo con tendencia estocástica



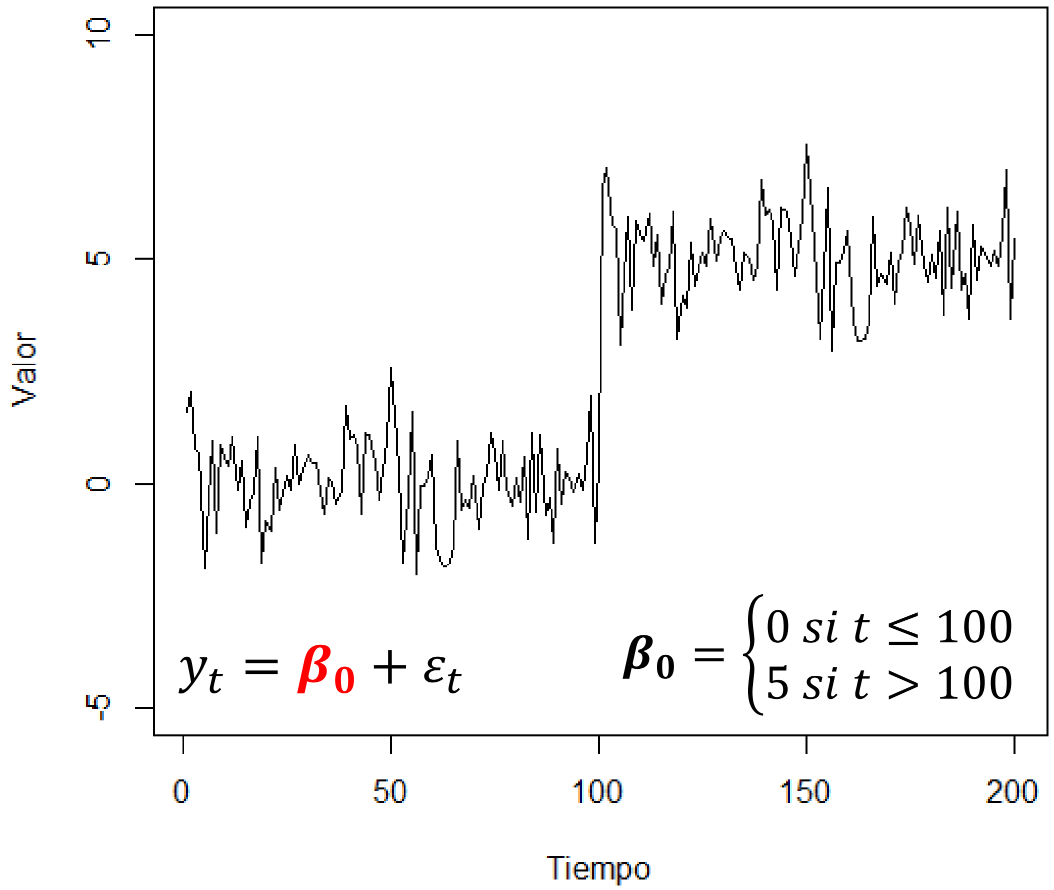
2. Tendencia estocástica

Otro ejemplo de serie con tendencia estocástica es la generada por un modelo "**random walk with drift**" **(paseo aleatorio con deriva)**.

A diferencia del modelo anterior, este tiene un valor de origen distinto a cero, y dada la persistencia que le da el componente autorregresivo, esto determina que la serie tienda a crecer si la deriva es positiva, y a bajar si la deriva es negativa. Por este motivo a veces puede confundirse con una tendencia determinística.

¿Cuáles son los motivos por los que una serie puede ser no estacionaria?

Serie de tiempo con quiebre estructural



3. Quiebre estructural

Cuando hay cambios discretos (en un momento del tiempo) o graduales (a lo largo de un período) en los parámetros poblacionales del proceso generador.

En series macroeconómicas, estos cambios suelen estar asociados a cambios en la política económica o en la estructura de la economía.

El quiebre puede significar un cambio en el promedio de la variable dependiente (reflejado en la constante) o en la relación entre las variables (reflejado en la pendiente de la relación entre ellas).

¿Cómo saber si una serie no tiene tendencia estocástica?

1. Función de autocorrelación (o correlograma)

Función de Autocorrelación

$$r_k = \frac{\sum_{t=k+1}^{T}(y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{T}(y_t - \bar{y})^2}$$

Covarianza entre $\boldsymbol{y_t}$ y $\boldsymbol{y_{t-k}}$

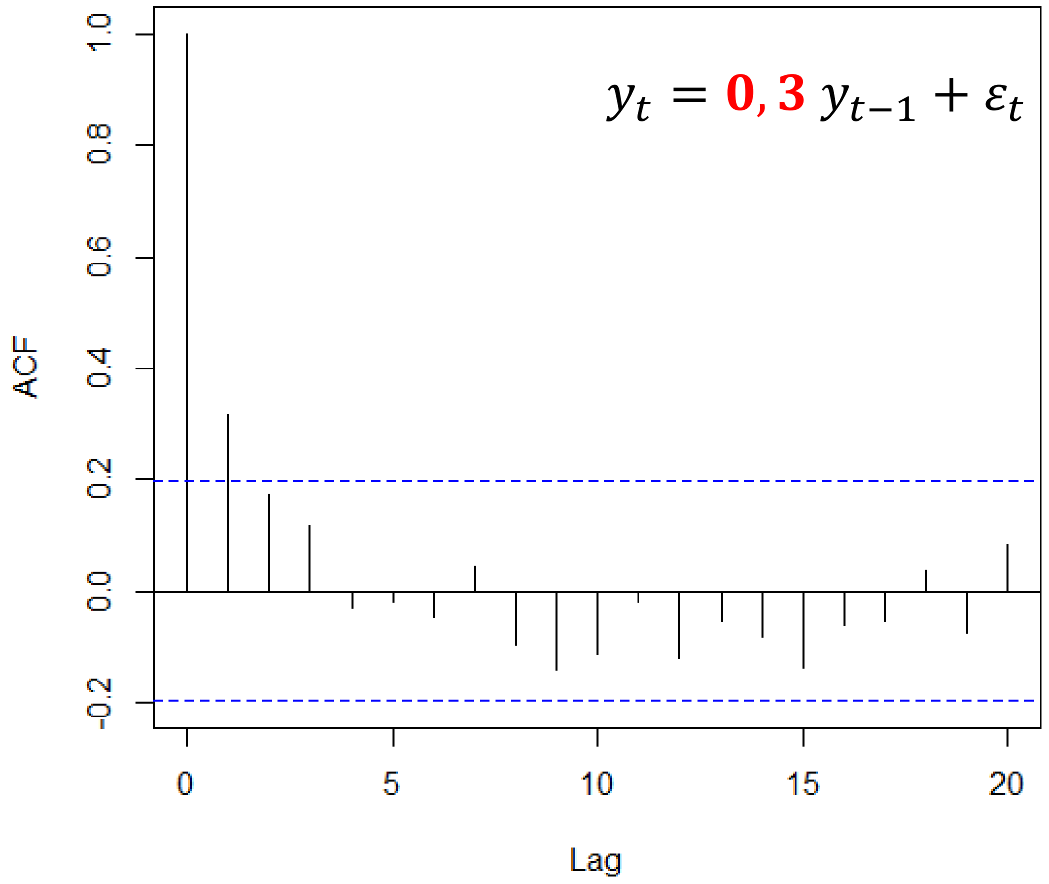
Varianza de $\boldsymbol{y_t}$

<u>Si la serie es estacionaria, la función de autocorrelación tiene que caer geométricamente</u>. Si el parámetro autorregresivo es positivo la caída es directa, y si es negativo realizará un movimiento oscilante hasta converger.

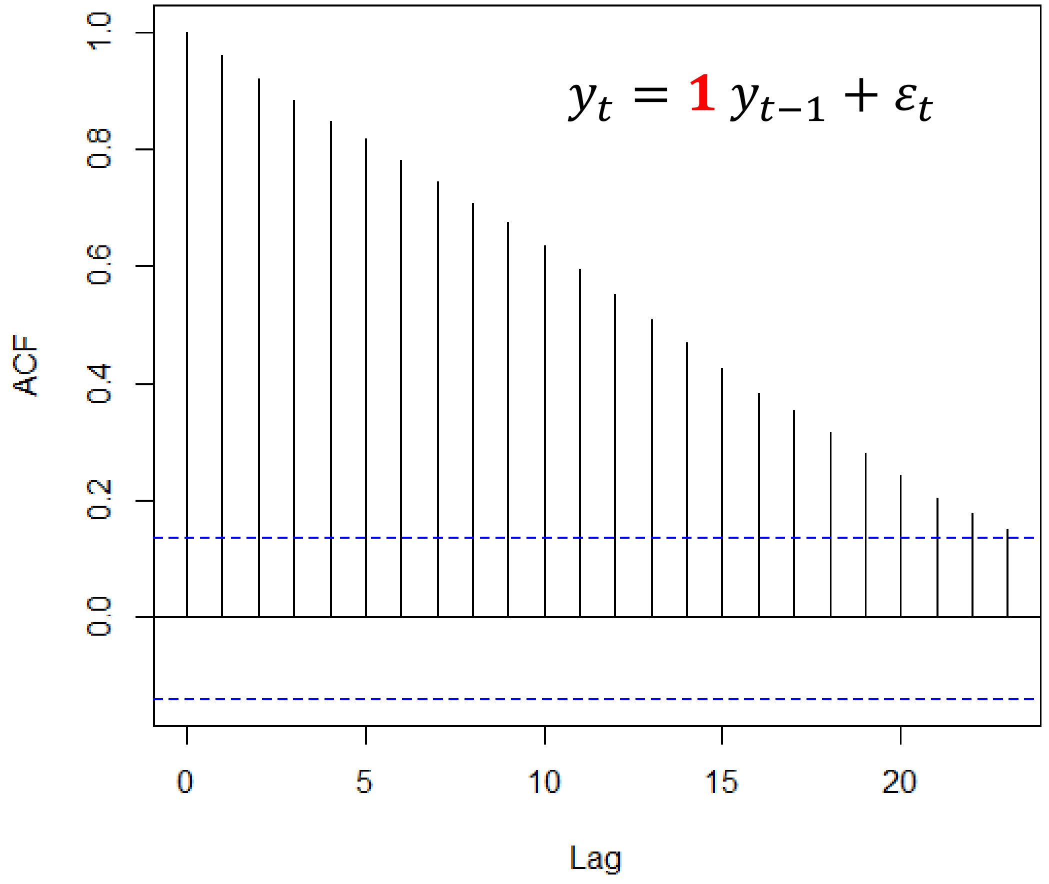
Concepto de dependencia débil. Una serie de tiempo estacionaria en covarianza es débilmente dependiente si la correlación entre y_t y y_{t+k} se vuelve cero con la "suficiente rapidez" cuando k tiende a infinito.

¿Cómo saber si una serie no tiene tendencia estocástica?

Correlograma: Ejemplo de AR(1) estacionario



Correlograma: Ejemplo de AR(1) no estacionario



¿Cómo saber si una serie es estacionaria?

2. Test de Dickey-Fuller

Un test formal para una tendencia estocástica fue propuesto por Dickey y Fuller (1979). Este test evalúa si el coeficiente autorregresivo es igual a 1.

Test sobre AR(1)

$$y_t = \beta_0 + \beta_1 y_{t-1} + e_t$$

Hipótesis nula: tiene raíz unitaria:

$$H_0\colon\ \beta_1 = 1$$

Hipótesis alternativa: no tiene raíz unitaria

$$H_A\colon\ |\beta_1| < 1$$

Reescritura más común del test

$$\Delta y_t = \beta_0 + \delta y_{t-1} + e_t$$

Donde: $\delta = \beta_1$-*1*

Hipótesis nula: tiene raíz unitaria

$$H_0\colon \delta = 0$$

Hipótesis alternativa: no tiene raíz unitaria

$$H_A\colon\ |\delta| < 0$$

¿Cómo saber si una serie es estacionaria?

3. Test de Dickey-Fuller aumentado

En el test de Dickey-Fuller aumentado se controla por los rezagos de Δy_t ante la posible presencia de autocorrelación serial. Según Wooldridge (2012), la cantidad de los rezagos (p) está determinada por la frecuencia de los datos. Para datos anuales se suelen considerar uno o dos rezagos, para datos trimestrales cuatro, y para mensuales 12 rezagos. No obstante, no se tienen que seguir reglas estrictas para la determinación de la cantidad de rezagos, y los programas en general incorporan algún método para su estimación automática.

$$\Delta y_t = \beta_0 + \delta y_{t-1} + \sum_{h=1}^{p} \phi_h \Delta y_{t-h} + e_t$$

Donde:
$\delta = \beta_1$-1
$p = cantidad\ de\ rezagos$

Hipótesis nula: tiene raíz unitaria

$$H_0: \delta = 0$$

Hipótesis alternativa: no tiene raíz unitaria

$$H_A: |\delta| < 0$$

¿Cómo saber si una serie es estacionaria?

Valores críticos asintóticos para prueba t de raíz unitaria: <u>sin tendencia de tiempo</u>

Nivel de significatividad	**1%**	**2.50%**	**5%**	**10%**
Valor crítico	-3.43	-3.12	-2.86	-2.57

¿Qué hacer cuando una serie es no estacionaria?

Curso de acción 1: si es integrada de orden 1 expresar la serie en diferencias (su variación)

Niveles logarítmicos

$$y_t = \beta_0 + \beta_1 x_t + e_t$$

Diferencia logarítmica

$$\Delta y_t = \beta_0 + \beta_1 \Delta x_t + e_t$$

Si la serie es integrada de orden 1 significa que la serie expresada en primera diferencia es estacionaria, por lo que una estrategia sencilla de implementar en ese caso es aplicar el operador diferencia a la variable explicada (o a ambas si fuera el caso de que las dos son integradas de orden 1).

¿Qué hacer cuando una serie es no estacionaria?

Curso de acción 2: controlar en la regresión por el componente que hace que sea no estacionaria

Con tendencia determinística

$$y_t = \beta_0 + \beta_1 x_t + \boldsymbol{\beta_2 t} + e_t$$

$\boldsymbol{t}$: la tendencia es una función no estocástica del tiempo.

Con quiebre estructural en la media

$$y_t = \beta_0 + \beta_1 x_t + \boldsymbol{\beta_2 \gamma} + e_t$$

$\boldsymbol{\gamma}$: vale 0 para un período, y 1 para el otro. Otro tipo de quiebre estructural es cuando lo que cambia es β_1.

¿Qué hacer cuando una serie es no estacionaria?

Curso de acción 3: evaluar si la variable explicativa y explicada comparten una misma tendencia, es decir, si están cointegradas. En ese caso podemos implementar la regresión de una variable sobre la otra sin aplicar ninguna transformación.

$$y_t = \beta_0 + \beta_1 x_t + e_t$$

$$e_t = y_t - \beta_0 - \beta_1 x_t +$$

Donde y_t y x_t son $I(1)$ y e_t es $I(0)$

Cuando x_t y y_t son integradas de orden 1 y hay una combinación lineal entre ellas tal que el residuo de esta combinación es integrado de orden 0 ($e_t = y_t - \beta_1 x_t$), x_t y y_t están cointegradas.

Esto significa que ambas series comparten la misma tendencia estocástica, y por tanto incluir a x_t en la regresión ya permite controlar por esta tendencia.

Observación

Cuando la variable explicada es integrada de orden 1 y aplicamos algunas de las estrategias anteriores para controlar por esta tendencia en la regresión, el ajuste medido por el R2 suele dar muy elevado.

Esto no necesariamente significa que elegimos a los mejores predictores, sino que refleja que una parte importante de la variabilidad se encuentra explicada por la tendencia que, explicita o implícitamente estamos incorporando entre las variables explicativas.

Una estrategia alternativa recomendada para evitar una lectura incorrecta del ajuste del modelo es, por ejemplo, restar la tendencia en el caso de las series que presentan tendencia determinística (ver Wooldridge, 2015).

¿Cómo saber si las series están cointegradas?

Realizamos el test de **Engle y Granger**. Este test consiste en evaluar si el residuo de una combinación lineal entre las dos variables (conocida con anterioridad o no) es estacionario.

Cuando no conocemos la forma de la relación lineal entre las variables se puede realizar en dos pasos:

1. Corremos la regresión de una variable sobre la otra y guardamos el residuo.
2. Realizamos un test de raíz unitaria sobre el residuo obtenido en el paso anterior. Si se rechaza la hipótesis nula de que ese residuo es integrado de orden 1, entonces tenemos evidencia a favor de que las series están cointegradas.

Relación entre dos variables integradas de orden 1 y cointegradas

- **Relación de largo plazo.** Es la que surge de la regresión de las dos variables con sus respectivas tendencias. Esto es, cuando corremos la regresión con las series en su nivel original.

$$y_t = \beta_0 + \beta_1 x_t + e_t$$

Este tipo de relación nos interesa cuando queremos obtener parámetros estructurales o hacer proyecciones de mediano-largo plazo

Relación entre dos variables integradas de orden 1 y cointegradas

- **Relación de corto plazo.** Es la que surge de la regresión de las dos variables habiendo extraído sus respectivas tendencias a partir del operador diferencia. Esta relación capta el efecto instantáneo (y no total) de una variable sobre la otra.

$$\Delta y_t = \beta_0 + \beta_1 \Delta x_t + e_t$$

Sin embargo, esta forma de expresar la relación de corto plazo no es correcta porque no incorpora la relación de largo plazo entre las variables: la variación en cada período de y_t depende no solo de cambios contemporáneos de x_t, sino también de cambios en períodos anteriores (x_{t-j}) que no alcanzaron a impactar de forma plena en y_t.

Modelo de corrección de errores

El problema presentado en la filmina anterior es lo que da lugar a la necesidad de incorporar un término corrector al equilibrio de largo plazo en la ecuación de la relación de corto plazo. Esto es lo que comúnmente se conoce como **modelo de corrección de errores**.

$$y_t = \beta_0 + \beta_1 x_t + e_t$$

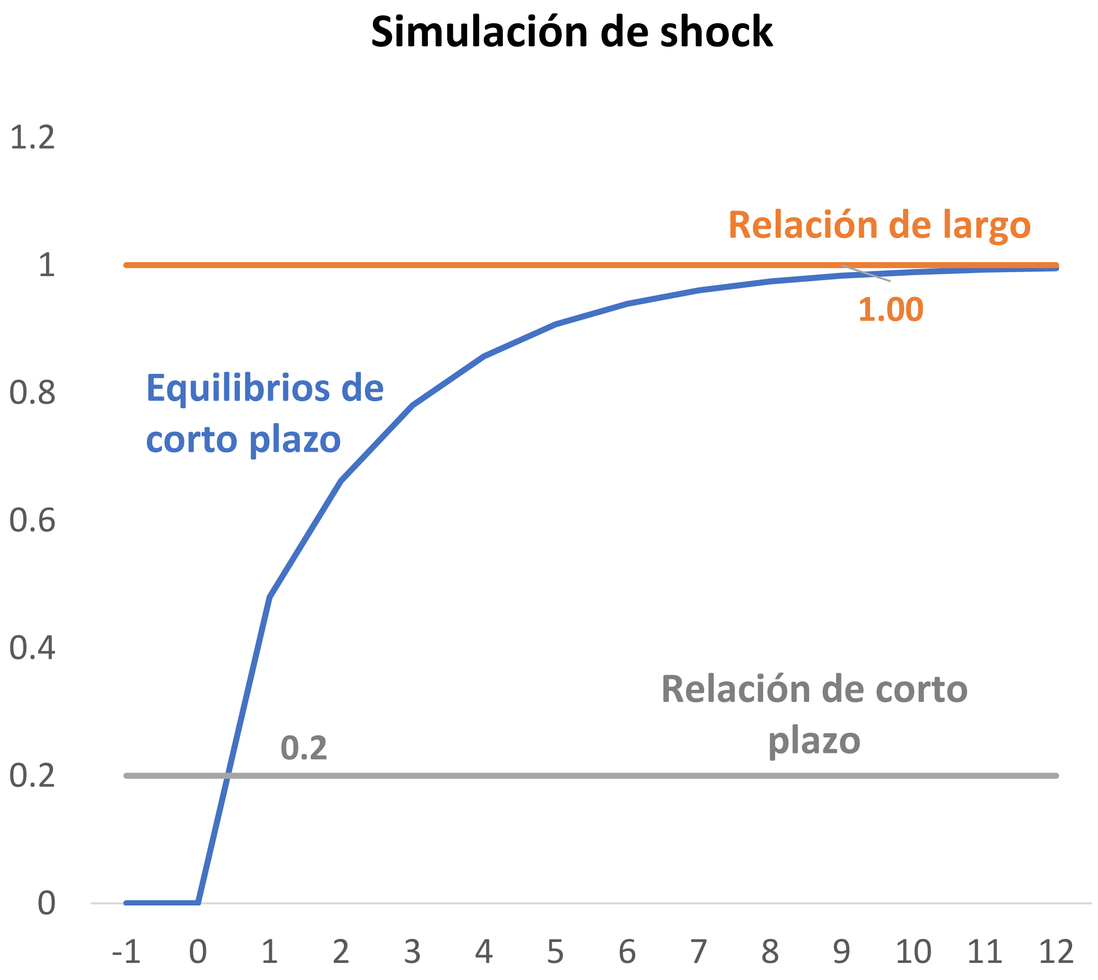
$$e_t = y_t - \hat{y}_t$$

$$\Delta y_t = \beta_0 + \beta_1 \Delta x_t \boxed{- \boldsymbol{\beta_2} e_{t-1}}$$

En cada período, y_t se mueve en función de cambios en x_t, pero también en función de desalineamientos de su equilibrio de largo plazo del período anterior.

$\boldsymbol{\beta_2}$ representa la velocidad a la cual ajusta a esa relación de largo plazo.

Modelo de corrección de errores



El coeficiente de largo plazo suele ser mayor en términos absolutos al de corto plazo.

La velocidad a la cual se alcanza el efecto pleno de una variable sobre la otra está dado por el coeficiente β_2. Este coeficiente siempre tiene que ser negativo.

$$\Delta y_t = \beta_0 + \beta_1 \Delta x_t - \beta_2 e_{t-1}$$

En resumen…

- Hay tres factores que pueden explicar que una serie sea no estacionaria: (i) **tendencia determinística**, (ii) **tendencia estocástica**, y (iii) **quiebre estructural**. Para saber si una serie tiene tendencia estocástica podemos realizar el test de Dickey-Fuller.

- **Cursos de acción frente a series no estacionarias**:

    - Si tiene **tendencia estocástica y es integrada de orden 1** la podemos expresar en primera diferencia.

    - Si tiene **tendencia determinística** incorporamos a esta como control en la regresión.

    - Si tiene **quiebre estructural** incorporamos una *dummy* en la regresión que valga cero para el período previo al cambio y uno para el posterior. Esta *dummy* puede incluirse como una constante o como un término de interacción, dependiendo el caso.

    - Si las **series están cointegradas** (comparten tendencia) podemos realizar la regresión de una sobre la otra sin ningún tipo de transformación.

En resumen…

- Para saber si dos series están cointegradas podemos realizar el **test de Engle y Granger**, que consiste en dos pasos:
    - Corremos la regresión de una variable sobre la otra y guardamos el residuo.
    - Realizamos un test de raíz unitaria sobre ese residuo. Si es estacionario, las series están cointegradas.
- Si nos interesa la relación de corto plazo de series cointegradas debemos incorporar un ajuste corrector al equilibrio. Esto es lo que conocemos como **modelo de corrección de errores**.

Bibliografía

- **Enders**, W. (2008). Applied econometric time series. John Wiley & Sons.
- **Hyndman**, R.J., & **Athanasopoulos**, G. (2018) Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2.
- **Wooldridge**, J. M. (2015). Introductory econometrics: A modern approach. Cengage learning.