Cointegración y corrección de errores

Econometría IV

Rodrigo Chang

Metodología de Engle y Granger

Introducción

¿Qué pasa si hacemos una regresión con dos variables que **no son estacionarias**?

1. **Riesgo:** Regresión Espuria (R^2 alto, t-stat significativos, pero sin sentido económico).
2. **Sin embargo,** si una combinación lineal de ellas es estacionaria, están **Cointegradas**.
3. **Solución:** La metodología de dos pasos de Engle y Granger (1987).

El Proceso resumido

La metodología consiste en dos etapas distintas:

- **Paso 1 (Largo Plazo):** Estimar la regresión estática y probar si los residuos son estacionarios.
- **Paso 2 (Corto Plazo):** Si hay cointegración, estimar el Modelo de Corrección de Errores (ECM) utilizando los residuos del paso 1.

Paso 1: La Ecuación de Largo Plazo

Supongamos dos variables I(1), x_t y y_t. Estimamos por Mínimos Cuadrados Ordinarios (MCO):

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

Guardamos los residuos estimados:

$$\hat{u}_t = y_t - \hat{\beta}_0 - \hat{\beta}_1 x_t$$

> **Nota:** En esta etapa, ignoramos los estadísticos t y el R^2 estándar, ya que la distribución asintótica no es normal.

Prueba de Cointegración (Test sobre Residuos)

Aplicamos la prueba Dickey-Fuller Aumentada (ADF) sobre $\hat{u}_t$.

Importante: Como $\hat{u}_t$ es un residuo estimado (tiene media cero), la prueba ADF se hace **sin constante y sin tendencia**.

$$\Delta\hat{u}_t = \gamma\hat{u}_{t-1} + \sum \delta_i \Delta\hat{u}_{t-i} + e_t$$

- $H_0: \gamma = 0$ (No estacionario / No cointegración)
- $H_1: \gamma < 0$ (Estacionario / Cointegración)

El Modelo de Corrección de Errores (ECM)

- Si encontramos cointegración, el Teorema de Representación de Granger garantiza que existe un mecanismo de corrección de errores.
- Estimamos la dinámica de corto plazo incluyendo el residuo rezagado ($\hat{u}_{t-1}$):

$$\Delta y_t = \alpha_1 \Delta x_t + \alpha \cdot \hat{u}_{t-1} + \epsilon_t$$

Donde α (coeficiente de ajuste) debe ser negativo y significativo.

Mecanismo de corrección de errores

$$\Delta y_t = \alpha_1 \Delta x_t + \alpha \left(y_{t-1} - \hat{\beta}_0 - \hat{\beta}_1 x_{t-1} \right) + \epsilon_t$$

- El término $\left(y_{t-1} - \hat{\beta}_0 - \hat{\beta}_1 x_{t-1} \right)$ representa el error o distancia al equilibrio ayer.
- Si el error fue positivo, y_{t-1} estaba muy alto y $\alpha < 0$, entonces $\Delta y_t < 0$, es decir, el modelo fuerza a y a bajar.

Prueba de Johansen

Motivación: ¿Por qué Johansen?

La metodología de Engle-Granger (dos pasos) es útil, pero tiene limitaciones importantes cuando trabajamos con más de dos variables:

Modelos de Corrección de Errores Vectoriales

El procedimiento de Johansen se basa en estimar un **Modelo de Corrección de Errores Vectorial (VECM)**:

El punto de partida es reparametrizar un modelo VAR en niveles como un VECM. Si partimos de un VAR de orden p:

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + u_t$$

Este se transforma en:

$$\Delta y_t = \Pi y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + u_t$$

- ¿A qué es igual la matriz Π?

$$\Pi = A_1 - I_K$$

Rango de Π

El núcleo del procedimiento de Johansen es analizar el rango (r) de la matriz Π. Existen tres casos posibles:

- $r = 0$: No existe co-integración. Las variables no comparten tendencias comunes y el modelo debe estimarse en primeras diferencias (VAR en diferencias).
- $r = K$ (rango completo): El proceso es estacionario en niveles (I(0)), por lo que no hay raíces unitarias y se puede usar un VAR estándar en niveles.

Rango de Π

- $0 < r < K$: Existen r relaciones de co-integración. En este caso, la matriz Π se descompone en el producto de dos matrices de rango r:

$$\Pi = \alpha\beta'$$

Test de Johansen en R

```r
# Estimación del VAR en niveles
var_levels <- VAR(datos_vec, type = "const", lag.max = 8, ic = "AIC")
# Orden sugerido por AIC
var_levels$p

# K = número de rezagos (basado en VARselect sobre niveles - 1)
johansen_test <- ca.jo(datos_vec,
    type = "trace",     # Usualmente "trace" es el estándar
    ecdet = "const",    # Constante en la relación de cointegración
    K = var_levels$p,   # Rezagos (del VAR en niveles, = 2)
    spec = "longrun")   # Especificación de largo plazo
```

Modelo VECM estimado

Se estima el modelo:

$$
\begin{aligned}
y_t &= A_1 y_{t-1} + A_2 y_{t-2} + u_t \\
y_t - y_{t-1} &= \left(A_1 - I_2\right) y_{t-1} + A_2 y_{t-2} + u_t \\
\Delta y_t &= \left(A_1 + A_2 - I_2\right) y_{t-1} - A_2 y_{t-1} + A_2 y_{t-2} + u_t \\
\Delta y_t &= -\left(I_2 - A_1 - A_2\right) y_{t-1} - A_2 \Delta y_{t-1} + u_t
\end{aligned}
$$

Finalmente,

$$
\Delta y_t = \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + u_t \, ,
$$

donde $\Pi = -\left(I_2 - A_1 - A_2\right)$ y $\Gamma_1 = -A_2$.

Resultados del Test (Salida de R)

Analizaremos la siguiente salida de `urca::ca.jo`:

```
Test type: trace statistic

Values of teststatistic and critical values of test:

          test 10pct  5pct  1pct
r <= 2 |  3.31  7.52  9.24 12.97
r <= 1 | 18.12 17.85 19.96 24.60
r = 0  | 56.79 32.00 34.91 41.07

Eigenvectors (salida de la primera columna):
y.l2: 1.000, z.l2: 1.035, w.l2: -1.016, const: 0.013

Weights W (salida de la primera columna):
y.d: -0.546, z.d: -0.165, w.d: -0.218
```

El Test de la Traza (1/2)

- Observamos la sección `Values of teststatistic`.
- La lógica del Test de la Traza es secuencial, empezando desde abajo ($r = 0$) hacia arriba.
- Hipótesis 1 ($H_0 : r = 0$): ¿Existen cero relaciones?
  - Estadístico: 56.79. Valor Crítico (5%): 34.91
  - Decisión: Como $56.79 > 34.91$, RECHAZAMOS la nula. Hay al menos una relación.

El Test de la Traza (2/2)

- Hipótesis 2 ($H_0 : r \leq 1$): ¿Existen como máximo 1 relación?
  - Estadístico: 18.12. Valor Crítico (5%): 19.96
  - Decisión: Como $18.12 < 19.96$, NO RECHAZAMOS la nula.
  - Resultado: Nos quedamos con $r = 1$. Por lo tanto, para el resto del análisis, solo nos importa la primera columna de las matrices siguientes.

La Ecuación de Largo Plazo (β)

Velocidad de Ajuste (α)

- Usamos la primera columna de `Weights W`.
- Estos valores indican qué tan rápido reacciona cada variable cuando se pierde el equilibrio.

$$\Delta y_t = \alpha\beta' y_{t-1} + \Gamma_1 \Delta y_{t-1} + u_t$$

$$\Delta y_t = -0.546(Error_{t-1}) + \dots$$

  - Signo Negativo: Indica estabilidad. Si y está por encima del equilibrio, debe bajar.
  - Magnitud (0.546): El sistema corrige el 54.6% del desequilibrio en un solo periodo. Es un ajuste rápido.

Resumen

- El test de la traza confirmó que las tres variables (y, z, w) se mueven juntas a largo plazo ($r = 1$).
- Identificamos que w impulsa a y positivamente, mientras que z lo hace negativamente.
- El sistema es estable, con y siendo la variable que reacciona más agresivamente (54%) para corregir cualquier desviación del equilibrio.

Ejercicio de interpretación

- ¿Cuántos vectores de cointegración sugiere la prueba de Johansen?

```
######################
# Johansen-Procedure #
######################

Values of teststatistic and critical values of test:

          test    10pct   5pct    1pct
r <= 2 |  4.12    7.52    9.24   12.97  <-- Paso 3
r <= 1 | 18.45   13.75   15.67   20.20  <-- Paso 2
r = 0  | 45.30   19.77   22.00   26.81  <-- Paso 1
```