Regresion con procesos no estacionarios

Tere García Muñoz (tgarciam@ugr.es)
Román Salmerón Gómez (romansg@ugr.es)
Alessio Gaggero (alessiogaggero@ugr.es)

Universidad de Granada
Departamento de Métodos Cuantitativos para la Economía y la Empresa

Tema 6– Econometría 3
Grado en Económicas

En Economía es (relativamente) habitual encontrarse con variables que muestran una tendencia a lo largo del tiempo.

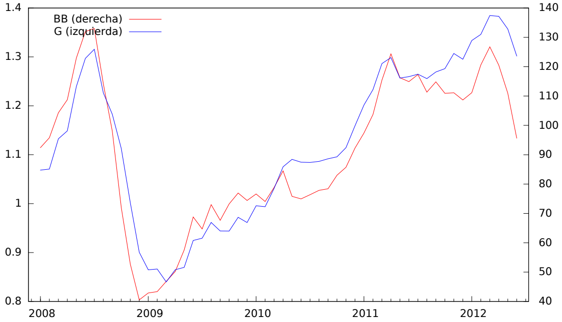

Figura: Precio del barril de brent (BB) y del gasóleo (G)

- Cuando series temporales relacionadas debido a una tendencia común, pero no hay una relación causal real entre ellas, se definen como regresión espuria.
- Regresión espuria pueden mostrar una fuerte correlación, pero no hay un vínculo causal. Esto puede llevar a conclusiones engañosas si no se aborda adecuadamente.
- Comprender la regresión espuria es crucial en econometría, ya que puede afectar la interpretación de resultados y conducir a conclusiones erróneas si no se tiene en cuenta adecuadamente.

- En estos casos, con variables no estacionarias, el modelo de regresión múltiple, $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, tiene buenas propiedades si las variables exógenas y endógenas evolucionan conjuntamente, de forma que los residuos del modelo sean estacionarios, aunque las variables presentes en el modelo no lo sean. Esta es la idea de cointegración.
- Es conocido que la tendencia a rechazar la hipótesis nula en los contrastes de significación individual de los coeficientes de las variables exógenas se debe a que cuando se tienen variables no estacionarias en el modelo de regresión el estadístico usado en este tipo de contrastes no se distribuye según una t de Student.
- Por tanto, el hecho de que una variable no sea estacionaria no es un problema menor, ya que puede conducir a una mala especificación del modelo (perturbación no estacionaria) y a establecer relaciones entre variables de forma errónea.
- A continuación se desarrollan técnicas que permitan detectar la existencia de una raíz unitaria en el comportamiento de una variable y se profundiza en el concepto de cointegración.

- Para analizar si un proceso tiene una raíz unitaria se aplican los contrastes de Dickey y Fuller.
- Considerando el caso más sencillo, el de un proceso AR(1), se estudian las tres situaciones siguientes donde ε_t es ruido blanco:
    1. Modelo sin componentes deterministas: $Y_t = \rho Y_{t-1} + \varepsilon_t$.
    2. Modelo con término constante: $Y_t = \alpha + \rho Y_{t-1} + \varepsilon_t$.
    3. Modelo con término constante y tendencia lineal: $Y_t = \alpha + \beta t + \rho Y_{t-1} + \varepsilon_t$.
- de forma que contrastar: $H_0 : \rho = 1$; $H_1 : \rho < 1$

- El estadístico del contraste es dado por:

$$t_{exp} = \frac{\hat{\rho}}{\sqrt{Var(\hat{\rho})}},$$

- Si éste es mayor (en valor absoluto) que el proporcionado por las tablas construidas DF se rechaza la hipótesis nula. Es decir, no habría raíz unitaria.
- A modo de ejemplo, la Tabla recoge los valores críticos comentados para distintos tamaños muestrales y niveles de significación para los modelos 1 a 3. Los programas informáticos especializados calculan estos valores y su p-valor asociado de forma rutinaria para cualquier tamaño.

	T = 25		T = 100	
Nivel de significación	1 %	5 %	1 %	5 %
Modelo (1)	-2.66	-1.95	-2.6	-1.95
Modelo (2)	-3.75	-3	-3.51	-2.89
Modelo (3)	-4.38	-3.6	-4.04	-3.45

Cuadro: Valores críticos de Dickey y Fuller para contrastes de raíz unitaria

- La hipótesis alternativa es de estacionariedad, es decir, $H_1 : \rho < 1$. Por tanto, el contraste es a una cola. Por ejemplo, para rechazar la hipótesis nula en el modelo (2) con 25 observaciones al 5 % de significación se ha de verificar que $|t_{exp}| > |-3|$.
- **Notar** Usando el estadístico t tradicional y aproximando mediante una normal tipificada, se rechazaría la hipótesis nula si $t_{exp} < -1{,}96$. Por tanto, al 5 % de significación se rechazaría H_0 con mayor frecuencia si no se usan los valores críticos proporcionados por Dickey y Fuller.
- Los valores críticos de las tablas referenciadas son obtenidos suponiendo que el proceso es un AR(1) y su término de error es incorrelado.

- Precisamente, con el objetivo de eliminar en lo posible la autocorrelación del término de error, se incorporan al modelo retardos de la variable en diferencias, es decir:

$$\nabla Y_t = \alpha + \beta t + \nu Y_{t-1} + \delta_1 \nabla Y_{t-1} + \delta_2 \nabla Y_{t-2} + \cdots + \delta_s \nabla Y_{t-s} + \epsilon_t,$$

- donde s representa normalmente la frecuencia con la que se obervan los datos, aunque no hay un criterio preestablecido. Este modelo se conoce como contraste de Dickey-Fuller ampliado (DFA).
- En este caso, para contrastar la existencia de una raíz unitaria en procesos autorregresivos de orden p superior a 1, se actuaría en el DFA como se ha indicado antes para contrastar si $\rho = 1$.

Regresión espúrea

- ¿Qué ocurre si se usan variables con raíces unitarias en un modelo de regresión lineal múltiple?
- Tal y como se ha comentado, entonces es posible que se obtengan conclusiones erróneas debido a la tendencia a rechazar la hipótesis nula en los contrastes de significación individual. Este problema que se conoce como regresión espúrea.
- Para ilustrar el fenómeno de la regresión espúrea, en este apartado se van a simular dos variables independientes no estacionarias y se va a estudiar la regresión simple de una en función de la otra prestando especial atención al contraste de significación individual.
- Sean $Y_t = Y_{t-1} + \xi_t$ y $X_t = X_{t-1} + \eta_t$ dos paseos aleatorios iniciados en cero y donde ξ_t, η_t son innovaciones independientes distribuidas según una normal de media 0 y varianza 1. Con estas características, se simulan 300 observaciones de las cuales se deshechan las 100 primeras para eliminar el efecto de los valores iniciales. En la Figura 2 se representan dichos paseos aleatorios.

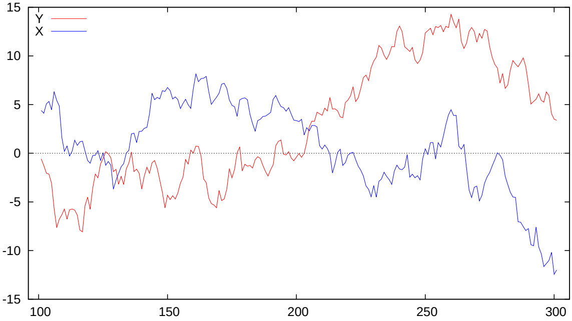

Figura: Representación gráfica de los paseos aleatorios independientes X e Y

- Si se estiman los modelos (1) a (3) para los procesos X e Y se obtiene que el valor experimental de la significación individual de ν son -0.5729, -0.3878 y -1.392 para X y -0.9522, -1.189 y -1.662 para Y. Tomando como orientativos los valores de la Tabla 1 se tiene que los valores experimentales son en todos los casos mayores que los críticos, por tanto, no se rechaza la hipótesis nula en todos los casos. Los procesos simulados tienen una raíz unitaria.
- Al analizar la regresión simple de Y sobre X es esperable que al realizar el contraste de significación individual del coeficiente de X no se rechace la hipótesis nula, sin embargo se obtiene que:

$$\widehat{Y}_t = \underset{(0.3809)}{3.48} - \underset{(0.084)}{0.74} \cdot X_t, \quad R^2 = 0{,}2769, \quad F_{exp} = 76{,}203,$$

- de forma que el coeficiente de X es significativamente distinto de cero al 5 % de significación, es decir, se concluiría que las variaciones de X influyen en Y cuando la realidad es que dichos procesos se han generado de forma totalmente independiente.

- Una posible solución consiste en usar las primeras diferencias de ambas variables, de esta forma se pretende obtener variables estacionarias (ver Figura 3) y se obtendría una interpretación de los coeficientes a corto plazo. En este caso se obtendría que:

$$\nabla \widehat{Y}_t = \underset{(0.0711)}{- \ 0.0125} \cdot \nabla X_t, \quad R^2 = 0{,}0001, \quad F_{exp} = 0{,}8606,$$

donde ahora el coeficiente de X no es significativamente distinto de cero al 5 % de significación.

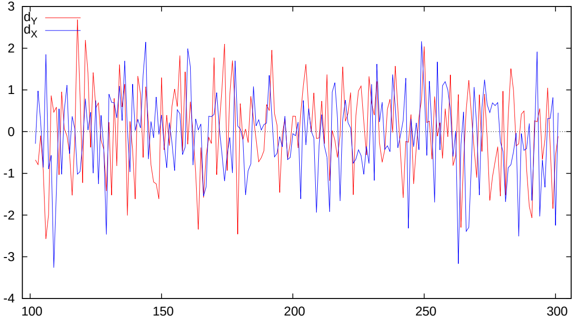

Figura: Primeras diferencias de los paseos aleatorios independientes X e Y

Estacionariedad y cointegración

- ¿Todas las regresiones entre variables no estacionarias pueden entenderse como espúreas?
- Si los residuos de la regresión son estacionarios aunque las variables que las componen no lo sean se tendría que la evolución temporal de ambas variables es estable, por lo que se entiende que existe entre ambas una relación a largo plazo.
- En tal caso se diría que ambas variables están cointegradas.
- Así, en la Figura 4 se tienen los residuos de las regresiones ajustadas en el apartado anterior.
- Para el primer caso, donde se parte de variables no estacionarias, se tiene que son claramente no estacionarios (izquierda) y, por tanto, X e Y no estarían cointegradas, mientras que sí lo son para el segundo donde se parte de variables estacionarias.

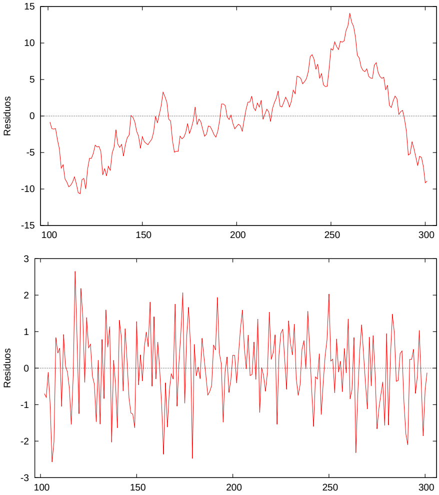

Figura: Residuos de las regresiones $Y_t = \alpha + \beta X_t + u_t$ y $\nabla Y_t = \beta \nabla X_t + \nabla u_t$

Definition (Variable integrada)

Una variable X_t es integrada de orden d, $X_t \sim I(d)$, si su diferencia de orden d admite una representación ARMA estacionaria e invertible. $\Diamond$

A partir de esta definición se tiene que:

- Si $X_t \sim I(0)$, entonces es estacionaria.
- Si $X_t \sim I(d)$, entonces $aX_t + b \sim I(d)$ con $a, b \in \mathbb{R}$ siendo $a \neq 0$.
- Si $X_t \sim I(d)$, $Y_t \sim I(s)$ con $d < s$, entonces $X_t + Y_t \sim I(s)$.

Estacionariedad y cointegración

Definition (Cointegración: dos variables)

Sean X_t e Y_t variables $I(d)$, se dicen que están cointegradas de orden (d, b), $CI(d, b)$, si existe una constante $a \in \mathbb{R}$, llamada parámetro de cointegración, tal que $Y_t - a \cdot X_t$ es $I(d - b)$ con $b > 0$.
En el caso que $d = b$, $Y_t - a \cdot X_t$ es estacionario. $\lozenge$

- Comprobar si las variables X e Y están cointegradas es equivalente a estudiar si $Y_t - aX_t$ tiene una raíz unitaria donde el parámetro de cointegración puede obtenerse a partir de la regresión $Y_t = \beta X_t + u_t$, es decir, $a = \widehat{\beta}$.
- En tal caso, $Y_t - \widehat{\beta} X_t$ correspondería a los residuos de la regresión y, por tanto, X e Y estarán cointegradas si los residuos de la regresión de Y sobre X son estacionarios.
- Para comprobar ésta cuestión simplemente hay que aplicar el contraste DFA sobre los residuos.

- Entonces, la importancia del concepto de cointegración consiste en que si se tienen, por ejemplo, dos variables X_t e Y_t cointegradas de órdenes (d, b) con $d = b$, entonces la estimación del modelo:

$$Y_t = \alpha + \beta X_t + u_t, \quad t = 1, 2, \ldots, T,$$

- tiene buenas propiedades ya que los residuos serán estacionarios y, además, la estimación de β coincide con la constante de cointegración.
- La generalización es inmediata en el caso de tener más de una variable exógena.

Definition (Cointegración)

Sea $\mathbf{Z}_t = (Z_{1t}, \ldots, Z_{kt})$ un vector formado por variables $I(d)$, se dicen que están cointegradas de orden (d, b), $CI(d, b)$, si existe un vector de números reales $\mathbf{a} = (a_1 \ldots, a_k)$, llamado vector de cointegración, tal que $\mathbf{a}^t \cdot \mathbf{Z}_t$ es $I(d - b)$ con $b > 0$. $\diamondsuit$