

Clase auxiliar # 10

Test de Dickey-Fuller (raíz unitaria), variables instrumentales y modelos de series de tiempo

Personaje de hoy: Felipe Garcés.

Participación especial: Cristian Aguayo

1. Test de Dickey-Fuller

Se presentan datos desde enero de 2000 sobre log cantidades producidas de limones en Chile Q . Hay una presunción de que cambios climáticos pudieran generar efectos permanentes en la producción. Por esto, se lleva a cabo tests aumentados de Dickey-Fuller, con los siguientes resultados:

1. Comente la siguiente afirmación: al examinar una regresión entre una o varias series de tiempo integradas de orden 1 ($I(1)$), se suele obtener un R^2 alto. Sin embargo el R^2 puede ser engañoso en este contexto porque el modelo más simple que podemos utilizar ya no corresponde, en este caso, al promedio de la variable.

Correcto.

Supongamos que los regresores no tienen poder predictivo sobre la variable dependiente.

En un modelo estacionario, el mejor predictor sería el promedio de las observaciones ($y_t = \bar{y} + u_t$), por lo que el coeficiente R^2 queda descrito de la forma tradicional:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y}_i)^2}$$

Sin embargo, en una serie de tiempo donde la variable dependiente es $I(1)$, el promedio muestral no será un buen predictor. En este caso, un “camino aleatorio” ($y_t = y_{t-1} + u_t$) podría ser el modelo más simple para explicar y_t .

Considerando esto, podemos formular un R^2 que –siendo menor– nos de una noción más razonable del poder explicativo del modelo:

$$R^2 = 1 - \frac{\sum_{t=1}^T (y_t - \hat{y}_t)^2}{\sum_{t=1}^T (y_t - y_{t-1})^2}$$

2. En clases vimos que el número de ahogados en accidentes de piscinas en EEUU (X) y el número de películas filmadas por año del Nicholas Cage (Y) son variables que están fuertemente correlacionadas. ¿Cómo podemos entender este fenómeno? ¿Qué métodos o ideas podemos utilizar para estudiar la veracidad de esta relación?

El fenómeno que correlaciona fuertemente variables que no están relacionadas se conoce como regresión espúrea o falsa. Se debe principalmente a la existencia de tendencias y/o fuerte persistencia muy común en series temporales. En clases vimos dos formas de evaluar si existe una relación espúrea:

- Redefiniendo el R^2 (como vimos en la parte anterior), asumiendo que el modelo básico en contextos de series de tiempo puede ser, por ejemplo, un camino aleatorio $Y_t = Y_{t-1} + \varepsilon_t$.

$$R^2 = 1 - \frac{\sum_{t=1}^T (y_t - \hat{y}_t)^2}{\sum_{t=1}^T (y_t - y_{t-1})^2}$$

Otra opción sería considerar como modelo básico uno con tendencia lineal ($Y_t = \alpha + \beta t + \varepsilon_t$).

$$R^2 = 1 - \frac{\sum_{t=1}^T (y_t - \hat{y}_t)^2}{\sum_{t=1}^T (y_t - [\hat{\alpha} + \hat{\beta}t])^2}$$

- Estudiar si existe cointegración: combinaciones lineales de las variables que tengan comportamiento estacionario. Por ejemplo, se estimaría el modelo explicando a Y por X y con sus residuos, se realizaría un tests de raíz unitaria.
3. Para esta pregunta utilizaremos la base de datos Auxiliar 09 - BD Test de Dickey-Fuller.xlsx, proporcionada por la Universidad de Illinois¹.
- a) Proponga un modelo con dinámica de ajuste sin tendencia ni constante, con un horizonte de hasta 4 períodos, que le permita llevar a cabo el test de Dickey-Fuller.

El modelo con dinámica de ajuste más genérico que hemos visto es el siguiente:

$$\Delta y_t = \lambda + \pi t + \delta y_{t-1} + \sum_{k=1}^K \Delta_{t-k} \eta_k \Delta y_{t-k} + u_t$$

Imponiendo las condiciones del enunciado tenemos...

$$\Delta y_t = \delta y_{t-1} + \eta_1 \Delta y_{t-1} + \eta_2 \Delta y_{t-2} + \eta_3 \Delta y_{t-3} + \eta_4 \Delta y_{t-4} + u_t$$

- b) Señale las características del test.

- **Hipótesis nula:** $\delta = 0$ (presencia de raíz unitaria).
- **Estadístico de prueba:** $t = \frac{\hat{\delta}}{\widehat{DS}(\hat{\delta})} \sim ADF(T = 72)$
- **Región de rechazo:** $RR = \{t < ADF(T = 72, 5\%) \approx -1,95\}$

- c) Ejecute el test de Dickey-Fuller en *Stata* con y sin el comando `dfuller`. Luego concluya a partir de los resultados del test.

Como se puede observar en las figuras 1 y 2, se rechaza la hipótesis nula de que exista raíz unitaria. Esto provee evidencia a favor de la estabilidad del sistema con un 95 % de confianza.

¹<http://www.econ.uiuc.edu/~econ472/tutorial9.html>

Augmented Dickey-Fuller test for unit root			Number of obs =		72
		Interpolated Dickey-Fuller			
	Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value	
Z(t)	1,944	-2,611	-1,950	-1,610	
D.y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
y					
L1.	,0070187	,0036098	1,94	0,056	-,0001866 ,0142239
LD.	,3695085	,1225645	3,01	0,004	,1248687 ,6141483
L2D.	-,001241	,130477	-0,01	0,992	-,2616742 ,2591922
L3D.	-,0439054	,1291756	-0,34	0,735	-,3017409 ,2139301
L4D.	-,0322862	,1212828	-0,27	0,791	-,2743676 ,2097953

Figura 1: Resultados del test de Dickey-Fuller aumentado sin usar el comando `dfuller`.

Source	SS	df	MS	Number of obs	=	72
				F(5, 67)	=	4,25
Model	457113,846	5	91422,7692	Prob > F	=	0,0020
Residual	1440055,15	67	21493,3605	R-squared	=	0,2409
				Adj R-squared	=	0,1843
Total	1897169	72	26349,5694	Root MSE	=	146,61

D.y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
y						
L1.	,0070187	,0036098	1,94	0,056	-,0001866	,0142239
LD.	,3695085	,1225645	3,01	0,004	,1248687	,6141483
L2D.	-,001241	,130477	-0,01	0,992	-,2616742	,2591922
L3D.	-,0439054	,1291756	-0,34	0,735	-,3017409	,2139301
L4D.	-,0322862	,1212828	-0,27	0,791	-,2743676	,2097953

Figura 2: Resultados del test de Dickey-Fuller aumentado usando el comando `dfuller`.

2. Variables instrumentales (Econometría, Examen 2017-1)

Esta pregunta se basa en el artículo *Can Women Have Children and a Career? IV Evidence from IVF Treatments* (AER, Junio 2017) de Lundborg, Plug, and Rasmussen. Dicho documento presenta evidencia de que las mujeres con hijos ganan menos y trabajan menos por el hecho de tener hijos, no porque sean más o menos productivas.

Para cada mujer $i = \{1, \dots, N\}$ en el año $t = \{1994, \dots, 2005\}$, se tienen los siguientes datos:

- $W_{it} \in R^{NT \times 1}$: salario.
- $C_{it} \in R^{NT \times K}$: características exógenas de las mujeres que miden productividad (edad, experiencia, educación, estado de salud física y mental, salarios pasados, etc).
- F_{it} : variable dummy que indica si la mujer tiene hijos o no.

- a) Suponga que las mujeres en Dinamarca tienen acceso a anticonceptivos para una planificación familiar. Argumente y establezca claramente el potencial problema de endogeneidad.

El hecho de que una mujer recurra a anticonceptivos es una variable no observada (contenida en u_t) que claramente está correlacionada con otras variables observadas del modelo, en particular con el hecho de tener hijos (F_t).

Es así como $Cov(F_t, u_t) \neq 0$, por lo que habría endogeneidad.

- b) Para las mujeres sin hijos se poseen datos sobre aquellas que recibieron tratamientos de fertilización in-vitro. Para este subconjunto de mujeres se conoce la dummy Z_i que es igual a uno si el primer intento de fertilización in vitro es exitoso y cero en caso contrario. ¿Qué condiciones debe cumplir Z_i para ser un instrumento válido? Sugiera formas de testear estas condiciones.

Un instrumento válido debe cumplir dos condiciones:

- **Relevancia:** $Cov(F_{it}, Z_i) \neq 0$. Se puede testear regresionando F en función de Z .
- **Exclusión:** $Cov(Z_{it}, u_{it}) = 0$. Se puede regresionar W en función de Z y chequear que Z solo afecta a W a través de F .

- c) Utilizando las variables C_{it} , W_{it} , F_{it} y Z_i , establezca claramente las estimaciones a realizar para encontrar el efecto causal de un hijo en el salario de la madre. Plantear las condiciones de momentos que implica este modelo de mínimos cuadrados en dos etapas.

$$F_{it} = \underbrace{C_{it}}_{1 \times K} \underbrace{\gamma}_{K \times 1} + \delta Z_{it} + v_{it}$$

$$W_{it} = \underbrace{C_{it}}_{1 \times K} \underbrace{\alpha}_{K \times 1} + \beta \hat{F}_{it} + u_{it}$$

3. Modelos AR(p) y MA(q)

Considere los siguientes procesos estocásticos en series de tiempo:

$$Y_t = \mu + \gamma Y_{t-1} + \varepsilon_t$$

$$W_t = \alpha + \varepsilon_t - \theta \varepsilon_{t-1}$$

1. Usando operadores de rezago, muestre que un modelo AR(1) puede representarse también como un modelo MA(∞).

$$\begin{aligned} Y_t &= \mu + \gamma Y_{t-1} + \varepsilon_t \\ &= \mu + \gamma(\mu + \gamma Y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= \mu + \gamma\mu + \gamma^2 Y_{t-2} + \varepsilon_t + \gamma\varepsilon_{t-1} \\ &= \mu + \gamma\mu + \gamma^2(\mu + \gamma Y_{t-3} + \varepsilon_{t-2}) + \varepsilon_t + \gamma\varepsilon_{t-1} \\ &= \mu + \gamma\mu + \gamma^2\mu + \gamma^3 Y_{t-3} + \varepsilon_t + \gamma\varepsilon_{t-1} + \gamma^2\varepsilon_{t-2} \end{aligned}$$

$$\begin{aligned}
&= \underbrace{\mu \sum_{i=0}^{\infty} \gamma^i + \lim_{j \rightarrow \infty} \gamma^j Y_{t-j}}_{\alpha'} + \sum_{k=0}^{\infty} \gamma^k \varepsilon_{t-k} \\
&= \alpha' + \sum_{k=0}^{\infty} \gamma^k \varepsilon_{t-k}
\end{aligned}$$

2. Usando operadores de rezago, muestre que un modelo MA(1) puede representarse también como un modelo AR(∞).

$$\begin{aligned}
W_t &= \alpha + \varepsilon_t + \theta \varepsilon_{t-1} \\
&= \alpha + \varepsilon_t + \theta(W_{t-1} - \alpha - \theta \varepsilon_{t-2}) \\
&= \alpha + \varepsilon_t + \theta W_{t-1} - \alpha \theta - \theta^2 \varepsilon_{t-2} \\
&= \alpha + \varepsilon_t + \theta W_{t-1} - \alpha \theta - \theta^2(W_{t-2} - \alpha - \theta \varepsilon_{t-3}) \\
&= \alpha(1 - \theta + \theta^2 - \theta^3) + (\theta W_{t-1} - \theta^2 W_{t-2} + \theta^3 W_{t-3}) + \varepsilon_t - \theta^4 \varepsilon_{t-4} \\
&= \underbrace{\alpha \sum_{j=0}^{\infty} (-\theta)^j}_{\mu'} + \sum_{j=0}^{\infty} \theta(-\theta)^j W_{t-1-j} + \underbrace{\varepsilon_t + \lim_{j \rightarrow \infty} \theta^j \varepsilon_{t-j}}_{\varepsilon'_t} \\
&= \mu' + \sum_{j=0}^{\infty} \theta(-\theta)^j W_{t-1-j} + \varepsilon'_t
\end{aligned}$$