IN4402 - Aplicaciones de Probabilidades y Estadística en Gestión
Profesor: Benjamín Villena Roldán
Profesor auxiliar: Ronald Leblebici Garo

INGENIERIA INDUSTRIAL
UNIVERSIDAD DE CHILE

Clase auxiliar # 9

Test de Breusch-Godfrey (autocorrelación de errores) y de Dickey-Fuller (raíz unitaria)

Canción de hoy: En el aire - Nonpalidece.

¡Recordar!

- **Test de Breusch-Godfrey (autocorrelación de los errores)**: sirve para testear autocorrelación de los errores. A continuación se describe cómo llevar a cabo el test.
    - Regresionar $Y = X\beta + U$ y guardar los resudos $\hat{U}$.
    - Regresionar $\hat{U}_t = \rho_1 \hat{U}_{t-1} + ... + \rho_p \hat{U}_{t-p} + V_t$.
    - Construir el estadístico $T \cdot R^2 \sim \chi^2_p$
    - Testear $H_0 : \rho_1 = ... = \rho_p = 0$ (no existencia de auto-correlación de los errores). Si $T \cdot R^2 > \chi^2_{p,1-\alpha}$, se rechaza la hipótesis nula.

- **Test de Dickey-Fuller (raíz unitaria):** sirve para testear si en un modelo AR(1) $\rho = 1$, con el fin de ver si es estable o no. Tiene distintas variantes, pero la idea es básicamente la misma siempre. Mostraremos el caso más sencillo:

$$Y_t = \rho Y_{t-1} + U_t$$
$$\Rightarrow \Delta Y_t \equiv Y_t - Y_{t-1} = \underbrace{(\rho - 1)}_{\delta} Y_{t-1} + U_t$$

  Testear $H_0 : \rho = 1$, será equivalente a testear la significancia individual de δ, con la salvedad de que $\hat{\delta}$ no sigue una distribución T-Student, sino que una denominada ADF(T) (Augmented Dickey-Fuller), que varía con el tamaño muestral T.

  Si el estadístico $t = \frac{\hat{\delta}}{\hat{DS}(\hat{\delta})}$ es tal que $t < ADF(T, \alpha)$, se rechaza la hipótesis nula y se concluye que $|\rho| < 1$, no habiendo raíz unitaria y siendo estable el modelo.

1. Test de Breusch-Godfrey y un poco más

El objetivo de esta pregunta es mostrar que la autocorrelación de los errores en una serie de tiempo puede generar endogeneidad. Sin embargo, no nos quedaremos solo en lo teórico, sino que nos basaremos en un ejemplo aplicado. Utilizaremos el archivo

Consideremos el siguiente modelo para la curva de Phillips vista (o por ver) en Macroeconomía:

$$\pi_t = \beta_0 + \beta_1 \pi_{t-1} + \alpha_0 d_t + u_t$$

Esta ecuación busca predecir el comportamiento de la inflación π_t, en un determinado país, en función de su desempleo d_t.

1. Formule el test de Breusch Godfrey suponiendo autocorrelación de orden 1 para los errores u_t.

   Suponer autocorrelación de orden 1, significa que podemos expresar u_t de la siguiente manera:

   $$u_t = \rho u_{t-1} + \varepsilon_t$$

   El test de Breusch-Godfrey debe tener las siguiente características:

   - **Hipótesis nula:** $\rho = 0$ (ausencia de autocorrelación de orden 1 entre los errores).
   - **Estadístico de prueba:** $T \cdot R^2 \sim \chi^2_1$
   - **Región de rechazo:** $RR = \{X^2 > \chi^2_{1,95\,\%} \approx 3,84\}$

2. Realice el test de Breusch-Godfrey con y sin ocupar el comando `estat bgodfrey`. ¿Qué concluye? Recuerde esta conclusión para la realización del resto del problema.

Source	SS	df	MS			
				Number of obs	=	302
				F(2, 299)	=	9885,81
Model	7174,32468	2	3587,16234	Prob > F	=	0,0000
Residual	108,495007	299	,362859554	R-squared	=	0,9851
				Adj R-squared	=	0,9850
Total	7282,81969	301	24,1954142	Root MSE	=	,60238

inflacion	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
desempleo	-,0395072	,0192159	-2,06	0,041	-,0773229	-,0016916
inflacion						
L1.	,9669459	,0069523	139,08	0,000	,9532641	,9806276
_cons	,4243536	,1601846	2,65	0,008	,1091216	,7395855

Figura 1: Resultado de regresionar el modelo $\pi_t = \beta_0 + \beta_1 \pi_{t-1} + \alpha_0 d_t + u_t$ en *Stata*.

Source	SS	df	MS			
				Number of obs	=	301
				F(1, 299)	=	10,61
Model	3,71043647	1	3,71043647	Prob > F	=	0,0013
Residual	104,554984	299	,349682222	R-squared	=	0,0343
				Adj R-squared	=	0,0310
Total	108,265421	300	,360884737	Root MSE	=	,59134

u	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
u						
L1.	,1850181	,0567987	3,26	0,001	,0732422	,2967939
_cons	-,0017863	,0340843	-0,05	0,958	-,0688618	,0652892

Figura 2: Resultado de regresionar el modelo $\hat{u}_t = \rho \hat{u}_{t-1} + \varepsilon_t$ en *Stata*.

Breusch-Godfrey LM test for autocorrelation

lags(p)	chi2	df	Prob > chi2
1	10,477	1	0,0012

H0: no serial correlation

Figura 3: Resultado de realizar el test de Breusch-Godfrey en *Stata* mediante el comando `estat bgodfrey`.

De la figura 3 se puede observar que se ha rechazado la hipótesis de ausencia de autocorrelación de los errores ($\rho = 0$) con un nivel de significancia del $5\,\%$. Por lo tanto, hay evidencia para afirmar que existe dicha autocorrelación.

Consecuentemente, a través de los resultados de la figura 2, se puede obtener el valor del estadístico $T \cdot R^2 = 303 \cdot 0{,}0343 = 10{,}3929 > \chi^2_1 \approx 3{,}84$

3. Calcule $Cov(\pi_{t-1}, u_t)$.

$$
\begin{gathered}
Cov(\pi_{t-1}, u_t) = \mathbb{E}(\pi_{t-1}u_t) - \mathbb{E}(\pi_{t-1})\mathbb{E}(u_t) \\
u_t = \rho u_{t-1} + \varepsilon_t \\
\Rightarrow (1-\rho L)u_t = \varepsilon_t \\
\Rightarrow \mathbb{E}(u_t) = \frac{1}{1-\rho}\mathbb{E}(\varepsilon_t) = 0 \\
\Rightarrow Cov(\pi_{t-1}, u_t) = \mathbb{E}(\pi_{t-1}u_t)
\end{gathered}
$$

$$
\begin{gathered}
\pi_{t-1} = \beta_0 + \beta_1\pi_{t-2} + \alpha_0 d_{t-1} + u_{t-1} \\
\Rightarrow \pi_{t-1}u_t = \beta_0 u_t + \beta_1\pi_{t-2}u_t + \alpha_0 d_{t-1}u_t + u_{t-1}u_t \\
\Rightarrow \mathbb{E}(\pi_{t-1}u_t) = \beta_1\mathbb{E}(\pi_{t-2}u_t) + \mathbb{E}(u_{t-1}u_t)
\end{gathered}
$$

$$
\begin{gathered}
u_t u_{t-1} = \rho u_{t-1}^2 + \varepsilon_t u_{t-1} \\
\Rightarrow \mathbb{E}(u_t u_{t-1}) = \rho\mathbb{E}(u_{t-1}^2) + \mathbb{E}(\varepsilon_t u_{t-1}) = \rho\mathbb{E}(u_{t-1}^2) + \mathbb{E}(\varepsilon_t)\mathbb{E}(u_{t-1}) = \rho\mathbb{E}(u_{t-1}^2) = \rho\mathbb{V}(u_{t-1})
\end{gathered}
$$

$$
\begin{gathered}
\mathbb{V}(u_{t-1}) = \mathbb{V}(\rho u_{t-2} + \varepsilon_t) \\
\Rightarrow \mathbb{V}(u_{t-1}) = \rho^2\mathbb{V}(u_{t-2}) + \mathbb{V}(\varepsilon_t) \\
\Rightarrow \mathbb{V}(u_{t-1}) = \frac{\sigma^2_\varepsilon}{1-\rho^2}
\end{gathered}
$$

$$\Rightarrow \mathbb{E}(u_t u_{t-1}) = \rho \mathbb{V}(u_{t-1})$$

$$\pi_{t-2}u_t = \rho\pi_{t-2}u_{t-1} + \pi_{t-2}\varepsilon_t$$
$$\mathbb{E}(\pi_{t-2}u_t) = \rho\mathbb{E}(\pi_{t-2}u_{t-1}) + \mathbb{E}(\pi_{t-2}\varepsilon_t)$$
$$\Rightarrow \mathbb{E}(\pi_{t-2}u_t) = \rho\mathbb{E}(\pi_{t-2}u_{t-1}) = \rho\mathbb{E}(\pi_{t-1}u_t) = \rho Cov(\pi_{t-1}, u_t)$$

$$Cov(\pi_{t-1}, u_t) = \beta_1\rho Cov(\pi_{t-1}, u_t) + \frac{\rho\sigma_\varepsilon^2}{1-\rho^2}$$
$$Cov(\pi_{t-1}, u_t)(1-\beta_1\rho) = \frac{\rho\sigma_\varepsilon^2}{1-\rho^2}$$
$$Cov(\pi_{t-1}, u_t) = \frac{\rho\sigma_\varepsilon^2}{(1-\beta_1\rho)(1-\rho^2)}$$

4. Explique por qué en este caso existe endogeneidad.

   Porque, a través del test de Breusch Godfrey, hemos rechazado la hipótesis de que $\rho = 0$. Luego tendremos que:

$$Cov(\pi_{t-1}, u_t) = \frac{\rho\sigma_\varepsilon^2}{(1-\beta_1\rho)(1-\rho^2)} \neq 0$$

   Por lo tanto, $\pi_{t-1} = 0$ resulta ser un regresor endógeno.

2. 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)	(2)	(3)
Var dep	$\Delta \log Q_t$	$\Delta \log Q_t$	$\Delta \log Q_t$
$\log Q_{t-1}$	-0.00370	-0.245	-0.331
	(0.00729)	(0.0465)	(0.0477)
$\Delta \log Q_{t-1}$			0.297
			(0.0703)
Constant		1.779	2.405
		(0.339)	(0.348)
Observations	172	172	171
R-squared	0.001	0.140	0.243
Valor crítico ADF (5 %)	-2.885	-1.654	-1.654

Figura 4: Desviaciones Estándar en Paréntesis.

1. ¿Qué puede concluir de esta tabla? ¿Existe evidencia de raíz unitaria? ¿Cómo lo interpreta?

   Necesitamos hacer el test ADF para poder examinar si existe evidencia estadística importante en contra de la hipótesis nula de raíz unitaria en la variable. Para ello, debemos construir los test t-Student de significancia individual y compararlos contra los valores críticos a ADF al $5\,\%$.

   - **Caso 1:** $t = \frac{0{,}0037}{0{,}00729} \approx -0,51 > 2,885 \rightarrow$ No se rechaza la hipótesis nula de raíz unitaria.
   - **Caso 2:** $t = \frac{0{,}245}{0{,}0465} \approx -5,27 < 1,654 \rightarrow$ Se rechaza la hipótesis nula de raíz unitaria.
   - **Caso 3:** $t = \frac{0{,}331}{0{,}0477} \approx -6,94 < 1,654 \rightarrow$ Se rechaza la hipótesis nula de raíz unitaria.

   Se concluye que considerar un modelo con constante (o con tendencia lineal), incluyendo o no el término $\Delta \log Q_{t-1}$, hace que la variable analizada sea estacionaria (se rechazaría la hipótesis de raíz unitaria).

2. Comente la siguiente afirmación: al examinar una regresión entre una o varias series de tiempo integradas de orden 1 $\left(I(1)\right)$, 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:

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

3. Para esta pregunta utilizaremos la base de datos `Auxiliar 09 - BD Test de Dickey-Fuller.xlsx`, proporcionada por la Universidad de Illinois[1].

[1] `http://www.econ.uiuc.edu/~econ472/tutorial9.html`

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:** $t = \frac{\hat{\delta}}{\hat{DS}(\hat{\delta})} \sim ADF(T = 72)$
- **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 5 y 6, 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.

Augmented Dickey-Fuller test for unit root — Number of obs = 72

	Test Statistic	Interpolated Dickey-Fuller 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 5: 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 6: Resultados del test de Dickey-Fuller aumentado usando el comando `dfuller`.