

Práctica de Aula PA-MD.1:

Un investigador dispone de una base de datos anuales, para el período de 1948 a 1993, de los siguientes índices agrarios de EEUU, todos ellos con base 1982¹:

output : producción agrícola (Rango 51 - 116).

labor : mano de obra agrícola (Rango 81 - 278).

land : superficie utilizada en la producción agrícola (Rango 89 - 102)

machines : maquinaria (duradera) (Rango 38 - 102)

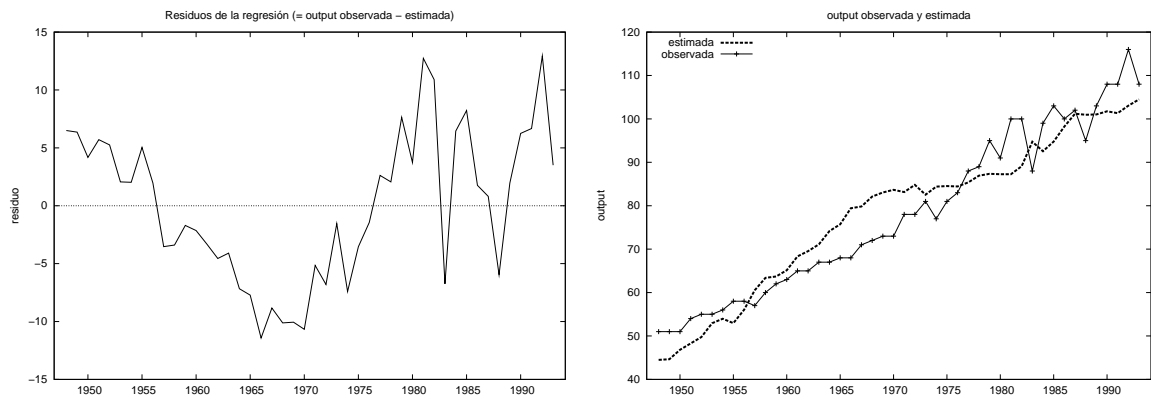
El objetivo del investigador es determinar la función de producción agraria, para ello especifica el siguiente modelo de regresión lineal:

$$output_t = \beta_1 + \beta_2 labor_t + \beta_3 land_t + \beta_4 machines_t + u_t \quad t = 1, \dots, T \quad (1)$$

en el que se considera que los regresores son no estocásticos. Los resultados obtenidos de la estimación MCO son los que se muestran a continuación:

$$\begin{aligned} \widehat{output}_t &= 181,201 - 0,307 labor_t - 0,517 land_t - 0,096 machines_t \\ (\widehat{des}(\hat{\beta}_{MCO})) & \quad (40,194) \quad (0,038) \quad (0,564) \quad (0,169) \\ R^2 &= 0,884 \quad DW = 0,612 \quad SCR = 1885,08 \quad T = 46 \end{aligned}$$

Figura 1: *Gráfico de residuos y de la serie gfr observada y ajustada*



1. Explica cómo se han calculado los residuos y para qué sirven los gráficos. Interpreta ambos gráficos y señala si existe alguna evidencia de que la perturbación del modelo no cumpla alguna de las hipótesis básicas, justificando tu respuesta.

¹Fichero data9-5.gdt disponible en gretl pestaña Ramanathan. Fuente: Ramanathan, Ramu (2002): *Introductory Econometrics with Applications*.

2. Realiza algún contraste basándote en la información disponible, para cualquier problema detectado en el apartado anterior. Explica detalladamente todos los elementos que intervengan.
3. Con respecto a los contrastes de significatividad individual de las variables explicativas del modelo (1):
 - a) ¿Es fiable realizarlos utilizando la información disponible? ¿Por qué?
 - b) ¿Sería posible llevarlos a cabo si no tuviésemos otra opción que la de estimar los coeficientes del modelo por MCO? Explica cómo lo harías en caso afirmativo.

Viendo los resultados obtenidos el investigador decide estimar el mismo modelo por el método de Cochrane-Orcutt (CO). A continuación se muestran los resultados obtenidos:

Estimaciones Cochrane–Orcutt utilizando las 45 observaciones 1949–1993
Variable dependiente: output
 $\hat{\rho} = 0.791585$

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	54,3902	30,3065	1,7947	0,0801
labor	−0,4046	0,0649	−6,2284	0,0000
land	1,0727	0,3741	2,8673	0,0065
machines	−0,2874	0,2000	−1,4372	0,1583

Estadísticos basados en los datos rho-diferenciados:

R^2	0.957590	R^2 corregido	0.954487
$F(3, 41)$	12.99460	Valor p (de F)	4.18e−06
$\hat{\rho}$	−0.184791	Durbin–Watson	2.339505

$$\widehat{V}(\hat{\beta}_{CO}) = \begin{bmatrix} 918,486 & 0,175515 & -9,59451 & -0,06293 \\ 0,175515 & 0,00422165 & -0,00963229 & 0,00270274 \\ -9,59451 & -0,00963229 & 0,139972 & -0,034905 \\ -0,06293 & 0,00270274 & -0,034905 & 0,0400013 \end{bmatrix}$$

4. ¿Cuándo estás dispuesto a aplicar este método de estimación? En particular, ¿consideras adecuado utilizar este método en las circunstancias actuales? Responde razonadamente.
5. Describe detalladamente cómo obtener las estimaciones de los coeficientes del modelo (1) utilizando el método del apartado anterior.
6. Con la información disponible, realiza el siguiente contraste $H_0 : \beta_2 = \beta_3$. Escribe la hipótesis nula, la alternativa, el estadístico de contraste junto con su distribución y realiza el contraste. ¿Cómo interpretas el resultado?

A continuación el investigador introduce un retardo de la variable endógena como variable explicativa en el modelo inicial, con la pretensión de recoger la influencia de la producción agrícola del año anterior:

$$output_t = \beta_1 + \beta_2 labor_t + \beta_3 land_t + \beta_4 machines_t + \beta_5 output_{t-1} + v_t \quad t = 2, \dots, T \quad (2)$$

Estimado el modelo por MCO se obtienen los siguientes resultados:

Estimaciones MCO utilizando las 45 observaciones 1949–1993
Variable dependiente: output

	Coefficiente	Desv. Típica	Estadístico t	Valor p
const	-26,6869	33,4166	-0,7986	0,4292
labor	-0,0868	0,0356	-2,4339	0,0195
land	0,6694	0,3649	1,8342	0,0741
machines	-0,1715	0,1013	-1,6929	0,0983
output_1	0,8535	0,0979	8,7126	0,0000
R^2	0.959224	R^2 corregido		0.955146
$F(4, 40)$	235.2394	Valor p (de F)		3.26e-27
$\hat{\rho}$	-0.299970	h de Durbin		-2.617899
$BG(1)$	6,199	Valor p (de $BG(1)$)		0.0128

7. Utilizando esta información, explica detalladamente la validez del siguiente estadístico de contraste

$$\frac{\hat{\beta}_{5,MCO} - 0}{\widehat{desv}(\hat{\beta}_{5,MCO})} \xrightarrow{H_0, d} N(0, 1)$$

para argumentar a favor de incluir en el modelo el retardo de la variable endógena como variable explicativa.

Alternativamente se ha obtenido la siguiente estimación consistente y asintóticamente eficiente de los coeficientes del modelo (2):

$$\begin{aligned} \underbrace{output_t - \hat{\rho} output_{t-1}}_{Q_t^*} &= -27,47 \underbrace{(1 - \hat{\rho})}_{X_t^*} - 0,058 \underbrace{(labor_t - \hat{\rho} labor_{t-1})}_{LB_t^*} \\ &+ 0,546 \underbrace{(land_t - \hat{\rho} land_{t-1})}_{LN_t^*} - 0,130 \underbrace{(machines_t - \hat{\rho} machines_{t-1})}_{MA_t^*} \\ &+ 0,925 \underbrace{(output_{t-1} - \hat{\rho} output_{t-2})}_{Q_{t-1}^*} + \hat{\epsilon}_t \end{aligned} \quad (3)$$

$R^2 = 0,976 \quad DW = 2,30$

siendo ϵ_t es un ruido blanco tal que $\epsilon_t = v_t - \rho v_{t-1}$ y v_t son las perturbaciones del modelo (2).

8. Completa y/o realiza lo siguiente:

- a) $\epsilon_t \sim \dots\dots (\quad , \quad)$
- b) ¿Cuál es el método de estimación que se ha utilizado?
- c) ¿Cuál es el estimador consistente de ρ empleado? Describe todos los elementos y las condiciones que te garantizan la consistencia del parámetro ρ estimado.
- d) Escribe la expresión matricial del estimador utilizado:

$$\begin{bmatrix} -27,47 \\ -0,058 \\ 0,546 \\ -0,130 \\ 0,925 \end{bmatrix} = \begin{bmatrix} \dots\dots\dots & \dots\dots\dots & \dots\dots\dots & \dots\dots\dots & \dots\dots\dots \\ \dots\dots\dots & \dots\dots\dots & \dots\dots\dots & \dots\dots\dots & \dots\dots\dots \\ \dots\dots\dots & \dots\dots\dots & \dots\dots\dots & \dots\dots\dots & \dots\dots\dots \\ \dots\dots\dots & \dots\dots\dots & \dots\dots\dots & \dots\dots\dots & \dots\dots\dots \\ \dots\dots\dots & \dots\dots\dots & \dots\dots\dots & \dots\dots\dots & \dots\dots\dots \end{bmatrix}^{-1} \begin{bmatrix} \dots\dots\dots \\ \dots\dots\dots \\ \dots\dots\dots \\ \dots\dots\dots \\ \dots\dots\dots \end{bmatrix}$$

9. Utilizando la información contenida en la ecuación (3), explica detalladamente la validez del siguiente estadístico de contraste

$$\frac{\hat{\beta}_5 - 0}{\widehat{des}(\hat{\beta}_5)} \xrightarrow{H_0, d} N(0, 1)$$

para argumentar a favor de incluir en el modelo el retardo de la variable endógena como variable explicativa. Finalmente, ¿incluirías dicha variable en el modelo?

10. ¿Cuál es el estimador óptimo, dada toda la información de que dispones, para los parámetros de la ecuación (2)? Razona tu respuesta detalladamente en relación a todas las alternativas posibles.