

ECONOMETRIA

Guía de Trabajos Prácticos

Año 2020

ÍNDICE TEMÁTICO

TP Nº 1: PROPIEDADES MATEMATICAS Y OPERACIONES CON MATRICES	3
TP Nº 2: EL MODELO DE REGRESIÓN SIMPLE	5
TP Nº 3: EL MODELO DE REGRESIÓN MÚLTIPLE: ESTIMACIÓN	8
TP Nº 4: ANÁLISIS DE REGRESIÓN MÚLTIPLE: INFERENCIA	10
TP Nº 5: MULTICOLINEALIDAD Y ERRORES DE ESPECIFICACIÓN	14
TP Nº 6: TRANSFORMACIONES DE LAS VARIABLES	16
TP Nº 7: ANÁLISIS DE REGRESIÓN MÚLTIPLE: PREDICCIÓN	20
TP Nº 8: VARIABLES FICTICIAS	22
TP Nº 9: HETEROCEDASTICIDAD	25
TP Nº 10: ESTIMACIÓN CON VARIABLES PROXY	28
TP Nº 11: VARIABLE INSTRUMENTAL Y MÍNIMOS CUADRADOS EN 2 ETAPAS.....	30
TP Nº 12: ESTABILIDAD DE LOS COEFICIENTES	33
TP Nº 13: TENDENCIA Y ESTACIONALIDAD EN SERIES TEMPORALES	35
TP Nº 14: AUTOCORRELACIÓN	38

TRABAJO PRÁCTICO Nº 1

PROPIEDADES MATEMATICAS Y OPERACIONES CON MATRICES

Bibliografía:

- Wooldridge (2005), *Introducción a la Econometría. Un enfoque moderno*. 2da. ed, Thomson. Apéndices A, B y D

ACTIVIDAD Nº 1:

- a) ¿Qué significa que una matriz sea simétrica?
- b) ¿Qué significa que una matriz sea idempotente?
- c) ¿Cuál es el dominio de la función logarítmica?

EJERCICIO Nº 1:

A partir de los datos suministrados, verifique numéricamente las siguientes propiedades:

- a) $\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n (x_i - \bar{x})y_i$
- b) $\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n (y_i - \bar{y})x_i$

Y	21	24	26	27	29	25	25	30	21	24	26	27
X	2,8	3,4	3	3,5	3,6	3	2,7	3,7	2,8	3,4	3	3,5

EJERCICIO Nº 2:

Para los datos del ejercicio Nº 1, resuelva y calcule:

- a) $Var(X) =$
- b) $Var(y) =$
- c) $Cov(XY) =$
- d) $Var(X + Y) =$
- e) $Var(X - Y) =$
- f) $Var(2X + 3Y) =$
- g) $Corr(XY) =$
- h) Compruebe que $Var(X) = sd(x)^2$

EJERCICIO Nº 3:

Sean X_1 , X_2 y X_3 las cantidades de pizzas pequeñas, medianas y grandes vendidas en una pizzería en un día determinado. Éstas son variables aleatorias cuyos valores esperados son $E(X_1) = 25$, $E(X_2) = 57$ y $E(X_3) = 30$. Si los precios son $Px_1 = \$5,5$, $Px_2 = \$7,6$ y $Px_3 = \$9,15$, calcule el ingreso esperado por día de la venta de pizza.

EJERCICIO Nº 4:

Para los datos del ejercicio Nº 1, resuelva y calcule:

- a) $\log(YX) =$
- b) $\log(3X) =$
- c) $\log\left(\frac{Y}{X}\right) =$
- d) $\log(X^2) =$

EJERCICIO Nº 5:

Dadas las siguientes matrices, resuelva:

$$A = \begin{bmatrix} 3 & 1 \\ 2 & -2 \end{bmatrix}$$

$$B = \begin{bmatrix} 1 & 4 \\ 5 & -1 \end{bmatrix}$$

- a) $A + B$
- b) $3A$
- c) $A \cdot B$
- d) $B \cdot A$
- e) $\det(A)$
- f) $\det(B)$
- g) A'
- h) B'
- i) $(A \cdot B)'$
- j) $B' A'$
- k) $\text{tr}(A)$
- l) $\text{tr}(B)$

m) A^{-1}

n) B^{-1}

o) $(A')^{-1}$

p) $(A^{-1})'$

q) Verifique la condición para que exista A^{-1}

EJERCICIO Nº 6:

Repita las operaciones del Ejercicio 5 para las siguientes matrices. Verifique sus cálculos en Stata.

$$\mathbf{A} = \begin{bmatrix} 3 & 2 & 1 \\ 0 & -1 & 2 \\ 3 & 1 & 0 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 1 & -2 & 2 \\ 0 & -1 & 3 \\ 0 & 1 & 4 \end{bmatrix}$$

TRABAJO PRÁCTICO Nº 2

EL MODELO DE REGRESIÓN SIMPLE

Bibliografía:

- Wooldridge (2005), *Introducción a la Econometría. Un enfoque moderno*. 2da. ed, Thomson. Cap. 2

ACTIVIDAD Nº 1:

Responda si las siguientes afirmaciones son verdaderas o falsas:

- Los coeficientes de regresión estiman parámetros poblacionales desconocidos.
- Los modelos econométricos son deterministas.
- La influencia de factores no observables no se capta en el modelo.
- Las relaciones estimadas sólo a veces pueden interpretarse como causales.

ACTIVIDAD Nº 2:

- ¿Qué es el error cuadrático medio?
- Busque un ejemplo de datos de corte transversal y otro de serie temporal relevantes para la economía Argentina.

PROBLEMAS:

Resuelva los problemas 2.1), 2.2), 2.5) y 2.6) del libro de Wooldridge.

EJERCICIO Nº 1:

Calcule la recta de regresión estimada entre las siguientes variables y verifique numéricamente las siguientes relaciones:

- $\sum_{i=1}^n (X_i - \bar{X})\hat{u}_i = \sum_{i=1}^n X_i\hat{u}_i = 0$
- $\sum_{i=1}^n \hat{u}_i = 0$
- Calcule el producto cruzado entre:
 - los residuos y los valores observados de Y ,
 - los residuos y los valores estimados de Y .
- Explique los resultados anteriores.
- Verifique las propiedades algebraicas de MCO.

Y	21	24	26	27	29	25	25	30	21	24	26	27
X	2,8	3,4	3	3,5	3,6	3	2,7	3,7	2,8	3,4	3	3,5

EJERCICIO Nº 2:

El archivo *GASTO_SALUD.dta* contiene los datos del gasto promedio en salud por persona en cada estado de los EE.UU. y el ingreso promedio per cápita por estado (datos anuales en miles de dólares).

- Calcule los estadísticos descriptivos básicos para ambas variables: media, mediana, mínimo, máximo y desvío estándar. Interprete.
- Formule el modelo teórico.
- Realice un gráfico de dispersión entre X e Y . Evalúe si es razonable suponer una relación lineal entre las variables.
- Proponga el signo esperado para la pendiente. Justifique.
- Estime e interprete los coeficientes de regresión. Formule el modelo estimado.
- Interprete el coeficiente de determinación (R^2).
- Calcule los residuos.

EJERCICIO Nº 3:

De una población de 1.000 Municipios que disponen de Telefonía Pública Rural, se ha seleccionado una muestra aleatoria simple de 36 municipios en los que se ha observado la renta anual disponible por habitante en miles de pesos (X) y el número mensual de llamadas a larga distancia (Y). Se desea analizar la serie para lo cual se le solicita:

- Determine si se trata de un análisis de serie de tiempo o de corte transversal.
- Represente gráficamente los datos.
- Proponga el signo esperado para los coeficientes de regresión.
- Estime e interprete los coeficientes de regresión por el método de MCO.
- Calcule los residuos.

Municipio	Cantidad llamadas	Ingreso disponible	Municipio	Cantidad llamadas	Ingreso disponible
1	4,2	25,0	19	3,5	22,9
2	4,0	24,1	20	2,3	18,6
3	3,2	21,0	21	1,4	15,4
4	1,1	13,8	22	2,9	21,2
5	2,7	19,6	23	3,8	25,3
6	3,1	19,0	24	1,6	12,2
7	2,5	18,4	25	3,2	21,7
8	3,6	22,2	26	3,3	23,0
9	4,5	24,8	27	1,9	16,2
10	1,7	15,3	28	2,6	18,7
11	0,9	14,2	29	3,5	23,1
12	2,4	17,6	30	1,7	16,7
13	1,9	16,8	31	2,4	17,9
14	3,2	22,1	32	5,1	26,4
15	1,5	13,5	33	2,8	21,7
16	3,9	24,0	34	2,5	18,5
17	4,1	23,7	35	4,2	24,9
18	2,7	19,1	36	3,3	20,0

EJERCICIO Nº 4:

Suponiendo que en la regresión de Y sobre X el coeficiente estimado de X es 1,2 determine si las siguientes afirmaciones son verdaderas:

- Tomando distintas muestras para la variable Y pueden obtenerse otras estimaciones.
- La distribución de estas estimaciones debe estar centrada en torno al valor 1,2.

EJERCICIO Nº 5:

Se quiere estimar un modelo que explique el ahorro generado (S) en función del tipo de interés (R) de la forma $S_i = \beta_0 + \beta_1 R_i + u_i$. Para estimar con mayor precisión beta, si pudiera elegir la muestra en diferentes períodos, ¿la elegiría durante un período de tiempo en el cual los tipos de interés fueran fluctuantes o durante un período de tiempo en el cual los tipos de interés fueran relativamente constantes? ¿Por qué?

EJERCICIO Nº 6:

Se han obtenido los siguientes resultados al estimar dos rectas de regresión:

$$\hat{y}_i = 1,72 + 0,37x_i; \quad R^2 = 0,71; \quad \text{valor} - p \hat{\beta}_1 = 0,03$$

$$\hat{y}_i = 2,37 + 0,81z_i; \quad R^2 = 0,22; \quad \text{valor} - p \hat{\beta}_1 = 0,35$$

- Sabiendo que las variables Z y X no están correlacionadas entre sí, y que $\bar{y} = 20$, indique cuáles serán las estimaciones de los coeficientes y el valor R^2 al estimar por MCO la siguiente ecuación de regresión: $y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + u_i$.
- ¿Se puede decir que la suma de cuadrados residual (SCR) del modelo es igual a la suma de SCR de los dos modelos anteriores? Justifique.

EJERCICIO Nº 7:

En una población, la verdadera relación entre X e Y viene dada por la ecuación: $y_i = 2 + 3x_i$. Suponga que los valores de X en una muestra de 10 observaciones son 1,2, ...,10. Los valores de los términos de perturbación se obtienen de forma aleatoria a partir de una población normal de media cero y varianza unitaria: 0,464; 0,06; 1,486; 1,022; 1,394; 0,906; -1,501; -0,69; 0,179; -1,372.

- a) Presente los 10 valores observados de X e Y .
- b) Estimar los coeficientes de regresión y sus errores estándar utilizando MCO, comparando estos resultados con los verdaderos valores.
- c) Contraste la hipótesis de existencia de relación entre X e Y .

TRABAJO PRÁCTICO Nº 3

EL MODELO DE REGRESIÓN MÚLTIPLE: ESTIMACIÓN

Bibliografía:

- Wooldridge (2005), *Introducción a la Econometría. Un enfoque moderno*. 2da. ed, Thomson. Cap. 3

ACTIVIDAD Nº 1:

- a) Indique en qué casos la regresión simple de Y sobre X_1 produce la misma estimación de MCO para el coeficiente de X_1 que la regresión de Y sobre X_1 y X_2 .

PROBLEMAS:

Resuelva los problemas 3.1), 3.3) y 3.9) del libro de Wooldridge.

EJERCICIO Nº 1:

Los siguientes datos corresponden a la cantidad demandada por distintos restaurantes de camarones frescos, medidos en kgs. (Y), el precio que paga (X) y precio de un bien relacionado (Z). Se le solicita:

- a) Formule el modelo teórico.
- b) Estime los coeficientes de regresión por el método de MCO e interprételes. Expresé el modelo estimado.
- c) Calcule el coeficiente de determinación (R^2).
- d) Compare el coeficiente anterior con los R^2 de ambas regresiones simples.
- e) Calcule los coeficientes de correlación simple. Interprete.
- f) Calcule los coeficientes de correlación parcial. Interprete.

Año	1	2	3	4	5	6	7	8	9	10	11	12	13
Cantidad demandada	800	560	550	520	480	510	515	535	530	520	520	510	540
Precio del bien	5	7,5	7,4	7,8	7,9	8	8,1	8	7,7	7,6	6,9	7	6,4
Precio de otro bien	15	13,6	12,8	12,5	11,9	11,5	11,6	11,8	11,3	11,3	11,6	11,9	12

EJERCICIO Nº 2:

Los datos del ejercicio *TRIGO.dta* permiten estimar un modelo de exportaciones de trigo ($SHIP$) en función del tipo de cambio real ($EXRATE$) y el precio por bushel ($PRICE$). Se le solicita:

- a) Deje expresado el modelo de regresión estimado.
- b) Interprete económicamente los coeficientes estimados.
- c) Verifique que $\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2 \hat{\delta}_1$.
- d) Obtenga los valores estimados del volumen mensual de exportaciones de trigo de acuerdo a la función obtenida anteriormente.
- e) ¿Cuál es la media aritmética de los residuos?
- f) Obtenga la matriz de var-cov de los estimadores del modelo.
- g) ¿Qué porcentaje de la variación de la variable dependiente es explicada por la regresión?
- h) ¿Cómo podría continuar el análisis?

EJERCICIO Nº 3:

- a) A partir de los siguientes datos, obtenga: $(X'X)^{-1}$ y $X'Y$ y estime los coeficientes de regresión en forma matricial.
- b) Proponga a qué variables representan Y_i , X_{1i} y X_{2i} .
- c) Construya la matriz de varianzas y covarianzas de los estimadores.

Y_i	64	71	53	67	55	58	77	57	56	51	76	68	70	63	62	60
X_{1i}	8	11	6	11	8	7	10	9	10	7	12	9	10	9	9	8
X_{2i}	4	9	2	1	9	7	4	7	2	8	3	4	2	9	1	8

EJERCICIO Nº 4:

Para los datos del ejercicio anterior, verifique numéricamente las siguientes relaciones:

- $\sum_{i=1}^n X_{1i} \hat{u}_i = \sum_{i=1}^n X_{2i} \hat{u}_i = 0$
- $\sum_{i=1}^n \hat{u}_i = 0$
- Calcule el producto cruzado entre:
 - los residuos y los valores observados de Y ,
 - los residuos y los valores estimados de Y .
- Explique los resultados anteriores.

EJERCICIO Nº 5:

- A partir de los siguientes datos, obtenga: $(X'X)^{-1}$ y $X'Y$ y estime los coeficientes de regresión en forma matricial.
- Proponga a qué variables representan Y_i , X_{1i} y X_{2i} .
- Construya la matriz de varianzas y covarianzas de los estimadores.

Y_i	192	213	159	201	165	174	231	171	168	153	228	204	210	189	186	180
X_{1i}	24	33	18	33	24	21	30	27	30	21	36	27	30	27	27	24
X_{2i}	12	27	6	3	27	21	12	21	6	24	9	12	6	27	3	24

TRABAJO PRÁCTICO Nº 4

ANÁLISIS DE REGRESIÓN MÚLTIPLE: INFERENCIA

Bibliografía:

- Wooldridge (2005), *Introducción a la Econometría. Un enfoque moderno*. 2da. ed, Thomson. Cap. 4

ACTIVIDAD Nº 1:

Responda si las siguientes afirmaciones son verdaderas o falsas:

- En una regresión simple, el estadístico t elevado al cuadrado es igual al estadístico F .
- Cuando la prueba t se efectúa para una hipótesis alternativa unilateral, el valor p asociado al estadístico es mayor que si la hipótesis alternativa es bilateral.
- Si se rechaza la prueba F para el contraste de restricciones múltiples, eso implica que todos los parámetros son distintos de cero.
- Una varianza del estimador grande implica un estimador menos preciso pero no necesariamente afecta la amplitud de los intervalos de confianza ni la exactitud de las pruebas de hipótesis.

ACTIVIDAD Nº 2:

- Defina error tipo I, error tipo II y potencia de una prueba.
- Explique la diferencia entre el nivel de significación y el valor- p .
- Suponga un modelo de regresión con dos variables explicativas ¿qué estadísticos puede utilizar para contrastar las siguientes hipótesis nulas? (i) $H_0: \beta_1 = 0$; (ii) $H_0: \beta_1 = \beta_2 = 0$.

PROBLEMAS:

Resuelva los problemas 4.3), 4.4), 4.5) y 4.8) del libro de Wooldridge.

EJERCICIO Nº 1:

Teniendo en cuenta que Y_i representa la cantidad demandada de un bien, X_{1i} su precio y X_{2i} el precio de un bien sustituto:

- Formule el modelo teórico.
- Calcule los coeficientes de correlación simple y los coeficientes de correlación parcial.
- Estime los coeficientes de regresión por el método de MCO e interprételos. Expresé el modelo estimado.
- Realice la prueba F . Interprete y grafique.
- Realice las pruebas t . Interprete y grafique.
- Calcule el coeficiente de determinación (R^2). Interprete su valor.
- Calcule los intervalos de confianza (al 95%) para los parámetros del modelo.

Y_i	55	70	90	100	90	105	80	110	125	115	130	130
X_{1i}	100	90	80	70	70	70	70	65	60	60	55	50
X_{2i}	5,5	6,3	7,2	7	6,3	7,4	5,6	7,2	7,5	6,9	7,2	6,5

EJERCICIO Nº 2: (tarea 5/5)

A partir de datos de 37 países sobre esperanza de vida ($lexp$), PBN per cápita ($gnppc$) y porcentaje de la población con acceso a agua potable ($safewater$) se estimó un modelo de regresión múltiple utilizando el software Stata. A partir de la siguiente salida de regresión, responda:

- ¿Es el modelo globalmente significativo?
- ¿El PBN per cápita y el acceso al agua potable explican la esperanza de vida al nacer? ¿Qué porcentaje de la esperanza de vida es explicada por estas dos variables conjuntamente?
- Con estos resultados, ¿es posible determinar qué variable tiene un mayor impacto sobre la esperanza de vida?

```
. webuse lifeexp.dta
(Life expectancy, 1998)
```

```
. reg lexp gnppc safewater
```

Source	SS	df	MS	Number of obs = 37		
Model	668.756311	2	334.378156	F(2, 34) = 38.94		
Residual	291.946392	34	8.58665858	Prob > F = 0.0000		
Total	960.702703	36	26.6861862	R-squared = 0.6961		
				Adj R-squared = 0.6782		
				Root MSE = 2.9303		

lexp	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
gnppc	.0001175	.0000671	1.75	0.089	-.0000189	.0002539
safewater	.188709	.0386174	4.89	0.000	.110229	.2671891
_cons	56.25072	2.631063	21.38	0.000	50.90375	61.59768

EJERCICIO Nº 3:

A partir de los datos del archivo WAGE2.dta se pretende estimar el impacto que tienen los años de escolaridad de los padres y la cantidad de hermanos sobre los años de escolaridad de los trabajadores hombres. Siendo *meduc* los años de escolaridad de la madre, *feduc* los años de escolaridad del padre, *sibs* la cantidad de hermanos y *educ* los años de escolaridad de la persona, se pide:

- Formule el modelo teórico.
- Proponga el signo esperado para las pendientes. Justifique.
- Calcule los estadísticos descriptivos básicos para todas las variables.
- Realice un gráfico de dispersión entre la variable *Y* y cada una de las *X*. Calcule los coeficientes de correlación simple. Interprete.
- Estime los coeficientes de regresión por el método de MCO e interpréte los. Expresé el modelo estimado.
- Realice las pruebas *F* y *t*. Interprete y grafique.
- Interprete el coeficiente de determinación (R^2).
- Calcule los intervalos de confianza (al 95%) para los parámetros del modelo.
- Manteniendo constantes *meduc* y *feduc*, ¿cuánto tiene que aumentar *sibs* para tener una reducción de un año en los años de escolaridad predichos? (es aceptable una respuesta en números no enteros).
- Suponga que el hombre A no tiene hermanos y que su padre y su madre tienen cada uno 12 años de escolaridad. El hombre B tiene dos hermanos y su padre tiene 16 años de escolaridad, mientras que su madre posee 14 años. Calcule los años predichos de escolaridad para cada uno.

EJERCICIO Nº 4:

A partir de la siguiente salida de regresión que se le presenta, se le solicita:

- ¿Es el modelo globalmente significativo?
- ¿Son las variables estadísticamente significativas?
- Deje expresado el modelo estimado.

ANOVA^a

Modelo	Suma de cuadrados	gl	Media cuadrática	F	Sig.
1 Regresión	109286389297	2	54643194649	898,947	,000 ^a
Residual	28630106139	471	60785787,981		
Total	137916495436	473			

a. Variables predictoras: (Constante), Años de educación, Salario inicial

b. Variable dependiente: Salario corriente

Coeficientes ^a								
		Coeficientes no estandarizados		Coeficientes estandarizados			Intervalo de confianza para B al 95%	
		B	Error típ.	Beta			Límite inferior	Límite superior
1	(Constante)	-7808,714	1753,860		-4,452	,000	-11255,073	-4362,355
	Salario inicial	1,673	,059	,771	28,423	,000	1,557	1,788
	Años de educación	1020,390	160,550	,172	6,356	,000	704,906	1335,874

a. Variable dependiente: Salario corriente

EJERCICIO Nº 5:

Con los datos del ejercicio 3 del TPNº 3, se le solicita:

- Obtenga la matriz de var-cov.
- Efectúe los contrastes t bajo la hipótesis nula de que los parámetros son nulos.

EJERCICIO Nº 5:

Se ha estimado el siguiente modelo de regresión simple, utilizando MCO:

$$\hat{V}_i = 1767,61 + 7,48 W_i; \quad R^2 = 0,80$$

Los estadísticos t para el contraste de los coeficientes resultan ser iguales a 1,4 para el término constante y 18,5 para la pendiente. Dado que el valor del estadístico para el contraste de significación lleva a no rechazar la hipótesis nula para el intercepto, se formula y estima el siguiente modelo:

$$\hat{V}_i = 11,71 W_i; \quad R^2 = 0,91$$

¿Está de acuerdo en optar por el segundo modelo en función al mayor R^2 ? Justifique.

EJERCICIO Nº 6:

El modelo siguiente puede usarse para estudiar si los gastos de campaña afectan los resultados de las elecciones:

$$votosA = \beta_0 + \beta_1 \log(gastosA) + \beta_2 \log(gastosB) + \beta_3 fortA + u$$

donde $votosA$ es el porcentaje de votos recibidos por el candidato A, $gastosA$ y $gastosB$ son los gastos de campaña del candidato A y del candidato B y $fortA$ es una medida de la fortaleza del partido del candidato A (el porcentaje de votos que obtuvo el partido A en la elección presidencial más reciente).

- ¿Cuál es la interpretación de β_1 ?
- En términos de los parámetros, establezca la hipótesis nula de que un aumento de 1% en los gastos de A es compensado por un aumento del 1% en los gastos de B.
- Estime el modelo dado usando los datos del archivo VOTO.dta e interprete. ¿Los gastos de A afectan el resultado de las elecciones? ¿Y los gastos de B? ¿Puede usar estos resultados para probar la hipótesis del inciso anterior?
- Estime el modelo que proporciona directamente el estadístico t para probar la hipótesis del inciso b). Concluya, utilizando una hipótesis alternativa de dos colas.

EJERCICIO Nº 7: (tarea 5/5)

A partir de un modelo de regresión múltiple con dos variables explicativas se desea modelar el comportamiento de la oferta de soja (medida a través de las toneladas cosechadas) como función del precio de la soja y de un índice agregado del precio de los granos a lo largo de un período de 16 años conforme al detalle presentado en la tabla.

Se solicita:

- Estime la ecuación por MCO. ¿Es el signo de los coeficientes el esperado de acuerdo a la teoría económica?
- Realice una evaluación del modelo, en particular:
 - la significatividad global del modelo;

- la significatividad estadística de los β ;
- el valor del coeficiente de determinación.

- Corra dos regresiones simples y repita el análisis anterior.
- Calcule el coeficiente de correlación entre las variables explicativas y concluya.

Tn cosechadas de soja	20	25	26	27	28	25	29	26	30	34	32	35	36	45	42	40
Precio de la soja	10	10,5	10,7	10,9	20	10,8	21	24	23	25	24	25	26	39	38	37
Índice de precios de granos	101	107	108	110,5	201,5	109	212	241	232	251,5	241,5	251	261	392	382	371,5

TRABAJO PRÁCTICO Nº 5

MULTICOLINEALIDAD Y ERRORES DE ESPECIFICACIÓN

Bibliografía:

- Wooldridge (2005), *Introducción a la Econometría. Un enfoque moderno*. 2da. ed, Thomson.
Cap. 4

ACTIVIDAD Nº 1:

Responda si las siguientes afirmaciones son verdaderas o falsas:

- a) La interpretación *ceteris paribus* sigue siendo posible ante la presencia de alta colinealidad entre las variables explicativas.
- b) Si una variable explicativa es combinación lineal de las restantes, la matriz $(X'X)^{-1}$ es singular y los coeficientes de regresión no pueden estimarse.
- c) Si la variable omitida del modelo no está correlacionada con la variable incluida en el modelo, los estimadores MCO siguen siendo insesgados.
- d) Aun cuando la variable omitida del modelo no está correlacionada con la variable incluida en el modelo, los estimadores MCO resultan inconsistentes.
- e) La omisión de una variable relevante hace que el estimador subestime al verdadero parámetro.

ACTIVIDAD Nº 2:

- a) Explique la relación entre el sesgo de los estimadores MCO y la covarianza entre las variables explicativas omitidas e incluidas en el modelo.
- b) ¿Cuáles son las consecuencias de la multicolinealidad sobre los estimadores MCO?
- c) ¿Se puede cometer algún error de especificación si debido a la existencia de multicolinealidad se decide eliminar una variable explicativa?

Multicolinealidad

EJERCICIO Nº 1:

Dadas las siguientes observaciones, estime, si es posible, la siguiente regresión:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$$

Y_i	11	24	35	42	25	23	29	45	18	19
X_{1i}	1	5	7	9	3	2	8	7	11	12
X_{2i}	3	11	15	19	7	5	17	15	23	25

En caso de no ser posible, explique por qué e indique cómo solucionaría este problema.

EJERCICIO Nº 2:

- a) Estime los coeficientes de regresión por MCO.
- b) Realice los cálculos que considere convenientes para detectar la existencia de multicolinealidad.
- c) ¿Es el signo de los coeficientes de regresión el esperado de acuerdo a la teoría económica?
- d) Proponga medidas para reducir la multicolinealidad.

Consumo	70	65	90	95	110	115	120	140	155	150
Ingreso	80	100	120	140	160	180	200	220	240	260
Riqueza	810	1009	1273	1425	1633	1876	2052	2201	2435	2686

EJERCICIO Nº 3:

- a) En base a los datos del archivo CEOSAL2.dta, estime mediante MCO un modelo que permita explicar el salario de los ejecutivos (SALARY) en función a las ventas de la empresa (SALES), el precio de mercado (MKTVAL) y la antigüedad de la persona en el puesto (CEOTEN).

- b) Añada la variable PROFITS (utilidades de la empresa) al modelo anterior. Interprete el nuevo modelo y compárelo con el anterior.
- c) Evalúe la existencia de multicolinealidad en este caso. En particular, calcule la matriz de correlación, los Factores de Inflación de Varianza y los Índices de Tolerancia.

Errores de especificación

EJERCICIO Nº 4:

Si el verdadero modelo de regresión es $Y_i = a + bX_i + u_i$ pero, por error, se estima $Y_i = c + dX_i + eZ_i + u_i$

- a) ¿Es $b = d$ y $e = 0$?
- b) ¿Es $\hat{b} = \hat{d}$ y $\hat{e} = 0$?

EJERCICIO Nº 5:

Supongamos que el verdadero modelo es $Y_i = \beta_1 X_i + u_i$, pero en lugar de especificar esta regresión a través del origen, se especifica el modelo usual con presencia de la ordenada al origen $Y_i = \alpha_0 + \alpha_1 X_i + v_i$

- a) Evaluar las consecuencias de cometer este tipo de error de especificación.
- b) Resolver el punto anterior previendo que la segunda ecuación conforma el verdadero modelo a estimar y la primera constituye el modelo realmente estimado. Analizar las consecuencias de ajustar el modelo mal especificado.

EJERCICIO Nº 6: (tomi marco)

En base a la siguiente información se solicitó estimar la elasticidad de sustitución de la mano de obra y de capital utilizando una función CES (Elasticidad de Sustitución Constante). Es decir:

$$\ln(V/L)_i = \ln \beta_1 + \beta_2 \ln W_i + u_i$$

Donde:

(V/L): Valor agregado por unidad de mano de obra.

L : Mano de obra.

W : Tasa de salario real.

Sin embargo, la ecuación presentada está basada en el supuesto de que existe competencia perfecta en el mercado de trabajo. Pero si la competencia es imperfecta, la fórmula correcta para el modelo es:

$$\ln(V/L)_i = \ln \beta_1 + \beta_2 \ln W_i + \beta_3 \ln \left(1 + \frac{1}{E} \right) + u_i$$

Donde "E" representa la elasticidad de la oferta de mano de obra.

- a) ¿Qué tipo de error de especificación está involucrado en la estimación original CES de la elasticidad de sustitución si de hecho el mercado laboral es imperfecto?
- b) ¿Cuáles son las consecuencias teóricas de este error sobre el estimador de β_2 , el parámetro de elasticidad de sustitución?

EJERCICIO Nº 4:

El archivo MORTALIDAD INFANTIL.dta contiene información sobre la mortalidad infantil (MI), el PBI per cápita (PBI) y la tasa de alfabetización femenina (TAF) de 64 países. En base a dicha información, estime los siguientes modelos:

$$1) M_i = \beta_0 + \beta_1 PBI_i + \beta_2 TAF_i + u_i$$

$$2) M_i = \beta_0 + \beta_1 PBI_i + u_i$$

- a) ¿Cuál es la mejor especificación? ¿Qué tipo de error de especificación se está cometiendo si se estima el modelo 2?
- b) Realice la regresión de la variable excluida sobre la incluida y calcule el sesgo.
- c) Realice la prueba RESET (Prueba del error de especificación en regresión).

TRABAJO PRÁCTICO Nº 6

TRANSFORMACIONES DE LAS VARIABLES

Bibliografía:

- Wooldridge (2005), *Introducción a la Econometría. Un enfoque moderno*. 2da. ed, Thomson.
Cap. 6 y Apéndice A

ACTIVIDAD Nº 1:

Responda:

- a) ¿Es el coeficiente de determinación invariable ante cambios en las unidades de las variables?
- b) ¿Qué motivos pueden llevar a que se modele el logaritmo de la variable en lugar de la variable en su escala original?
- c) ¿Por qué las funciones del tipo Cobb-Douglas constituyen un ejemplo de modelos intrínsecamente lineales?

PROBLEMAS:

Resuelva el problema 6.3) del libro de Wooldridge.

EJERCICIO Nº 1:

A partir de los datos suministrados y dados $c_1 = 2$ y $c_2 = 3$, complete la tabla. Concluya.

Y	21	24	26	27	29	25	25	30	21	24	26	27
X	2,8	3,4	3	3,5	3,6	3	2,7	3,7	2,8	3,4	3	3,5

variable dependiente	variable explicativa	constante	pendiente	relación beta 0	relación beta 1
Y	X	7,29902	5,661765	creciente	creciente
c1 Y	X	14,59804	11,32353	creciente	creciente
Y	c2 X	7,299017	1,887255	creciente	creciente
c1 Y	c2 X	14,59803	3,77451	creciente	creciente
c1 + Y	X	9,29902	5,661765	creciente	creciente
Y	c2 + X	-9,686268	5,661764	decreciente	creciente
log(Y)	X	2,514242	0,2236551	creciente	creciente
log(c1 Y)	X	3,207389	0,2236551	creciente	creciente
Y	log(X)	4,827341	17,7869	creciente	creciente
Y	log(c2 X)	-14,71355	17,78689	decreciente	creciente

EJERCICIO Nº 2:

A partir de los siguientes datos, se le solicita:

- a) Corra una regresión suponiendo que ambas variables están medidas en miles de \$.
- b) Repita el ejercicio calculando a Y en \$ (X en miles de \$). Compare los coeficientes estimados con los obtenidos en el inciso a).
- c) Repita el ejercicio calculando a X en \$ (Y en miles de \$). Compare los coeficientes estimados con los obtenidos en el inciso a).
- d) ¿Se modificó el coeficiente de determinación?
- e) ¿Se modificaron los estadísticos t? Explique.

Y	55	70	90	100	90	105	80	110	125	115	130	130
X	100	90	80	70	70	70	70	65	60	60	55	50

EJERCICIO N° 3:

Con los datos del archivo MORTALIDAD INFANTIL.xls, evalúe el impacto que posee el PBI *per cápita* y la tasa de alfabetización femenina sobre la tasa de mortalidad infantil en los 64 países de la muestra. Evalúe la significatividad global del modelo y de cada uno de los coeficientes. Repita el análisis con las variables estandarizadas. Interprete y concluya: ¿Qué variable posee mayor impacto sobre la reducción de la mortalidad infantil?

EJERCICIO N° 4:

Los siguientes datos corresponden a la nómina de empleados de una empresa de venta de electrodomésticos. El gerente de personal sabe que los empleados aumentan su productividad (medida en términos de ventas realizadas) con la experiencia o antigüedad en la empresa, aunque observa que los empleados más antiguos presentan menos iniciativa y motivación.

- Estime un modelo de regresión que permita evaluar la hipótesis del gerente acerca del efecto diferencial de la antigüedad sobre el desempeño de los vendedores.
- ¿Cuál es el impacto de un año más de antigüedad?
- ¿A partir de qué momento se produce el efecto “desmotivación”?

Vendedor	Remuneración por comisiones	Antigüedad	Antigüedad ²
A	2000	6	36
B	2200	6	36
C	2500	7	49
D	3000	14	196
E	3000	15	225
F	2800	13	169
G	4000	9	81
H	5000	9	81
I	6000	10	100
J	7000	11	121
K	8000	10	100
L	8500	10	100
M	9000	10	100
N	2900	13	169
O	2700	14	196
P	3000	12	144
Q	2000	6	36
R	2500	5	25
S	3000	14	196

EJERCICIO N° 5:

Se sabe que las variables Y y X se encuentran relacionadas a través de una función $f(\cdot)$, que, en principio, puede tomar las siguientes formas:

$$\text{Modelo lineal: } Y_i = \beta_0 + \beta_1 X_i + u_i$$

$$\text{Modelo log - log: } \ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + u_i$$

$$\text{Modelo cuadrático: } Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + u_i$$

Utilizando los datos del DATOS1.dta, se le solicita:

- Estime los tres modelos por MCO. Discuta los resultados obtenidos en términos de la significación individual y conjunta del modelo, así como su ajuste.
- ¿Cuál es la variación esperada de Y_i ante un cambio unitario en X_i ? Interprete bajo cada una de las formulaciones propuestas y compare los resultados obtenidos.
- Comparando las especificaciones lineal y cuadrática, ¿cuál considera más apropiada? Para ello compare los valores del criterio de información de Akaike (AIC).

EJERCICIO Nº 6:

Aplicando una transformación doble logarítmica al siguiente modelo $Y_i = AL_i^{\beta_1}K_i^{\beta_2}u_i$ a fin de obtener linealidad en los parámetros:

- Estime una función de producción Cobb-Douglas.
- Interprete los coeficientes estimados.
- ¿Podría haber estimado el modelo por MCO sin transformar las variables? Explique.

Producción bruta en tn	85	96	56	23	87	80	62	45	40	75
Consumo eléctrico en kw	200000	225000	135000	85000	200000	190000	160000	125000	115000	175000
Horas de trabajo	2500	2700	1800	500	2600	2250	2000	1600	1400	2100

EJERCICIO Nº 7:

En una aplicación de la función de producción Cobb-Douglas se obtuvieron los siguientes resultados:

$$\ln(\hat{Y}_i) = 2,3542 + \frac{0,9576}{(0,3022)} \ln(X_{1i}) + \frac{0,8242}{(0,3571)} \ln(X_{2i})$$
$$R^2 = 0,8432 \quad gl = 12$$

Donde: (Y) representa la producción, (X_{1i}) representa los insumos requeridos de trabajo y (X_{2i}) representa los insumos requeridos de capital. Los coeficientes de los insumos de trabajo y capital de la ecuación anterior suministran las elasticidades producto con respecto al trabajo y al capital, respectivamente. Se le solicita probar la hipótesis nula que establece que las elasticidades son individualmente iguales a 1.

EJERCICIO Nº 8:

Sea la siguiente función de producción Cobb-Douglas $Q_i = AX_{1i}^{\alpha_1}X_{2i}^{\alpha_2}X_{3i}^{\alpha_3}u_i$, en la que Q es la producción de la i-ésima empresa de electricidad de la muestra, X_1 es el consumo de factor trabajo, X_2 es el consumo de factor capital, X_3 es el consumo de combustible y A_i es un parámetro que recoge otras fuentes de diferencias (no observables) entre la eficiencia productiva de las empresas. El grado de los rendimientos a escala (r) viene dado por la suma de los exponentes. Utilizando los datos del archivo *PRODUCCION1.dta*, se le solicita:

- Estime la forma linealizada del modelo y comente los resultados (significación individual de los parámetros, significación conjunta del modelo y bondad del ajuste).
- Interprete los estimadores obtenidos.
- Obtenga una estimación de los rendimientos a escala. Dé su definición económica.

EJERCICIO Nº 9:

El modelo de la gravedad del comercio mundial establece una analogía con la ley de la gravedad de Newton: igual que la atracción de la gravedad entre dos objetos es proporcional al producto de sus masas y disminuye con la distancia, el comercio entre dos países cualquiera es, *ceteris paribus*, proporcional al producto de sus PBI y disminuye con la distancia (geográfica, cultural, comercial, etc.). La ecuación de gravedad a estimar propone que:

$$T_{ij} = \frac{AY_{ai}Y_{bj}}{D_i}$$

Responda:

- ¿Es el modelo intrínsecamente lineal?

- b) Proponga la ecuación a ser estimada por mínimos cuadrados ordinarios.
- c) Proponga un modelo a estimar que operacionalice el modelo de gravedad.
- d) Krugman (2006: 18) señala que “Las estimaciones habituales afirman que un incremento de la distancia de un 1% entre dos países se asocia con una disminución de entre el 0,7% y el 1% del comercio entre dichos países”. A partir de esa afirmación, indique a qué coeficiente del modelo se está refiriendo y cuál sería su valor.

TRABAJO PRÁCTICO Nº 7

ANÁLISIS DE REGRESIÓN MÚLTIPLE: PREDICCIÓN

Bibliografía:

- Wooldridge (2005), *Introducción a la Econometría. Un enfoque moderno*. 2da. ed, Thomson. Cap. 6

ACTIVIDAD Nº 1:

Responda si las siguientes afirmaciones son verdaderas o falsas:

- a) Las predicciones están sujetas a variaciones de muestreo.
- b) El intervalo de confianza es mayor para una predicción individual que para una predicción promedio.
- c) El intervalo de confianza de la predicción es más estrecho en el rango de X cercano a la media de las X (en comparación con una predicción para valores de X alejados de su media).

EJERCICIO Nº 1:

A partir de la siguiente serie de corte transversal, se le solicita:

- a) Estime el modelo y efectúe una predicción promedio (puntual) para el caso en que $X_0 = 16.000$, reformulando el modelo tal que $X' = X - 16000$. Obtenga el intervalo de confianza a partir de la estimación del intercepto.
- b) Realice predicciones (puntuales) para los siguientes niveles de ingreso: 8.000; 11.400; 12.800; 13.600; 14.200; 16.000 y 20.000. Calcule los intervalos de confianza para las predicciones individuales y promedio. Grafique las bandas de confianza.
- c) ¿Dónde es más estrecho el intervalo de confianza? Explique.

Ahorro	3200	3500	4200	3900	4500	4400	5000	4700
Ingreso	10800	12300	12500	13200	14000	15200	15600	15500

EJERCICIO Nº 2:

A partir de los siguientes datos:

Y_i	64	71	53	67	55	58	77	57	56	51	76	68	70	63	62	60
X_{1i}	8	11	6	11	8	7	10	9	10	7	12	9	10	9	9	8
X_{2i}	4	9	2	1	9	7	4	7	2	8	3	4	2	9	1	8

Efectúe la predicción promedio para $X_1 = 12$ y $X_2 = 4$. Recuerde que la varianza del pronóstico se calcula como $\hat{\sigma}_F^2 = \hat{\sigma}^2(X_0'(X_0'X)^{-1}X_0)$. Compare el resultado anterior con la predicción individual, en la cual $\hat{\sigma}_F^2 = \hat{\sigma}^2(1 + X_0'(X_0'X)^{-1}X_0)$, siendo $\hat{\sigma}^2 = 28,85$.

EJERCICIO Nº 3:

La variable Y es función de tres variables explicativas que se relacionan de la siguiente manera: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$. Utilizando los datos contenidos en la tabla:

- a) Estime el modelo.
- b) Prediga Y cuando $X_1 = 1$; y $X_2 = -3$ y $X_3 = -1$. Compare con el valor observado, ¿por qué no son iguales?
- c) ¿Presentan los datos evidencia suficiente para indicar que X_3 proporciona información para la explicación de Y?
- d) Encuentre un intervalo de confianza del 95% para el valor esperado de Y dados $X_1 = 1$; y $X_2 = -3$ y $X_3 = -1$ (predicción promedio).

- e) Encuentre un intervalo de predicción del 95% para Y usando los valores del inciso anterior (predicción individual).
- f) Compare ambos intervalos de confianza y concluya.

Y_i	1	0	0	1	2	3	3
X_{1i}	-3	-2	-1	0	1	2	3
X_{2i}	5	0	-3	-4	-3	0	5
X_{3i}	-1	1	1	0	-1	1	1

EJERCICIO N° 4: (Ejercicio 6.5 Wooldridge)

Con los datos del archivo *HPRICE1.dta*, se le solicita:

- a) Estime el modelo:
- $$\log(\text{price}) = \beta_0 + \beta_1 \log(\text{lotsize}) + \beta_2 \log(\text{sqrft}) + \beta_3 \text{bdrms} + u$$
- Reporte los resultados en la forma habitual de MCO.
- b) Determine el valor que se predice para el $\log(\text{price})$, cuando $\text{lotsize} = 20.000$, $\text{sqrft} = 2.500$ y $\text{bdrms} = 4$. Empleando los métodos de la sección 6.4, halle el valor que se predice para price a estos mismos valores de las variables explicativas.
- c) Si se trata de explicar la variación de price , diga si prefiere el modelo del inciso a) o el modelo:
- $$\text{price} = \beta_0 + \beta_1 \text{lotsize} + \beta_2 \text{sqrft} + \beta_3 \text{bdrms} + u$$

TRABAJO PRÁCTICO Nº 8

VARIABLES FICTICIAS

Bibliografía:

- Wooldridge (2005), *Introducción a la Econometría. Un enfoque moderno*. 2da. ed, Thomson. Cap. 7

ACTIVIDAD Nº 1:

Responda si las siguientes afirmaciones son verdaderas o falsas:

- a) Solamente las variables explicativas pueden ser ficticias.
- b) Si existen n categorías, el modelo puede incluir n variables ficticias.
- c) El analista elige qué categoría tomar como base de comparación (*baseline*).
- d) Las variables ficticias pueden representar variables categóricas o variables cuantitativas categorizadas.

ACTIVIDAD Nº 2:

- a) Explique el significado de las esperanzas condicionales cuando el modelo incorpora variables ficticias.
- b) Defina categoría base.

PROBLEMAS:

Resuelva los problemas 7.3), 7.4) y 7.5) del libro de Wooldridge.

EJERCICIO Nº 1:

El archivo *Profesores.dta* contiene información sobre el salario de un grupo de profesores universitarios clasificados según su género (Hombre y Mujer), su rango (Profesor, JTP y Ayudante), el tipo de universidad (A: pública; B: privada) y su antigüedad.

- a) Compare el salario medio de los profesores según género.
- b) Estime el siguiente modelo teórico $Y_i = \beta_0 + \delta_0 D_{1i} + u_i$ tomando como variable binaria al género. Interprete los coeficientes de regresión y calcule las esperanzas condicionales.
- c) Realice el contraste t para determinar la significatividad de la variable cualitativa.
- d) Compare los resultados obtenidos con los de un modelo sin constante que incluya a todas las categorías.
- e) Compare los resultados obtenidos con un modelo cuya categoría base sea "Mujer".

EJERCICIO Nº 2:

Con los datos del archivo *Profesores.dta*, estime un modelo que permita evaluar el impacto del rango o categoría del profesor sobre su salario: $Y_i = \beta_0 + \delta_0 D_{1i} + \delta_1 D_{2i} + u_i$.

- a) Halle los coeficientes de regresión e interprételos. ¿Cuál es la categoría base?
- b) Calcule las esperanzas condicionales considerando que se trata de una variable cualitativa con más de dos categorías.
- c) Calcule las esperanzas condicionales con otro ejemplo en el que se trate de dos variables cualitativas. Proponga un ejemplo.

EJERCICIO Nº 3:

Se desea estimar el modelo $Y_i = \beta_0 + \beta_1 X_i + \delta_0 D_i + \delta_1 X_i D_i + u_i$. A partir de las siguientes observaciones:

- a) Halle los coeficientes de regresión e interprételos.
- b) Calcule las esperanzas condicionales e interprételas estadísticamente. Grafique.

Costo total	98	100	103	105	80	87	94	113	116	118	121	123	126	128
Cantidad producida	0,79	0,8	0,82	0,82	0,93	0,95	0,96	0,88	0,88	0,9	0,93	0,94	0,96	0,97

Empresa Estatal	0	0	0	0	1	1	1	0	0	0	0	0	0	0
$X_i * D_i$														

EJERCICIO Nº 4:

A usted lo contratan para estimar un modelo que detecte discriminación por sexo y por procedencia étnica en el mercado de trabajo. El modelo debería demostrar si el grado de discriminación sexual disminuye a medida que los años de experiencia aumentan. ¿Qué resultados deben arrojar las estimaciones para no rechazar esta hipótesis?

EJERCICIO Nº 5:

En una regresión la variable ficticia tiene asignado el par (1, 0), en ese orden, y arroja como resultados los coeficientes (α, β) ambos positivos, indicando la influencia de la discriminación racial en la selección de puestos de trabajo (α) y además que el grado de discriminación aumenta con los años de experiencia (β). Si ahora se corre el mismo modelo pero se invierte el orden de los valores de la variable ficticia (0, 1):

- ¿Qué signo se espera que tengan los coeficientes?
- ¿Cómo se interpretan?

EJERCICIO Nº 6:

Se desea estimar una regresión por tramos para una empresa comercial que paga comisiones por ventas. La empresa tiene dos estructuras de comisiones según las ventas sean menores o mayores a \$5.500. Dado el modelo $Y_i = \beta_0 + \beta_1 X_i + \delta_1 (X_i - X^*) D_i + u_i$, donde:

Y_i = comisiones

X_i = facturación

$X^* = 5500$

$Z_i = X_i - X^*$

$D_i = 1$ si $X_i > X^*$

$D_i = 0$ si $X_i \leq X^*$

$W_i = Z_i D_i$

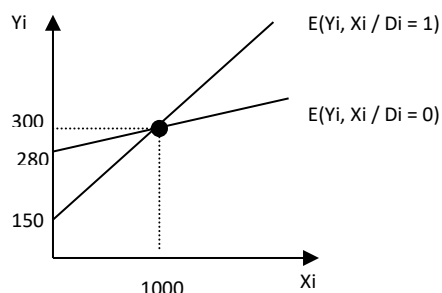
- Halle los coeficientes de regresión e interpréte los.
- Calcule las esperanzas condicionales.
- Grafique sin escalas.

Comisiones	256	414	634	778	1003	1839	2081	2423	2734	2914
Facturación	1000	2000	3000	4000	5000	6000	7000	8000	9000	10000
Z_i										
W_i										

EJERCICIO Nº 7:

Sean: Y_i = ventas diarias; X_i = gastos en publicidad; $D_i = 1$ el negocio está localizado adentro en un shopping y $D_i = 0$ si el negocio está afuera. Teniendo en cuenta los valores del gráfico, se le solicita:

- Deje planteado el modelo estimado (calcule el valor de los coeficientes estimados desde el gráfico).
- Interprete la esperanza condicional correspondiente a $D_i = 1$.



EJERCICIO Nº 8:

En base a los datos del archivo *WAGE2.dta* se estimó un modelo para explicar el salario de los trabajadores de la industria textil de EE.UU. en base a su nivel de instrucción formal, los años de experiencia, antigüedad y su edad (modelo A). Dado que se desea contrastar si el mismo describe por igual el salario de los trabajadores casados y solteros, se estimaron dos modelos adicionales: uno para casados (modelo B) y otro para solteros (modelo C). Los resultados del análisis ANOVA de estos modelos se presentan a continuación.

	Modelo A		Modelo B		Modelo C	
	gl	Suma de Cuadrados	gl	Suma de Cuadrados	gl	Suma de Cuadrados
educ	1	16340645	1	877377	1	16547080
exper	1	4406379	1	86	1	4142020
tenure	1	1531171	1	674457	1	974591
age	1	651149	1	24194	1	636677
Residuals	930	129786826	95	10085373	830	115905419

- En base a estos datos, realice el test de Chow sobre el modelo A y concluya.
- ¿Qué limitación puede tener el test de Chow en este caso?
- Alternativamente, formule el modelo teórico a estimar para probar la misma hipótesis nula utilizando la variable binaria *married*. Enuncie la hipótesis nula, construya el estadístico de contraste y concluya.
- Estime los siguientes modelos y elija el que considere más apropiado:
 - mod1: $wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + \beta_4 age$
 - mod2: $wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + \beta_4 age + \beta_5 married$
 - mod3: $wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + \beta_4 age + \beta_5 married + \beta_6 educ * married + \beta_7 exper * married + \beta_8 tenure * married + \beta_9 age * married$
 - mod1_ln: $\ln(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + \beta_4 age$
 - mod2: $\ln(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + \beta_4 age + \beta_5 married$
 - mod3: $\ln(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + \beta_4 age + \beta_5 married + \beta_6 educ * married + \beta_7 exper * married + \beta_8 tenure * married + \beta_9 age * married$

EJERCICIO Nº 9:

El archivo *IMPACTO.dta* contiene información recabada con posterioridad a un curso de capacitación del Ministerio de Trabajo. La base del ministerio contiene información sobre el nivel de educación y los años de experiencia de un grupo de participantes y no participantes, así como el salario que perciben a partir del 6º mes de finalizado el curso.

- Formule un modelo que permita evaluar el impacto de la participación en dicho curso en el salario.
- Estime el modelo por MCO y mencione cuál es el impacto promedio de la participación en el salario. ¿En qué casos podría sospechar que existe un sesgo?
- Estime el modelo utilizando como variable a explicar el logaritmo del salario. Interprete los coeficientes y el impacto del programa.
- Si existen capacidades no observadas en los trabajadores (habilidad, proactividad, rapidez, etc.) y usted estima esta ecuación por MCO. ¿Qué puede decir del posible sesgo en el coeficiente que mide el impacto del programa en los siguientes casos?
 - Caso 1: los trabajadores con menos capacidades no observadas tienen más probabilidades de ser elegidos para participar en el programa.
 - Caso 2: la participación en el programa es abierta, de manera que los trabajadores con mayores capacidades no observadas participan en mayor medida.
 - Caso 3: la participación es resultado de un sorteo y todas las personas sorteadas toman el curso.

TRABAJO PRÁCTICO Nº 9

HETEROCEDASTICIDAD

Bibliografía:

- Wooldridge (2005), *Introducción a la Econometría. Un enfoque moderno*. 2da. ed, Thomson. Cap. 8

ACTIVIDAD Nº 1:

Responda si las siguientes afirmaciones son verdaderas o falsas:

- Los EMC resultan sesgados ante la existencia de heterocedasticidad.
- Los EMC son consistentes ante la existencia de heterocedasticidad.
- El método de mínimos cuadrados ponderados permite obtener estimadores eficientes pero requiere más información que la estimación de errores robustos.

ACTIVIDAD Nº 2:

- ¿Cuáles son las consecuencias de la heterocedasticidad?
- ¿Qué patrón debe exhibir el gráfico de residuos contra el valor predicho para sospechar la existencia de heterocedasticidad?
- Proponga ejemplos de modelos a estimar donde sospeche que la existencia de heterocedasticidad.

PROBLEMAS:

Resuelva el problema 8.2) del libro de Wooldridge y el ejercicio en computadora 8.1).

EJERCICIO Nº 1: Detección de heterocedasticidad

El archivo VOTE1.dta contiene datos de 173 distritos electorales. A partir de ellos es posible estudiar si los gastos de campaña afectan los resultados de las elecciones a partir del modelo:

$$votosA = \beta_0 + \beta_1 \ln(gastosA) + \beta_2 \ln(gastosB) + \beta_3 votospartA + \beta_4 joven + u$$

donde *votosA* es el porcentaje de votos recibidos por el candidato A, *gastosA* y *gastosB* son los gastos de campaña del candidato A y del candidato B, *votospartA* es una medida de la fortaleza del partido del candidato A (el porcentaje de votos que obtuvo el partido A en la elección presidencial más reciente) y *joven* es una variable dummy que es igual a 1 si ese distrito posee una proporción de votante jóvenes superior a la media nacional.

- Regrese los residuos de una regresión por MCO sobre todas las variables independientes y explique por qué el R^2 es igual a 0 en este caso.
- Grafique los residuos de una regresión por MCO contra los valores predichos. ¿Qué se observa en este gráfico?
- Efectúe el contraste de Breusch-Pagan.
- Realice el caso especial del contraste de White.
- Efectúe el contraste de White sólo con cuadrados. Concluya: ¿Qué tan fuerte es la evidencia de heterocedasticidad?
- Estime la regresión original con errores robustos a heterocedasticidad. ¿Existen grandes diferencias con respecto a los errores estándar originales? Compare la significatividad individual de las variables en ambos casos.
- Verifique que los errores robustos pueden calcularse en Excel a partir de la fórmula:

$$\widehat{Var}(\beta_j) = \frac{\sum r_{ij}^2 \hat{u}_i^2}{SRC_j^2}$$

- ¿Vale la pena calcular los errores robustos a heterocedasticidad en este caso? Explique.

EJERCICIO Nº 2: Heterocedasticidad conocida (MCP)

- A partir de los siguientes datos, grafique los residuos al cuadrado contra los valores de X . ¿Qué relación se observa entre ambos? ¿qué supuesto acerca del comportamiento de la varianza propondría para corregir la heterocedasticidad en este caso?
- Estime los coeficientes de regresión por el método de MCP, dado el supuesto: $\sigma_i^2 = \sigma^2 X_i^2$.
- Los estimadores obtenidos, ¿son insesgados? ¿son eficientes?
- ¿Cómo se interpreta el R^2 en este caso?
- ¿Qué sucede si la función de varianza está mal especificada?
- Considere que la función de varianza estimada pueda estar mal especificada y calcule los errores estándar robustos a los estimadores de MCP del inciso b). ¿Varían mucho los errores estándar?

Fallas registradas	Horas de uso
160	2000
160	2000
180	2000
200	2000
200	4000
210	2000
220	2000
220	4000
230	2000
230	4000
250	2000
300	4000
300	6000
300	6000
310	4000
340	4000
350	4000
400	6000
450	6000
450	6000

EJERCICIO Nº 3: (Ejercicio 8.1 Wooldridge)

Considere el modelo siguiente para explicar el comportamiento del sueño:

$$\text{sleep} = \beta_0 + \beta_1 \text{totwrk} + \beta_2 \text{educ} + \beta_3 \text{age} + \beta_4 \text{age}^2 + \beta_5 \text{yngkid} + \beta_6 \text{male} + u$$

- Plantee un modelo que permita que la varianza de u difiera entre hombres (male) y mujeres. La varianza no debe depender de otros factores.
- Emplee datos del archivo SLEEP75.dta para estimar los parámetros de la ecuación de heterocedasticidad. (Tiene que estimar primero la ecuación para sleep mediante MCO para obtener los residuos). ¿Es la varianza estimada de u mayor para los hombres o para las mujeres?
- ¿Es la varianza de u diferente estadísticamente para hombres y para mujeres?

EJERCICIO Nº 4:

El análisis gráfico de los residuos correspondientes al modelo: $y_i = \beta_0 + \beta_1 x_i + u_i$, indica que la varianza se incrementa con la inversa de x_i , siendo $\hat{y}_i \sqrt{x_i} = 0,121\sqrt{x_i} + 3,803(x_i \sqrt{x_i})$ la ecuación estimada. Se le solicita:

- Grafique el posible patrón exhibido por los residuos.
- Indique el supuesto utilizado para corregir la heterocedasticidad.
- Deje planteado el modelo en variables originales.

EJERCICIO N° 5: Estimación de la función de heterocedasticidad (MCGF)

Con los datos del archivo WAGE1.dta, estime un modelo que explique la remuneración por hora de los empleados (*wage*) a partir de sus años de instrucción (*educ*), los años de experiencia (*exper*) y los años de antigüedad en el empleo actual (*tenure*).

- ¿Es de esperar que exista heterocedasticidad en este modelo? ¿Por qué?
- Realice las pruebas que considere necesarias para evaluar la existencia de heterocedasticidad.
- Estime el modelo tomando como variable dependiente el logaritmo del salario. ¿Qué conclusiones puede obtener?
- Aplique el método de MCG factibles para corregir la heterocedasticidad (utilice la formulación original de la variable dependiente). Estime el ponderador a utilizar en base al siguiente supuesto acerca del comportamiento de la varianza: $Var(u|x) = \sigma^2 \exp(\delta_0 + \delta_1 x_1 + \delta_2 x_2 + \dots + \delta_k x_k)$.
- Los estimadores MCGF obtenidos, ¿son insesgados? ¿son consistentes? ¿son eficientes?

EJERCICIO N°6:

- Proponga distintas especificaciones del modelo, incorporando términos cuadráticos y/o interacciones. Compruebe la presencia de heteroscedasticidad y estime con errores estándar válidos.
- Repita el ejercicio anterior transformando en logaritmos a las variables expresadas en niveles. Compruebe la presencia de heteroscedasticidad y estime con errores estándar válidos.
- Presente una especificación que Ud considere adecuada para modelar un rubro de gastos en particular, tal que los errores estándar estimados sean válidos.

EJERCICIO N° 7: (EJERCICIO PARA COMPUTADORA 8.8)

Utilizando los datos del archivo GPA1.dta:

- Estime por MCO un modelo que relacione el promedio de calificaciones en la universidad del estudiante i ($colGPA_i$) con el promedio de calificaciones de ese estudiante en la escuela secundaria ($hsGPA_i$), calificación en el examen de ingreso (ACT_i), la cantidad de clases a las que faltó ($skipped_i$) y una variable dummy que indica si el estudiante posee computadora personal (PC_i).
- Calcule el caso especial de la prueba de White para heterocedasticidad. En la regresión de \hat{u}_i^2 sobre $colGPA_i$ y $colGPA_i^2$, obtenga los valores ajustados, es decir, \hat{h}_i .
- Verifique que los valores ajustados del inciso b) sean todos estrictamente positivos. Después obtenga las estimaciones de MCP usando como ponderadores $1/\hat{h}_i$. Compare las estimaciones de MCP con las correspondientes estimaciones de MCO para el efecto de número de clases perdidas y el efecto de poseer una computadora. ¿Qué puede decir de su significancia estadística?
- En la estimación de MCP del inciso c), obtenga los errores estándar robustos a heterocedasticidad. En otras palabras, considere la posibilidad de que la función de varianza estimada en el inciso ii) pueda estar mal especificada. ¿Varían mucho los errores estándar con respecto a los del inciso iii)?

TRABAJO PRÁCTICO Nº 10

ESTIMACIÓN CON VARIABLES PROXY

Bibliografía:

- Wooldridge (2005), *Introducción a la Econometría. Un enfoque moderno*. 2da. ed, Thomson. Cap. 9.

ACTIVIDAD Nº 1:

Responda si las siguientes afirmaciones son verdaderas o falsas:

- Una variable binaria no puede utilizarse como variable proxy.
- En fenómenos con un importante componente inercial no es conveniente incluir la variable dependiente rezagada como variable explicativa.
- Variable proxy es un método de estimación distinto a MCO.

ACTIVIDAD Nº 2:

- ¿Qué es una variable proxy?
- ¿Qué requisitos debe cumplir una variable proxy para proporcionar estimadores consistentes? ¿Qué sucede si no los cumple?
- Busque un ejemplo de variable proxy en un paper aplicado de econometría. Comente.

PROBLEMAS:

Resuelva el problema 9.2) del libro de Wooldridge.

EJERCICIO Nº 1:

A partir de los datos del archivo *Impacto.dta*, se desea estimar el impacto de un curso de capacitación del Ministerio de Trabajo, a través del siguiente modelo:

$$\text{salario} = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{partic} + u$$

Donde *SALARIO* es el monto salarial percibido a partir del 6º mes de finalizado el curso, *EDUC* los años de instrucción del individuo, *EXPER* los años de experiencia laboral y *PARTIC* una variable binaria que indica si participó o no del curso.

- Estime el modelo por MCO y mencione cuál es el impacto promedio de la participación en el curso en el salario. ¿En qué casos podría sospechar que existe un sesgo?
- Proponga un modelo que incluya la variable no observada *CAPACIDAD*. ¿Es posible estimarlo? ¿Qué variables proxy propone para controlar el efecto de la capacidad individual sobre el salario?
- Estime el modelo inicial incluyendo como variable proxy el coeficiente intelectual (*IQ*). Compare los resultados con el modelo inicial y concluya.
- ¿El estimador de *PARTICIP* es consistente? ¿y el de *CAPACIDAD*?

EJERCICIO Nº 2:

Se desea evaluar el impacto del desempleo (*DESEMP*) y del gasto en seguridad (*GASTOSEG*) sobre la cantidad de delitos cada 1000 habitantes (*DELITO*) en 46 ciudades argentinas. Para ello se dispone de los datos del año 2005 en el archivo *DELITOS.dta*. Se pide:

- Estime el modelo $\log(\text{delito}) = \beta_0 + \beta_1 \text{desemp} + \beta_2 \log(\text{gastoseg}) + u$. Interprete los resultados.
- Incorpore como variable independiente la tasa de delincuencia del año 2003 a fin de controlar los factores no observables urbanos que afectan la delincuencia y que pueden estar correlacionados con el gasto corriente en seguridad pública.
- Compare ambos modelos y extraiga conclusiones.

- d) Pruebe la hipótesis nula de que el parámetro de $\log(\text{DELITO_03})$ es igual a uno contra la alternativa de dos colas. Concluya.
- e) Repita los incisos anteriores empleando los errores estándar robustos a heterocedasticidad y analice brevemente cualquier diferencia notable.

TRABAJO PRÁCTICO Nº 11

VARIABLE INSTRUMENTAL Y MÍNIMOS CUADRADOS EN 2 ETAPAS

Bibliografía:

- Wooldridge (2005), *Introducción a la Econometría. Un enfoque moderno*. 2da. ed, Thomson. Cap. 9 y 15.

ACTIVIDAD Nº 1:

Responda si las siguientes afirmaciones son verdaderas o falsas:

- Cuando existen errores de medición en la variable a explicar los Estimadores de MCO son sesgados pero consistentes.
- Es posible comprobar empíricamente la condición de exogeneidad del instrumento, pero no la de relevancia.

ACTIVIDAD Nº 2:

- Enumere causas posibles para la existencia de endogeneidad.
- ¿Cuál es la consecuencia sobre las propiedades de los EMC de la presencia de endogeneidad?
- Explique el método de Mínimos Cuadrados en 2 Etapas.

PROBLEMAS:

Resuelva los problemas 15.1) 15.5) 15.7) 15.8) y 15.10) del libro de Wooldridge.

EJERCICIO Nº 1: Regresión simple

A partir de los datos del archivo Variable Instrumental.dta:

- Estime el siguiente modelo por MCO: $\log(\text{salario}) = \beta_0 + \beta_1 \text{educ} + u$.
- ¿Los estimadores son consistentes? ¿Es posible sospechar la existencia de endogeneidad?
- Estime el modelo con el método de variable instrumental, utilizando como instrumento los años de educación de la madre. Obtenga el error estándar para el coeficiente de *EDUC*, evalúe la significatividad estadística del parámetro y calcule el intervalo de confianza correspondiente.
- Estime el modelo con el método de MC2E, de manera manual (estimando ambas etapas por MCO). Compare las estimaciones con las obtenidas en el punto c).
- Estime el modelo con el método de MC2E, utilizando Stata. Compare las estimaciones con las obtenidas en los puntos c) y d).
- ¿Qué sucede si usted cuenta con más de una variable instrumental? ¿Qué método utilizaría y qué resultados obtiene?
- ¿Detecta evidencia de endogeneidad en este caso particular? Realice el test de Hausman y explique.

EJERCICIO Nº 2: Regresión múltiple

Repita el análisis realizado en el ejercicio anterior para el siguiente modelo:

$$\log(\text{salario}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + u$$

Recuerde que los estimadores de VI pueden obtenerse con la fórmula: $\hat{\beta}^{IV} = (Z'X)^{-1}(Z'Y)$ y que la matriz de varianzas y covarianzas se calcula como: $\hat{\sigma}^2(X'Z(Z'Z)^{-1}Z'X)^{-1}$.

EJERCICIO Nº 3:

Card (1995) (CARD.dta) estimó el impacto de educación sobre el salario de un grupo de hombres en 1976 a partir del siguiente modelo teórico:

$$\log(\text{salario}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{exper}^2 + \beta_4 \text{raza} + \beta_5 \text{capital} + \beta_6 \text{sur} + u$$

donde *RAZA* es una variable binaria para raza negra, *CAPITAL* una variable binaria por vivir en un área metropolitana y *SUR* una variable binaria por vivir en el sur. También se incluyen un conjunto completo de variables binarias regionales y una variable binaria de área metropolitana para el lugar donde el hombre vivía en 1966.

- ¿Por qué la variable *EDUC* podría ser endógena en este caso?
- Estime la ecuación por MCO y explique cuáles son las consecuencias de la heterogeneidad inobservable sobre los EMCO.
- Card utilizó como variable instrumental para *EDUC* una variable binaria que era igual a 1 si la persona había crecido cerca de una universidad con carreras de 4 años (*CERCANÍA4*). Verifique el supuesto de relevancia del instrumento.
- Argumente sobre el cumplimiento del supuesto de exogeneidad del instrumento. Realice el Ejercicio para computadora 15.3 de Wooldridge para profundizar sobre este punto.
- Estime la ecuación por MC2E utilizando el instrumento propuesto por Card y compare los coeficientes estimados con los EMC del punto a). Interprete el coeficiente asociado a *EDUC* y realice una estimación por intervalos para ese parámetro con ambos métodos de estimación.
- Obtenga los residuales de la forma reducida y utilícelos para probar si *EDUC* es exógena. Es decir, determine si la diferencia entre MCO y VI es estadísticamente significativa.

EJERCICIO Nº 4: Errores en las variables bajo el supuesto de errores clásico (ECV)

Los datos contenidos en la tabla corresponden a una explotación de tomates. Dado que existen sospechas de que se han cometido errores en la registración del consumo de fertilizantes, se propone utilizar como variable instrumental el fertilizante comprado, sin tener en cuenta los remanentes no utilizados. La elección se debe a que es muy probable que dicha variable se encuentre altamente correlacionada con los litros de fertilizante utilizados y no correlacionada con los errores del modelo ni con el error de medición.

- Estimar la ecuación por mínimos cuadrados. ¿Los estimadores son consistentes?
- Construir la variable instrumental y verificar el cumplimiento del supuesto de relevancia.
- Re-estimar la ecuación con el método de variable instrumental.
- Comparar las dos estimaciones y comentar brevemente los resultados.
- ¿Cuáles hubieran sido las consecuencias si el error de medición estuviera en las toneladas de tomates cosechadas?

Tn de tomates cosechadas	Litros de fertilizante utilizados	Litros de fertilizante adquiridos
60,0	35,1	38,7
63,4	37,3	41,3
68,2	41,0	45,6
78,0	44,9	50,1
84,7	46,5	51,9
90,6	50,3	56,3
98,2	53,5	60,1
101,7	52,8	52,9
102,7	55,9	62,8
108,3	63,0	71,0
124,7	73,0	82,7
157,9	84,8	96,4
158,2	86,6	98,5
170,2	98,8	112,6
180,0	110,8	126,5
198,0	124,7	142,7

EJERCICIO Nº 5:

Los datos del archivo *Vinos.dta* corresponden a 75 bodegas de California. Con ellas se quiere estimar la siguiente función de producción:

$$Q = \beta_1 + \beta_2 M + \beta_3 K + \beta_4 L + e$$

Donde Q es un la cantidad de producción mensual (en toneladas), M es una variable que refleja la eficiencia del management, K es una medida del insumo capital y L del trabajo.

- Teniendo en cuenta que no existe una medida fiable de la eficiencia de la gerencia, estime por MCO la siguiente función de producción simplificada: $Q = \beta_0 + \beta_1 K + \beta_2 L + u$. Realice los contrastes necesarios para detectar heterocedasticidad.
- ¿Cuáles son las consecuencias sobre los estimadores MCO de la imposibilidad de medir la eficiencia de la gerencia?
- Utilice los años de experiencia (*Exper*) de cada gerente general como variable proxy. Interprete los coeficientes estimados.

Suponga ahora que el modelo a estimar es:

$$Q = \beta_0 + \beta_1 K + \beta_2 L + \beta_3 Exper + u$$

- Se cree que la experiencia puede estar correlacionada con el término de error. Reestime la ecuación con el método de mínimos cuadrados en dos etapas empleando la variable edad (*age*) como instrumento. Deje expresada la ecuación estimada, los errores estándar de cada coeficiente e interprete el coeficiente estimado correspondiente a β_3 .
- Realice el test de Hausman y concluya.
- Obtenga las predicciones, con sus respectivos IC, para bodegas que tengan el promedio de L y K, y que tengan manager con:
 - 10 años de experiencia
 - 20 años de experiencia
 - 30 años de experiencia.

EJERCICIO Nº 6: (Ejercicio 15.3 Wooldridge)

La ecuación que se estimó en el ejemplo 15.4 del libro se puede escribir como:

$$\ln(\text{SALARIO}) = \beta_0 + \beta_1 EDUC + \beta_2 EXPER + \dots + u$$

- Con el fin de que VI sean consistentes, la VI para educ, nearc4, no debe estar correlacionada con u. ¿Estaría nearc4 correlacionada con cuestiones en el término de error, como la capacidad inobservable? Explique.
- Para una submuestra de hombres en el archivo *Card2.dta* se dispone de puntuaciones de IQ. Realice una regresión de IQ sobre nearc4 para revisar si el promedio de las puntuaciones IQ varía si el hombre se crio cerca de una universidad con programas de cuatro años. ¿Qué se concluye?
- Realice una regresión de IQ sobre nearc4, smsa66 y las variables binarias regionales de 1966, reg662,..., reg669. ¿Están IQ y nearc4 relacionadas después de haber descontado el efecto parcial de las variables binarias geográficas? Compare esto con los resultados del inciso b).
- A partir de los incisos b) y c), qué concluye acerca de la importancia de controlar smsa66 y las variables binarias regionales de 1966 en la ecuación del log(wage)?

TRABAJO PRÁCTICO Nº 12

ESTABILIDAD DE LOS COEFICIENTES

Bibliografía:

- Wooldridge (2005), *Introducción a la Econometría. Un enfoque moderno*. 2da. ed, Thomson. Cap. 7

ACTIVIDAD Nº 1:

Responda si las siguientes afirmaciones son verdaderas o falsas:

- El contraste de Chow puede aplicarse si existe heterocedasticidad en el modelo.
- La prueba de Chow puede usarse con datos de corte transversal y datos de series temporales.

ACTIVIDAD Nº 2:

- Formule la H_0 de la prueba de Chow con datos de corte transversal y con datos de series de tiempo.
- Explique qué consecuencias implica que los coeficientes no sean estables.
- ¿Podrán ser estables los coeficientes de un modelo macroeconómico que se estime para Argentina con datos de 1990 a 2019? Justifique.

EJERCICIO 1:

El archivo *wage.dta* contiene datos de la economía estadounidense, donde *price* es el índice mensual de precios y *wages* es el índice mensual de salario por hora. Se le solicita:

- Transforme las variables *wages* y *price* en series de tiempo.
- Grafique las series en niveles y en primeras diferencias. ¿Qué representan los datos en niveles y los datos en diferencias? ¿Cómo se comportan?
- Estime el modelo $\text{diff}(Wages_t) = \beta_0 + \beta_1 \text{diff}(price_t) + u_t$.
- Aplique la prueba de Chow para evaluar si hay un cambio estructural a partir de $t = 150$. Concluya.

EJERCICIO Nº 2:

En un modelo de regresión simple con 160 observaciones, al correr la regresión con todos los datos, la suma de los errores al cuadrado fue igual a 5140,83. Cuando se corrió la regresión con las primeras 80 observaciones el valor hallado fue 1970,66 mientras que el correspondiente a la regresión con la segunda mitad de los datos fue 1680,71. Aplicando el contraste de Chow, responda: ¿son los coeficientes estables? (deje expresados todos los cálculos que justifiquen su respuesta).

EJERCICIO Nº 3:

Realice el contraste de Chow de estabilidad de los coeficientes – con la H_0 : Existe estabilidad de los coeficientes – para el siguiente caso, con la finalidad de evaluar si la publicidad en contra del consumo de alcohol, vigente desde el octavo período, impactó sobre las ventas de gaseosas.

Venta de gaseosas en mill. u\$s	Precio de las gaseosas	Índice de precio de las bebidas alcohólicas
6	9	100
8	10	108
8	8	110
7	7	75
7	10	80
12	5	120
9	5	98
8	5	85
9	6	96

10	8	95
10	7	110
11	4	120
9	9	89
10	5	102
11	6	114

EJERCICIO N° 4:

Se ha estimado la ecuación $\hat{Y}_t = 2,417 + 0,72X_t$ con datos trimestrales desde 1963 hasta 1972 inclusive, siendo el $R^2 = 0,68$ y la $SRC = 109,24$. Se calcularon separadamente dos regresiones más para los períodos del 1º trim. 1963 hasta el 3º trim. 1966 con $SRC = 33,04$ y del 4º trim. 1966 hasta el 4º trim. 1972 con $SCR = 68,19$. Se le solicita contrastar la proposición de que a partir del 4º trim. 1966 se ha producido un cambio estructural.

EJERCICIO N° 5:

El archivo EJERCICIO 5.dta contiene datos trimestrales de consumo e ingreso de un promedio de 15 países de Medio Oriente en el período 1980-1990. El modelo de regresión que se trata de estimar es el siguiente:

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 W_t + u_t$$

Donde: X_t representa la renta disponible y W_t representa una variable ficticia correspondiente al período de guerra entre Irán e Irak, que toma valor 1 para los años 1980-1987 y cero en los demás años.

- Efectuar el análisis de regresión del modelo planteado e interpretar los coeficientes de regresión.
- Calcular las esperanzas condicionales e interpretar los resultados obtenidos.
- ¿Puede considerarse que el fenómeno bélico haya afectado sensiblemente el consumo de estos países? De ser así, explique las razones desde la perspectiva de un economista.
- ¿Se ha producido un cambio estructural a partir de la finalización de la guerra?

EJERCICIO N° 6:

En base al modelo anterior, evaluar si el efecto del exceso de oferta de petróleo que contribuyó a deprimir los precios del crudo a inicio de los 80 ha afectado sensiblemente el consumo de estos países. Tal fenómeno se ha presentado en los primeros 4 años de la década analizada (es decir, desde 1980 a 1983 inclusive). En consecuencia el modelo que se propone evaluar consistiría en:

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 W_t + \beta_3 Z_t + u_t$$

Donde: X_t representa la renta disponible y W_t representa una variable ficticia correspondiente al período de guerra entre Irán e Irak, que toma valor 1 para los años 1980-1987 y cero en los demás años y Z_t representa la variable ficticia correspondiente al período de precios deprimidos del petróleo (1 desde 1980 a 1983 y 0 en el resto del período muestral).

- Estime el modelo teórico e interprete los coeficientes de regresión.
 - Calcule las esperanzas condicionales.
 - ¿Puede considerarse que el fenómeno bélico siga afectando sensiblemente el consumo de estos países? ¿Cómo afectan la introducción del fenómeno de exceso de oferta de petróleo al modelo original?
- Obtenga conclusiones al respecto.

TRABAJO PRÁCTICO Nº 13

TENDENCIA Y ESTACIONALIDAD EN SERIES TEMPORALES

Bibliografía:

- Wooldridge (2005), *Introducción a la Econometría. Un enfoque moderno*. 2da. ed, Thomson.
Cap. 10

ACTIVIDAD Nº 1:

Responda si las siguientes afirmaciones son verdaderas o falsas:

- a) Cuando se introduce la variable tiempo, se consigue corregir de tendencia sólo a la variable dependiente.
- b) El tiempo como variable explicativa también permite captar la estacionalidad de la serie.
- c) Sólo puede modelarse la tendencia lineal.

ACTIVIDAD Nº 2:

- a) Explique a qué se denomina regresión espuria.
- b) ¿Qué significa que una serie sea estacionaria?
- c) ¿Por qué es importante eliminar la tendencia y la estacionalidad de una serie antes de utilizarla en un modelo de regresión?

PROBLEMAS:

Resuelva los problemas 10.1), 10.4) Y 10.5) del libro de Wooldridge.

EJERCICIO Nº 1:

A partir del archivo EJERCICIO1.dta:

- a) Grafique las variables contra el tiempo.
- b) Estime un modelo que explique el gasto en bienes no durables (Cnd) en función al ingreso disponible (Yd) y a la población (Pob). Estime los coeficientes de regresión sin considerar y considerando a la variable tiempo.
- c) Extraiga la tendencia (lineal) de todas las variables y estime los coeficientes de regresión. Compare los resultados.
- d) Si sólo las variables explicativas tuvieran una tendencia (crecimiento o decrecimiento en el tiempo), ¿tiene sentido incluir la variable tiempo en la regresión? ¿Por qué?

EJERCICIO Nº 2:

- a) Genere dos secuencias de números aleatorios (n=25) y estime una función de regresión entre ambas. (Recuerde copiar y pegar sólo los valores de los números aleatorios para evitar que se modifiquen) ¿Esperaría que el modelo fuera globalmente significativo?
- b) Introduzca una columna con una tendencia del 25% acumulativo y aplique dicha tendencia a los datos generados anteriormente.
- c) Estime una ecuación de regresión entre ambas series una tendencia común. ¿Qué deduce de los resultados de la salida de regresión? ¿Es posible que la variable “independiente” explique a la “dependiente” en este contexto?
- d) Concluya sobre la importancia de extraer la tendencia a los datos.

n	Nº a	Nº b	Tendencia	a con tendencia	b con tendencia
1	aleatorio	aleatorio	1		
2	aleatorio	aleatorio	1,25		
3	aleatorio	aleatorio	1,56		
...					
25	aleatorio	aleatorio	211,76		

EJERCICIO N°3:

Los datos del archivo *dinero.dta* corresponden a variables del mercado de dinero de Dinamarca entre 1974 y 1987, en particular: la oferta real de dinero, el ingreso real y la tasa de interés. Responda:

- Mediante un análisis gráfico, analice si las variables presentan estacionalidad.
- Evalúe la existencia de estacionalidad mediante una prueba formal (análisis de significación individual y conjunta de variables ficticias trimestrales).
- Desestacionalice mediante variables ficticias trimestrales y estime un modelo que explique la oferta monetaria en función del ingreso real y la tasa de interés de los bonos del tesoro.
- Extraiga la estacionalidad de todas las variables y estime nuevamente los coeficientes de regresión. Compare ambos modelos.
- Calcule las esperanzas condicionales para el primer y el último trimestre del año. Interprete.

EJERCICIO N° 4:

A partir del archivo *Consumo.dta*:

- Grafique las variables contra el tiempo (series anuales, comienzan en 1959).
- Estime un modelo que explique el gasto en consumo en bienes no durables (Cnd) en función al ingreso disponible (Yd) y a la población (Pob). Estime los coeficientes de regresión sin considerar y considerando a la variable tiempo.
- Extraiga la tendencia (lineal) de todas las variables y estime los coeficientes de regresión. Compare los resultados.
- Realice el análisis anterior incorporando una tendencia cuadrática ¿Se observan cambios en los resultados?
- Si sólo las variables explicativas tuvieran una tendencia (crecimiento o decrecimiento en el tiempo), ¿tiene sentido incluir la variable tiempo en la regresión? ¿Por qué?

EJERCICIO N° 5:

Utilizando los datos del ejercicio anterior, veamos otro tratamiento a los datos. Si bien en esta instancia no aplicaremos pruebas estadísticas, nos iremos aproximando a la idea de estacionariedad. Para ello se le solicita:

- Grafique cada serie en niveles y en primeras diferencias. ¿En qué escala parecen ser las series estacionarias?
- Vuelva a estimar el modelo en primeras diferencias, donde ahora las variables representan variaciones respecto del período anterior.
- Analice la normalidad de los residuos del modelo.

EJERCICIO N° 6: (Ejercicio 10.2Wooldridge)

A partir de los datos *barium.dta*, que corresponden a los empleados por Krupp y Pollard (1996) para analizar los efectos de las demandas antidumping de las industrias químicas estadounidenses, estime la ecuación:

$$\log(\text{chnimp}) = \beta_0 + \beta_1 \log(\text{chempi}) + \beta_2 \log(\text{gas}) + \beta_3 \log(\text{rtwex}) + \beta_4 \text{befile6} + \beta_5 \text{affile6} + \beta_6 \text{afdec6} + u$$

Donde *chnimp* es el volumen de importación de cloruro de bario de China, *chempi* es un índice de producción química (que controla la demanda general de cloruro de bario), *gas* es el volumen de producción de combustible, *rtwex* es un índice del tipo de cambio. Además, se incorporan 3 variables dummies: *befile6* que vale 1 para los seis meses anteriores a la demanda, *affile6* indica seis meses posteriores a la demanda y *afdec6* indica los seis meses posteriores al fallo positivo.

- Agregue una tendencia lineal ¿qué sucede con las variables explicativas del modelo?
- Realice una prueba de significatividad conjunta de todas las variables, salvo la tendencia ¿qué conclusión obtiene de este resultado?
- Incorpore al modelo variables binarias mensuales y determine si hay estacionalidad.

EJERCICIO Nº 7: (Ejercicio 10.5 Wooldridge)

Emplee los datos *ezanders.dta*. Estos datos se refieren al desempleo mensual (*uclms*) del trabajo de Anderson Township en Indiana, de enero de 1980 a noviembre de 1988. En 1984, una zona empresarial se ubicó en Anderson.

- Realice la regresión de $\log(uclms)$ sobre una tendencia lineal y 11 variables binarias mensuales. Interprete el coeficiente de la tendencia. ¿Hay evidencia de estacionalidad en el desempleo?
- Agregue *ez*, una variable dummy igual a 1 que indica los meses en los que Anderson tuvo una zona industrial. La zona industrial disminuyó el desempleo, ¿en cuánto? Emplee la fórmula (7.10) del libro para obtener la variación exacta.

EJERCICIO Nº 8: (Ejercicio 10.6 Wooldridge)

Con los datos *fertil3.dta* estime la ecuación de fertilidad:

$$gfr = \beta_0 + \beta_1 pe + \beta_2 ww2 + \beta_3 pill + u$$

Donde *gfr* es la tasa de fertilidad (cantidad de niños nacidos por cada 1000 mujeres en edad de concebir); *pe* es el valor en dólares de la exención personal de impuestos por cada hijo; *ww2* es una variable dummy que indica el período de guerra (1941 a 1945) y *pill* es una variable dummy que es igual a 1 a partir del momento de introducción de la píldora anticonceptiva en EE.UU. (año 1963).

- Agregue una tendencia cuadrática, ¿resulta significativa?
- Extraiga la tendencia cuadrática a la variable *gfr* y realice la regresión de esta variable sobre las variables empleadas anteriormente (incluidas t y t^2) ¿Qué conclusión obtiene?
- Incorpore un término cúbico de tendencia (t^3) a la ecuación de fertilidad, ¿resulta significativo?

TRABAJO PRÁCTICO Nº 14

AUTOCORRELACIÓN

Bibliografía:

- Wooldridge (2005), *Introducción a la Econometría. Un enfoque moderno*. 2da. ed, Thomson. Cap. 12

ACTIVIDAD Nº 1:

Responda si las siguientes afirmaciones son verdaderas o falsas:

- a) El uso del estadístico d de Durbin-Watson requiere el cumplimiento de los supuestos del MLC.
- b) Si los errores están autocorrelacionados, los contrastes t y F no son válidos.

ACTIVIDAD Nº 2:

- a) ¿Cuáles son las consecuencias de la autocorrelación sobre los EMCO?
- b) Mencione las medidas gráficas utilizadas para detectar la existencia de autocorrelación.
- c) Explique el concepto de serie temporal débilmente dependiente.
- d) ¿Qué condiciones debe cumplir una variable para ser “estrictamente exógena”?

PROBLEMAS:

Resuelva los problemas 12.2) y 12.5) del libro de Wooldridge.

EJERCICIO Nº 1:

La base de datos TRAFICO.dta contiene 108 observaciones mensuales sobre accidentes automovilísticos, leyes de tránsito y otras variables para California de 1981 a 1989. Responda:

- a) ¿En qué mes y en qué año entró en vigor la ley del cinturón de seguridad? ¿Cuándo aumentó el límite de velocidad en rutas a 65 millas por hora?
- b) Haga una regresión de la variable $\log(totacc)$ sobre una tendencia lineal en el tiempo y 11 variables binarias mensuales, usando enero como mes base. Interprete la estimación del coeficiente de tendencia. ¿Diría que hay estacionalidad en el total de accidentes?
- c) Añada a la regresión del inciso b) las variables $finde$, $desem$, ley_vel y ley_cint . Comente el coeficiente de la variable de desempleo. ¿Su signo y magnitud tienen sentido?
- d) Interprete los coeficientes de ley_vel y ley_cint . ¿Los efectos estimados son los esperados?
- e) La variable $prcfat$ es el porcentaje de accidentes fatales (sobre el total de accidentes). Realice un análisis descriptivo de dicha variable e interprete los principales resultados.
- f) Realice la regresión del inciso c) pero usando $prcfat$ como variable dependiente. Comente los efectos estimados.
- g) Pruebe si los errores de la regresión del inciso f) tienen correlación serial AR (1), mediante: i) pruebas gráficas; ii) contraste de significatividad de p ; iii) criterio de Durbin Watson.
- h) Estime el modelo utilizando el método de cuasi-diferencias.
- i) Estime el modelo usando la estimación iterativa de Cochrane-Orcutt y de Prais-Winsten.

EJERCICIO Nº 2:

La base de datos Microsoft.dta contiene 55 observaciones cuatrimestrales sobre ingresos (revenues), gastos de publicidad y marketing (marketing), y gastos de investigación y desarrollo (idexpenditure) de Microsoft para Q1 de 1986 a Q3 de 2000. A partir de ello responda:

- a) Grafique las series contra el tiempo. ¿Considera oportuno incorporar un componente de tendencia?
- b) Estime el efecto del gasto publicitario sobre los ingresos de Microsoft, incorporando una tendencia cuadrática y las variables necesarias para quitar la estacionalidad al modelo.
- c) Pruebe si los errores de la regresión anterior tienen correlación serial AR (1), mediante pruebas gráficas y formales.
- d) Estime el modelo mediante Mínimos Cuadrados Generalizados Factibles (MCGF). En particular, utilice los métodos de: i) cuasi-diferencias; ii) Cochrane-Orcutt y iii) Prais-Winsten. Compare los resultados

obtenidos con la estimación por MCO, ¿qué sucede con los estimadores? ¿y con los errores estándares?

- e) Incorpore al modelo el gasto en I+D rezagada 4 periodos. ¿Qué pasa con los resultados del modelo, en términos de los parámetros estimados y los residuos? ¿A qué se pueden deber estos resultados?

EJERCICIO Nº 3:

En base a los datos del ejercicio anterior:

- a) Si no hubiese corregido tendencia y estacionalidad, ¿podría haber modelado las series en niveles? Para responder, grafique las series y juzgue la estacionariedad de las mismas.
- b) Una alternativa sería modelarla en primeras diferencias, aun cuando la interpretación de los coeficientes sería otra. Calcule y grafique las primeras diferencias de las series.

EJERCICIO Nº 4: (Adaptado del ejercicio 5.15 de Baltagi, 2008)

En la base de datos *Consumo.dta* se encuentran los datos de ingreso y consumo reales personales (dólares de 2000) de EEUU para el periodo 1959-2007. En base a ello:

- a) Estime la relación entre el consumo y el ingreso, incluyendo una variable de tendencia.
- b) Pruebe la existencia de correlación serial de primer orden por medio de pruebas gráficas y formales.
- c) Realice la estimación por el procedimiento de Cochrane-Orcutt.
- d) Realice la estimación por el procedimiento de Prais-Winsten.
- e) Calcule los errores estándares robustos a la autocorrelación en el modelo del inciso a.

EJERCICIOS Nº 5:

- a) En base a los datos del archivo *Inversion.dta*, estime los coeficientes de regresión aplicando el Enfoque de Koyck.
- b) Deje expresada la ecuación de regresión estimada con la transformación de Koyck.
- c) Deje expresada la ecuación de regresión estimada en variables originales, suponiendo infinitos rezagos.
- d) Deje expresada la ecuación de regresión estimada en variables originales suponiendo 3 rezagos.
- e) Calcule el multiplicador de impacto y de largo plazo de X sobre Y estimado para este modelo.
- f) Calcule e interprete el retardo mediano.

EJERCICIO Nº 6: (Ejercicio 12.9 Wooldridge)

El archivo *FISH.dta* contiene 97 observaciones de precios y cantidades diarias sobre los precios del pescado en el mercado de Nueva York. Utilice la variable $\log(\text{avgprc})$ el logaritmo del precio promedio del pescado, como la variable dependiente.

- a) Realice la regresión de $\log(\text{avgprc})$ sobre cuatro variables binarias (mon, tues, wed, thurs), con el día viernes como base. Incluya una tendencia lineal en el tiempo. ¿Existe evidencia de que los precios varían de manera sistemática a lo largo de la semana?
- b) Añada las variables *wave2* y *wave3*, que son mediciones de la altura de las olas durante varios días pasados. ¿Estas variables son individualmente significativas? Describa un mecanismo mediante el cual el mar tempestuoso aumente el precio del pescado.
- c) ¿Qué ocurrió con la tendencia en el tiempo cuando *wave2* y *wave3* se añadieron a la regresión? ¿Qué debe estar pasando?
- d) Explique por qué se supone que todas las variables explicativas son estrictamente exógenas.
- e) Pruebe los errores para la correlación serial AR(1).
- f) Obtenga las estimaciones de PraisWinsten para el modelo estimado en el inciso b). ¿Son *wave2* y *wave3* conjuntamente significativas en términos estadísticos?

BIBLIOGRAFÍA

Básica:

- Wooldridge, J. (2010), *Introducción a la econometría. Un enfoque moderno*. 4ª edición. Madrid: Thomson Editores Spain.

Complementaria:

- Berndt, E. (1996), *The practice of econometrics: classic and contemporary*. Addison Wesley, Reading Massachusetts.
- Draper, N. and Smith, H. (1998), *Applied regression analysis*. 3rd ed., John Wiley & Sons, USA.
- Greene, W. (2000), *Análisis Económico*. 3ª edición, Prentice-Hall.
- Gujarati, D. (2004), *Econometría*. 4ª edición, México: Mc Graw Hill.
- Stock, J. y Watson, M. (2012), *Introducción a la Econometría*. 3ª edición, Pearson, Madrid.
- Verbeek, M. (2008), *A Guide to Modern Econometrics*. John Wiley & Sons, Sussex, England.