PUBLICACIONES DE 4º CURSO

Curso: 4º

Grado: Economía

Asignatura: ECONOMETRÍA III

ENUCIADOS EJERCICIOS PARTE 2

Profesores: Antonio Aznar y Mª Isabel Ayuda

Departamento de ANÁLISIS ECONÓMICO

Curso Académico 2015/16

Facultad de
Economía y Empresa
Universidad Zaragoza

Ejercicio 2.1

Una regresión del promedio de los ingresos salariales semanales (ISM, medidos en dólares) sobre la edad (medida en años), utiliza una muestra aleatoria de trabajadores con estudios universitarios a tiempo completo entre 25 y 65 años de edad, y obtiene lo siguiente:

$$\widehat{ISM} = 969,7 + 9,6\ Edad, \quad R^2 = 0.023, \quad ESR = 624,1$$

1. Explique qué significan los valores de los coeficientes estimados.
2. El error estándar de la regresión (ESR) es 624,1, ¿Cuáles son las unidades de medida?
3. ¿Cuáles son las unidades de medida del R^2?
4. ¿Cuáles son los ingresos salariales previstos por la regresión por un trabajador de 25 años de edad? ¿Y para uno de 45 años de edad?
5. ¿Será fiable la regresión en sus predicciones para un trabajador de 99 años de edad? ¿Por qué o por qué no?
6. Teniendo en cuenta lo que sabe sobre la distribución de los ingresos, ¿crees que es posible que la distribución de los errores de la regresión sea normal? (Pista: piensa si la distribución es simétrica o asimétrica, ¿cuál es el menor valor de los ingresos? ¿es compatible con la normal?).
7. El promedia de edad de la muestra es de 41,6 años, ¿Cuál es el valor medio muestral de ISM?

Ejercicio 2.2 (Ejercicio 5.2 Stock-Watson)

Supóngase que un investigador, con datos salariales sobre 250 trabajadores y 280 trabajadoras seleccionados aleatoriamente, estima la regresión MCO:

$$\widehat{Salario} = \underset{(0,23)}{12,52} + \underset{(0,36)}{2,12}\ Masculino, \quad R^2 = 0,06, \quad ESR = 4,2$$

Donde *Salario* se mide en dólares por hora y *Masculino* es una variable binaria que toma valor 1 si la persona es hombre y 0 si es mujer. Defina la brecha salarial por género como la diferencia de ingresos salariales medios entre hombres y mujeres.

1. ¿Cuál es la brecha de género estimada?
2. ¿Es la brecha de género significativamente distinta de cero? (Calcule el p-valor para el contraste de la hipótesis nula de que no existe brecha de género?
3. Construya un intervalo de confianza del 95% para la brecha de género?
4. En la muestra ¿cuál es el salario medio de las mujeres? ¿y el de los hombres?
5. Otro investigador utiliza estos mismos datos pero realiza la regresión de la variable *Salario* sobre la variable *Femenino*, una variable que toma valor 1 si la persona es mujer y 0 si es hombre. ¿Cuáles son las estimaciones de esta regresión?

Ejercicio 2.3 (Ejercicio 5.5 Stock-Watson)

En la década de 1980, Tennessee llevó a cabo un experimento en el que los estudiantes de guardería fueron asignados aleatoriamente a clases de distintos tamaños, normal o pequeños, realizándose a final de curso unos exámenes o pruebas estandarizados. (Las clases normales eran aproximadamente de 24 estudiantes, y las clases pequeñas de aproximadamente 15). Supóngase que, en la población, las pruebas estandarizadas arrojan una puntuación media de 925 puntos y una desviación típica de 75 puntos. Sea *ClasePequeña* una variable binaria igual a 1 si el estudiante es asignado a un clase pequeña e igual a 0 en otro caso. Una regresión de *CalificacionExamen* sobre *ClasePequeña* proporciona estos resultados:

$$\widehat{CalificacionExamen} = \underset{(1,6)}{918,0} + \underset{(2,5)}{13,9}\ ClasePequeña, \quad R^2 = 0,01, \quad ESR = 74,6$$

1. ¿Mejoran las clases pequeñas los resultados de la prueba? ¿En cuánto? ¿Es grande el efecto? Explíquelo.
2. ¿Es estadísticamente significativo el efecto estimado del tamaño de las clases sobre las calificaciones obtenidas? Realice un contraste al 5% de nivel de significación.
3. Construye un intervalo de confianza al 99% para el efecto de *ClasePequeña* sobre las calificaciones.

Ejercicio 2.4 (Ejercicio 6.1, 6.2, 6.3 y 6,4 Stock-Watson)

Utilizando datos de 1998 de la Encuesta Actualizada de Población (CPS) que consta de 4000 trabajadores a tiempo completo durante todo el año se han estimado varios modelos. El mayor grado educativo alcanzado por cada trabajador es o un diploma de escuela secundaria o un título de licenciatura. El rango de edades de los trabajadores oscila entre los 25 y los 34 años. La base de datos, asimismo, contiene información sobre la región del país donde reside la persona, el estado civil y el número de hijos. A los efectos de esto ejercicios, sean:

IMH = ingresos medios por hora (en dólares de 1998)

Universidad = variable binaria (1 si tiene estudios universitarios y 0 si titulado escuela secundaria).

Femenino = variable binaria (1 si es mujer, 0 si es hombre).

Edad = edad (en años)

Noroeste = variable binaria (1 si Región = Noroeste, 0 en caso contrario)

Centro-Oeste = variable binaria (1 si Región = Centro-Oeste, 0 en caso contrario)

Sur = variable binaria (1 si Región = Sur, 0 en caso contrario)

Oeste = variable binaria (1 si Región = Oeste, 0 en caso contrario)

Variable dependiente: Ingresos salariales medios por hora			
Regresor	**(1)**	**(2)**	**(3)**
Universidad	5,46	5,48	5,44
Femenino	-2,64	-2,62	-2,62
Edad		0,29	0,29
Noroeste			0,69
Centro-Oeste			0,60
Sur			-0,27
Constante	12,69	4,40	3,75
Estadísticos de resumen			
ESR	6,27	6,22	6,21
R^2	0,176	0,190	0,194
$\bar{R}^2$			

1. Calcule el $\bar{R}^2$ de cada una de las regresiones.
2. Utilizando los datos de la regresión de la columna (1)
   - **A)** ¿Ganan más los trabajadores con títulos universitarios en promedio que los trabajadores con tan solo grado de secundaria? ¿Cuánto más?
   - **B)** ¿Ganan los hombres más que las mujeres en promedio?¿Cuánto más?
3. Utilizando los resultados de la regresión de la columna (2):

   A) ¿Es la edad un determinante importante para los ingresos? Explíquelo.

   B) Sally ese una mujer titulada universitaria de 29 años de edad. Betsy es una mujer titulada universitaria de 34 años. Prediga los ingresos de ambas.

4. Utilizando los datos de la regresión de la columna (3):

A) ¿Parece que existen diferencias regionales importantes?

B) ¿Por qué se ha omitido la variable Oeste de la regresión? ¿Qué pasaría si se incluyese?

C) Juanita es una mujer universitaria de 28 años de la región del Sur, Jennifer es una mujer de 28 años titulada universitaria de la región Centro-Oeste. Calcule la diferencia esperada entre los ingresos de Juanita y los de Jennifer.

Ejercicio 2.5 (Ejercicio 7.1, 7.2, 7.3 y 7.4 Stock-Watson)

Con los datos de la base de datos del ejercicio anterior se han obtenido los siguientes resultados:

Variable dependiente: Ingresos salariales medios por hora			
Regresor	**(1)**	**(2)**	**(3)**
Universidad	5,46 (0,21)	5,48 (0,21)	5,44 (0,21)
Femenino	-2,64 (0,20)	-2,62 (0,20)	-2,62 (0,20)
Edad		0,29 (0,04)	0,29 (0,04)
Noroeste			0,69 (0,30)
Centro-Oeste			0,60 (0,28)
Sur			-0,27 (0,26)
Constante	12,69 (0,14)	4,40 (1,05)	3,75 (1,06)
Estadísticos de resumen			
Estadístico F efectos regionales = 0			6,10
ESR	6,27	6,22	6,21
R^2	0,176	0,190	0,194
$\bar{R}^2$			

1. Añadir *(5%) y ** (1%) a la tabla anterior para indicar la significación estadística de los coeficientes.
2. Utilizando los resultados de la regresión de la columna (1):
   a) ¿Es estadísticamente significativa la diferencia estimada para esta regresión entre los ingresos salariales de los graduados universitarios y los graduados en enseñanza secundaria al nivel del 5%? Construya un intervalo de confianza al 95% para esta diferencia.

b) ¿es estadísticamente significativa la diferencia estimada por esta regresión entre los ingresos salariales de hombres y mujeres al nivel del 5%? Construya un intervalo de confianza al 95% para esta diferencia.

3. Utilizando los resultados de la regresión de la columna (2):
   a) ¿Es la edad un factor importante de los ingresos salariales?. Utilice un contraste estadístico apropiado y/o un intervalo de confianza para explicar la respuesta.
   b) Sally es una graduada universitaria de 29 años. Betsy es una mujer de 34 años graduada universitaria. Construya un intervalo de confianza al 95% para la diferencia esperada entre los ingresos salariales.
4. Utilizando los resultados de la regresión de la columna (3):
   a) ¿Parece que existan diferencias regionales importantes? Utilice un contraste adecuado para explicar su respuesta.
   b) Juanita es una mujer de 28 años graduada universitaria de la región Sur. Molly es una mujer graduada universitaria de 28 años de la región Oeste. Jennifer es una mujer de 28 años graduada universitaria de la región Centro-Oeste.
      i) Construya un intervalo de confianza al 95% para la diferencia de los ingresos esperados de Juanita y de Molly.
      ii) Explique cómo se construiría un intervalo de confianza al 95% para la diferencia entre los ingresos esperados de Juanita y Jennifer.

Ejercicio 2.6 (Ejercicio 9.3 Stock-Watson)

Los economistas laborales descubrieron un resultado empírico desconcertante por medio de un estudio sobre los determinantes de los ingresos salariales de las mujeres. Utilizando mujeres empleadas aleatoriamente seleccionadas, realizaron una regresión de los ingresos salariales sobre el número de hijos de las mujeres y un conjunto de variables de control ¿edad, educación, ocupación, etc.?. Hallaron que las mujeres con más hijos tenían salarios más altos, teniendo en cuenta todos estos factores. Explique como la selección muestral podría ser la causa de este resultado. Este enigma empírico motivó la investigación de James Heckman sobre la selección muestral que le llevó al Premio Nobel de economía 2000].

Ejercicio 2.7 (Ejercicio 9.5 Stock-Watson)

La demanda de un bien está dada por $Q = \beta_0 + \beta_1 P + u$, donde *Q* expresa la cantidad demandada, *P* expresa el precio y *u* expresa otros factores distintos del precio que determinan la cantidad. La oferta de un bien está dada por $Q = \gamma_0 + \gamma_1 P + v,$, donde *v* expresa los factores distintos del precio que influyen en la oferta. Suponga que tanto *u* como *v* tienen media igual a cero y varianzas σ_u^2 y σ_v^2, respectivamente. Y están mutuamente incorreladas.

1. Resuelva las dos ecuaciones simultáneas para mostrar cómo *Q* y *P* dependen de *u* y de *v*.
2. Obtenga las medias de *P* y *Q*.
3. Obtenga la varianza de *Q*, la varianza de *P* y la covarianza entre *Q* y *P*.
4. Se obtiene una muestra aleatoria de observaciones de (Q_i, P_i) y se hace la regresión de *Q* sobre *P*. Supóngase que la muestra es muy grande.
   I) Utilice las respuestas de 1. Y 2 para obtener los coeficientes de la regresión.
   II) Un investigador utiliza la pendiente como una estimación de la pendiente de la función de demanda. ¿Es la pendiente estimada demasiado grande o demasiado pequeña?

Ejercicio 2.8 (Ejercicio 9.6 Stock-Watson)

Supóngase que n = 100, las observaciones i.i.d. sobre (Y_i, X_i) dan lugar a los siguientes resultados:

$$\hat{Y}_i = \underset{(15,1)}{32,1} + \underset{(12,2)}{66,8} X_i, \quad ESR = 15,1, \quad R^2 = 0,81$$

Otro investigador está interesado en la misma regresión, pero comete un error al introducir los datos: introduce dos veces cada observación.

1. Mediante estas 200 observaciones ¿Qué resultados arroja su regresión?

$$\hat{Y}_i = \underset{(\_\_)}{\_\_\_} + \underset{(\_\_\_)}{\_\_\_\_} X_i, \quad ESR = \_\_\_\_, \quad R^2 = \_\_\_\_$$

2. ¿Qué requisitos se violan para la validez interna?

Ejercicio 2.9

En el marco del modelo $y_t = \alpha + \beta x_t + u_t$ en donde *u* cumple las hipótesis ideales y también se cumple que $\frac{1}{T}\sum x_t, \sum \frac{1}{x_t}, \frac{1}{T}\sum \frac{1}{x_t^2}$ convergen a constantes conforme crece el tamaño muestral, se definen los dos estimadores siguientes:

$$\hat{\beta}_1 = \frac{\sum y_t - T\alpha}{\sum x_t} \quad \text{y} \quad \hat{\beta}_2 = \frac{\sum \frac{y_t}{x_t}}{T}$$

Obtener la media y varianza de ambos estimadores y demostrar que son consistentes suponiendo que el regresor no es estocástico.

Ejercicio 2.10

En el marco del modelo $y_t = \beta x_t + u_t$, en donde *u* cumple las hipótesis ideales y suponemos que el regresor no es estocástico, se definen dos estimadores de β : el MCO y el siguiente estimador

$$\widehat{\beta}_1 = \frac{\sum y_t}{\sum x_t}$$

1. Obtener la media y varianza de ambos estimadores e indicar las condiciones que debe cumplir el regresor para que ambos estimadores sean consistentes.
2. Definir los residuos correspondientes a cada estimador y, para la observación t-ésima, derivar la media y varianza del correspondiente residuo. Demostrar que la suma de cuadrados de los residuos MCO es menor que la suma de los cuadrados de los residuos que corresponden a $\widehat{\beta}_1$.

cumple que $\sum x_t \hat{u}_t = 0$.

Ejercicio 2.11

Considerar el siguiente modelo:

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + u_t$$

donde se supone que las variables están medidas en desviaciones respecto a sus medias, que suponemos regresores no estocásticos y la perturbación *u* cumple las hipótesis ideales.

1. Expresar las varianzas de los estimadores MCO de β en función de las varianzas muestrales de x_1 y x_2 $(S_1^2 \ y \ S_2^2)$, respectivamente y del coeficiente de correlación entre ambas variables (r_{12}).
2. Conociendo que σ^2=1, S_1^2=2, T=8, $\hat{\beta}_1$(MCO)=2,se pide determinar la región crítica para contrastar la hipótesis nula $H_0{:}\beta_1 = 0$ frente a la alternativa de que es diferente de cero para aquellos casos en que $r_{12}^2 = 0{,}02$ y $r_{12}^2 = 0{,}99$. Utilizar un tamaño del error tipo 1 igual a 0,05.
3. Obtener el tamaño del error tipo 2 para los dos casos contemplados en el apartado anterior suponiendo β_1=1,8.

Ejercicio 2.12

Considerar los dos modelos siguientes y considerando que son regresores no estocásticos:

M1: $y_t = \beta_1 x_{1t} + u_{1t}$

M2: $y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + u_{2t}$

Para estimar β_1 se definen dos estimadores MCO, uno a partir de M1 y otro a partir de M2, suponiendo que genera los datos M2, se pide:

1. Escribir la forma que toman ambos estimadores y derivar el sesgo y la varianza de ambos. Hallar el cociente de las dos varianzas y comentar de qué depende el resultado.
2. Escribir los estimadores MCO de las varianzas de los dos estimadores de β_1 y derivar los sesgos de los mismos.

Ejercicio 2.13

Considerar los dos modelos siguientes, suponiendo regresores no estocásticos:

M1: $y_t = \beta_1 x_{1t} + u_{1t}$

M2: $y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + u_{2t}$

Se dispone de la siguiente información

$$\left(X'X\right) = \begin{pmatrix} \sum x_{1t}^2 & \sum x_{1t}x_{2t} \\ \sum x_{2t}x_{1t} & \sum x_{2t}^2 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \qquad X'y = \begin{pmatrix} \sum x_{1t}y_t \\ \sum x_{2t}y_t \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \qquad y'y = 10 \qquad T = 10$$

1. Estimar β_1 con MCO, utilizando ambos modelos. Estimar insesgadamente la varianza de los dos estimadores. Obtener las propiedades (media y varianza) del estimador restringido de β_2.
2. Se va a discriminar entre los dos modelos utilizando el estimador máximo verosímil de la varianza de ambos estimadores. Escribir la forma que adoptaría la regla de discriminación si lo escribimos como un contraste F derivando el punto crítico implícito. Aplicar esa regla y determinar el modelo que sería rechazado en este caso. Comparar este punto crítico con el que corresponde al criterio SBIC.

Ejercicio 2.14

Considerar los dos modelos siguientes:

M1: $y_t = \beta_1 x_{1t} + u_{1t}$

M2: $y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + u_{2t}$

Para estimar β_1 se definen dos estimadores MCO, uno a partir de M1 y otro a partir de M2. Suponiendo que genera los datos M2, se pide discriminar entre los dos modelos utilizando el estimador máximo verosímil de la varianza de ambos estimadores. Determinar la región crítica y obtener el factor de parsimonia correspondiente a la regla de decisión.

Ejercicio 2.15

Considerar los dos modelos siguientes:

M1: $y_t = \beta_1 x_{1t} + u_{1t}$

M2: $y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + u_{2t}$

Sea X la matriz T$\times$2 de las observaciones de las dos variables. Se dispone de la siguiente información:

$$\bar{y} = 4; \quad y'y = \sum y_t^2 = 200; \quad T=10, \ \sum \hat{u}_{1t}^2 = 40; \sum \hat{u}_{2t}^2 = 10$$

Se pide:

1. Para los criterios $\bar{R}^2, AIC$ calcular sus valores para el modelo amplio y derivar y calcular el factor de parsimonia y el punto crítico implícito para esos dos criterios y el de los Multiplicadores de Lagrange.
2. Escribir la región crítica que corresponde a cada criterio y utilizando la información numérica aportada indicar si rechazaría la hipótesis nula.