

PUBLICACIONES DE 4º CURSO

Curso: 4º

Grado: Economía

Asignatura: ECONOMETRÍA III

ENUCIADOS EJERCICIOS EMPÍRICOS PARTE 2

Profesores: Antonio Aznar y Mª Isabel Ayuda

Departamento de ANÁLISIS ECONÓMICO

Curso Académico 2015/16



**Facultad de
Economía y Empresa
Universidad Zaragoza**

EJERCICIOS EMPÍRICOS PARTE 2

Ejercicio empírico 2.1 (E4.3 y E5.2 Stock-Watson)

Basándose en la información contenida en la base de datos **CollegeDistance.xls** cuyos datos están descritos en **CollegeDistance_DataDescription**, se pretende investigar la relación entre el número de años de educación completados por adultos jóvenes (**ed**) y la distancia de la escuela secundaria de cada estudiante a la universidad más próxima (**dist**). (La proximidad a la universidad reduce el coste de la educación, por lo que los estudiantes que viven más cerca de una universidad deberían, en promedio, completar más años de educación superior). Para ello, se pide:

1. Realice una regresión de los años completados de educación (*ed*) sobre la distancia del instituto a la universidad más cercana (*dist*), medida en decenas de millas. Interprete la estimación de la pendiente.
2. ¿Explica la distancia a la universidad una proporción grande de la varianza de los años de educación?
3. ¿Cuál es la desviación típica de la regresión, ESR?
4. Suponiendo que se cumplen las hipótesis básicas, ¿Es el coeficiente estimado de la pendiente estadísticamente significativo para un nivel de significación del 1% y del 5%?
5. Construya un intervalo de confianza al 95% para el coeficiente de la pendiente.
6. Si la distancia del instituto de Bob a la universidad más cercana es de 20 millas. Prediga los años de educación completados por Bob utilizando la regresión estimada. ¿Cómo cambiaría la predicción si la distancia fuese de 10 millas?
7. Contraste las hipótesis de homocedasticidad (contrastes de White, Glejser y Goldfeld-Quandt) y normalidad (contraste de Jarque-Bera). Si se incumple alguna de las hipótesis ¿cuál cree que podría ser el problema?
8. Realice la regresión utilizando sólo los datos para las mujeres.
9. Realice la regresión utilizando sólo los datos para los hombres.
10. ¿Es diferente el efecto de la distancia sobre los años completados de educación para los hombres y para las mujeres?

Ejercicio empírico 2.2 (E6.2 y E7.3 Stock-Watson)

Utilizando los mismos datos que en el ejercicio anterior:

1. Realice la regresión anterior pero incluyendo algunas variables de control sobre las características del estudiante, su familia y sobre el mercado laboral local. En concreto utilice como variables explicativas adicionales las variables *bytest*, *female*, *black*, *hispanic*, *incomehi*, *ownhome*, *dadColl*, *momColl*, *cue80* y *stwmfg80*. ¿Cuál es el efecto estimado de la variable *Dist* sobre *ed*?
2. ¿Diría que La regresión del apartado 1 presenta sesgo de estimación?

3. Compare el ajuste de las dos regresiones.
4. El valor del coeficiente de la variable *dadcoll* (ficticia que toma valor 1 si el padre del estudiante estudió en el mismo instituto) es positivo. ¿Qué mide este coeficiente? Interpretar los coeficientes de las variables *cue80* (tasas de paro del condado en 1980) y *stwmfg80* (salario por hora medio en el estado en 1980) e indicar si son significativos al 5%.
5. Si Bob es un hombre negro, su instituto estaba situado a 20 millas de la universidad más cercana, su calificación en la prueba (*bytest*) fue de 58, su renta familiar fue de 26.000\$, su familia poseía una casa, su madre estudio en la universidad y su padre no estudio en la universidad, la tasa de paro de su condado era de 7,5% y el promedio del salario por hora en su estado era de 9,75\$. Estime el número de años de estudio completados por Bob utilizando la segunda regresión.
6. Jim tiene las mismas características que Bob, salvo que su escuela secundaria estaba a 40 millas de la universidad más cercana. Estime los años completados por Jim en la segunda regresión.
7. Contrastar si en promedio los años universitarios completados por una persona se incrementarían en aproximadamente 0,15 años si la distancia a la universidad más cercana se redujera en 20 millas.
8. Contraste las hipótesis de homocedasticidad y normalidad. Si se incumple alguna de las hipótesis ¿cuál cree que podría ser el problema?
9. Según la respuesta del apartado anterior, ¿Cree que los contrastes de significatividad del apartado 4 serían válidos? ¿Cómo los realizaría?

Ejercicio empírico 2.3 (E8.3 y E9.3 Stock-Watson)

Con los datos de los ejercicios anteriores, se pide:

1. Realice una regresión de $\ln(ed)$ sobre las variables explicativas del ejercicio 2.2. Si la variable *dist* aumenta de 2 a 3 (de 20 millas a 30 millas) ¿cuánto se espera que cambien los años de educación? ¿Cuánto se espera que cambien los años de ed si la variable *dist* cambia de 6 a 7 (de 60 a 70 millas)?
2. Realice una regresión de la variable *ED* sobre las variables *dist*, $dist^2$, *bytest*, *tuition*, *female*, *black*, *hispanic*, *incomehi*, *ownhome*, *dadcoll*, *momcoll*, *cue80* y *stwmfg80*. Si la variable *dist* aumenta de 2 a 3 (de 20 millas a 30 millas), ¿Cuánto se espera que aumenten los años de educación? ¿Cuánto se espera que cambien los años de educación si la variable *dist* cambia de 6 a 7 (de 60 a 70 millas)?
3. Añada el término de interacción *dadcoll* x *momcoll* a la regresión anterior, ¿qué mide el coeficiente del término de interacción?

4. ¿Existe alguna evidencia de que el efecto de la variable *dist* sobre la variable *ed* dependa de la renta de la familia? ¿Cuál es el efecto de la variable *dist* en la variable *ed* en las familias de renta alta y en las de renta baja? ¿Es esta diferencia significativa?
5. Discuta la validez interna de las regresiones anteriores. Incluya una discusión sobre el posible sesgo de variable omitida, error de especificación de la forma funcional, errores en las variables, causalidad simultánea e inconsistencia de los errores estándar MCO.

Ejercicio empírico 2.4

El fichero “Salarios-Educacion.xls” tiene mil observaciones correspondientes a diferentes individuos con información sobre salarios horarios (SAL), educación en años (EDUC), experiencia en años (EXPER), horas trabajadas a la semana (HRSWK), MARRIED (= 1 si casado), FEMALE (= 1 si mujer), METRO (= 1 si vive en zona metropolitana), MID(= 1 si vive en el medio oeste), SOUTH (= 1 si vive en el sur), WEST(= 1 si vive en el oeste), BLACK(= 1 si es negro) y ASIAN (= 1 si es asiático) para el año 2008. Se pide:

- 1) Discutir las características de los datos de *SAL* y *EDUC*. Estadísticos e histograma.
- 2) Estimar la regresión lineal $SAL = \beta_1 + \beta_2 EDUC + u$ y comentar los resultados.
- 3) Obtener los residuos MCO y hacer un gráfico de estos residuos con la variable *EDUC*. ¿Se observa alguna pauta relevante?. Si *EDUC* no es estocástica y se cumplen las hipótesis vistas en teoría, ¿cabría esperar alguna pauta de comportamiento conjunto?
- 4) Estimar regresiones separadas para hombres, mujeres, asiáticos y blancos. Comparar los resultados.
- 5) Estimar la regresión cuadrática $SAL = \beta_1 + \beta_2 EDUC^2 + u$ y comentar los resultados. Estimar el efecto marginal de un año adicional de educación sobre los salarios para una persona con 12 años de educación y para otra con 16 años de educación. Comparar estos resultados con los obtenidos en 2).
- 6) Construir un histograma de $\ln(SAL)$. Comparar la forma de este histograma con la correspondiente a *SAL*. ¿Cuál de ellas parece más simétrica y ajustada a una campana de Gauss?
- 7) Estimar la regresión $\ln(SAL) = \beta_1 + \beta_2 EDUC + u$. Estimar el efecto marginal de un año adicional de educación sobre los salarios para una persona con 12 años de educación y para otra con 16 años de educación. Comparar estos resultados con los obtenidos en 2) y 5).
- 8) Contrastar la significatividad estadística del coeficiente de *EDUC* en 2) utilizando un nivel de significación del 5%. Con el mismo nivel de significación obtener una estimación por intervalo de ese coeficiente.

- 9) Contrastar la hipótesis nula de que $\beta_1 \geq 0$ frente a la alternativa de que es menor que cero, utilizando un nivel de significación del 1%.
- 10) Estimar el modelo $\ln(SAL) = \beta_1 + \beta_2 EDUC + \beta_3 EXPER + \beta_4 HRSWK + u$ e interpretar las estimaciones obtenidas para los tres parámetros. ¿Son, desde un punto de vista estadístico, significativamente diferentes de cero?. Los apartados que siguen se refieren a este modelo general si no se dice lo contrario.
- 11) Contrastar la hipótesis nula de que el efecto de un año adicional de educación es, al menos, de un crecimiento del salario del 10% frente a la hipótesis alternativa de que es menos del 10%.
- 12) Definir un intervalo de confianza del 99% para el cambio porcentual de los salarios consecuencia de aumentar una hora el trabajo semanal.
- 13) Reestimar el modelo en 2) con las variables adicionales $EXPER, EDUC \times EXPER, EDUC^2, EXPER^2$. Comentar los resultados.
- 14) Para el nuevo modelo, encontrar expresiones para el efecto marginal $\partial(SAL)/\partial EDUC$ y $\partial(SAL)/\partial EXPER$
- 15) Estimar el efecto marginal $\partial(SAL)/\partial EDUC$ para dos trabajadores Hill and Wendy; Hill tiene 16 años de educación y 2 años de experiencia, mientras que Wendy tiene 12 años de educación y 2 años de experiencia. ¿Qué puede decirse acerca del efecto marginal de la educación conforme esta crece?
- 16) Contrastar considerando como hipótesis alternativa que el efecto marginal de la educación es mayor en Hill que en Wendy.
- 17) Definir un intervalo de confianza del 95% para el efecto marginal en el salario consecuencia de un año adicional de educación para alguien con 16 años de educación y 2 años de experiencia.
- 18) En el modelo en 10), contrastar la hipótesis conjunta de que los coeficientes de las variables EXPER y HRSWK son cero.
- 19) Considerar el modelo

$$\ln(SAL) = \beta_1 + \beta_2 EDUC + \beta_3 BLACK + \beta_4 FEMALE + \beta_5 (BLACK \times FEMALE) + \beta_6 SOUTH + \beta_7 MID + \beta_8 WEST + u$$
 Contrastar si la interacción BLACK y FEMALE es estadísticamente significativa. Contrastar, también, que no hay efecto regional.
- 20) Estimar el siguiente modelo con MCO y desviaciones típicas robustas a la heterocedasticidad

$$\ln(SAL) = \beta_1 + \beta_2 EDUC + \beta_3 EXPER + \beta_4 EXPER^2 + \beta_5 (EXPER \times EDUC) + u$$
- 21) Añadir la variable MARRIED al modelo en 20) y reestimarlos. Manteniendo constantes la educación y la experiencia, ¿tienen salarios más altos los trabajadores casados? Utilizando un nivel de significación del 1% contrastar la hipótesis nula de que los salarios de los casados son menores que los de los solteros, frente a la alternativa de que son mayores.

- 22)** Hacer un gráfico con los residuos del modelo en 20) contra los dos valores de MARRIED. ¿Hay evidencia de heterocedasticidad?
- 23)** Estimar el modelo en 20) dos veces-la primera, utilizando sólo la información para los casados y la segunda solo la información para los solteros. Contrastar la heterocedasticidad con Goldfeld-Quandt utilizando el nivel del 5%.
- 24)** Hacer gráficos de los residuos del modelo en 20) contra EDUC y contra EXPER. ¿Qué sugieren?
- 25)** Utilizando el contraste de Breusch-Pagan contrastar la heterocedasticidad suponiendo que la varianza depende de EDUC, EXPER y MARRIED. ¿A qué conclusión se llega?