ECONOMETRIA II

Mínimos Cuadrados en dos Etapas

Oferta de mano de obra de mujeres trabajadoras y casadas en el Perú

Contenido

Dedicatoria

Este proyecto es el resultado del esfuerzo conjunto de todos los que formamos el grupo de trabajo. Por esto agradezco a nuestro profesor, Dr. SAMANAMUD LOYOLA, OSCAR FRANCISCO, mis compañeros y mi persona, quienes a lo largo de este tiempo han puesto a prueba sus capacidades y conocimientos en el desarrollo de este nuevo plan estratégico de negocios el cual ha finalizado llenando todas nuestras expectativas. A nuestros padres quienes a lo largo de toda nuestra vida han apoyado y motivado mi formación académica, creyeron en mi en todo momento y no dudaron de mis habilidades. A mis profesores a quienes les debo gran parte de mis conocimientos, gracias a su paciencia y enseñanza y finalmente un eterno agradecimiento a esta prestigiosa universidad la cual abrió abre sus puertas a jóvenes como nosotros, preparándonos para un futuro competitivo y formándonos como personas de bien.

Introducción

En temas anteriores se mostró que el método de variables instrumentales puede resolver dos tipos de problemas: las variables omitidas y el error de medición. En términos conceptuales estos problemas son sencillos. En el caso de las variables omitidas, existe una variable (o más de una) que se desearía mantener fija al estimar el efecto ceteris paribus de una o más de las variables explicativas observadas. En el caso del error de medición, lo que interesa estimar es el efecto de ciertas variables explicativas sobre y, pero que se midieron una o más variables de manera equivocada. En ambos casos, se podría estimar el parámetro de interés mediante MCO si fuera posible recabar mejores datos. Otra forma importante de endogeneidad de las variables explicativas es la simultaneidad. Esta surge cuando una o más de las variables explicativas se determina juntamente con la variable dependiente, por lo general, a través de un mecanismo de equilibrio. En este trabajo se estudiará un método para estimar mediante dos fases. Aunque un análisis completo de este tema rebasa el alcance de esta pequeña aplicación para fines prácticos se estudiaran un modelo realizado por unas variables simples con una muestra obtenida por el INEI.

El método más importante para estimar modelos en dos fases es el de variables instrumentales.

Método de mínimos cuadrados en dos etapas (MC2E)

Trata el problema de la endogeneidad de una o más variables explicativas en un modelo de regresión múltiple.
Su principal objetivo es evitar que una o más variables explicativas endógenas de un modelo estén correlacionadas con el término de error y poder realizar estimaciones eficientes de mínimos cuadrados ordinarios (MCO) sobre el modelo inicial. Las herramientas para utilizar son variables instrumentales (VI), modelos estructurales y ecuaciones reducidas.

En otras palabras, MC2E nos ayuda a realizar una estimación con garantías cuando una o más variables explicativas endógenas están correlacionadas con el término de error y hay exclusión de variables explicativas exógenas. MC2E hace referencia al procedimiento a seguir para tratar este problema de endogeneidad.

- En la primera etapa se aplica un "filtro" para eliminar la correlación con el término de error.
- En la segunda etapa se obtienen los valores ajustados a partir de los cuales se pueden realizar buenas estimaciones MCO sobre la forma reducida del modelo original.

El modelo estructural

Un modelo estructural representa una ecuación donde se pretende medir la relación causal entre las variables y la atención se centra en los regresores (β_j). El Modelo 1 es una regresión lineal múltiple con dos variables explicativas: Y_2 y Z_1.

$$Modelo\,1 \rightarrow Y_1 = \beta_0 + \beta_1 Y_2 + \beta_2 Z_1 + u_1$$

Las variables explicativas pueden dividirse en dos tipos: variables explicativas endógenas y variables explicativas exógenas. En el Modelo 1, la variable explicativa endógena es Z_1 y la variable explicativa exógena es Y_2. La variable endógena viene dada por el modelo (es el resultado del modelo) y está correlacionada con u_1. La variable exógena la tomamos como dada (es necesaria para que el modelo expulse un resultado) y no está correlacionada con u1.

Procedimiento MC2E

En lo que sigue vamos a explicar detalladamente el procedimiento para realizar una estimación a través del método de mínimos cuadrados en dos etapas.

Primera etapa

a) Paso uno. –

Suponemos que tenemos dos variables explicativas exógenas que están excluidas en el Modelo 1, siendo Z_2 y Z_3. Recordemos que ya tenemos una variable explicativa exógena en el Modelo 1, Z_1, por tanto, en total ahora tendremos tres variables explicativas exógenas: Z_1, Z_2y Z_3

Las restricciones de exclusión son:

- Z_2y Z_3 no aparezcan en el Modelo 1, por tanto, que estén excluidas.
- Z_2y Z_3no estén correlacionadas con el error.

b) Paso dos. –

Tenemos que obtener la ecuación en la forma reducida para Y_2 . Para ello, sustituimos:

- La variable endógena Y_1 por Y_2 .
- Los regresores β_1 por π_1.
- El error u_1 por v_1 .

La forma reducida para Y_2 del Modelo 1 es:

$Y_2=\pi_0+\pi_1 Z_1+\pi_2 Z_2+\pi_3 Z_3+v_2$

En el caso de que Z_2y Z_3 estén correlacionadas con Y_2 , se podría utilizar el método de Variables Instrumentales (VI) pero terminaríamos con dos estimadores de VI y en tal caso los dos estimadores serían ineficientes o imprecisos. Decimos que un estimador es más eficiente o preciso cuánto más pequeña sea su varianza. El estimador más eficiente sería el que tenga la mínima varianza posible.

c) Paso tres. -

Suponemos que la combinación lineal anterior es la mejor Variable Instrumental (VI), denominamos Y_2* para Y_2 y quitamos el error (v_2) de la ecuación:

$Y_2 * ¿ \pi_0 + \pi_1 Z_1 + \pi_2 Z_2 + \pi_3 Z_3 + v_2 \forall \pi_2 \neq 0 \pi_3 \neq 0$

Segunda etapa

d) Paso cuatro

Realizamos la estimación MCO sobre la forma reducida del Modelo 1 anterior y obtenemos los valores ajustados (los representamos con el acento circunflejo "^"). El valor ajustado es la versión estimada de $Y_2 * ¿$ que a su vez no está correlacionada con u_1.

$\widehat{Y}_2 = \hat{\pi}_0 + \widehat{\pi_1} Z_1 + \widehat{\pi_2} Z_2 + \widehat{\pi_3} Z_3$

Resumen del proceso

Primera etapa:

Realizar regresión en el modelo del circunflejo (punto 4) donde justamente se obtienen los valores ajustados. Este valor ajustado es la versión estimada de $Y_2 * ¿$ y, por tanto, no está correlacionada con el error u_1. La idea es aplicar un filtro de no correlación del valor ajustado con el error u_1.

Segunda etapa:

Realizar regresión MCO sobre la forma reducida del Modelo 1 (punto 2) y obtenemos los valores ajustados. Dado que se utiliza el valor ajustado y no el valor original (Y_2) que no cunda el pánico si las estimaciones de MC2E no coinciden con las estimaciones de MCO sobre la forma reducida del Modelo 1.

Aplicación del modelo de Mínimos cuadros en dos etapas

Oferta de mano de obra de mujeres trabajadoras y casadas en el Perú

Para ilustrar la cuestión de la identificación, considere la oferta de mano de obra de trabajadoras casadas que ya forman parte del mercado laboral. En lugar de la función de la demanda, se escribe la oferta salarial en función de las horas y las variables de productividad acostumbradas. Con la condición de equilibrio impuesta, las dos ecuaciones estructurales son:

$$horas = \alpha 1 \log Salario + \beta 10 + \beta 11 educ + \beta 12 edad + \beta 13 kidslt6 + \beta 14 nwifeinc + u1$$

$$horas = \alpha_1 \log(Salario) + \beta_{10} + \beta_{11} educ + \beta_{12} edad + \beta_{13} kidslt6 + \beta_{14} nwifeinc + u_1 \ldots\ldots(1)$$

y

$$\log(Salario) = \alpha_2 horas + \beta_{20} + \beta_{21} educ + \beta_{22} exper + \beta_{23} exper^2 + u_2 \ldots\ldots\ldots\ldots(2)$$

Una vez planteado nuestro modelo, debemos identificar si es apropiado o no para utiliza una estimación en dos etapas

Identificación de ecuaciones simultaneas "método del orden"

K – k ≥ m – 1

Donde:

K: número de variables predeterminadas en el sistema

K: número de variables predeterminadas solo en la ecuación

M: número de variables endógenas en el sistema

M: número de variables endógenas en la ecuación

$$horas = \alpha_1 \log(Salario) + \beta_{10} + \beta_{11} educ + \beta_{12} edad + \beta_{13} kidslt6 + \beta_{14} nwifeinc + u_1 \ldots\ldots(1)$$

$$\log(Salario) = \alpha_2 horas + \beta_{20} + \beta_{21} educ + \beta_{22} exper + \beta_{23} exper^2 + u_2 \ldots\ldots\ldots\ldots(2)$$

Para el caso de la primera ecuación

tenemos:

2 variables endógenas:

- horas
- salario

6 variables exógenas

- educación
- edad
- edad de los niños menores de 6 años
- ingreso del esposo
- experiencia
- experiencia al cuadrado

para el caso de la primera ecuación tenemos:

$$6 - 4 \geq 2 - 1$$

$$2 \geq 1$$

- ➢ Está sobre identificada la primera ecuación.

para el caso de la segunda ecuación tenemos:

$$6 - 3 \geq 2 - 1$$

$$2 \geq 1$$

- ➢ Está sobre identificada la segunda ecuación.

Como podemos observar nuestro modelo está sobre identificado y por ello que el mejor método para poder estimar nuestro modelo es aplicando mínimos cuadrados en dos etapas.

Descripción del modelo

La variable age es la edad de la mujer, kidslt6 es el número de niños que tienen menos de seis años, nwifeinc es el ingreso no salarial de la mujer (que incluye los ingresos de su cónyuge) y educ
(educación) y exper (experiencia) son los anos de educación y experiencia previa, respectivamente. Se supone que todas las variables, salvo hours y log(wage) son exogenas. La forma funcional en este sistema, donde hours aparece en forma nivelada, pero wage (salario) está en forma logarítmica, es común en economía laboral.

La primera ecuación es la función de la oferta y satisface la condición de orden, pues las dos variables exógenas, exper y exper^2, se omitieron en la ecuación de oferta de mano de obra. Estas restricciones de exclusión son supuestos cruciales: se supone que, una vez que se controlan el salario, la educación, la edad, el número de niños pequeños y otros ingresos, la experiencia pasada no tiene efecto sobre la oferta actual de mano de obra. Desde luego, este supuesto es muy cuestionable, pero se utiliza aquí a modo de ilustración.

Dadas las ecuaciones (1) y (2), la condición de rango para identificar la primera ecuación es que al menos exper o exper^2 tenga un coeficiente diferente de cero en la ecuación (2). Si β_{22} = 0 y β_{23} = 0 no existen variables exógenas que aparezcan en la segunda ecuación y que tampoco aparezca en la primera (educ aparece en ambas). Se puede decir que la condición de rango para la identificación de (1) es equivalente en términos en la forma reducida de log(wage), que es

$$\log(wage) = \pi_{20} + \pi_{21} educ + \pi_{22} age + \pi_{23} kidslt\,6 + \pi_{24} nwifeinc + \pi_{25} exper + \pi_{26} exper^2 + v_2 \ldots\ldots\ldots\ldots (3)$$

La ecuación de la oferta salarial, (2), se identifica si al menos una de las variables age, kidslt6, o nwifeinc tiene un coeficiente diferente de cero en (1). Esto es idéntico a suponer que la forma reducida de hours, que tiene la misma forma que el lado derecho de (3), depende de al menos una de las variables age, kidslt6 o nwifeinc. Cuando se especifica la ecuación de la oferta salarial, se supone que ni age, kidslt6 o nwifeinc tienen efecto alguno en el salario ofrecido, una vez que se ha dado cuenta de las horas, la educación y la experiencia. Estos serían supuestos pobres si las variables tienen, de alguna manera, efectos directos en la productividad, o si las mujeres padecieran discriminación basada en su edad o en el número de hijos pequeños.

Estimación del modelo con mínimos cuadrados (sin aplicar dos etapas)

Al aplicar una estimación de mínimos cuadrados obtuvimos el siguiente resultado (véase el anexo 1):

A. Primera ecuación:

$$Hours = 1523.775 - 2.046796\, Lwage - 6.621869\, Edcu + 0.562254\, Age - 328.8584\, kidslt\,6 - 5.918459\, nwifei$$

Al intentar interpretar nuestro modelo estimado por MCO nos dice que el impacto que tiene el salario ante un incremento de un 1% nos haría ofrecer menos dos horas, es decir no ofrecería 2 horas de trabajo siendo una mujer casada.

B. Segunda ecuación:

$$\log(Salario) = -5.65\,horas - 0.461995 + 0.106214\,duc + 0.044703\,exper - 0.000859\,exper^2$$

Pero estimar por Mínimos Cuadrados Ordinarios cuando tenemos un problema de simultaneidad procuraría tener estimadores insesgados lo cual invalidaría nuestro modelo por completo.

Estimación del modelo mínimos cuadrados en dos etapas

Al realizar las correcciones a nuestro modelo procedemos a estimar nuestro modelo en dos etapas y obtenemos los siguientes resultados (véase el anexo 2). Se utilizan los datos de las mujeres trabajadoras casadas en extraída del Instituto Nacional de Estadística e Informática para estimar la ecuación de oferta de mano de obra (1) mediante MC2E. El conjunto completo de variables instrumentales incluye a educ, age, kidslt6, nwifeinc, exper y exper2. La curva estimada de oferta de mano de obra es:

$$Hours = 2225.662 + 1639.556\,Lwage - 183.7513\,Educ - 7.806092\,Age - 198.1543\,kidslt6 - 10.16959\,nwifei$$

la cual muestra que la curva de oferta de mano de obra tiene pendiente positiva. A simple vista podemos interpretar que al aumentar en 1% el salario, las mujeres están dispuestas en aumentar las horas trabajas en1639.556 aproximadamente. El coeficiente estimado en log(wage) tiene la siguiente interpretación, todos los factores constantes, $\Delta\widehat{hours} \approx$ 16.4(%Δwage). Las elasticidades de la oferta de mano de obra se pueden calcular si se multiplican ambos lados de esta última ecuación por 100/hours:

$$100 * \left(\frac{\Delta\widehat{hours}}{hours}\right) \approx \left(\frac{1{,}640}{hours}\right) ¿$$

lo cual implica que la elasticidad de la oferta de mano de obra (respecto al salario) simplemente es de 1,640/hours. La elasticidad no es constante en este modelo, pues la variable dependiente es hours y no log(hours). Al promedio de horas trabajadas, 1,303, la elasticidad estimada es 1,640/1,303≈1.26, lo cual implica un incremento mayor que 1% en las horas trabajadas dado un incremento de 1% en el salario. Esta es una elasticidad estimada grande. A más horas, la elasticidad será menor; a menos horas, como con hours = 800, la elasticidad será superior a dos.

Para fines comparativos, cuando la primera función se estima mediante MCO, el coeficiente de log(wage) es =2.05, lo cual implica que no hay efecto en las horas trabajadas. Cuando se agregan los residuales de la forma reducida $\hat{v}_2$ a la ecuación y se estima mediante MCO, el estadistico t en $\hat{v}_2$ es -6.61, lo cual es muy significativo y por tanto log(wage) parece ser endógeno.

También para fines comparativos estimaremos la oferta salarial (segunda función) mediante mínimos cuadrados en dos etapas. Al realizarlo obtuvimos que nuestro resultado fue el siguiente (véase el anexo 3):

$$\log(Salario) = 0.000126\,horas - 0.655725 + 0.110330\,educ + 0.034582\,exper - 0.000706\,\beta_{23}\,exper^2$$

Esto difiere de las ecuaciones de salario anteriores en que *hours* se incluye como variable explicativa y MC2E se utiliza para dar cuenta de la endogeneidad de *hours* (y se supone que *educ* y *exper* son exógenas). El coeficiente en *hours* es estadísticamente insignificante, lo cual significa que no hay evidencia de que la oferta salarial aumente con las horas trabajadas. Los demás coeficientes son similares a los que se obtuvieron al eliminar *hours* y estimar la ecuación mediante MCO.

Anexos

a) anexo

Equation: UNTITLED Workfile: DATA_SAMANAMUD::Mroz\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: HOURS
Method: Least Squares
Date: 06/16/19 Time: 21:23
Sample (adjusted): 1 428
Included observations: 428 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1523.775	305.5755	4.986574	0.0000
LWAGE	-2.046796	54.88015	-0.037296	0.9703
EDUC	-6.621869	18.11627	-0.365521	0.7149
AGE	0.562254	5.140012	0.109388	0.9129
KIDSLT6	-328.8584	101.4573	-3.241348	0.0013
NWIFEINC	-5.918459	3.683341	-1.606818	0.1088

R-squared	0.036106	Mean dependent var	1302.930
Adjusted R-squared	0.024686	S.D. dependent var	776.2744
S.E. of regression	766.6331	Akaike info criterion	16.13581
Sum squared resid	2.48E+08	Schwarz criterion	16.19272
Log likelihood	-3447.064	Hannan-Quinn criter.	16.15829
F-statistic	3.161515	Durbin-Watson stat	2.028217
Prob(F-statistic)	0.008173		

Equation: UNTITLED Workfile: DATA_SAMANAMUD::Mroz\

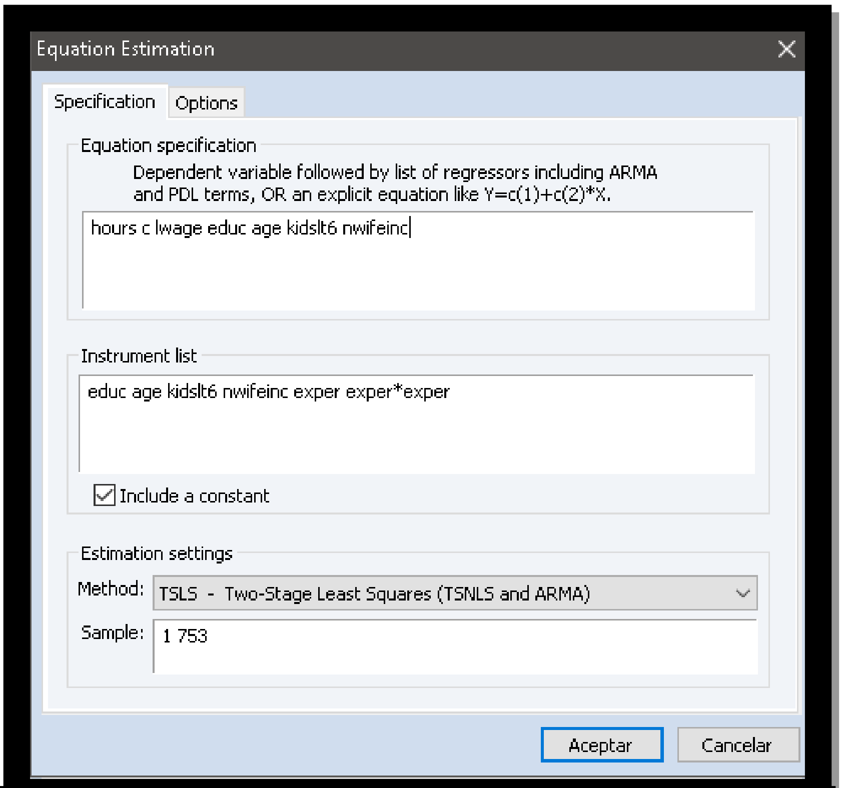
View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

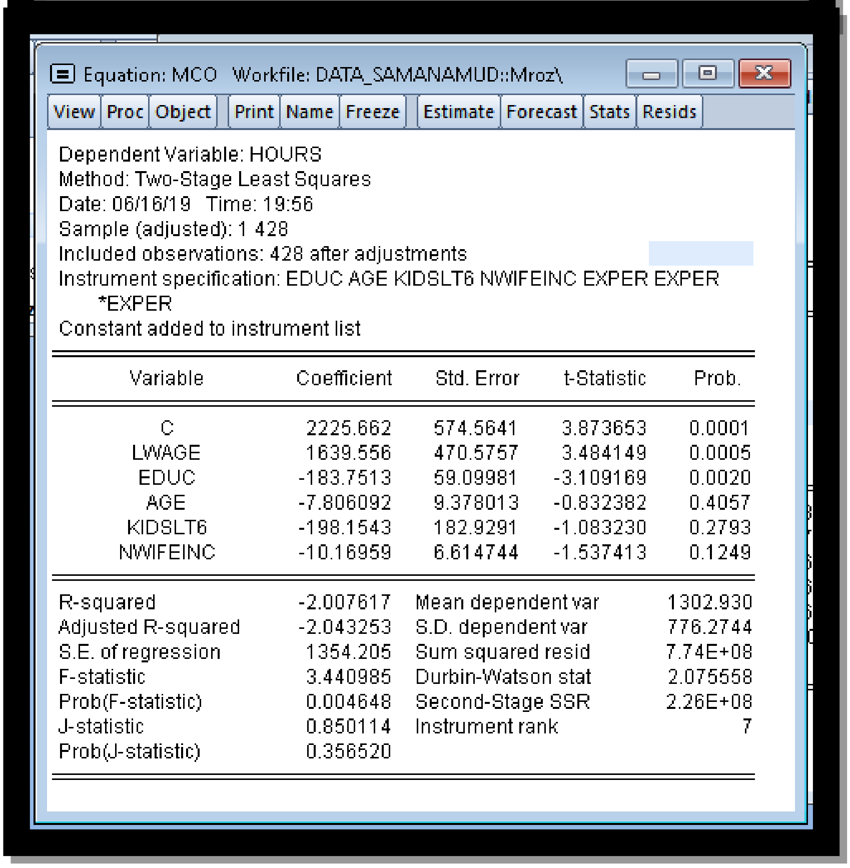
Dependent Variable: LWAGE
Method: Least Squares
Date: 06/17/19 Time: 00:16
Sample (adjusted): 1 428
Included observations: 428 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.461995	0.203848	-2.266376	0.0239
HOURS	-5.65E-05	4.38E-05	-1.291586	0.1972
EDUC	0.106214	0.014170	7.495785	0.0000
EXPER	0.044703	0.013387	3.339321	0.0009
EXPER^2	-0.000859	0.000395	-2.175517	0.0301

R-squared	0.160133	Mean dependent var	1.190173
Adjusted R-squared	0.152191	S.D. dependent var	0.723198
S.E. of regression	0.665896	Akaike info criterion	2.036246
Sum squared resid	187.5654	Schwarz criterion	2.083666
Log likelihood	-430.7567	Hannan-Quinn criter.	2.054974
F-statistic	20.16273	Durbin-Watson stat	1.946394
Prob(F-statistic)	0.000000		

b) Anexo 2



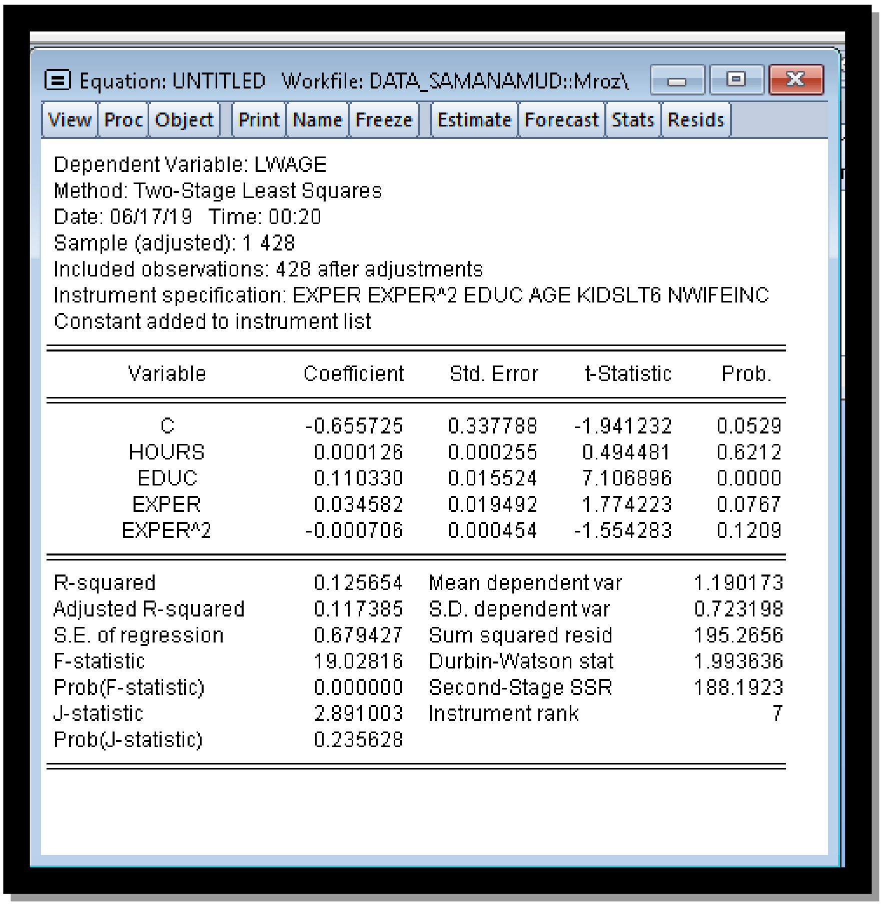


Dependent Variable: HOURS
Method: Two-Stage Least Squares
Date: 06/16/19 Time: 19:56
Sample (adjusted): 1 428
Included observations: 428 after adjustments
Instrument specification: EDUC AGE KIDSLT6 NWIFEINC EXPER EXPER *EXPER
Constant added to instrument list

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2225.662	574.5641	3.873653	0.0001
LWAGE	1639.556	470.5757	3.484149	0.0005
EDUC	-183.7513	59.09981	-3.109169	0.0020
AGE	-7.806092	9.378013	-0.832382	0.4057
KIDSLT6	-198.1543	182.9291	-1.083230	0.2793
NWIFEINC	-10.16959	6.614744	-1.537413	0.1249

R-squared	-2.007617	Mean dependent var	1302.930
Adjusted R-squared	-2.043253	S.D. dependent var	776.2744
S.E. of regression	1354.205	Sum squared resid	7.74E+08
F-statistic	3.440985	Durbin-Watson stat	2.075558
Prob(F-statistic)	0.004648	Second-Stage SSR	2.26E+08
J-statistic	0.850114	Instrument rank	7
Prob(J-statistic)	0.356520		

c) Anexo 3



Equation: UNTITLED Workfile: DATA_SAMANAMUD::Mroz\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: LWAGE
Method: Two-Stage Least Squares
Date: 06/17/19 Time: 00:20
Sample (adjusted): 1 428
Included observations: 428 after adjustments
Instrument specification: EXPER EXPER^2 EDUC AGE KIDSLT6 NWIFEINC
Constant added to instrument list

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.655725	0.337788	-1.941232	0.0529
HOURS	0.000126	0.000255	0.494481	0.6212
EDUC	0.110330	0.015524	7.106896	0.0000
EXPER	0.034582	0.019492	1.774223	0.0767
EXPER^2	-0.000706	0.000454	-1.554283	0.1209

R-squared	0.125654	Mean dependent var	1.190173
Adjusted R-squared	0.117385	S.D. dependent var	0.723198
S.E. of regression	0.679427	Sum squared resid	195.2656
F-statistic	19.02816	Durbin-Watson stat	1.993636
Prob(F-statistic)	0.000000	Second-Stage SSR	188.1923
J-statistic	2.891003	Instrument rank	7
Prob(J-statistic)	0.235628		

Bibliografía

- Carrascal, U., Y. González y B. Rodríguez. Análisis Econométrico con Eviews . Ra-Ma, Madrid, 2001.
- Goldberger, A. Introducción a la Econometría. Ariel Economía, Barcelona, 2001.
- Greene, W. Análisis Econométrico. Prentice Hall, Madrid, 1998.
- Gujarati, D. Econometría. McGraw-Hill, México, 2003.
- Johnston, J. y J. Dinardo. Métodos de Econometría. Vicens Vives, Barcelona, 2001.
- Pena J.B., J. Estavillo, E. Galindo, M. J. Leceta y M. M. Zamora. Cien Ejercicios de Econometría.
- Pirámide, Madrid, 1999.
- Pérez López, C. Problemas resueltos de Econometría, Thomson, Madrid, 2006.
- Pulido, A. y J. Pérez, 2001, Modelos Econométricos. Guia para la Elaboración de Modelos
- Econométricos con Eviews. Pirámide, Madrid.
- Trívez, J. Introducción a la Econometría. Pirámide, Madrid, 2004.