Econometria Aplicada Clase 5 Método Variables Instrumentales

Edinson Tolentino email: edinson.tolentino@gmail.com

Twitter: @edutoleraymondi

Educate Peru

30 de marzo de 2024

Contenido



Introducción

Se termino todo en MCO? Variables Endogenas Metodo de Variable Instrumental (VI)

Analisis

Pregunta 1

Pregunta 2

Pregunta 3

Sesgo en muestras finitas

Pregunta 4

Pregunta 5

Pregunta 6

Appendix

Roadmap



Introducción

Se termino todo en MCO? Variables Endogenas Metodo de Variable Instrumental (VI)

Analisis

Pregunta :

Pregunta 2

Pregunta 3

Sesgo en muestras finitas

Pregunta -

Pregunta 5

Pregunta 6

Appendix





La metodología de Variable Instrumental (VI) busca la identificación causal que fue originalmente desarrollada por Philip Greeen Wrigth (1861-1934), matematico estadounidense y economista



- La metodología de Variable Instrumental (VI) busca la identificación causal que fue originalmente desarrollada por Philip Greeen Wrigth (1861-1934), matematico estadounidense y economista
- ► El autor trabajo como un esperto especial dentro de la Comisión de Tarifas en USA durante 1920s



- La metodología de Variable Instrumental (VI) busca la identificación causal que fue originalmente desarrollada por Philip Greeen Wrigth (1861-1934), matematico estadounidense y economista
- El autor trabajo como un esperto especial dentro de la Comisión de Tarifas en USA durante 1920s
- El autor estuvo preocupado con la identificación causal de los efectos de precios dado la relación de demanda y oferta, donde comprendio que no solo bastaba una relación entre la cantidad y precio

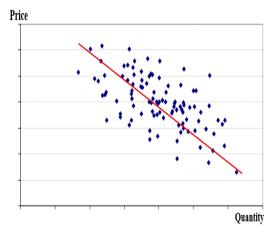


- La metodología de Variable Instrumental (VI) busca la identificación causal que fue originalmente desarrollada por Philip Greeen Wrigth (1861-1934), matematico estadounidense y economista
- El autor trabajo como un esperto especial dentro de la Comisión de Tarifas en USA durante 1920s
- El autor estuvo preocupado con la identificación causal de los efectos de precios dado la relación de demanda y oferta, donde comprendio que no solo bastaba una relación entre la cantidad y precio
- En el apendeci de su trabajo en 1928, the tariff on Animal and Vegetable Oils describe la derivación del estimador de IV

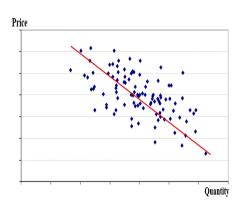




► Como identificar los parametros de una relación de demanda?





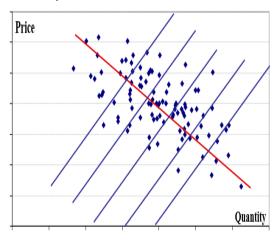


- ¿Cómo identificar los parametros de la relación de la demanda?
- La identificación de los parametros de la demanda a través de factores que cambien la oferta
- Por el otro lado, los parametros de la relación de oferta a través de factores de cambios de demanda
- Estos cambios o factores llegan a ser conocidos como variables de identificación o intrumentos de identificación



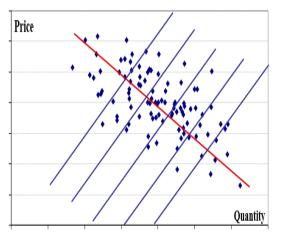


► Como identificar los parametros de una relación de demanda?





Como identificar los parametros de una relación de demanda?



A través de variables que cambian la relación de oferta



- La fundamentación de Cowles dentro de su investigación economica fue establecida en los Estados Unidos en 1932 con la teoria y medición
- El enfasis original fue la estimación de ecuaciones simultaneas para el modelo en una economía , en orden de la obtención de politicas económicas
- Uno de sus grandes contribuciones en 1940s fue el desarrollo más sistemático de las condiciones y marco teórico para el uso del estimador IV
- ► En terminos generales lo que se conoce como el uso de la metodología IV



► Recordando el modelo matricial de MCO :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + + \beta_n X_n + \mu$$



▶ Recordando el modelo matricial de MCO :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + + \beta_n X_n + \mu$$

lacktriangle La metodologia de MCO utilzia supuestos que descansan en los μ



► Recordando el modelo matricial de MCO :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + + \beta_n X_n + \mu$$

- lacktriangle La metodologia de MCO utilzia supuestos que descansan en los μ
- Es asi que tenemos los siguientes supuestos :



▶ Recordando el modelo matricial de MCO :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + + \beta_n X_n + \mu$$

- lacktriangle La metodologia de MCO utilzia supuestos que descansan en los μ
- Es asi que tenemos los siguientes supuestos :
 - 1. En promedio los errores deben sumar cero , expresado a traves de la notación $E\left[\mu\right]=0$



▶ Recordando el modelo matricial de MCO :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + + \beta_n X_n + \mu$$

- lacktriangle La metodologia de MCO utilzia supuestos que descansan en los μ
- Es asi que tenemos los siguientes supuestos :
 - 1. En promedio los errores deben sumar cero , expresado a traves de la notación $E\left[\mu\right]=0$
 - 2. La varianza de los errores se mantiene constante, notación: $\mathit{Var}\left(\mu\right)=\sigma^{2}$



▶ Recordando el modelo matricial de MCO :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_n X_n + \mu$$

- lacktriangle La metodologia de MCO utilzia supuestos que descansan en los μ
- Es asi que tenemos los siguientes supuestos :
 - 1. En promedio los errores deben sumar cero , expresado a traves de la notación $E\left[\mu\right]=0$
 - 2. La varianza de los errores se mantiene constante, notación: $\mathit{Var}\left(\mu\right)=\sigma^{2}$
 - 3. Exogenidad de los errores: es decir que los errores no tengan ninguna relación con las variables explicativas, expresado a traves de la notación: $E\left[\mu,X\right]=0$



Recordando el modelo matricial de MCO :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + + \beta_n X_n + \mu$$

- lacktriangle La metodologia de MCO utilzia supuestos que descansan en los μ
- Es asi que tenemos los siguientes supuestos :
 - 1. En promedio los errores deben sumar cero , expresado a traves de la notación $E\left[\mu\right]=0$
 - 2. La varianza de los errores se mantiene constante, notación: $\mathit{Var}\left(\mu\right)=\sigma^{2}$
 - 3. Exogenidad de los errores: es decir que los errores no tengan ninguna relación con las variables explicativas, expresado a traves de la notación: $E\left[\mu,X\right]=0$
- Es sobre este ultimo supuesto, cuando no se cuemple la exogenidad de los errores (es decir: $E\left[\mu,X\right] \neq 0$), sobre el cual se utilizará la metodología conocida como variable instrumental

Regresores endogenos: $E[\mu, X] \neq 0$



Regresores endogenos: $E[\mu, X] \neq 0$



$$\widehat{\beta}_{MCO} = (X'X)^{-1}X'Y$$

Regresores endogenos: $E[\mu, X] \neq 0$



$$\widehat{\beta}_{MCO} = \left(X'X \right)^{-1} X'Y$$

$$E\left[\widehat{\beta}_{MCO}\right] = \beta$$



▶ Una de las consecuencias de la $E[\mu, X] = 0$, descansa en el supuesto de insesgadez de los parametros, lo recuerda?

$$\widehat{\beta}_{MCO} = (X'X)^{-1}X'Y$$

$$E\left[\widehat{\beta}_{MCO}\right] = \beta$$

Recuerden que la al no cumplirse el supuesto de exogenidad $(E[\mu,X]\neq 0)$, tendremos que el estimador de MCO se vuelve sesgado.



$$\widehat{\beta}_{MCO} = (X'X)^{-1}X'Y$$

$$E\left[\widehat{\beta}_{MCO}\right] = \beta$$

- Recuerden que la al no cumplirse el supuesto de exogenidad $(E[\mu,X]\neq 0)$, tendremos que el estimador de MCO se vuelve sesgado.
- Entonces cuando la variable explicativa esta relacionada con el termino de error diremos que dicha variable es endogena





En este ejemplo nosotros estamos interesados en medir el efecto de los años de educación sobre los salarios de una persona. Es decir cuando influye la eduación en los ingresos que percibe una persona.



- En este ejemplo nosotros estamos interesados en medir el efecto de los años de educación sobre los salarios de una persona. Es decir cuando influye la eduación en los ingresos que percibe una persona.
- Entonces la pregunta seria, ¿Cuál seria la mejor forma de capturar (es decir medir el $\hat{\beta}$) dicho efecto?, ¿Que modelo lineal podria sugerir?:



- En este ejemplo nosotros estamos interesados en medir el efecto de los años de educación sobre los salarios de una persona. Es decir cuando influye la eduación en los ingresos que percibe una persona.
- Entonces la pregunta seria, ¿Cuál seria la mejor forma de capturar (es decir medir el $\widehat{\beta}$) dicho efecto?, ¿Que modelo lineal podria sugerir?:

$$log(Salario_i) = \beta_0 + \beta_1 Educa_i + \varepsilon_i$$



- En este ejemplo nosotros estamos interesados en medir el efecto de los años de educación sobre los salarios de una persona. Es decir cuando influye la eduación en los ingresos que percibe una persona.
- Entonces la pregunta seria, ¿Cuál seria la mejor forma de capturar (es decir medir el $\hat{\beta}$) dicho efecto?, ¿Que modelo lineal podria sugerir?:

$$log(Salario_i) = \beta_0 + \beta_1 Educa_i + \varepsilon_i$$

¿Que otras variables pensaria usted que pueden influir en ladeterminanción del salario de una persona muy aparte de la eduación?



- En este ejemplo nosotros estamos interesados en medir el efecto de los años de educación sobre los salarios de una persona. Es decir cuando influye la eduación en los ingresos que percibe una persona.
- Entonces la pregunta seria, ¿Cuál seria la mejor forma de capturar (es decir medir el $\hat{\beta}$) dicho efecto?, ¿Que modelo lineal podria sugerir?:

$$log(Salario_i) = \beta_0 + \beta_1 Educa_i + \varepsilon_i$$

- ¿Que otras variables pensaria usted que pueden influir en ladeterminanción del salario de una persona muy aparte de la eduación?
 - 1. La eduación del padre y/o madre



- En este ejemplo nosotros estamos interesados en medir el efecto de los años de educación sobre los salarios de una persona. Es decir cuando influye la eduación en los ingresos que percibe una persona.
- Entonces la pregunta seria, ¿Cuál seria la mejor forma de capturar (es decir medir el $\hat{\beta}$) dicho efecto?, ¿Que modelo lineal podria sugerir?:

$$log(Salario_i) = \beta_0 + \beta_1 Educa_i + \varepsilon_i$$

- ¿Que otras variables pensaria usted que pueden influir en ladeterminanción del salario de una persona muy aparte de la eduación?
 - 1. La eduación del padre y/o madre
 - 2. Las habilidades y/o capacidades de la persona



- En este ejemplo nosotros estamos interesados en medir el efecto de los años de educación sobre los salarios de una persona. Es decir cuando influye la eduación en los ingresos que percibe una persona.
- Entonces la pregunta seria, ¿Cuál seria la mejor forma de capturar (es decir medir el $\hat{\beta}$) dicho efecto?, ¿Que modelo lineal podria sugerir?:

$$log(Salario_i) = \beta_0 + \beta_1 Educa_i + \varepsilon_i$$

- ¿Que otras variables pensaria usted que pueden influir en ladeterminanción del salario de una persona muy aparte de la eduación?
 - 1. La eduación del padre y/o madre
 - 2. Las habilidades y/o capacidades de la persona
- Entonces nuestro modelo planteado sería:

$$log(Salario_i) = \beta_0 + \beta_1 Educa_i + \beta_2 Habilidades_i + \varepsilon_i$$



- En este ejemplo nosotros estamos interesados en medir el efecto de los años de educación sobre los salarios de una persona. Es decir cuando influye la eduación en los ingresos que percibe una persona.
- Entonces la pregunta seria, ¿Cuál seria la mejor forma de capturar (es decir medir el $\hat{\beta}$) dicho efecto?, ¿Que modelo lineal podria sugerir?:

$$log(Salario_i) = \beta_0 + \beta_1 Educa_i + \varepsilon_i$$

- ¿Que otras variables pensaria usted que pueden influir en ladeterminanción del salario de una persona muy aparte de la eduación?
 - 1. La eduación del padre y/o madre
 - 2. Las habilidades y/o capacidades de la persona
- Entonces nuestro modelo planteado sería:

$$log(Salario_i) = \beta_0 + \beta_1 Educa_i + \beta_2 Habilidades_i + \varepsilon_i$$

Sin embargo existe un problema?, identifica cual es?





► La habilidad o capacidad de una persona, no es una variable medible en cada persona, por lo cual nuestro modelo seria:

$$log(Salario_i) = \beta_0 + \beta_1 Educa_i + v_i$$



La habilidad o capacidad de una persona, no es una variable medible en cada persona, por lo cual nuestro modelo seria:

$$log(Salario_i) = \beta_0 + \beta_1 Educa_i + v_i$$

Donde v_i esta correlacionado con una variable explicativa (en este caso Educa_i, ¿Porqué?), es decir tenemos que:

$$E[v_i, Educa_i] \neq 0$$



► La habilidad o capacidad de una persona, no es una variable medible en cada persona, por lo cual nuestro modelo seria:

$$log(Salario_i) = \beta_0 + \beta_1 Educa_i + v_i$$

Donde v_i esta correlacionado con una variable explicativa (en este caso Educa_i, ¿Porqué?), es decir tenemos que:

$$E[v_i, Educa_i] \neq 0$$

Esto originado porque:

$$E[Educa_i, Habilidades_i] \neq 0$$



La habilidad o capacidad de una persona, no es una variable medible en cada persona, por lo cual nuestro modelo seria:

$$log(Salario_i) = \beta_0 + \beta_1 Educa_i + v_i$$

▶ Donde v_i esta correlacionado con una variable explicativa (en este caso Educa_i, ¿Porqué?), es decir tenemos que:

$$E[v_i, Educa_i] \neq 0$$

Esto originado porque:

$$E[Educa_i, Habilidades_i] \neq 0$$

Las primeras consecuencias son trasladas a los parametros del modelo, en particular el parametro de nuestro interes sera: $\hat{\beta}_1$, el cual ya no cumple el supuesto de Insesgadez (al ser $E\left[Educa_i,v_i\right] \neq 0$).



► Siendo el modelo

$$Y = X\beta + \mu$$

▶ donde tenemos el problema siguiente: $E[X, \mu] \neq 0$, por tanto el metodo de variables instrumentales sera:



► Siendo el modelo

$$Y = X\beta + \mu$$

▶ donde tenemos el problema siguiente: $E[X, \mu] \neq 0$, por tanto el metodo de variables instrumentales sera:

$$egin{array}{cccc} {\sf Y} & \longrightarrow & {\sf X} & \longrightarrow & \mu \ & & \downarrow & & \ & {\sf Z} & & & \end{array}$$

Entonces nuestro instrumento (Z) debe cumplir con los siguientes supuestos:



► Siendo el modelo

$$Y = X\beta + \mu$$

▶ donde tenemos el problema siguiente: $E[X, \mu] \neq 0$, por tanto el metodo de variables instrumentales sera:

$$egin{array}{cccc} {\sf Y} & \longrightarrow & {\sf X} & \longrightarrow & \mu \ & & \downarrow & & \ & {\sf Z} & & & \end{array}$$

- Entonces nuestro instrumento (Z) debe cumplir con los siguientes supuestos:
 - 1. Condición de relevancia: $E[Z'X] \neq 0$



Siendo el modelo

$$Y = X\beta + \mu$$

▶ donde tenemos el problema siguiente: $E[X, \mu] \neq 0$, por tanto el metodo de variables instrumentales sera:

$$egin{array}{cccc} {\sf Y} & \longrightarrow & {\sf X} & \longrightarrow & \mu \ & & \downarrow & & \ & {\sf Z} & & & \end{array}$$

- Entonces nuestro instrumento (Z) debe cumplir con los siguientes supuestos:

 - 1. Condición de relevancia: $E[Z'X] \neq 0$ 2. Condición de exogenidad: $E[Z'\mu] = 0$



► Teniendo la ecuación estructural:

$$y = x\beta + \mu \tag{1}$$



► Teniendo la ecuación estructural:

$$y = x\beta + \mu \tag{1}$$

ightharpoonup Multiplicandola por Z', obtenemos

$$z'y = z'x\beta + z'\mu \tag{2}$$

lacktriangle Entonces tenemos que si deseamos el valor de eta



► Teniendo la ecuación estructural:

$$y = x\beta + \mu \tag{1}$$

ightharpoonup Multiplicandola por Z', obtenemos

$$z'y = z'x\beta + z'\mu \tag{2}$$

ightharpoonup Entonces tenemos que si deseamos el valor de eta

$$z'x\beta = z'y + z'\mu$$



► Teniendo la ecuación estructural:

$$y = x\beta + \mu \tag{1}$$

ightharpoonup Multiplicandola por Z', obtenemos

$$z'y = z'x\beta + z'\mu \tag{2}$$

ightharpoonup Entonces tenemos que si deseamos el valor de β

$$z'x\beta = z'y + z'\mu$$

$$\beta = E(z'y)(z'x)^{-1} + E(z'\mu)$$



► Teniendo la ecuación estructural:

$$y = x\beta + \mu \tag{1}$$

ightharpoonup Multiplicandola por Z', obtenemos

$$z'y = z'x\beta + z'\mu \tag{2}$$

ightharpoonup Entonces tenemos que si deseamos el valor de eta

$$z'x\beta = z'y + z'\mu$$

$$\beta = E(z'y)(z'x)^{-1} + E(z'\mu)$$

$$\beta = E \left(z'x \right)^{-1} E \left(z'y \right)$$

▶ Entonces el estimador de variables instrumental sera:



Teniendo la ecuación estructural:

$$y = x\beta + \mu \tag{1}$$

ightharpoonup Multiplicandola por Z', obtenemos

$$z'y = z'x\beta + z'\mu \tag{2}$$

ightharpoonup Entonces tenemos que si deseamos el valor de eta

$$z'x\beta = z'y + z'\mu$$

$$\beta = E(z'y)(z'x)^{-1} + E(z'\mu)$$

$$\beta = E(z'x)^{-1} E(z'y)$$

Entonces el estimador de variables instrumental sera:

$$\beta_{IV} = \left[\frac{1}{n} \sum_{i=1}^{n} z_i' x_i \right]^{-1} \left[\frac{1}{n} \sum_{i=1}^{n} z_i' y_i \right] \equiv (Z'X)^{-1} Z'Y$$
 (3)

Roadmap



Introducción

Se termino todo en MCO? Variables Endogenas Metodo de Variable Instrumental (VI)

Analisis

Pregunta :

Pregunta 2

Pregunta 3

Sesgo en muestras finitas

Pregunta -

Pregunta 5

Pregunta 6

Appendix

Descripción de Variables



Variable	Description	
logwage	The natural log of the individual's hourly wage.	
female	A dummy variable adopting a value of 1 if the individual is female and 0 if male.	
educ	The number of years in post-compulsory education.	
test11r	The score achieved by the individual in a reading test at age 11 years.	
test11m	The score achieved by the individual in a maths test at age 11 years.	
Mumso	A dummy variable adopting a value of 1 if the individual's mother stayed on in post-compulsory education and 0 otherwise.	
Dadso	A dummy variable adopting a value of 1 if the individual's father stayed on in post-compulsory education and 0 otherwise.	

Descripción de Variables



Variable	Description		
logwage	The natural log of the individual's hourly wage.		
female	A dummy variable adopting a value of 1 if the individual is female and 0 if male.		
educ	The number of years in post-compulsory education.		
test11r	The score achieved by the individual in a reading test at age 11 years.		
test11m	The score achieved by the individual in a maths test at age 11 years.		
mumso	A dummy variable adopting a value of 1 if the individual's mother stayed on in post-compulsory education and 0 otherwise.		
dadso	A dummy variable adopting a value of 1 if the individual's father stayed on in post-compulsory education and 0 otherwise.		



- La informacion contiene un 50 % de jefe de hogares de mujeres , en promedio.
- ▶ En promedio, los individuos en la muestra presentan 3.5 años de educacion

Cuadro: Estadisticas descriptivas

	Personas	Promedio	Mediana	Min.	Max.	Std
logwage	1000	1.81	1.76	0.00	4.54	0
female	1000	0.50	1.00	0.00	1.00	0
educ	1000	3.46	3.00	0.00	8.00	2
test11r	1000	47.72	45.71	0.00	100.00	17
test11m	1000	44.84	42.50	0.00	100.00	26
mumso	1000	0.27	0.00	0.00	1.00	0
dadso	1000	0.22	0.00	0.00	1.00	0

Fuente: ncds. Elaboracion: Autor

Roadmap



Introducción

Se termino todo en MCO? Variables Endogenas Metodo de Variable Instrumental (VI)

Analisis

Pregunta 1

Pregunta 2

Pregunta 3

Sesgo en muestras finitas

Pregunta -

Pregunta 5

Pregunta 6

Appendix



- Objetivo de la Investigación deseamos medir la identificación causal de los efectos de educación sobre los salarios
- ▶ Se propone la sigueinte ecuación de salarios usando MCO:

$$log(wage_i) = \alpha + \beta educ_i + \pi female_i + \mu_i$$

- Cuales son problemas del regresos asociaciados a la variable educ
- Especialmente, se tiene un potencial problema de

$$E\left(educ_{i},\mu_{i}\right)\neq0$$

Mecanismo IV



► Ecuacion (1)

$$\log(wage_i) = \alpha + \beta educ_i + \pi female_i + \mu_i$$

- ► Resultados en (→ stata)
- ► Resultados en (► Python)

Cuadro: Ecuación (1)

	Log(wage)
educ	0.076***
	(0.006)
female	-0.295***
Terriare	(0.028)
	(0.020)
_cons	1.695***
	(0.030)
Observaciones	1000
R ²	0.222
К	0.222

Fuente: ncds. Elaboracion: Autor ***, **, * denote statistical significance at the 1 %, 5 % and 10 % levels res-

pectively for zero.

Roadmap



Introducción

Se termino todo en MCO? Variables Endogenas Metodo de Variable Instrumental (VI)

Analisis

Pregunta 1

Pregunta 2

Pregunta 3

Sesgo en muestras finitas

Pregunta

Pregunta 5

Pregunta 6

Appendix

Mecanismo IV



- Las posibles rasones para el uso de metodo de variables instrumenales o IV seran:
 - Sesgo de habilidad: El estimador de educ podria estar sobre estimado
 - Medición de error: El estimador de educ podria estar sub estimado



 La identificación de instrumentos necesitan ser satisfechos a través de las siguientes condiciones de momentos

$$E\left(test11r_{i},educ_{i}\right) \neq 0$$

$$E\left(test11m_i, educ_i\right) \neq 0$$

$$E\left(dadso_{i}, educ_{i}\right) \neq 0$$

$$E\left(mumso_{i}, educ_{i}\right) \neq 0$$

► Relevancia de Instrumentos o restricción de relevancia



La identificación de instrumentos tambien necesita satisfacer las siguientes condiciones de momentos:

$$E\left(test11r_{i},\mu_{i}
ight)=0$$
 $E\left(test11m_{i},\mu_{i}
ight)=0$
 $E\left(dadso_{i},\mu_{i}
ight)=0$
 $E\left(mumso_{i},\mu_{i}
ight)=0$

 Ortogonalidad de Instrumentos (O instrumentos exogenos o restricción de exclusion)

Mecanismo IV



Por tanto:

Relevancia-Instrumentos + Ortogonalidad-Instrumentos = Instrumentos Valido

Sin embargo y mas importante aún, necesitamos que sea plausible la narrativa (teoria o historia) del uso de este instrumento.



- ► Realice la estimación de la siguiente relación por OLS
 - $\textit{edu}_i = \phi_0 + \phi_1 \textit{female}_i + \phi_2 \textit{testr}_i + \phi_3 \textit{testm}_i + \phi_4 \textit{mumso}_i + \phi_5 \textit{dadso}_i + \epsilon_i$
- ¿Qué tipo de ecuación se tiene?



Esta es una ecuación en su forma reducida



- Esta es una ecuación en su forma reducida
- ▶ Se realiza una predicción de la variable $\overrightarrow{educ_i}$ desde la ecuación propuesta



- Esta es una ecuación en su forma reducida
- ► Se realiza una predicción de la variable educ; desde la ecuación propuesta
- Luego del cual, dicha preducción sera usada en la estimación de MCO

$$log(wage_i) = \alpha + \widehat{\beta educ_i} + \pi female_i + \mu_i$$



- Esta es una ecuación en su forma reducida
- Se realiza una predicción de la variable educ; desde la ecuación propuesta
- Luego del cual, dicha preducción sera usada en la estimación de MCO

$$log(wage_i) = \alpha + \widehat{\beta educ_i} + \pi female_i + \mu_i$$

Para este caso $\widehat{educ_i}$ es ahora una variable independiente de μ_i , entonces el estimador de OLS es consistente

Mecanismo IV



Cuadro: Metodo IV

	Y=Educ	Y = log(wage)
female	-0.056	-0.283***
	(0.117)	(0.029)
test11r	0.032***	
testili	(0.005)	
	(0.003)	
test11m	0.031***	
	(0.003)	
	0.567***	
mumso		
	(0.143)	
dadso	0.539***	
	(0.153)	
	, ,	
educhat_x		0.110***
		(0.010)
cons	0.304	1.571***
	(0.185)	(0.042)
		. ,
Observaciones	1000	1000
R^2	0.383	0.197

Fuente: ncds.

Elaboracion: Autor

***, **, * denote statistical significance at the 1%, 5% and 10% levels respectively for zero.

Mecanismo IV



Cuadro: Metodo IV

	MCO	IV
educ	0.076***	0.110***
	(0.006)	(0.010)
female	-0.295***	-0.283***
	(0.028)	(0.029)
_cons	1.695***	1.571***
	(0.030)	(0.042)
Observaciones	1000	1000
R ²	0.222	0.198

Fuente: ncds.

Elaboracion: Autor ***, **, * denote statistical significance at the $1\,\%,\,5\,\%$ and $10\,\%$ levels respectively for zero.

Roadmap



Introducción

Se termino todo en MCO? Variables Endogenas Metodo de Variable Instrumental (VI)

Analisis

Pregunta 1

Pregunta 2

Pregunta 3

Sesgo en muestras finitas

Pregunta

Pregunta 5

Pregunta 6

Appendix

Condición de Relevancia





¿Cómo se puede realizar una prueba (test) para la relevancia de un instrumento?



¿Cómo se puede realizar una prueba (test) para la relevancia de un instrumento?

$$edu_i = \phi_0 + \phi_1$$
 female $_i + \phi_2$ test $r_i + \phi_3$ test $m_i + \phi_4$ mums $o_i + \phi_5$ dads $o_i + \varepsilon_i$



¿Cómo se puede realizar una prueba (test) para la relevancia de un instrumento?

$$\textit{edu}_{\textit{i}} = \phi_0 + \phi_1 \textit{female}_{\textit{i}} + \phi_2 \textit{testr}_{\textit{i}} + \phi_3 \textit{testm}_{\textit{i}} + \phi_4 \textit{mumso}_{\textit{i}} + \phi_5 \textit{dadso}_{\textit{i}} + \epsilon_{\textit{i}}$$

- Dicha ecuación describe la ecuación en su forma reducida
- ▶ La proposición del test en su forma reducida de la ecuación educ establece el análisis de la relevancia del instrumento si:

$$H_0: \phi_2 = \phi_3 = \phi_4 = \phi_5 = 0$$
 versus $H_a: H_o$ no es verdad



► Resultados en (► stata)

Cuadro: Forma Reducida

	1ra Etapa
female	-0.056
	(0.117)
test11r	0.032***
	(0.005)
test11m	0.031***
	(0.003)
mumso	0.567***
	(0.143)
dadso	0.539***
	(0.153)
_cons	0.304
	(0.185)
Observaciones	1000
R^2	0.383
Test-F	153.6788851488095
	·

Fuente: ncds. Elaboracion: Autor ***, **, * denote statistical significance at the 1%, 5% and 10% levels respectively for zero.

Roadmap



Introducción

Se termino todo en MCO? Variables Endogenas Metodo de Variable Instrumental (VI)

Analisis

Pregunta 1

Pregunta 2

Pregunta 3

Sesgo en muestras finitas

Pregunta 4

Pregunta 5

Pregunta 6

Appendix



- Por que es tan importante la relevancia de un instrumento?
- ► Hahn & Hausman (2005) demuestran que para muestras finitas puede existir un sesgo en el estimador IV , el cual puede ser expresado como:

$$E\left[\widehat{\beta}_{IV} - \beta\right] = \frac{Lp\left(1 - R^2\right)}{nR^2}$$

- **D** Donde β es el verdadero valor del parametros, y :
 - L = el numero de instrumentos identificados
 - $\rho = corr(\mu_i, \varepsilon_i)$ la fuerza de la endogenidad
 - $R^2 = el R^2$ de la forma reducida de la ecuación
 - > n= es el numero de observaciones





 Sesgo en muestras finitas: existen tres ideas proporcionales para este expresión



- Sesgo en muestras finitas: existen tres ideas proporcionales para este expresión
- ► En muestras finitas el sesgo puede incrementarse por:
 - Numero de instrumentos
 - ► Tamaño de la endogenidad
 - Debilidad de los instrumentos , medido como:

$$(1 - R^2)$$





Asumiento la fuerza de la endogenidad de 0.5

$$IV - Finite - Smaple - Bias = \frac{4x0.5x(1 - 0.3856)}{1000x0.3856} = 0.0032$$

- Este es un sesgo bastante insignificante, por tanto, en este caso es probable que se exagere dada la fuerza de la endogeneidad asumida aquí.
- Asimismo, el uso de la condición de relevancia o instrumento relevante asegura que el estimador IV converge hacia una distribución normal asintótica.

Roadmap



Introducción

Se termino todo en MCO? Variables Endogenas Metodo de Variable Instrumental (VI)

Analisis

Pregunta 1

Pregunta 2

Pregunta 3

Sesgo en muestras finitas

Pregunta 4

Pregunta!

Pregunta 6

Appendix





Los residuos de la forma estructural $(\widehat{\mu}_{iv})$, ¿Completa las condiciones de un instrumento?



- Los residuos de la forma estructural $(\widehat{\mu}_{iv})$, ¿Completa las condiciones de un instrumento?
- Usando la información obtenida, se realizará el test relevante del mismo



- Los residuos de la forma estructural $(\widehat{\mu}_{iv})$, ¿Completa las condiciones de un instrumento?
- Usando la información obtenida, se realizará el test relevante del mismo
- ¿Qué concluye?



► Resultados del metodo de Variable Instrumentales

Cuadro: Metodo IV

	Logwage
educ	0.110***
	(0.010)
female	-0.283***
remare	(0.029)
	, ,
_cons	1.571***
	(0.042)
Observaciones	1000
R^2	0.198

Fuente: ncds.

Elaboracion: Autor ***, **, * denote statistical significance at the 1%, 5% and 10% levels respectively for zero.

► Resultados en (► stata)





 Una regresión auxiliar es usada para el calulo del Test de Sargan , para poder testear la sobre-identificación

$$\widehat{\mu}_{iv} = \gamma_0 + \gamma_1$$
 female; $+ \gamma_2$ testr; $+ \gamma_3$ testm;

$$\gamma_4$$
mumso $_i + \gamma_5$ dadso $_i + \varepsilon_i$

Donde la Hipotesisi Nula sera:

$$H_0: \gamma_2 = \gamma_3 = \gamma_4 = \gamma_5 = 0$$

 $H_a: H_0$ no es verdad

El tes de Sargan sera calculad usando el Multiplicador de Lagrange (LM) sera calculado como:

$$Sargan = nxR^2 \sim \chi_r^2$$

- Donde *n* es el tamaño de la muestra
- $ightharpoonup R^2$ es el R^2 ajustado de la regresión auxiliar
- Por ultimo, r es el nñumero de restricciones en la sobre- identificación



- La columna (1) evidencia la inclusion de todos los instrumentos (stata)
- La columna (2) se excluye la variable dadso como instrumento (stata)

Cuadro: Test de Sargan

	μ̂_iv	$\widehat{\mu}^r$ _iv
female	-0.002	-0.001
	(0.029)	(0.029)
test11r	-0.000	0.000
	(0.001)	(0.001)
test11m	-0.001	-0.000
	(0.001)	(0.001)
mumso	0.016	0.043
	(0.035)	(0.034)
dadso	0.101***	
	(0.038)	
_cons	0.028	0.006
	(0.046)	(0.045)
Observaciones	1000	1000
R^2	0.00412	-0.00219

Fuente: BD5.

Elaboracion: Autor

***, **, * denote statistical significance at the 1%, 5% and 10% levels respectively for zero.





 El Test Sargan es potencialmente debil cuando existe un gran número de instrumentos usados



- El Test Sargan es potencialmente debil cuando existe un gran número de instrumentos usados
- Por tanto, es importante no sobre instrumentalizar porque podriamos incurrir en los terminos de implicancias de muestras finitas



- El Test Sargan es potencialmente debil cuando existe un gran número de instrumentos usados
- Por tanto, es importante no sobre instrumentalizar porque podriamos incurrir en los terminos de implicancias de muestras finitas
- Sin embargo, poder encontrar un buen instrumento (o instrumenos fuertes) no es una tarea facil

Roadmap



Introducción

Se termino todo en MCO? Variables Endogenas Metodo de Variable Instrumental (VI)

Analisis

Pregunta 1

Pregunta 2

Pregunta 3

Sesgo en muestras finitas

Pregunta 4

Pregunta 5

Pregunta 6

Appendix





► El test de Wu-hausman se puede estimar de la siguiente forma (dado el ejemplo)

$$logwage_i = \alpha + \beta educ_i + \pi female_i + \phi \widehat{\epsilon}_i + \nu_i$$



► El test de Wu-hausman se puede estimar de la siguiente forma (dado el ejemplo)

$$logwage_i = \alpha + \beta educ_i + \pi female_i + \phi \hat{\varepsilon}_i + v_i$$

Donde $\widehat{\varepsilon}_i$ son los residuos de OLS de la forma reducida (ecuación donde \widehat{educ}_i es la variable dependiente)



 El test de Wu-hausman se puede estimar de la siguiente forma (dado el ejemplo)

$$logwage_i = \alpha + \beta educ_i + \pi female_i + \phi \hat{\varepsilon}_i + v_i$$

- Donde $\widehat{\varepsilon}_i$ son los residuos de OLS de la forma reducida (ecuación donde \widehat{educ}_i es la variable dependiente)
- ► Entonces:

$$H_0: \phi = 0$$
 (exogeneidad de *educ*)

$$H_a: \phi \neq 0$$
 (endogeneidad de $educ$)



Cuadro: Test-Wu Hausman

	Logwage
educ	0.106***
	(0.010)
female	-0.284***
remaie	(0.028)
	(0.020)
ehat	-0.048***
	(0.012)
	, ,
_cons	1.585***
	(0.041)
Observaciones	1000
R ²	0.233
	0.233

Fuente: ncds. Elaboracion: Autor ***, **, * denote statistical significance at the 1%, 5% and 10% levels respectively for zero.



Cuadro: Test-Wu Hausman

	Logwage
educ	0.106***
	(0.010)
female	-0.284***
Terriale	
	(0.028)
ehat	-0.048***
	(0.012)
	(0.012)
_cons	1.585***
	(0.041)
Observaciones	1000
R^2	0.233

Fuente: ncds. Elaboracion: Autor ***, ** * denote statistical significance at the 1%, 5% and 10% levels respectively for zero. ▶ Dado el t − value de la variable ehat en la regresión



Cuadro: Test-Wu Hausman

	Logwage
educ	0.106***
	(0.010)
female	-0.284***
remaie	
	(0.028)
ehat	-0.048***
ciide	
	(0.012)
_cons	1.585***
	(0.041)
	(0.041)
Observaciones	1000
R^2	0.233

Fuente: ncds. Elaboracion: Autor ***, ** * denote statistical significance at the 1%, 5% and 10% levels respectively for zero.

- ▶ Dado el t − value de la variable ehat en la regresión
- ► Se puede observar que el test de exogeneidad de Wu-hausman, se rechaza la hipotesis nula



Cuadro: Test-Wii Hausman

	Logwage
educ	0.106***
	(0.010)
female	-0.284***
remale	
	(0.028)
ehat	-0.048***
	(0.012)
	(0.012)
_cons	1.585***
	(0.041)
	(3.0.1)
Observaciones	1000
R^2	0.233

Fuente: ncds Elaboracion: Autor ***, **, * denote statistical significance at the 1%, 5% and 10% levels respectively for zero.

- \triangleright Dado el t value de la variable ehat en la regresión
- Se puede observar que el test de exogeneidad de Wu-hausman, se rechaza la hipotesis nula
- ► Resultados en → stata
- Resultados en

Roadmap



Introducción

Se termino todo en MCO? Variables Endogenas Metodo de Variable Instrumental (VI)

Analisis

Pregunta 1

Pregunta 2

Pregunta 3

Sesgo en muestras finitas

Pregunta -

Pregunta 5

Pregunta 6

Appendix

Resultados



Table 1: OLS and IV Estimates of Wage Equation

	OLS	IV
Constant	1.69489***	1.5854***
	(0.0298)	(0.0417)
Educ	0.0761***	0.1062***
	(0.0061)	(0.0100)
Female	-0.2949***	-0.2843***
	(0.0283)	(0.0288)
Adjusted-R ²	0.2221	0.2029
Sample Size	1000	1000
F-test of Instruments	n/a	F(3, 995)
(i.e., relevance)		=198.51
		p-value=0.00
Sargan Test	n/a	$\chi_{2}^{2} = 1.81$
(i.e., orthogonality)		p-value=0.40
Hausman-Wu Exogeneity	n/a	t = 3.87
Test (based on the auxiliary		p-value=0.00
regression approach)		

Notes: where ***, ** and * denote statistical significance at the 0.01, 0.05 & 0.10 levels respectively.



► Resultados en ► stata

Cuadro: OLS and IV estimates

	Retornos a la Educacion	
	OLS	IV
educ	0.08***	0.11***
	(0.01)	(0.01)
female	-0.29***	-0.28***
	(0.03)	(0.03)
Constant	1.69***	1.57***
	(0.03)	(0.04)
Adj. R ²	0.2	0.2
F-Test of Instrument		
Hansen J test (p-value)		0.03
Controls	\checkmark	\checkmark
Region FE	\checkmark	\checkmark
Observations	1000	1000

Source: BD5. Elaboration: Author Note: authors

Roadmap



Introducción

Se termino todo en MCO? Variables Endogenas Metodo de Variable Instrumental (VI)

Analisis

Pregunta :

Pregunta 2

Pregunta 3

Sesgo en muestras finitas

Pregunta -

Pregunta 5

Pregunta 6

Appendix

Modelo OLS: Stata



						1,000
E7 2640E4		20 6200755				
						0.0000
198.736498	997	.199334501	. R-sqı	iared		0.2237
			· Adj F	-square		0.2221
255.998449	999	.256254703	Root	MSE		.44647
Coef.	Std. Err.		P> t	[95%	Conf.	Interval]
.0761051	.0060637	12.55	0.000	.0642	061	.088004
2948813	.028318	-10.41	0.000	3504	511	2393116
1.69489	.0297845	56.91	0.000	1.636		1.753337
	57.261951 198.736498 255.998449 Coef.	57.261951 2 198.736498 997 255.998449 999 Coef. Std. Err0761051 .0060637	57.261951 2 28.6309755 198.736498 997 .199334501 255.998449 999 .256254703 Coef. Std. Err. t .0761051 .0060637 12.55	F(2,	57.261951 2 28.6309755 Prob > F 198.736498 997 .199334501 R-squared 255.998449 999 .256254703 Root MSE Coef. Std. Err. t P> t [95% 0.0761051 .0066637 12.55 0.000 .0642	F(2, 997) =

→ Back

MODELO OLS: Python



OLS Regression Results

Dep. Variable:	logwage	R-squared:	0.224
Model:	OLS	Adj. R-squared:	0.222
Method:	Least Squares	F-statistic:	145.8
Date:	Sat, 30 Mar 2024	Prob (F-statistic):	2.93e-56
Time:	19:16:19	Log-Likelihood:	-611.05
No. Observations:	1000	AIC:	1228.
Df Residuals:	997	BIC:	1243.
Df Model:	2		
Covariance Type:	HC1		

	coef	std err	z	P> z	[0.025	0.975]	
const	1.6949	0.032	53.272	0.000	1.633	1.757	
numberfemale	-0.2949	0.029	-10.325	0.000	-0.351	-0.239	
numbereduc	0.0761	0.006	11.721	0.000	0.063	0.089	
Omnibus:		292.664	Durbin-Wat	son:		2.005	
Prob(Omnibus):		0.000	Jarque-Ber	ra (JB):	252	2525.046	
Skew:		1.086	Prob(JB):			0.00	
Kurtosis:		10.475	Cond. No.			10.9	

Notes:

[1] Standard Errors are heteroscedasticity robust (HC1)

→ Back

SUPUESTO DE RELEVANCIA: relacion lineal



. reg educ female test11r test11m mumso dadso

Source	SS	df	MS	Number of obs	=	1,000 124.78
Model	2102.50375	5	420.50075	Prob > F	=	0.0000
Residual	3349.81525		3.37003546	R-squared	=	0.3856
Nebidudi	3313101020	331		•	=	0.3825
Total	5452.319	999	5.45777678	Root MSE	=	1.8358

educ	Coef.	Std. Err.	t	P> t	[95% Conf.	Interval]
female	0561978	.1168101	-0.48	0.631	2854206	.1730249
test11r	.0320939	.0048545	6.61	0.000	.0225677	.0416201
test11m	.0307488	.0032841	9.36	0.000	.0243042	.0371935
mumso	.5666817	.1427254	3.97	0.000	.2866039	.8467595
dadso	.5393084	.1533118	3.52	0.000	.2384564	.8401604
_cons	.3042332	.1852539	1.64	0.101	0593005	.6677668

F(4, 994) = 153.68 ← Test de relevancia (condicion relevancia)!

Prob > F = 0.0000



METODO DE IV: Stata



. ivreg logwage (educ = test11r test11m mumso dadso) female

Instrumental variables (2SLS) regression

Source	SS	df	MS	Number of obs	=	
				F(2, 997)	=	123.92
Model	51.0081859	2	25.5040929	Prob > F	=	0.0000
Residual	204.990263	997	.205607084	R-squared	=	0.1993
				Adj R-squared	=	0.1976
Total	255.998449	999	.256254703	Root MSE	=	.45344

logwage	Coef.	Std. Err.	t	P> t	[95% Conf.	. Interval]
educ female	2829424	.0099624	-9.79	0.000	.090519 3396377	.1296184
_cons	1.571405	.0415415	37.83	0.000	1.489886	1.652923

Instrumented: educ

Instruments: female test11r test11m mumso dadso



Sargan TEST : ortogonalidad de los instrumentos



. reg uhat_iv female test11r test11m mumso dadso

Source	SS	df	MS		er of obs	=	1,000
Model	1.8662855	5	.3732571		994) > F	=	1.83 0.1050
		•			-		
Residual	203.12398	994	.204350081	K-50	uared	=	0.0091
				Adj	R-squared	=	0.0041
Total	204.990266	999	.205195461	Root	MSE	=	. 45205
uhat_iv	Coef.	Std. Err.	t	P> t	[95% Co	nf.	Interval]
-					•		
female	002093	.0287641	-0.07	0.942	058538	4	.0543523
test11r	0004051	.0011954	-0.34	0.735	002750	8	.0019407
test11m	0007511	.0008087	-0.93	0.353	002338	1	.0008359
mumso	.0155974	.0351457	0.44	0.657	053370	В	.0845656
dadso	.1011149	.0377525	2.68	0.008	.027031	1	.1751987
uuubo							

- dadso es significativo (sobre la ecuación auxiliar realizada)
- Por lo tanto, el instrumento dadso es ortogonal sobre el error estructural de la ecuación de log salario (ecuación estructural)
- La hipotesis nula de no ortoganilidad de la identificación de instrumentos es rechazada
- Por tanto , el estimador de IV sera inconsistente

SARGAN TEST: fuerte instrumento



 Se elimina de la regresión anterior la variable dadso y se reestima su resultado

. reg uhat_iv2 female test11r test11m mumso

Source	SS	df	MS	Number of obs	=	1,000
				F(4, 995)	=	0.45
Model	.370239427	4	.092559857	Prob > F	=	0.7702
Residual	203.281997	995	.204303515	R-squared	=	0.0018
				Adj R-squared	=	-0.0022
Total	203.652237	999	.203856093	Root MSE	=	.452

uha	t_iv2	Coef.	Std. Err.	t	P> t	[95% Conf.	Interval]
f	emale	0012984	.0287605	-0.05	0.964	0577367	.0551399
te	stllr	.0000927	.0011874	0.08	0.938	0022374	.0024228
te	st11m	0004747	.0008066	-0.59	0.556	0020574	.001108
	mumso	.0429821	.0339108	1.27	0.205	0235629	.1095271
	_cons	.0057801	.0454928	0.13	0.899	0834927	.0950529

- ► El Sargan Test sera igual a 1.8179983
- ightharpoonup Su probabilidad o p-value sera igual a 0.40292729
- La hipótesis nula de la ortogonalidad ahora se mantiene.

₩ Back



EDÚCATE PERÚ CONSULTORES

Dado el t-value de la variable en la regresión



- Dado el t-value de la variable en la regresión
- Se puede observar que el test de exogeneidad de Wu-hausman, se rechaza la hipotesis nula
 - . reg logwage educ female ehat

Source	SS	df	MS	Number of obs	=	1,000
				F(3, 996)	=	102.09
Model	60.2041496	3	20.0680499	Prob > F	=	0.0000
Residual	195.794299	996	.196580622	R-squared	=	0.2352
				Adj R-squared	=	0.2329
Total	255.998449	999	.256254703	Root MSE	=	.44337

logwage	Coef.	Std. Err.	t	P> t	[95% Conf	. Interval]
educ	.1062169	.0098408	10.79	0.000	.0869058	.1255281
female	2842964	.0282545	-10.06	0.000	3397416	2288512
ehat	0481346	.012442	-3.87	0.000	0725502	023719
_cons	1.585409	.0409354	38.73	0.000	1.505079	1.665738

. test ehat

(1) ehat = 0



```
R 4.3.2 · C:/Users/et396/Dropbox/Docencia/Educate/Econometria/S5/Data/
> # Pregunta 5
> m6 <- lm(logwage ~ educ + female + rerror, data= base)
> summarv(m6)
Call:
lm(formula = logwage ~ educ + female + rerror, data = base)
Residuals:
    Min
            10 Median
                                     Max
-2.18621 -0.24722 -0.01708 0.20367 2.65413
Coefficients:
           Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.571405 0.040523 38.778 < 2e-16 ***
          0.110069 0.009718 11.326 < 2e-16 ***
educ
female -0.282942 0.028183 -10.039 < 2e-16 ***
rerror -0.054968 0.012363 -4.446 9.73e-06 ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1
Residual standard error: 0.4423 on 996 degrees of freedom
Multiple R-squared: 0.2388. Adjusted R-squared: 0.2365
F-statistic: 104.1 on 3 and 996 DF. p-value: < 2.2e-16
```

Resultados en Stata





IV (2SLS) estimation

Estimates efficient for homoskedasticity only Statistics consistent for homoskedasticity only

		d) SS = red) SS = =	255.9984486 3531.418075 203.6522387			F(2, 997) Prob > F Centered R2 Uncentered R2 Root MSE	= 119.36 = 0.0000 = 0.2045	
10	ogwage	Coef.	Std. Err.	z	P> z	[95% Conf.	Interval]	
	educ female _cons	.1062169 2842964 1.585409	.0100163 .0287582 .0416651	10.60 -9.89 38.05	0.000 0.000 0.000	.0865854 3406615 1.503747	.1258484 2279313 1.667071	
Underio	dentific	ation test (Anderson cand	n. corr		tistic): i-sq(3) P-val =	374.424 0.0000	
			Cragg-Donald itical values	10% m 10% m 20% m 30% m	aximal] aximal] aximal] aximal]	V relative bia V relative bia V relative bia V relative bia	s 9.08 s 6.46 s 5.39	
Source:	: Stock-	Yogo (2005).	Reproduced	15% m 20% m 25% m	aximal] aximal] aximal] aximal] ission.	V size V size	22.30 12.83 9.54 7.80	These relate to the sampling distribution of the IV estimator
Sargan	statist	<u>ic</u> (overiden	tification te	st of a		uments): i-sq(2) P-val =	1.818 0.4029	\leftarrow This is the <u>Sargan</u> test we computed manually

1000

Instrumented: educ Included instruments: female

Excluded instruments: test11r test11m mumso
Note: 'dadso' is no longer an identifying instrument here.

Number of obs -

. ***Compute the Wu-Hausman test with STATA command . ivendog

Tests of endogeneity of: educ HO: Regressor is exogenous