**TP 10: Proxy**

ACTIVIDAD Nº 1: Responda si las siguientes afirmaciones son verdaderas o falsas:

a) Una variable binaria no puede utilizarse como variable proxy.

 b) En fenómenos con un importante componente inercial no es conveniente incluir la variable dependiente rezagada como variable explicativa.

 c) Variable proxy es un método de estimación distinto a MCO.

ACTIVIDAD Nº 2:

a) ¿Qué es una variable proxy?

b) ¿Qué requisitos debe cumplir una variable proxy para proporcionar estimadores consistentes? ¿Qué sucede si no los cumple?

 c) Busque un ejemplo de variable proxy en un paper aplicado de econometría. Comente.

Actividad 1

a) Falso. Una variable proxy puede ser dummy

b) Falso.Cuando hay un gran componente inercial, es conveniente incluir la variable dependiente rezagada porque me permite controlar todos aquellos valores pasados. Estas variable se puede utilizar como proxy.

c)

Actividad 2

a)  La variable proxy busca una variable que sea muy parecida a la variable no observable, de forma tal de sacarla del error y meterla en el modelo.

b) -La variable proxy tiene que ser una variable relacionada con la variable omitida. El segundo requisito es que la variable que voy a utilizar como proxy, no pertenezca al modelo porque si pertenece al modelo, tengo dos variables omitidas. Si no se cumplen los requisitos, voy a tener malas estimaciones de los parámetros que me interesan.

c) **Measuring Living Standards With Proxy Variables.**

**Ejercicio 2**

. rename desemp desempleo

. rename lndelito\_03 logdelito03

. rename lndelito\_05 logdelito05

. rename gasto\_05 gastoseguridad

**a) Estime el modelo logdelito05 = B0 + B1\*desempleo + B2\*loggastoseguridad + u. Interprete los resultados.**

.

. gen loggastoseguridad = log(gastoseguridad)

.

. reg logdelito05 desempleo loggastoseguridad

Source | SS df MS Number of obs = 46

-------------+------------------------------ F( 2, 43) = 14.86

Model | 1.94636344 2 .973181721 Prob > F = 0.0000

Residual | 2.81560529 43 .065479193 R-squared = 0.4087

-------------+------------------------------ Adj R-squared = 0.3812

Total | 4.76196873 45 .105821527 Root MSE = .25589

-----------------------------------------------------------------------------------

logdelito05 | Coef. Std. Err. t P>|t| [95% Conf. Interval]

------------------+----------------------------------------------------------------

desempleo | .0499093 .0116802 4.27 0.000 .026354 .0734646

loggastoseguridad | .2292523 .1169056 1.96 0.056 -.0065103 .465015

\_cons | 2.126935 .7837878 2.71 0.010 .5462758 3.707593

-----------------------------------------------------------------------------------

Todas las variables son individualmente significativas a un nivel de significancia de incluso el 1%, y por ende también hay significatividad conjunta/global.

Ante un aumento en un 1% en el gasto en seguridad, se espera que en promedio el delito aumente en un 0,22%, siendo la variable estadísticamente significativa a un alfa del 10%. Esto, que a priori parece no tener sentido puede ser que haya variables omitidas por las que debamos controlar para que no nos afecte el coeficiente de estimación, que no nos lo sesgue. Es decir, deberíamos incorporar al modelo aquellos factores correlacionados con el gasto en seguridad que afectan al nivel de delitos.

.

.

**b) Incorpore como variable independiente la tasa de independencia del año 2003 a fin de controlar los factores no observables urbanos que afectan la delincuencia y que puedan estar correlacionados con el gasto corriente en seguridad pública.**

Para mejorar el modelo e impedir tener un beta de gasto de seguridad sesgado, debemos incorporar factores urbanis que están correlacionados con él y que afectan al nivel de delito. Como muchos de estos factores son no observables no medibles, podemos usar como variable proxy de ellos a la tasa de delicuencia de un año anterior, como el 2003.

Es decir, que el modelo que nos gustaría analizar sería

. logdelito05 = B0 + B1\*desempleo + B2\*loggastoseguridad +B3\*factoresurbanos +u

El tema es que la variable factores urbanos son no observables, de modo que utilizamos la variable proxy logdelito03 en su lugar, tal que:

factoresurbanos = s0 + s3\*logdelito03 + v1

considerando que v1 no está correlacionado con desempleo ni loggastoseguridad.

Así, incorporamos la proxy al modelo:

logdelito05 = B0 + B1\*desempleo + B2\*loggastoseguridad + b3\*(s0+s3\*logdelito03) + u

==> logdelito05 = (B0+s0\*B3) + B1\*desempleo + B2\*loggastoseguridad + B3\*s3\*logdelito03 + (B3\*v1+u)

Nótese que B1 y B2 son los mismos que los del modelo original. El coeficiente asociado a logdelito es más complejo, pero en realidad no nos interesa; solo incorporamos logdelito03 como proxy para controlar por factores urbanos no observables que de no estar incorporados en el modelo, sesga las estimaciones.

. reg logdelito05 desempleo loggastoseguridad logdelito03

Source | SS df MS Number of obs = 46

-------------+------------------------------ F( 3, 42) = 36.36

Model | 3.43802145 3 1.14600715 Prob > F = 0.0000

Residual | 1.32394727 42 .031522554 R-squared = 0.7220

-------------+------------------------------ Adj R-squared = 0.7021

Total | 4.76196873 45 .105821527 Root MSE = .17755

-----------------------------------------------------------------------------------

logdelito05 | Coef. Std. Err. t P>|t| [95% Conf. Interval]

------------------+----------------------------------------------------------------

desempleo | .0250502 .0088734 2.82 0.007 .007143 .0429574

loggastoseguridad | -.0838224 .0930095 -0.90 0.373 -.2715232 .1038783

logdelito03 | 1.007941 .1465249 6.88 0.000 .7122422 1.30364

\_cons | .1569209 .6146202 0.26 0.800 -1.083433 1.397275

-----------------------------------------------------------------------------------

.

**c) Compare ambos modelos y extraiga conclusiones**

Como vemos, al controlar por factores urbanos empleando a logdelito03 como proxy, corregimos el sesgo del B2: el gasto de seguridad ahora impacta negativamente sobre el nivel de delito: un aumento en un 1% del gasto en seguridad se espera que en promedio reduzca el delito en 0,08%. Contrariamente dado que estaban sesgados los coeficientes por la variable omitida, en el modelo original que corrimos el coeficiente era positivo, lo cual no tiene sentido.

**d) Repita los incisos anteriores empleando los errores estándar robustos a la heteroscedasticidad y analice brevemente cualquier diferencia notable.**

.

. reg logdelito05 desempleo loggastoseguridad logdelito03

Source | SS df MS Number of obs = 46

-------------+------------------------------ F( 3, 42) = 36.36

Model | 3.43802145 3 1.14600715 Prob > F = 0.0000

Residual | 1.32394727 42 .031522554 R-squared = 0.7220

-------------+------------------------------ Adj R-squared = 0.7021

Total | 4.76196873 45 .105821527 Root MSE = .17755

-----------------------------------------------------------------------------------

logdelito05 | Coef. Std. Err. t P>|t| [95% Conf. Interval]

------------------+----------------------------------------------------------------

desempleo | .0250502 .0088734 2.82 0.007 .007143 .0429574

loggastoseguridad | -.0838224 .0930095 -0.90 0.373 -.2715232 .1038783

logdelito03 | 1.007941 .1465249 6.88 0.000 .7122422 1.30364

\_cons | .1569209 .6146202 0.26 0.800 -1.083433 1.397275

-----------------------------------------------------------------------------------

.

.

. reg logdelito05 desempleo loggastoseguridad logdelito03, r

Linear regression Number of obs = 46

F( 3, 42) = 54.51

Prob > F = 0.0000

R-squared = 0.7220

Root MSE = .17755

-----------------------------------------------------------------------------------

| Robust

logdelito05 | Coef. Std. Err. t P>|t| [95% Conf. Interval]

------------------+----------------------------------------------------------------

desempleo | .0250502 .0079361 3.16 0.003 .0090345 .0410659

loggastoseguridad | -.0838224 .1009027 -0.83 0.411 -.2874524 .1198076

logdelito03 | 1.007941 .1364331 7.39 0.000 .7326082 1.283275

\_cons | .1569209 .6573468 0.24 0.812 -1.169659 1.4835

-----------------------------------------------------------------------------------

.

.

. reg logdelito05 desempleo loggastoseguridad logdelito03

Source | SS df MS Number of obs = 46

-------------+------------------------------ F( 3, 42) = 36.36

Model | 3.43802145 3 1.14600715 Prob > F = 0.0000

Residual | 1.32394727 42 .031522554 R-squared = 0.7220

-------------+------------------------------ Adj R-squared = 0.7021

Total | 4.76196873 45 .105821527 Root MSE = .17755

-----------------------------------------------------------------------------------

logdelito05 | Coef. Std. Err. t P>|t| [95% Conf. Interval]

------------------+----------------------------------------------------------------

desempleo | .0250502 .0088734 2.82 0.007 .007143 .0429574

loggastoseguridad | -.0838224 .0930095 -0.90 0.373 -.2715232 .1038783

logdelito03 | 1.007941 .1465249 6.88 0.000 .7122422 1.30364

\_cons | .1569209 .6146202 0.26 0.800 -1.083433 1.397275

-----------------------------------------------------------------------------------

.

. estat hettes

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity

Ho: Constant variance

Variables: fitted values of logdelito05

chi2(1) = 0.23

Prob > chi2 = 0.6324

No podemos rechazar la hipótesis nula: no podemos decir que la varianza de los errores no sea constante; hay homoscedasticidad.

**TRABAJO PRÁCTICO Nº11**

**–**

**Ejercicio 5**

A) Excluimos la variable M (eficiencia del management) para correr la siguiente regresión:

*reg q k l  
predict u, r*

- Test de White:

*estat hettes* Detectamos heterocedasticidad, porque rechaza H0, ya que 0.0138<0.05

- Comprobación gráfica:

*scatter u k   
scatter u l*

Podemos notar cierta variabilidad en los residuales para los valores que toman las variables independientes, por lo que intuitivamente inferiríamos que por la forma hay heterocedasticidad.

- Test de Breusch-Pagan:

*gen u2 = u^2  
reg u2 k l   
scalar bp = e(N)e(r2)  
di "Estadistico de prueba "bp  
di "Valor critico " invchi2tail(e(df\_r), 0.05)*

(B) La principal consecuencia sobre los estimadores resulta en el SESGO que se genera por omitir una variable, a pesar de que esta es NO observable.

(C) Agregamos una variable proxy:

*reg q k l exper*

En base a la salida de la regresión anterior, podemos inferir lo siguiente:

* Por cada punto porcentual que se incrementa el insumo capital (K), manteniendo el resto constante, el Output (q) se incrementa en promedio 43 puntos porcentuales. Siendo el índice de insumo capital un variable significativa con un estádistico "t" de 3.73
* Por cada punto porcentual que se incrementa el insumo trabajo (L), para el resto de variables constantes, el Output (q) se incrementa en promedio 23.91 puntos porcentuales, siendo el índice de insumo laboral una variable significativa al 0.019 de valor p.
* Por cada unidad que aumenta los años de experiencia (exper), ceteris paribus, el nivel de output (q) se ve afectado en promedio 14 puntos porcentuales positivamente.

(D) Método de mínimos cuadrados en dos etapas:

*ivregress 2sls q k l (exper=age), first*

Los errores estándar son 0.2145; 0.15; 0.117 y 2.64 respectivamente

La interpretación del coeficiente B3 es que por cada unidad adicional que se incrementa la variable independiente de años de experiencia, para el resto constante, el output aumenta 51 puntos porcentuales en promedio... siendo una variable significativa a un pvalor 0.017<0.05

(E)

*estat endogenous*

De la salida del test de Haussman podemos concluir la siguiente apreciación:

* Para el estadístico chi2, debemos rechazar H0 ya que 0.0278<0.05
* Para el estadístico F, debemos también rechazar H0 porque 0.313<0.05

Por ende, en síntesis, rechazamos la H0 de que la variable experiencia sea exógena, esto implica que estamos ante un caso de endogeneidad.

(F) Computamos la predicción para los valores de experiencia 102030, seteando el intervalo con un step de 10 unidades. El comando vsquish es para eliminar espacios blancos en la salida del test y que se compacte más la tabla.

*margins, at(exper=(10(10)30)) vsquish*

* Para la variable experiencia determinada en 10, tenemos un resultado en la variable explicativa de 7.647, con un intervalo de confianza al 95% de 5.858181; 9.436752
* Para el caso que se fija la variable experiencia en 20, tenemos un resultado en la Y de 12.768, con un intervalo de confianza del 95% de 10.09244; 15.44453
* Cuando fijamos, una vez más, la variable independiente experiencia con un valor de 30, obtenemos un resultado de la variable dependiente de 17.889, con un intervalo de confianza de 95% de 11.07223; 24.70678