

Parte práctica

1. *El Archivo “Agresiones.dta”¹ Contiene información sobre agresiones a nivel municipal en Colombia². A su vez, encontrarán información de características generales de cada municipio. En este ejercicio buscaremos encontrar cual es el efecto de las agresiones sobre la actividad económica municipal. Para ello, contará con un vector de variables de control que contiene características físicas del municipio, el Índice de Pobreza Monetaria (IPM).*

Los modelos que se proponen son los siguientes:

- a. *Estime los modelos 1 y 2. Interprete los resultados (significancia individual y global, signos y magnitud de los coeficientes estimados).*

Regresion 1 MCO	
VARIABLES	(1) Logaritmo del PIB total
Ponderación entre lesiones, hurtos y riñas	0.00301*** (0.000787)
Índice de pobreza multidimensional total	-0.0282*** (0.00274)
Distancia lineal a la capital del Departamento - km (Kilómetros)	0.000676 (0.000623)
Distancia lineal a Bogotá - km (Kilómetros)	5.11e-05 (0.000217)
Área oficial municipio DANE - km ² (kilómetros cuadrados)	1.28e-06 (1.01e-05)
Altura del municipio - MSNM (Metros Sobre el Nivel del Mar)	-0.000376* (4.10e-05)
Constant	13.40*** (0.205)
Observations	780
R-squared	0.198

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Regresion 2 MCO	
VARIABLES	(1) Logaritmo del PIB total
Logaritmo de agresiones	0.165*** (0.0267)
Índice de pobreza multidimensional total	-0.0263*** (0.00272)
Distancia lineal a la capital del Departamento - km (Kilómetros)	0.000654 (0.000612)
Distancia lineal a Bogotá - km (Kilómetros)	-8.78e-05 (0.000215)
Área oficial municipio DANE - km ² (kilómetros cuadrados)	1.31e-06 (9.98e-06)
Altura del municipio - MSNM (Metros Sobre el Nivel del Mar)	-0.000396*** (4.06e-05)
Constant	12.95*** (0.222)
Observations	780
R-squared	0.221

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Modelo 1:

Con respecto a la significancia individual, esta se puede probar mediante el p-valor. Con un nivel de significancia del 1%, las variables cuyo p-valor es menor a 0.01, son individualmente significativas. Es decir, las variables Agresiones, IPM, y Altura. Se puede afirmar entonces que estas variables explican los cambios porcentuales del PIB (variable dependiente).

Con respecto a la significancia global, esta se puede probar mediante el p-valor en la distribución F. En los resultados de la regresión, se puede observar que este valor es 0, por lo que se puede afirmar que las variables del modelo tienen significancia conjunta y explican los cambios porcentuales del PIB (variables dependiente).

De las variables significativas, se pueden sacar interpretaciones de su signo y coeficiente. En primer lugar, un aumento en una unidad de Agresión, aumenta en 3.01% el PIB, manteniendo todo lo demás constante. En segundo lugar, un aumento en una unidad del Índice de Pobreza Multidimensional, disminuye en 2.82% el PIB, manteniendo todo lo demás constante. En tercer lugar, un aumento en una unidad de la altura del municipio, disminuye en 0.0376% el PIB, manteniendo todo lo demás constante.

Además tiene un r^2 de 0.1981, por lo que se puede afirmar que el 19.81% de la varianza del logaritmo del PIB está explicado por variaciones de las variables independientes.

Modelo 2:

Con respecto a la significancia individual, esta se puede probar mediante el p-valor. Con un nivel de significancia del 1%, las variables cuyo p-valor es menor a 0.01, son individualmente significativas. Es decir, las variables Inagresiones, IPM, y Altura. Se puede afirmar entonces que estas variables explican los cambios porcentuales del PIB (variable dependiente).

Con respecto a la significancia global, esta se puede probar mediante el p-valor en la distribución F. En los resultados de la regresión, se puede observar que también este valor es 0, por lo que se puede afirmar que las variables del modelo tienen significancia conjunta y explican los cambios porcentuales del PIB (variables dependiente).

De las variables significativas, se pueden sacar interpretaciones de su signo y coeficiente. En primer lugar, un aumento en uno por ciento de las agresiones, aumenta en 0.00165% el PIB, manteniendo todo lo demás constante. En segundo lugar, un aumento en una unidad del Índice de Pobreza Multidimensional, disminuye en 2.63% el PIB, manteniendo todo lo demás constante. En tercer lugar, un aumento en una unidad de la altura del municipio, disminuye en 0.0396% el PIB, manteniendo todo lo demás constante. Además, tiene un r^2 de 0.2214, por lo que se puede afirmar que el 22.14% de la varianza del logaritmo del PIB está explicado por variaciones de las variables independientes.

b. Realice la prueba J de Davidson y MacKinnon para comparar el modelo log-lin y log-log. Interprete los resultados y escoja el mejor modelo que usted considere.

Regresion 2 Prueba J	
VARIABLES	(1) Logaritmo del PIB total
Ponderación entre lesiones, hurtos y robos	-0.00304** (0.00138)
Indice de pobreza multidimensional total	0.0136 (0.00832)
Distancia lineal a la capital del Departamento - km (Kilómetros)	-0.000463 (0.000649)
Distancia lineal a Bogotá - km (Kilómetros)	6.85e-05 (0.000213)
Área oficial municipio DANE - km ² (kilómetros cuadrados)	-1.10e-06 (9.97e-06)
Altura del municipio - MSNM (Metros Sobre el Nivel del Mar)	0.000207* (0.000117)
Fitted values	1.525*** (0.287)
Constant	-6.925* (3.833)
Observations	780
R-squared	0.226

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Teniendo en cuenta que los dos modelos incluyen las mismas variables (la única diferencia es que en el 2 hay una variable con logaritmo) se puede determinar que un modelo esta dentro del otro, por lo que se usa la prueba para modelos anidados. Las tablas muestran los resultados de la regresión al estimar y_A & y_B (es decir, la variable dependiente de ambos modelos). Con esto, podemos hacer la prueba de hipótesis para determinar si hay sesgo de selección, para poder concluir cuál es el mejor modelo para analizar. En esta prueba de hipótesis, H_0 =No hay sesgo de especificación, H_a =Sí hay sesgo de especificación. Teniendo en cuenta que cuando incluimos y_B en el modelo 1 el valor de t es mayor al t crítico ($2.21 > 1.96$) determinamos que se rechaza la hipótesis nula. De igual manera, al introducir y_A en el modelo 2, determinamos que el valor t de la variable es mayor al t crítico ($5.31 > 1.96$), por lo tanto se rechaza la hipótesis nula. Por lo tanto, concluimos que ambos modelos tienen sesgo de especificación, con 5% de significancia.

Dados los resultados anteriores, los dos modelos son incorrectos, por lo que para escoger uno, es necesario analizar el tipo de funciones de cada uno de los modelos (es decir, con una variable Agresiones lineal, y una variable Agresiones logarítmica). En este caso, escogemos el modelo log-log, dado que si aumentamos cada vez más la muestra, la función logarítmica termina convergiendo a una distribución normal, lo cual es una mejor aproximación a los resultados poblacionales. Además, al analizar el p-valor de la variable agresiones en cada una de las regresiones, la variable con el logaritmo termina siendo más significativa que la variable por niveles.

- c. *Utilice la prueba RESET de Ramsey para verificar que el modelo que haya escogido no tenga componentes omitidos.*

Regresion Prueba RESET	
(1)	
VARIABLES	Logaritmo del PIB total
Logaritmo de agresiones	0.165*** (0.0267)
Indice de pobreza multidimensional total	-0.0263*** (0.00272)
Distancia lineal a la capital del Departamento - km (Kilómetros)	0.000654 (0.000612)
Distancia lineal a Bogotá - km (Kilómetros)	-8.78e-05 (0.000215)
Área oficial municipio DANE - km ² (kilómetros cuadrados)	1.31e-06 (9.98e-06)
Altura del municipio - MSNM (Metros Sobre el Nivel del Mar)	-0.000396*** (4.06e-05)
Constant	12.95*** (0.222)
Observations	780
R-squared	0.221
Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1	

Realizamos la prueba RESET de Ramsey para evaluar la posibilidad de que se estuvieran omitiendo variables que fueran relevantes para explicar las variaciones porcentuales del PIB. Teniendo en cuenta que H_0 indica que no hay variables omitidas y H_a indica que si hay variables omitidas, concluimos que rechazamos la hipótesis nula dado que el p-valor de la distribución F es mayor al nivel de significancia ($0.0241 < 0.05$). Por lo tanto se determina que si existen variables omitidas.

- d. *Usted considera que la medida de pobreza como Necesidades Básicas Insatisfechas puede afectar el efecto que usted quiere estimar. Realice la prueba del Multiplicador de Lagrange para determinar si es una variable relevante para el modelo.*

Regresion Prueba ML	
VARIABLES	(1) Residuals
Logaritmo de agresiones	-0.00447 (0.0267)
Indice de pobreza multidimensional total	0.00561 (0.00384)
Distancia lineal a la capital del Departamento - km (Kilómetros)	0.000192 (0.000618)
Distancia lineal a Bogotá - km (Kilómetros)	9.40e-05 (0.000219)
Área oficial municipio DANE - km ² (kilómetros cuadrados)	3.04e-06 (1.01e-05)
Altura del municipio - MSNM (Metros Sobre el Nivel del Mar)	-2.93e-06 (4.05e-05)
NBI - Necesidades básicas Insatisfechas	-0.00624* (0.00303)
Constant	-0.143 (0.232)
Observations	780
R-squared	0.005

Standard errors in parentheses
 *** p<0.01, ** p<0.05, * p<0.1

Realizamos la prueba del Multiplicador de Lagrange para evaluar la posibilidad de que la variable NBI fuera relevante para explicar las variaciones porcentuales del PIB. Teniendo en cuenta que H0 indica que NBI no es relevante para el modelo y Ha indica que NBI es relevante para el modelo, concluimos que se rechaza la hipótesis nula dado que el p-valor de la distribución F es menor al nivel de significancia ($0.0388 < 0.05$). Por lo tanto se determina que NBI es relevante para el modelo.

e. Usted analizó la variable agresiones y se dio cuenta que puede haber endogeneidad en esta variable. Explique por qué puede haber endogeneidad

La variable Agresiones presenta endogeneidad porque está correlacionada con el error, es decir, con variables no observables que no están medidas en el modelo pero que pueden explicar la variable dependiente. Por ejemplo, se puede afirmar que dentro del error está la presencia de conflicto armado en los municipios. Esto puede explicar tanto las variaciones porcentuales del PIB (ya que la violencia genera problemas de confianza y por ende disminuye la inversión extranjera, lo que disminuye este indicador) como las Agresiones (ya que la presencia del conflicto aumenta estas agresiones). Por lo tanto, y teniendo en cuenta que la variable Agresiones está relacionada con el error, se puede afirmar que esta variable tiene problema de endogeneidad.

f. Una compañera suya le recomienda agregar a su modelo la variable agresiones per cápita la cual pondera las agresiones por población. Explique los beneficios o perjuicios de usar esta variable.

Consideramos que incluir la variable per cápita en el modelo puede tener ventajas en la estimación. Las variables per cápita permiten analizar el valor para el individuo promedio de la población. Por lo tanto, es posible analizar de qué manera un individuo con agresiones en promedio de los municipios puede afectar al crecimiento porcentual del PIB. Esta nueva variable reflejaría la magnitud del número de agresiones en cada

municipio para un habitante promedio, dado que un municipio con una población pequeña puede tener un impacto distinto al de un municipio con población mayor y por ende, el análisis se encontraría sesgado al tamaño de la población.

Sin embargo, la inclusión de esta variable en nuestro modelo también tiene desventajas. Al fin y al cabo, esta nueva variable sería un promedio del número de agresiones en cada municipio. Por lo tanto, no tiene en cuenta posibles desigualdades en el número de agresiones dentro de los municipios.

g. Proponga el mejor modelo para analizar esta relación e interprete los resultados obtenidos

Regresion	
VARIABLES	(1) Logaritmo del PIB total
Logaritmo de agresiones	0.412*** (0.0292)
Indice de pobreza multidimensional total	-0.0139*** (0.00343)
Distancia lineal a la capital del Departamento - km (Kilómetros)	0.000286 (0.000548)
Distancia lineal a Bogotá - km (Kilómetros)	-0.000418** (0.000196)
Área oficial municipio DANE - km ² (kilómetros cuadrados)	1.13e-06 (8.91e-06)
Altura del municipio - MSNM (Metros Sobre el Nivel del Mar)	-0.000307*** (3.64e-05)
NBI - Necesidades básicas Insatisfechas	-0.00726*** (0.00268)
agresiones divididas entre la población	-99.83*** (6.787)
Constant	12.05*** (0.212)
Observations	780
R-squared	0.395

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

En el modelo 2 (el que se está analizando) tiene un r² de 0.226, por lo que se puede afirmar que el 22.6% de la varianza del logaritmo del PIB está explicado por variaciones de las variables independientes. Al agregar la variable Agresionespercap, el r² es 0.3954, por lo que el 39.54% de la varianza del PIB está explicado por variaciones de las variables independientes. Por lo tanto, se puede afirmar que el modelo incluyendo la variable per cápita es mejor, pues explica en mayor magnitud las variaciones de la variable dependiente.

La regresión estimada en este literal, incluye la variable per cápita y la NBI, que ambas son individualmente significativas (dado que su p-valor < 0.01), por lo que ayudan a explicar la variable del PIB con un nivel de significancia del 1%.

Con respecto a la significancia global, esta se puede probar mediante el p-valor en la distribución F. En los resultados de la regresión, se puede observar que este valor es 0, por lo

que se puede afirmar que las variables del modelo tienen significancia conjunta y explican los cambios porcentuales del PIB (variables dependiente).

2. *Suponga que usted desea estimar el número de visitas (visit) de una página web, a partir del tiempo que consumen las personas en internet (time) y de la cantidad de anuncios publicitarios que reciben vía e-mail a través de otros sitios web (ad). Adicionalmente, al creador de la página web le interesa conocer el sexo de sus visitantes. Para esto, usted debe utilizar la información contenida en la base de datos “website.dta”*

a. Estime el modelo por MCO, y comente los resultados

Regresion MCO	
	(1)
VARIABLES	Visits to website
Advertisements	0.800*** (0.0517)
Female	-0.0468 (0.210)
Time on internet (hrs.)	0.830*** (0.0437)
Constant	0.692*** (0.201)
Observations	500
R-squared	0.677
Standard errors in parentheses	
*** p<0.01, ** p<0.05, * p<0.1	

Con respecto a la significancia individual, esta se puede probar mediante el p-valor. Con un nivel de significancia del 1%, las variables cuyo p-valor es menor a 0.01, son individualmente significativas. Es decir, la variables Advertisements y Time. Se puede afirmar entonces que estas variables explican el número de visitas de la página web (variable dependiente).

Con respecto a la significancia global, esta se puede probar mediante el p-valor en la distribución F. En los resultados de la regresión, se puede observar que este valor es 0, por lo que se puede afirmar que las variables del modelo tienen significancia conjunta y explican las visitas de la página web (variable dependiente).

De las variables significativas, se pueden sacar interpretaciones de su signo y coeficiente. En primer lugar, en promedio un aumento en una unidad de propagandas, aumenta en 0.8 el número de visitas de la página web, manteniendo todo lo demás constante. En segundo lugar, un aumento de una hora en internet, aumenta en 0.83 el número de visitas de la página web, manteniendo todo lo demás constante. Además tiene un r^2 de 0.677, por lo que se puede afirmar que el 67.7% de la varianza del logaritmo del PIB está explicado por variaciones de las variables independientes.

- b. Algunos consideran que el tiempo en internet es una variable endógena, explique con sus propias palabras porque puede existir endogeneidad en la variable time.***

La variable “Tiempo” presenta endogeneidad porque está correlacionada con el error, es decir, con variables no observables que no están medidas en el modelo pero que pueden explicar la variable dependiente. Se puede suponer que dentro del error está la variable de “horas de trabajo”, ya que a pesar de explicar “y” (pues entre más horas de trabajo menos tiempo disponible hay para visitar la página web) no está incluida en el modelo. Sin embargo, esta variable de trabajo también está relacionada con la variable explicativa de “tiempo”, ya que una persona con trabajo destina menos tiempo a estar en internet. En conclusión, la variable “tiempo” es endógena dado que se relaciona tanto con la variable explicativa como con el error.

- c. Un investigador propone dos instrumentos, el primero corresponde al tiempo que una persona usa el celular (phone) y el segundo es el tiempo que gasta en interactuar con amigos, familia que vive fuera de la ciudad (frfam). ¿Serían estos instrumentos válidos y relevantes? Justifique intuitivamente y basándose en los supuestos de un buen instrumento.***

Un buen instrumento debe cumplir con dos condiciones, relevancia y validez. Con respecto a la primera condición, los dos instrumentos propuestos (phone y frfam) están correlacionados con la variable endógena (time) ya que el tiempo que se dedica a estar con la familia y amigos disminuye el tiempo disponible para estar en internet, y el tiempo que se dedica a estar en el celular aumenta el tiempo dentro de internet. Por lo tanto, los dos instrumentos propuestos son relevantes. En segundo lugar, hay algunas variables no observables dentro del error que se pueden correlacionar con los instrumentos, por lo que no se cumple la condición de exogeneidad. Por ejemplo, si dentro del error está la variable “horas de trabajo” (dado que entre más horas de trabajo menos tiempo disponible hay para visitar la página web, por lo que esta variable explica nuestra “y” pero no está dentro del modelo), esta está correlacionada con los instrumentos. Lo anterior porque entre más horas de trabajo, se le va a poder dedicar menos tiempo a pasar tiempo en familia y en el celular. En conclusión a pesar de que los instrumentos son relevantes, no son exógenos pues están correlacionados con el error del modelo y no resuelven el problema de endogeneidad.

- d. Realice la estimación con el instrumento (phone), comente los resultados de la primera y la segunda etapa. ¿Es razonable pensar que no se presenta el problema de instrumentos débiles? Adicionalmente, realice las pruebas de significancia global y parcial, y el test para probar relevancia. ¿Cómo probaría si el instrumento es exógeno y realícelo?***

IV Segunda etapa	
VARIABLES	(1) Visits to website
Time on internet (hrs.)	0.853*** (0.144)
Advertisements	0.788*** (0.0862)
Female	-0.0463 (0.209)
Constant	0.643* (0.354)
Observations	500
R-squared	0.677
Standard errors in parentheses	
*** p<0.01, ** p<0.05, * p<0.1	

IV Primera etapa	
VARIABLES	(1) Time on internet (hrs.)
Advertisements	0.269*** (0.0567)
Female	-0.0371 (0.206)
Time on phone (hrs.)	0.408*** (0.0578)
Constant	1.346*** (0.207)
Observations	500
R-squared	0.256
Standard errors in parentheses	
*** p<0.01, ** p<0.05, * p<0.1	

Instrumentos débiles:

Para comprobar si un instrumento es débil o no, se debe hacer la prueba F. En este caso:

$$F(3, 496) = 29.90$$

Ya que el F calculado > 10 , se puede afirmar que Phone es un instrumento fuerte.

Pruebas de significancia:

Primero se analizarán los resultados de la primera etapa.

Con respecto a la significancia individual, esta se puede probar mediante el p-valor. Con un nivel de significancia del 1%, las variables cuyo p-valor es menor a 0.01, son individualmente significativas. Es decir, la variables Advertisements y el Phone en internet. Se puede afirmar entonces que estas variables explican las visitas a la página web (variable dependiente).

Con respecto a la significancia global, esta se puede probar mediante el p-valor en la distribución F. En los resultados de la regresión, se puede observar que este valor es 0, por lo que se puede afirmar que las variables del modelo tienen significancia conjunta y explican los cambios porcentuales del PIB (variables dependiente).

Ahora, analizando la segunda etapa.

Con respecto a la significancia individual, esta se puede probar mediante el p-valor. Con un nivel de significancia del 1%, las variables cuyo p-valor es menor a 0.01, son individualmente significativas. Es decir, la variables Advertisements y el Tiempo en internet. Se puede afirmar entonces que estas variables explican las visitas a la página web (variable dependiente).

Con respecto a la significancia global, esta se puede probar mediante el p-valor en la distribución Chi cuadrado. En los resultados de la regresión, se puede observar que este valor es 0, por lo que se puede afirmar que las variables del modelo tienen significancia conjunta y explican los cambios porcentuales del PIB (variables dependiente).

Test de relevancia

Relevancia IV	
(1)	
VARIABLES	Time on internet (hrs.)
Advertisements	0.269*** (0.0567)
Female	-0.0371 (0.206)
Time on phone (hrs.)	0.408*** (0.0578)
Constant	1.346*** (0.207)
Observations	500
R-squared	0.256

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Para saber la relevancia de un instrumento se hace la “prueba de relevancia”. En la tabla superior se muestran los resultados de esta prueba, se puede observar que el p-valor de la distribución F es igual a cero, por lo que se puede afirmar que Phone es un instrumento relevante.

En este caso, que es exactamente identificado (tenemos mismo número de instrumentos y variables endógenas) no se puede hacer una prueba empírica para probar la exogeneidad del instrumento, solo se puede hacer esto de manera teórica (como se hizo en el numeral c).

e. Realice la estimación con los dos instrumentos, comente los resultados de la primera etapa y segunda etapa. ¿Es razonable pensar que no se presenta el problema de instrumentos débiles? Realice la prueba de relevancia de los instrumentos. ¿Cómo probaría si los instrumentos son exógenos y realícelo?

IV Segunda etapa	
(1)	
VARIABLES	Visits to website
Time on internet (hrs.)	0.838*** (0.121)
Advertisements	0.795*** (0.0769)
Female	-0.0466 (0.209)
Constant	0.674** (0.313)
Observations	500
R-squared	0.677

Standard errors in parentheses
 *** p<0.01, ** p<0.05, * p<0.1

Pruebas de significancia:
 de la primera etapa.

Primero se analizarán los resultados

Con respecto a la significancia individual, esta se puede probar mediante el p-valor. Con un nivel de significancia del 1%, las variables cuyo p-valor es menor a 0.01, son individualmente significativas. Es decir, la variables Advertisements, Phone y Tiempo en Familia. Se puede afirmar entonces que estas variables explican las visitas a la página web (variable dependiente).

Con respecto a la significancia global, esta se puede probar mediante el p-valor en la distribución F. En los resultados de la regresión, se puede observar que este valor es 0, por lo que se puede afirmar que las variables del modelo tienen significancia conjunta y explican los cambios porcentuales del PIB (variables dependiente).

Ahora, analizando la segunda etapa.

Con respecto a la significancia individual, esta se puede probar mediante el p-valor. Con un nivel de significancia del 1%, las variables cuyo p-valor es menor a 0.01, son individualmente significativas. Es decir, la variables Advertisements y el Tiempo en internet. Se puede afirmar entonces que estas variables explican las visitas a la página web (variable dependiente).

Con respecto a la significancia global, esta se puede probar mediante el p-valor en la distribución Chi cuadrado. En los resultados de la regresión, se puede observar que este valor es 0, por lo que se puede afirmar que las variables del modelo tienen significancia conjunta y explican los cambios porcentuales del PIB (variables dependiente).

Instrumentos débiles:

Para comprobar si un instrumento es débil o no, se debe hacer la prueba F. En este caso:

$$F(4, 495) = 29.3$$

Ya que el F calculado > 10, se puede afirmar que Phone y Frfam son instrumentos fuertes.

Test de relevancia

Relevancia IV	
VARIABLES	(1) Time on internet (hrs.)
Advertisements	0.166*** (0.0599)
Female	-0.0522 (0.202)
Time with friends and out of town family (hrs.)	0.276*** (0.0603)
Time on phone (hrs.)	0.291*** (0.0623)
Constant	0.935*** (0.222)
Observations	500
R-squared	0.286

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Para saber la relevancia de un instrumento se hace la “prueba de relevancia”. En la tabla superior se muestran los resultados de esta prueba, se puede observar que el p-valor de la distribución F es igual a 0, por lo que se puede afirmar que Phone es un instrumento relevante

Prueba de exogeneidad/validez

Validez IV	
VARIABLES	(1) Residuals
Time on phone (hrs.)	0.0103 (0.0648)
Time with friends and out of town family (hrs.)	-0.0101 (0.0628)
Advertisements	0.000299 (0.0624)
Female	0.000339 (0.210)
Constant	0.00352 (0.231)
Observations	500
R-squared	0.000

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Para saber si los instrumentos son exógenos, se hace la prueba de validez (que solo es posible realizarla en modelos sobre especificados). En la tabla se muestran los resultados de esta prueba, y se observa que el p-valor es igual a 0.84. Ya que este valor es mayor al nivel de significancia 0.05, no se rechaza la hipótesis nula y por lo tanto se puede afirmar que los instrumentos NO son exógenos/válidos.

f. Compare los resultados obtenidos en los numerales d y e. ¿Cuál modelo elegiría?

Ya que ambos instrumentos, a pesar de que son relevantes y fuertes, son endógenos/inválidos, tener un instrumento de este tipo es peor que no tenerlo. Por esto, elegimos el modelo con un solo instrumento, dado que esto tiene una varianza menor que si incluyéramos dos instrumentos endógenos (es mejor tener un instrumento invalido que tener dos). Sin embargo, recomendamos buscar otro instrumento que sí cumpla con las condiciones y características de un instrumento bueno, para que de esta manera se corrija adecuadamente el problema de endogeneidad.

Anexo: Do File

**TALLER 1*

cls

**Punto 1*

cd "/Users/valeria/Desktop/202310/Econometría 2"

use "Agresiones.dta", clear

**a) Estimar los modelos*

**Modelo 1*

reg lnpiib_total agresiones IPM discapital disbogota areaoficialkm2 altura

outreg2 using "Regresiones1punto1a.doc", label title(Regresion 1 MCO) replace

**Modelo 2*

reg lnpiib_total lnagresiones IPM discapital disbogota areaoficialkm2 altura

outreg2 using "Regresiones2punto1a.doc", label title(Regresion 2 MCO) replace

**b) Prueba J de Davidson y Mackinnon*

**Prueba de Modelos Anidados*

reg lnpiib_total agresiones IPM discapital disbogota areaoficialkm2 altura

predict yA,

reg lnpiib_total lnagresiones IPM discapital disbogota areaoficialkm2 altura yA

outreg2 using "Regresiones1punto1b.doc", label title(Regresion 1 Prueba J) replace

test yA

reg lnpiib_total lnagresiones IPM discapital disbogota areaoficialkm2 altura

predict yB

reg lnpiib_total agresiones IPM discapital disbogota areaoficialkm2 altura yB

outreg2 using "Regresiones2punto1b.doc", label title(Regresion 2 Prueba J) replace

ttest yB==0

**c) Prueba RESET de Ramsey*

reg lnpiib_total lnagresiones IPM discapital disbogota areaoficialkm2 altura

estat ovtest

outreg2 using "Regresionesresetpunto1c.doc", label title(Regresion Prueba RESET) replace

**Si hay variables omitidas, rechazo h_0*

**d) Prueba del Multiplicador de Lagrange*

**Prueba para variable omitida*

reg lnpiib_total lnagresiones IPM discapital disbogota areaoficialkm2 altura

predict r_2, residual

reg r_2 lnagresiones IPM discapital disbogota areaoficialkm2 altura nbi

*scalar ML=e(N)*e(r2)*

scalar prob=chi2tail(1,ML)

scalar list ML prob

outreg2 using "RegresionesMLpunto1d.doc", label title(Regresion Prueba ML) replace

**Hay sesgo de especificación por variable omitida. NBI si es relevante para el modelo*

**e) Endogeneidad en la variable agresiones*

**f) Beneficios o perjuicios de usar agresionespercap en el modelo*

**g) Proponga el mejor modelo para analizar esta relación e interprete*

*reg lnpiib_total lnagresiones IPM discapital disbogota areaoficialkm2 altura nbi
agresiones_percap*

outreg2 using "Regresionespunto1g.doc", label title(Regresion) replace

**Punto 2*

cls

clear all

use "website.dta", clear

**a) Estimar modelo por MCO*

reg visits ad female time

outreg2 using "RegresionIpunto2a.doc", label title(Regresion MCO) replace

**b) Explicar endogeneidad de variable TIME*

**c) Justificar si los instrumentos son relevantes y válidos*

**d) Realice estimación con el instrumento Phone*

ivregress2 2sls visits (time=phone) ad female, first

est restore first

outreg2 using "RegresionIVprimeraetapa.doc", label title(IV Primera etapa) replace

ivregress 2sls visits (time=phone) ad female, first

outreg2 using "RegresionIVsegundaetapa.doc", label title(IV Segunda etapa) replace

**Test para probar relevancia*

reg time ad female phone

**Se realiza la prueba de significancia conjunta*

test phone

outreg2 using "Regresionpunto2d.doc", label title(Relevancia IV) replace

**Test para probar exogeneidad*

regress visits ad female time

est store mco

ivregress 2sls visits (time= phone) ad female

est store iv

hausman iv mco, sigmamore

**Test instrumentos débiles*

reg time phone ad female,r

**e) Estimacion con los dos instrumentos*

ivregress2 2sls visits (time=phone frfam) ad female, first

est restore first

outreg2 using "RegresionIV2primeraetapa.doc", label title(IV Primera etapa) replace

ivregress 2sls visits (time=phone frfam) ad female, first

outreg2 using "RegresionIV2segundaetapa.doc", label title(IV Segunda etapa) replace

**Test instrumentos débiles*

reg time phone frfam ad female,r

**Test para probar relevancia*

reg time ad female frfam phone

**Se realiza la prueba de significancia conjunta*

test frfam phone

outreg2 using "Regresionpunto2e.doc", label title(Relevancia IV) replace

**Test de validez*

ivreg visits (time= phone frfam) ad female

predict res, residuals

reg res phone frfam ad female

*scalar ml=e(N)*e(r2)*

scalar prob=chi2tail(1,ml)

scalar list ml prob

outreg2 using "Regresionvalidezpunto2e.doc", label title(Validez IV) replace

**f) Comparar resultados y escoger mejor modelo*