

Análisis de Regresión y Causalidad 5

Estadística Urbana

Sebastián Espinoza Rojas

Universidad de Chile
Primavera 2017

- Para solucionar el problema de endogeneidad (a partir de cualquiera de sus fuentes) se utiliza el método de variables instrumentales.
- El método de variable instrumental se basa en la idea de encontrar otra variable que permita rescatar la parte exógena de X para poder establecer una relación con Y .
- El método consiste en:
 - 1 Estimar una regresión del tipo $X_i = Z_i * \gamma + u_i$.
 - 2 Calcular el \hat{X}_i , que representa la parte de X explicada por Z , que sería $\hat{X}_i = Z_i * \hat{\gamma}$.
 - 3 Regresionar $Y_i = \beta * \hat{X}_i + \varepsilon_i$

- Sea un instrumento Z , para que sea un instrumento apropiado, debe cumplir con:
 - 1 Ser exógeno (no estar relacionado directamente con Y).
 - 2 Ser relevante (explicar a X).
 - 3 Tener una relación fuerte (explicar a X no de una forma débil).

Variables Instrumentales 3/5

Ejemplo 1/2

- Supongamos estamos analizando cómo influye la estigmatización del barrio en el que reside una persona sobre la probabilidad de ser discriminado laboralmente. Asuma que tiene una variable que da cuenta si la persona ha sido o no discriminada (D), y otra que señala si la persona vive en un barrio estigmatizado o no (E).

Si se estima:

$$D_i = \beta_0 + \beta_1 * E_i + \varepsilon_i$$

¿ β_1 podrá ser interpretado como un efecto causal?

- ¿Cómo se comporta E_i ?

Variables Instrumentales 4/5

Ejemplo 2/2

- Si queremos instrumentalizar E_i entonces debemos buscar una variable que se relacione con E_i pero no directamente con D_i .
- Analicemos las siguientes variables:
 - 1 Escolaridad de la persona.
 - 2 Una dummy asociada a ser moreno.
 - 3 Escolaridad de los padres.

- 1 En STATA el comando que permite estimar con variables instrumentales es *ivregress*.
- 2 Si se tiene una variable dependiente y , una variable independiente exógena x_1 , una variable endógena x_2 y un instrumento z . La sintaxis es

ivregress 2sls y x1 (x2=z), r.

- Puede ocurrir que el efecto de alguna política o algún tratamiento dependa (sea mayor o menor) condicional a las características de las personas.
- Por ejemplo, si se quiere evaluar el efecto en las horas dedicadas a actividades físicas de una comunidad a raíz de la construcción de una plaza se podría pensar que ese efecto no será uniforme para toda la población y quizás va a afectar más a los niños (los niños aprovecharán más la plaza).

- Para poder incorporar al modelo estos efectos se deben agregar interactivas, esto es, multiplicaciones entre las variables:

$$H_i = \beta_0 + \beta_1 * Plaza_i + \beta_2 * Niño_i + \beta_3 * Plaza_i * Niño_i + \varepsilon_i$$

- En el ejemplo anterior el efecto marginal de la colocación de la plaza será $\beta_1 + \beta_3 * Niño$. Así, si la persona es niño, el efecto será $\beta_1 + \beta_3$; si no lo es, será β_1 solamente.

- La misma lógica aplicaría para otras variables como tiempo. Si por ejemplo, se piensa que la construcción de la plaza no tiene un efecto inmediato, si no que tiene un efecto en el año 2 de existir, se podría estimar:

$$H_i = \beta_0 + \beta_1 * Plaza_i + \beta_2 * Tiempo_i + \beta_3 * Plaza_i * Tiempo + \varepsilon_i$$

Donde *Tiempo* es una dummy que toma el valor de 1 si es el año 2 y 0 si no. Luego, β_1 es el efecto de la plaza cuando es el año 1, y $\beta_1 + \beta_3$ es el efecto de la plaza cuando es el año 2. Así, β_3 representa cuántas horas más le dedica una persona a raíz de la construcción de la plaza en el tiempo 2 respecto del tiempo 1.

Modelos de Probabilidad

- Si se estima via MCO cuando la variable dependiente es binaria (se le denomina Modelos de Probabilidad Lineal, MPL) se puede incurrir en una serie de problemas. Uno de los principales es el de predicciones fuera de muestra (calcula probabilidades mayor que 1).
- Para no incurrir en esos problemas se estima utilizando modelos Probit o Logit (o logísticos). La idea base es que desarrolla un modelo particular para variables binarias que permite incorporar su particularidad.
- Dado lo anterior, se considera que el efecto marginal de las variables no es lineal, si no que depende del valor de las variables en todos los puntos. Además, la tabla de stata no reporta automáticamente los efectos marginales, y hace falta solicitarlos para que los muestre.
- En términos generales, y si se tiene una gran cantidad de observaciones, ambos métodos son muy similares en términos de sus resultados.