

Taller de STATA:Tópicos en Econometría

Samuel Alarcon

ILADES/Universidad Alberto Hurtado
Maestría en Economía/MAE

9 de Agosto de 2018

Índice

- 1 Variables Instrumentales 2 SLS
- 2 Datos de Panel
- 3 Experimento Aleatorio
- 4 Introducción a la Regresión Discontinua

Existen tres sesgos en la estimaci3n por MCO

- Sesgo por omisi3n de variables
- Sesgo por causalidad simult3nea
- Sesgo por errores en las variables

La regresi3n VI puede eliminar los anteriores sesgos. Este m3todo debe cumplir la condiciones : Relaci3n entre la variables y el instrumento

Suponga el siguiente modelo:

$$y_i = \beta_0 + \beta_1 x_i + u_i \quad (1)$$

Supongamos que la variable independiente es endógena:

$$\text{Cov}(x_i, u_i) \neq 0 \quad (2)$$

$$E(x_i, u_i) \neq 0 \quad (3)$$

Para no generar un sesgo necesitamos un instrumento z_i que cumpla dos supuestos:

- Exogeneidad del Instrumento : $\text{Cov}(z_i, u_i) = 0$
- Relevancia del Instrumento : $\text{Cov}(z_i, x_i) \neq 0$

Aplicación en base al curso de econometría con Ramiro de Elejalde

Using Geographic Variation in College Proximity to Estimate the Return to Schooling, in Aspects of Labour Market Behavior by David Card (1995)

- Busca estimar los retornos de la educaci3n.
- Una relaci3n causal entre ganancia y educaci3n necesita un instrumento ex3geno.
- Se estudia la proximidad a un centro educativo como instrumento de la educaci3n
- Diferencias geogr3ficas en acceder a centros educativos como una fuente de variaci3n ex3gena.

Utilizando la base de datos : CARD.dta, estimar el siguiente modelo:

$$\log(\text{wage}_i) = \beta_0 + \beta_1 \text{educ}_i + \beta_2 \text{exper}_i + \beta_3 \text{exper}_i^2 + \beta_4 \text{black}_i \quad (4)$$

$$+ \beta_5 \text{south}_i + \beta_6 \text{smsa}_i + \beta_7 \text{smsa66}_i + \beta_8 \text{reg}_i \quad (5)$$

- Estime la ecuaci3n de demanda por MCO. Interprete el resultado de la estimaci3n.
- use la variable de un conjunto de universidad de 4 periodos a cierta distancia como instrumento de la educaci3n.
- use la variable de un conjunto de universidad de 4 periodos y dos periodos a cierta distancia como instrumento de la educaci3n.

Aplicación en Base al curso de econometría II de Marcela Peticar

Estimating the Economic Model of Crime with Panel Data by
Christopher Cornwell and William N. Trumbull

- El paper tiene como objetivo demostrar la incidencia que tiene el endurecimiento de la justicia e instrumentos de castigo en actividades criminales a nivel de condados en Carolina del Norte.
- Los autores tratan las dos fuentes posibles de endogeneidad, por omisión de variable y simultaneidad con un modelo de efecto fijos en dos etapas.

- La Base de datos CRIME4.dta fue obtenida del repertorio de datos estadísticos de Jeffrey M. Wooldridge quien la obtuvo del profesor Cornwell. Es un panel de datos de 90 condados de 1981-1987 para Carolina del Norte (Estados Unidos).
- El conjunto de datos contiene 630 observaciones y 59 variables.

Los autores proponen el siguiente modelo:

$$R_{it} = X_{it}^l \beta + P_{it}^l \gamma + \alpha_i + e_{it} \quad (6)$$

R_{it} es el ratio de crímenes.

X_{it}^l es un vector de variables que controlan por el retorno relativo de las actividades ilegales, estas variables son: wcon, wtuc, wtrd, wfir, wser, wser, wmfng, wfed, wsta, wloc, polpc, urban, density, west, central, pctymle, pctmin80.

P_{it}^l es un set de variables que representan los desincentivos del crimen: prbarr, prbconv, prbpris, avgsen.

α_i es un efecto por condado no observado que podría estar correlacionado con los sets de variables independientes. e_{it} es el error de la regresión.

A partir del modelo anterior se puede definir un modelo Between:

$$R_i = X_i' \beta + P_i' \gamma + \alpha_i + e_i \quad (7)$$

Donde las variables están expresadas en medias de condados:

$R_i = \frac{\sum_t R_{it}}{T}$. Asimismo, se puede plantear un modelo del tipo within:

$$\overline{R}_{it} = \overline{X}_{it}' \beta + \overline{P}_{it}' \gamma + \overline{e}_{it} \quad (8)$$

Donde los datos se encuentran en desviaciones con respecto de la media: $\overline{R}_{it} = R_{it} - R_i = R_{it} - \frac{\sum_t R_{it}}{T}$

Veamos un ejemplo mas:

La data llamada wagepan.raw proviene del paper de Vella y Verbeek (1998). Este archivo contiene datos para 545 hombres, que trabajan durante 1980-1987. Considere el siguiente modelo.

$$\log(\text{wage}_{it}) = \theta_t + \beta_1 \text{educ}_i + \beta_2 \text{black}_i + \beta_3 \text{hispanic}_i + \beta_4 \text{exper}_{it} \quad (9)$$

$$+ \beta_5 \text{exper}_{it}^2 + \beta_6 \text{married}_{it} + \beta_7 \text{union}_{it} + c_i + u_{it} \quad (10)$$

- Estime el modelo usando el estimador de POLS.
- Estime el modelo usando el estimador de efectos aleatorios.
- Ahora estime el modelo usando el estimador de efectos fijos.
- Ahora agregue interacciones entre las dummies temporales y la variable educacion. Estime el modelo por FE.

Aplicaci3n en base al curso de pol3ticas sociales con Eugenio
Guiolito

Incentives Work: Getting Teachers to Come to School by Duflo et.
al (2010)

Construcción del experimento:

- Las escuelas bajo tratamiento fueron provistas por cámaras fotográficas para controlar la asistencia del profesor
- Se le paga un extra por cada día adicional
- Se eligieron 120 escuelas, donde 57 aleatorias fueron tratadas.
- Asistencia de maestros y alumnos con chequeos aleatorios mensuales
- Fecha de anuncio del experimento fue el 24 de agosto de 2003

- a) Vamos a Replicar el impacto del tratamiento sobre la asistencia de los profesores (panel A) y sobre la asistencia de los estudiantes (panel B) de la Tabla No 1 .
- b) Replicamos también todos los resultados de la Tabla No 2 respecto a la asistencia de los profesores.

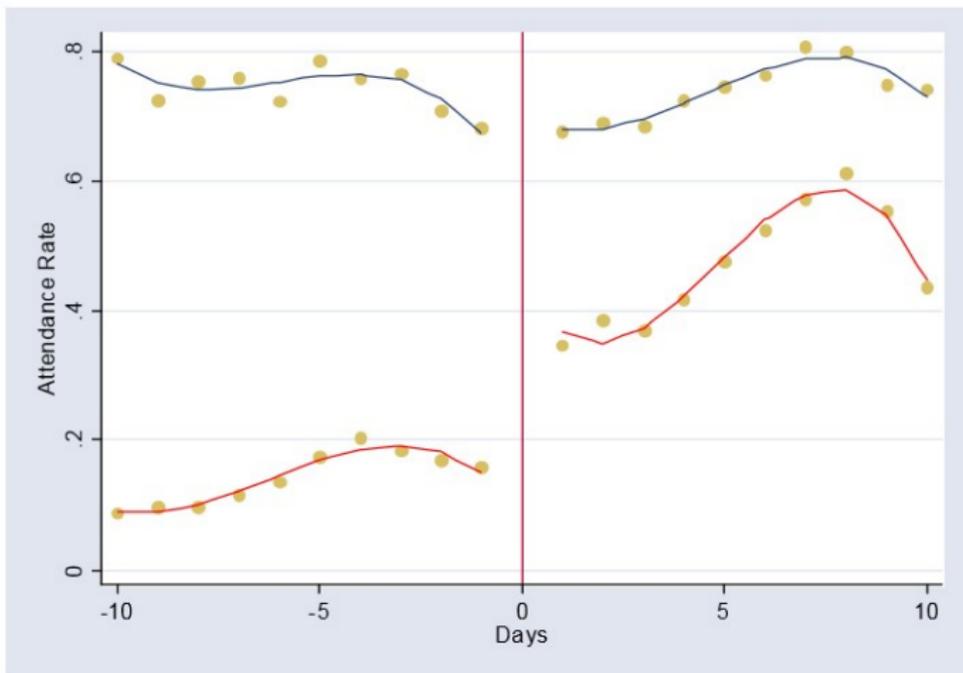
- Son experimentos aleatorios creados por una regla natural o artificial
- Supuesto Principal: Independencia Condicional.

Aplicación en base al curso de políticas sociales con Eugenio
Guiolito

Incentives Work: Getting Teachers to Come to School by Duflo et.
al (2010)

Aca el codigo aun

El resultado debería ser:



Note: (1) The top lines represent the months in which the teacher is in the money, while the bottom lines represent the months in which the teacher not in the money. (2) The estimation includes a third order polynomial of days on the left and