

Datos Panel

Oscar Galvez-Soriano

University of Houston
Department of Economics

November, 2021

1 ¿Qué sabemos hasta ahora?

2 Datos Panel

- Efectos Fijos
- Pooled and Between
- Efectos Aleatorios

¿Qué sabemos hasta ahora?

- RQ: ¿Cuál es el efecto del tratamiento D en el resultado y ?
- Sólo observamos a las personas tratadas y sus resultados $y_{D=1}$, pero para responder la pregunta necesitamos observar $y_{D=0}$
- ¿Y si usamos a las personas no-tratadas? El problema es que los no tratados son muy diferentes a los tratados
- Si tanto tratados como no-tratados se seleccionan aleatoriamente, en promedio serán iguales y, en consecuencia, comparables
- Aún controlando por covariantes NO podemos hablar de un ATE

Datos Panel

- Tenemos y_{it} , x_{it} por cada individuo i observado en el tiempo t
- Llamémosle N al número de individuos y T al número de observaciones en el tiempo
- Para la mayoría de los datos en Microeconomía, N es muy grande respecto a T . La teoría asintótica proviene del hecho de que $N \rightarrow \infty$
- La variación por individuo en el tiempo se conoce como “within variation”
- La variación entre individuos para un momento en el tiempo se conoce como “between variation”

Datos Panel

Diferentes tipos de estimadores usan diferentes fuentes de variación:

- Estimador de efectos fijos. Usa solo “within variation”
- “Between estimator”. Usa solo “between variation”
- Estimador “pooled OLS” y el estimador de efectos aleatorios. Usan ambos tipos de variación
- La heteroscedasticidad y la correlación serial son problemas típicos de datos panel

Modelo de efectos lineales no observables

Suponga que el modelo verdadero es:

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}$$

- $\mathbf{x}_{it} = (x_{it1}, x_{it2}, \dots, x_{itK})$ contiene variables que cambian en el tiempo pero no entre individuos, variables que cambian entre individuos pero no en el tiempo y variables que cambian entre individuos y en el tiempo
- c_i son características no observables del individuo que son constantes en el tiempo
- Nos preocupa que c_i esté correlacionado con \mathbf{x}_{it} . En este caso usar efectos fijos

Efectos Fijos

Si tuviéramos solo datos de corte transversal, β no estaría identificado porque no podemos separar el efecto de c_i del efecto de x_{it} . Supongamos ahora que solo hay dos periodos ($t = 1$ y $t = 2$). Es posible eliminar los problemas causados por la variable no observable al diferenciar los datos:

$$(y_{i2} - y_{i1}) = (\mathbf{x}_{i2} - \mathbf{x}_{i1})\beta + (c_i - c_i) + (u_{i2} - u_{i1})$$

$$(y_{i2} - y_{i1}) = (\mathbf{x}_{i2} - \mathbf{x}_{i1})\beta + (u_{i2} - u_{i1})$$

Efectos Fijos

También podríamos estimar nuestro modelo de interés directamente, sin transformar las series.

- Genere variables dummy por individuo ($d1_i, d2_i, \dots, dN_i$).
- Haga la regresión de y_{it} con x_{it} y con $d1_i, d2_i, \dots, dN_i$.
- En modelos lineales, el controlar por dummies individuales implica un control de efectos específicos de cada individuo

Efectos Fijos

Con una estimación de EF podemos controlar por todas las variables que no varían en el tiempo, pero sí entre individuos, sin la necesidad de especificarlas o medirlas. Sin embargo, podría haber otras fuentes de sesgo por variables omitidas:

- Variables que cambian en el tiempo pero no entre individuos
- Variables que cambian entre individuos y en el tiempo

Es decir, con una estimación por EF, no podemos recuperar los efectos de las variables que NO cambian entre individuos, pero sí en el tiempo

Efectos Fijos

En STATA podemos usar FE de muchas maneras. Supongamos que nuestra variable que identifica los individuos se llama “id_var”:

- 1 Usando el comando “xtreg, fe”
- 2 Usando el comando “areg”. Por ejemplo: “areg y x, absorb(id_var)”
- 3 Usando el comando “xi”. Por ejemplo: “xi: reg y x i.id_var”
- 4 También podríamos generar las variables dummies “a mano”

Efectos Fijos

La base de datos pertenece a un documento de investigación en proceso. Citar como:

Gálvez-Soriano, Oscar (2021). “A note on the international coffee price pass-through into the Mexican market”
Unpublished Manuscript.

Pooled and between estimators

Pooled estimator

$$y_{it} = x_{it}\beta + c_i + u_{it}$$

Between estimator

$$\bar{y}_i = \bar{x}_i\beta + c_i + \bar{u}_i$$

Ahora hay solo una observación por individuo. No estamos usando la variación en el tiempo

Efectos Aleatorios

- Los efectos aleatorios (EA) son un promedio ponderado de el estimador “within” y el estimador “between”
- Dado que el estimador EA utiliza parte de la variación del corte transversal, si c_i estuviera correlacionada con x_{it} , el estimador NO sería consistente
- La prueba de Hausman nos permite decidir si es conveniente utilizar EF o EA. Bajo la hipótesis nula, ambos estimadores son iguales. Bajo la hipótesis alterna ambos estimadores son diferentes
- Si no pueden rechazar la H_0 , entonces es válido reportar el estimador con EA. Sin embargo, siempre será válido reportar con EF