Datos en Panel

Walter Sosa-Escudero

Variables binarias para categorias multiples

Y (salarios), X (educacion), este, centro, oeste

$$Y_i = \beta_1 + \beta_2 X_i + \delta_1 D_{1i} + \delta_2 D_{2i} + u_i$$

$D_{1i} = 1[\text{este}], \;\; D_{2i} = 1[\text{centro}]$

$$\begin{aligned} Y_i|\text{oeste} &= \beta_1 + \beta_2 X_i + u_i \\ Y_i|\text{este} &= \beta_1 + \delta_1 + \beta_2 X_i + u_i \\ Y_i|\text{centro} &= \beta_1 + \delta_2 + \beta_2 X_i + u_i \end{aligned}$$

- Categoria base: oeste.
- Efectos diferenciales: $\delta_1 =$ este/oeste, $\delta_2 =$ centro/oeste.
- centro/este?

$D_{3i} = 1[\text{oeste}]$

$$Y_i = \beta_1 + \beta_2 X_i + \delta_1 D_{1i} + \delta_2 D_{2i} + \delta_3 D_{3i} + u_i$$

Problema?

En general

- Trampa de variables binarias: si hay un intercepto, para J categorias, incluir $J-1$ variables binarias.
- Intercepto y $J-1$ variables binarias da los mismos efectos diferenciales que sin intercepto y J variables binarias

$$Y_i = \beta_2 X_i + \delta_1 D_{1i} + \delta_2 D_{2i} + \delta_3 D_{3i} + u_i$$

Verificar, como ejercicio.

- Idea: misma pendiente, distintos interceptos para cada categoria. Generalizacion: diferentes pendientes.

Por que paneles?

Ejemplo (Cronwell y Trumbull): Determinantes del crimen

- $y = g(I)$, $y =$ crimen, $I =$ justicia criminal.
- Corte transversal: (y_i, I_i) para regiones $i = 1, \ldots, n$
- I es 'importante'
- Critica: I captura el efecto de otros factores regionales. Sesgo por variable omitida. Control

Paneles: solucion al problema de sesgo, sin agregar otras variables.

Modelo simple para paneles

$$y_{it} = x'_{it}\beta + u_{it}$$

$$u_{it} = \mu_i + \delta_t + \epsilon_{it}$$

$i = 1, \ldots, N, \;\; t = 1, \ldots, T$. x_{it}, a vector de K variables explicativas.

- Panel: doble variabilidad
- u_{it}
- μ_i efectos no observables que varian solo por individuo.
- SPG, $\delta_t = 0$

Efectos fijos

$$y_{it} = x_{it}'\beta + \mu_i + \epsilon_{it}$$

Efecto fijo: Estima β y μ_i como parametros $K + (N-1)$

Modelo lineal con un intercepto para cada individuo:

$$y_{it} = \underbrace{\mu_i + \beta_1} + \beta_2\ x_{2,it} + \cdots + \beta_K\ x_{K,it} + \epsilon_{it}$$

Agregar $N-1$ variables binarias por individuo, o sin intercepto y N binarias.

En terminos matriciales

$$Y = X\beta + D\mu + u$$

$Y_{NT\times 1}$, $X_{NT\times K}$ sin intercepto. D, matriz de N binarias por individuo.

$$1_N \equiv \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \qquad D = \begin{bmatrix} 1_N & 0 & \cdots & 0 \\ 0 & 1_N & 0 & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1_N \end{bmatrix}_{NT\times N}$$

Escribamos el modelo como:

$$Y = X\beta + D\mu + u = \dot{X}\delta + u$$

con $\dot{X} \equiv [X \;\; D]$ y $\delta \equiv [\beta' \;\; \mu']'$.

El *estimador de efectos fijos* es:

$$\hat{\delta}_{EF} = \begin{pmatrix} \hat{\beta}_{EF} \\ \hat{\mu}_{EF} \end{pmatrix} = (\dot{X}'\dot{X})^{-1}\dot{X}'Y.$$

MCO mas N dummies ('LSDV').

Efectos fijos y la transformacion 'within'

Por el TFWL

$$\hat{\beta} = (X^{*'} X^{*})^{-1} X^{*'} Y^{*}$$

con $X^* \equiv M_D X$ y $Y^* \equiv M_D Y$. $M_D = I - D(D'D)^{-1}D'$ es la matriz que genera errores de regresar las X y las Y en D.

Es facil mostrar (hacerlo!) que

$$X^*_{it} = X_{it} - \bar{X}_i$$

Transfomacion within: controlar por efectos fijos es identico a restar las medias individuales.

El TFWL sugiere dos forma identicas de obtener el estimador de efectos fijos

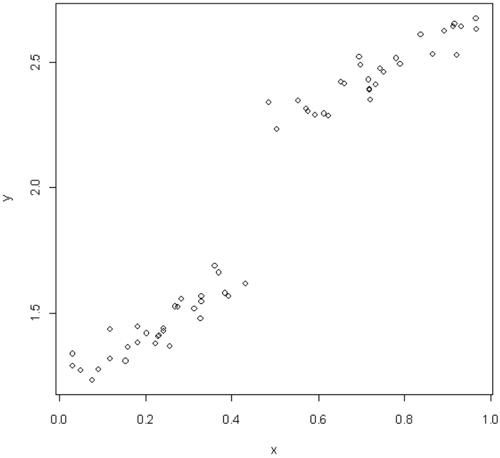
- Regresar Y on X and D.
- Primero restar las medias individuales y luego regresar.

Conceptualmente relevante (nexo con VI, ya veremos)

Efectos fijos y heterogeneidad no observable

- $\hat{\beta}_{FE}$ insesgado, independientemente de si X correlacionadad con D (FE controla for D).
- Si la variable omitida en nuestro ejempo varia solo por individuo, es como si hubiesemos controlado por ella aun sin observarla.
- Intuicion: la transformacion within elimina cualquier variable que varia a nivel individual, observada o no.

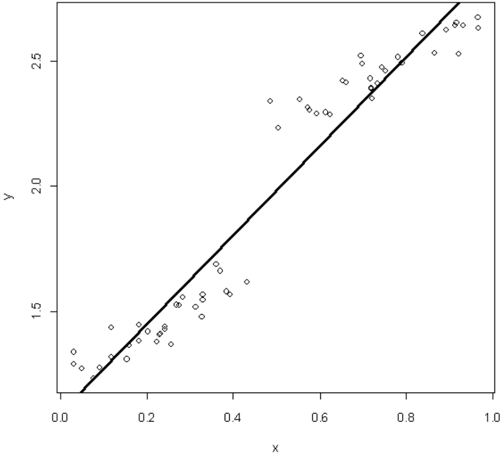
Panel como control: representacion grafica



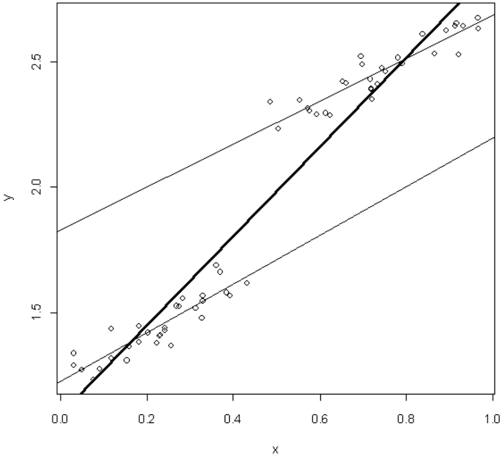
Verbalizacion

- Y = crimen.
- X = ineficiencia del sistema judicial.
- Dos regiones
- Determinante del crimen no observado, que varia por region y positivamente correlacionado con la ineficiencia judicial (densidad?).

MCO



Efectos fijos



Tres estimadores

$$y_{it} = x_{it}\beta + \mu_i + \epsilon_{it}$$

Within: $y^*_{it} = x^*_{it}\beta + \epsilon^*_{it}$

- Controla por μ_i. Varacion N, T. Insesgado siempre

Between: $\bar{y}_i = \bar{x}_i + \mu_i + \bar{\epsilon}_i$

- Omite μ_i. Varacion N. Insesgado solo si $\mu \not\angle \epsilon$

MCO de y_{it} en x_{it}?

Table 3.—Results from Estimation
(standard errors in parentheses)

	Between	Within	2SLS (fixed effects)	2SLS (no fixed effects)
CONSTANT	−2.097			−3.719
	(2.822)			(8.189)
P_A	−0.648	−0.355	−0.455	−0.507
	(0.088)	(0.032)	(0.618)	(0.251)
P_C	−0.528	−0.282	−0.336	−0.530
	(0.067)	(0.021)	(0.371)	(0.110)
P_P	0.297	−0.173	−0.196	0.200
	(0.231)	(0.032)	(0.200)	(0.343)
S	−0.236	−0.00245	−0.0298	−0.218
	(0.174)	(0.02612)	(0.0300)	(0.185)
POLICE	0.364	0.413	0.504	0.419
	(0.060)	(0.027)	(0.617)	(0.218)
DENSITY	0.168	0.414	0.291	0.226

Probs de arresto, condena, prision

Diferencias-en-diferencias

Efecto de salarios minimos en empleo (Card, Krueger (1994)).

- Salario minimo (MW) reduce el empleo?
- Empleo en McDonalds antes y despues de cambios en MW? Confunde MW con otros cambios temporales.
- Empleo en McDonalds, mismo periodo, estados diferentes con diferentes politicas de MW ? Confounde MW con determinantes regionales del empleo.

Contexto hiper-simple

- Unidad de analisis: restaurant i en el estado s en el periodo t.
- Variable de interes Y_{its} : empleo en el restaurant its.
- Dos periodos: $t = 1, 2$
- Dos estados: A, B.
- N restaurants por estado.
- MW, politica a nivel *estado*.
- Estado A no cambia la politica de MW. Solo B lo hace, en el periodo 2.
- $D_{ist} = 1$ si el estado s cambia MW en t, 0 si no.
- Notar que $D_{ist} = 1$ iff $s = B$ y $t = 2$.

Modelo aditivo:

$$Y_{its} = \gamma_s + \lambda_t + \beta D_{ist} + \epsilon_{ist}$$

β es el parametro de interes: efecto de MW controlando por factores regionales (γ_s) y temporales (λ_t).

Diferencias-en-diferencias

$$Y_{its} = \gamma_s + \lambda_t + \beta D_{ist} + \epsilon_{ist}$$

Notar

$$\begin{aligned} E(Y|B,2) - E(Y|B,1) &= \lambda_2 - \lambda_1 + \beta \\ E(Y|A,2) - E(Y|A,1) &= \lambda_2 - \lambda_1 \end{aligned}$$

Restando

$$\begin{array}{ccccc} \Big[E(Y|B,2) - E(Y|B,1)\Big] & - & \Big[E(Y|A,2) - E(Y|A,1)\Big] & = & \beta \\ \text{Cambio en B} & - & \text{Cambio en A} & = & \beta \end{array}$$

$\hat{\beta}$ = Average change in B − Average change in A

Dif-in-dif via paneles

$$Y_{its} = \gamma_s + \lambda_t + \beta D_{ist} + \epsilon_{ist}$$

- $DB_{ist} = 1$ sii i esta en el estado B
- $D2_{ist} = 0$ sii $t = 2$.
- Por contruccion $D_{ist} = DB_{ist} \times D2_{ist}$ (el cambio ocurre solo en el estado B y el periodo 2).

Reemplazando:

$$Y_{its} = \gamma_s + \lambda_t + \beta(DB_{ist} \times D2_{ist}) + \epsilon_{ist}$$

$$Y_{its} = \gamma_s + \lambda_t + \beta(DB_{ist} \times D2_{ist}) + \epsilon_{ist}$$

- Es un *panel* de N restaurants en 2 regions y 2 periods, con efectos fijos regionales y por periodos.
- Regresar Y_{its} en 1) dummy for region B, 2) dummy por periodo 2) 'interacciones' entre ambas.
- Parametro de interes: β.

Comments

- Panel? errores estandar y tests de hipotesis.
- 'Tendencia comun': crucial. Ambos estados comparten λ_t: la evolucion del empleo es identica en ambos estados. 'Tratamiento' (cambio en MW) solo implica salirse de la tendendia comun. Testeable
- No correlacion serial: supuesto clave para inferencia. Clustered standard errors, Bertrand, Duflo, y Mullainathan (2004).

Random effects estimator

Same model

$$Y = X\beta + D\mu + \epsilon$$

If seen as a random variable $D\mu$ es orthogonal to X, and if $E(\mu_i|X) = 0$, then the OLS estimator that regresses Y on X is unbiased.

That is, if $D\mu$ is orthogonal to X, omitting the dummy variables does not bias OLS.

Fixed or random?

Careful. It is a matter of treatment/estimation.

$$Y = X\beta + D\mu + \epsilon$$

Fixed effects (controls for $D\mu$)

$$Y = X\beta + D\mu + \epsilon$$

Random effects (treats $D\mu$ as an omitted variable)

$$Y = X\beta + D\mu + \epsilon$$

$$Y = X\beta + D\mu + \epsilon$$

$$Y = X\beta + u, \quad u \equiv D\mu + \epsilon$$

Problem: u does not satisfy the classical assumptions, even when $D\mu$ and ϵ do.

Simple proof: assume classical assumptions separately (zero expected value, no serial correlation/heteroskedasticity). Also $D\mu$ and ϵ uncorrelated. Then

$$\begin{aligned} V(u) &= V(D\mu + \epsilon) \\ &= DV(\mu)D' + V(\epsilon) \\ &= \sigma_\mu^2 DD' + \sigma_\epsilon^2 I_{NT}, \end{aligned}$$

certainly non-spherical.

Intuition: $u_{it} = \mu_i + \epsilon_{it}$

- Trivially, u_{it} is correlated with $u_{i,t-1}$ since both 'share' μ_i: the persistent presence of μ_i implies that random effects induce serial correlation.
- Though not biased, OLS is inefficient (in the sense discussed in class).
- Efficient? GLS.

Generalized Least Squares

Supongamos que valen todos los supuestos clasicos, pero que

$$V(u) = \sigma^2 \Omega,$$

con Ω cualquier matriz positiva (permitimos heterocedastiticad y/o correlacion serial). Perdimos Gauss/Markov

Resultado (Aitken): bajo todos los supuestos clasicos y cuando $V(u) = \sigma^2 \Omega$

$$\hat{\beta}_{gls} \equiv (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$$

es un estimador de varianza minima en la clase de estimadores lineales e insesgados.

Prueba: Caso X no aleatoria. En primer lugar, si Ω es positiva, existe $C_{n\times n}$ invertible tal que $C'C = \Omega^{-1}$. Partiendo de $Y = X\beta + u$, premultipliquemos ambos lados por C, para obtener

$$Y^* = X^*\beta + u^* \tag{1}$$

con $Z^* = CZ$, $Z = Y, X, u$. El modelo transformado (1) tiene los mismos coeficientes desconocidos que el modelo original, y es facil mostrar que satisface todos los supuestos clasicos (chequear). En particular

$$V(u^*) = E(u^*u^{*'}) = C\sigma^2\Omega C' = \sigma^2 C\left[(C'C)^{-1}\right]C' = \sigma^2 I_n$$

Entonces, por el Teorema de Gauss/Markov aplicado al modelo transformado, el estimador de minima varianza en la clase de los lineales e insesgados es

$$\hat{\beta}_{gls} = (X^{*'}X^*)^{-1}X^{*'}Y^* = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$$

GLS 'factible'

$$\hat{\beta}_{fgls} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}Y$$

en donde $\hat{\Omega}$ es cualquier estimador consistente de Ω.

GLS random effects

Recall

$$\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$$

In our case

$$V(u) = \sigma_\mu^2 DD' + \sigma_\epsilon^2 I_{NT} \equiv \Omega(\theta)$$

with $\theta' = (\sigma_\mu^2, \sigma_\epsilon^2)'$

- FGLS requires to estimate θ first (variance components).
- Random effects estimator: GLS for random effects.

Test de Hausman

$$Y = X\beta + D\mu + \epsilon$$

- $X \perp D\mu$: MCO, EF, BE,RE todos insesgados para β. RE (por que).
- $X \not\perp D\mu$: solo EF insesgado β.

Test de Hausman

$$H_0 : X \perp D\mu, \quad H_A : X \not\perp D\mu$$

Test de Hausman: under H_0

$$H = (\hat{\beta}_{BE} - \hat{\beta}_{EF})'(\Omega_{EF} - \Omega_{BE})^{-1}(\hat{\beta}_{BE} - \hat{\beta}_{EF}) \;\; \sim \;\; \chi^2(K)$$

Rechazar si H es significativamente alto.
Hausman/Taylor (1981): es lo mismo reemplazar BE con RE.

Intuicion: bajo H_0, $\hat{\beta}_{BE}$ y $\hat{\beta}_{EF}$ consistentes, H chico. Bajo H_A, soloi $\hat{\beta}_{EF}$ consistente, H grande.