Panel Data

Walter Sosa-Escudero

- A quick introduction to standard panel methods.
- Mostly and application of FWL and GLS theory.

Why panels?

Example (Cronwell y Trumbull): Determinants of crime

- $y = g(I)$, $y =$ crime, $I =$ criminal justice variables.
- Cross sectoin: (y_i, I_i) for several regions $i = 1, \ldots, n$
- I is 'important'
- Criticism: I captures the effect of other regional effects.

- There is an omitted variable, likely related to I.
- OLS that regressesa y on I is biased.
- Solution? ‘Control’ for this omitted variable.

Panels to the rescue: a feasible solution without using other variables

A simple linear panel model

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

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

$i = 1, \ldots, N, \;\; t = 1, \ldots, T.$ x_{it}, a vector of K explanatory variables, including a constant.

μ_i captures the effect of non-measured effects that vary by individuals only. ϵ_{it} is the standard errror term.

Fixed effects estimation

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

Estimates β and μ_i as extra parameters.

It can be seen as a standard linear model where each individual has its own intercept:

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

Add $N-1$ individual level dummy variables.

In matrix terms

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

Y is $NT \times 1$, X is $NT \times K$, X includes and intercept.
D is a matrix of $N-1$ individual level dummy variables.

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

Write the model as:

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

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

Then, the *fixed effects estimator* is:

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

Just the OLS estimator adding $N-1$ individual level dummies.

Fixed effects and the within transformation

If the interest is in estimating β, by the Frisch-Waugh-Lovell result, we can do

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

where $X^{*} \equiv M_D X$, and $M_D = I - D(D'D)^{-1}D'$, the matrix that projects X on the orthogonal complement of D. Y^{*} is defined accordingly.

As a simple exercise, show

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

That is, getting rid of the dummies in the first step amounts to substracting individual level means. This is the within tranformation.

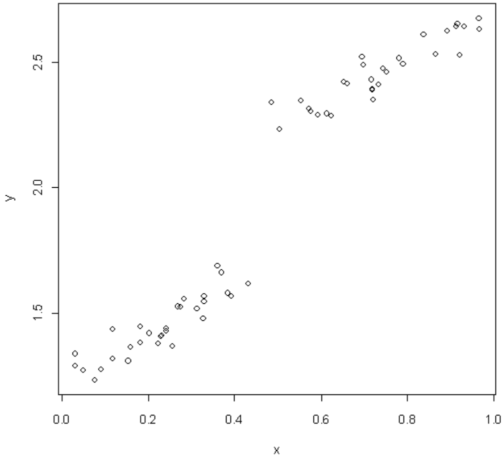
The FWL Theorem suggests two forms of obtaining the fixed effects estimator

1. Regress Y on X and D.
2. First demean all data with respect to individual means. Then do OLS with demeaned data

Fixed effects and unobserved heterogeneity

- $\hat{\beta}_{FE}$ is unbiased, independently of whether X is correlated with D (FE controls for D).
- If the ommited variable in our example varies only at the individual level, it is as if we had controlled for it without observing it.
- Intuition: the within transformation kills every variable that varies at the individual level only, observed or not.

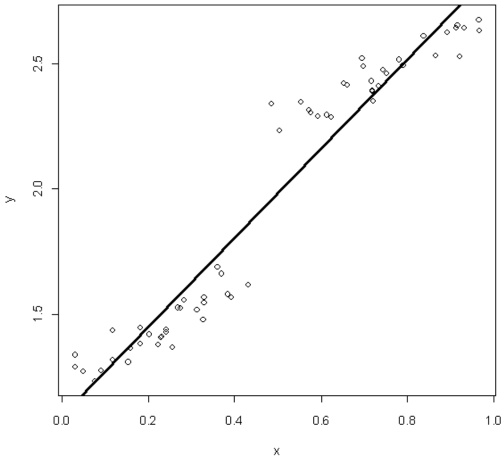
Biases: graphical representation



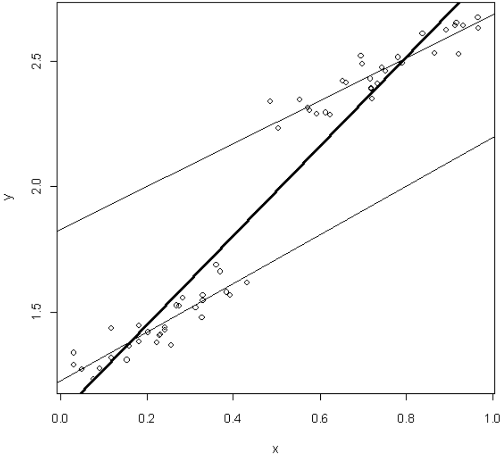
Verbalization

- Y = crime rate.
- X = ineficiency of judicial system (more inefficiency, more crime).
- Two regions
- An omitted crime determinant that varies only by region and positively correlated with judicial inefficienty (population density?).

OLS



Fixed effects



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.

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.

Summary

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

- $X \perp D\mu$: OLS, FE, RE are unbiased for β. RE is efficient.
- $X \neg\perp D\mu$: only EF is unbiased for β.
- Most practitioners use FE.

Hausman test

$H_0 : X \perp D\mu$, $H_A : X \;\neg\perp\; D\mu$

Hausman Test: under H_0

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

Reject if H is significantly high.

Intuition: under H_0, $\hat{\beta}_{EA}$ y $\hat{\beta}_{EF}$ are consistent, H should be small. Under H_A, only $\hat{\beta}_{EF}$ is consistent, H should be lasrge.

Panels and impact evaluation

Motivation: effect of minimum wages on employment (Card, Krueger (1994)).

- Intuition: minimum wage (MW) reduces employment.
- Compare employment in McDonalds before and after MW wage changes? Confounds MW effect with temporal changes.
- Compare employment in McDonalds, same period, two states with different MW policies? Confounds MW with regional determinants.

A very simple context

- Unit of analysis: restaurant i in state s in period t.
- Variable of interest: Y_{its} : employment in restaurant its.
- Two periods: $t = 1, 2$
- Two states: A, B.
- N restaurants per state.
- MW is a state policy.
- State A does not change MW. Only B does it, in period 2.
- $D_{ist} = 1$ if state s changes MW in t, 0 otherwise.
- Note that in our case $D_{ist} = 1$ iff $s = B$ y $t = 2$.

Additive structure:

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

- β is the key parameter: effect of MW controlling for regional (γ_s) and temporal (λ_t) factors that confound MW in the determination of employment.
- We will assume that given γ_s and λ_t, D_{st} is exogenous.

Estimation 1: 'Differences in differences'

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

Note

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

Substracting

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

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

Estimation 2: Panels

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

- $DB_{ist} = 1$ iff i is in state B
- $D2_{ist} = 0$ iff $t = 2$.
- Then, by construction $D_{ist} = DB_{ist} \times D2_{ist}$ (change occurs only in state B and in period 2).

Replacing:

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

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

- It is a *panel* of N restaurants in 2 regions and 2 periods, with regional and period specific fixed effects.
- Regress Y_{its} on 1) dummy for region B, 2) dummy for period 2) 'interaction' between both.
- The parameter of interest is the coefficient of the interaction term.

Comments

- Panel estimation facilitates computation of standard errors and hypothesis tests.
- 'Common trends': crucial. Both states 'share' λ_t: the temporal evolution of employment in both states is identical. 'Treatment' (MW change) implies departing away from this common trend.
- No serial correlation: key assumption for inference. See Bertrand, Duflo, y Mullainathan (2004) on cluster correlation.