Econometrics

Panel Data

Paul P. Momtaz

The Anderson School
UCLA

Outline

The Cross-Sectional Problem

The problem with cross sectional data (solved by panel data):
Assume true causal relationship

$$y_{it} = \beta_1 + \beta_2 l_{it} + \beta_3 m_i + \varepsilon_{it} \qquad m_i \text{ is time invariant, unobservable}$$

Assume only cross-sectional data available to estimate:

$$y_i = \beta_1 + \beta_2 l_i = \eta_i$$

-

$$\begin{aligned}\mathbb{E}[y_i|l] &= \beta_1 + \beta_2 l_i + \mathbb{E}[\eta_i|l] \\ &= \beta_1 + \beta_2 l_i + \beta_3 \mathbb{E}[m_i|l]\end{aligned}$$

-

$$\begin{aligned}\mathbb{E}[m_i|l] &= \lambda_1 + \lambda_2 l_i \\ \mathbb{E}[y_i|l] &= \underbrace{(\beta_1 + \beta_3\lambda_1)}_{b_1} + \underbrace{(\beta_2 + \beta_3\lambda_2)}_{b_2} l_i\end{aligned}$$

b_2 biased estimate of β_2

- If $\beta_3 > 0, \beta_2$ over-estimated if $\lambda_2 > 0$ or underestimated if $\lambda_2 < 0$

General Model

$$\begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} i & 0 & \dots & 0 \\ 0 & i & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & i \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_N \end{bmatrix} + \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \beta + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{bmatrix}$$

Where, i is the T-dimensional column vector with ones, y_i and x_i are T times obs for i , ε_i is T disturbances and a_i is FE for i

Outline

Fixed Effects (or Within) Estimators

Least-Suares Dummy Variable (LSDV) Model

$$y_{it} = d_1\alpha_1 + \cdots + d_N\alpha_N + x_{it}\beta + \varepsilon_{it} \qquad \text{or} \qquad Y = D\alpha + X\beta + \varepsilon$$

Where D_i is $NT \times T$ matrix, α is $N \times 1$, X is $NT \times x$, β is $x \times 1$, ε is $NT \times 1$

$\Rightarrow$Very simple, but computationally infeasible (often times N too large)

Outline

Fixed Effects (or Within) Estimators

Analysis of Covariance

Idea: use partitioned regression:

1. $y \sim D$ and get residuals y^*.
2. $x \sim D$ and retrieve residuals x^*.
3. $y^* \sim x^*$ to obtain unbiased β

Fixed Effects (or Within) Estimators

Analysis of Covariance

Mechanics:

- Projection matrix $M = D(D'D)^{-1}D'$.
- Pre-multiply any vector z by M gives least squares prediction of z given D.
- Partialling out matrix $M = I - D(D'D)^{-1}D'$
- Pre-multiply any z by M gives least squares residuals of regression $z \sim D$.
- Note $MD = 0, MM = M, DD = D$ (idempotent)
- $My = MD\alpha + Mx\beta + M\varepsilon \Rightarrow y^* = x^*\beta + \varepsilon^*$

Fixed Effects (or Within) Estimators

Analysis of Covariance

Note

$$D = \begin{bmatrix} i & 0 & \dots & 0 \\ 0 & i & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & i \end{bmatrix}, \qquad M = I - D(D'D)^{-1}D' = \begin{bmatrix} \overline{M} & 0 & \dots & 0 \\ 0 & \overline{M} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \overline{M} \end{bmatrix}$$

$$\overline{M} = I_T - \frac{1}{T} ii', \qquad \overline{M}z = z - \overline{z}_i$$

Hence

$$\text{My}{=}\text{MD}\alpha + Mx\beta + M\varepsilon = y^* = x^*\beta + \varepsilon^* = [y_{it} - \overline{y}_i] = [x_{it} - \overline{x}_i] + [\varepsilon_{it} - \overline{\varepsilon}_i]$$

$\Rightarrow$Unbiased and consistent $b_{FE} = [X'MX]^{-1}[X'MY]$, $a_i \overline{y}_i - \overline{x}_i b_{FE}$

Outline

The Cross-Sectional Problem

General Model

Within Estimators

Least-Suares Dummy Variable Model

Analysis of Covariance

First Difference

FE Estimators and Lagged Dependent Variables

Between Estimators

Comparison of Estimators

Random Effects Estimator

First Difference

Only two time observation for each i

$y_{i1} - y_{i2} = [x_{i1} - x_{i2}]\beta + \varepsilon_{i1} - \varepsilon_{i2}$ by OLS where β is *FE* estimator

Difference-in-differences Strategies

Outline

FE Estimators and Lagged Dependent Variables
Problem

$$y_{it} = \alpha_i + y_{it-1}\rho + x_{it}\beta + \varepsilon_{it}$$

problem: y_{it-1} predetermined but not strictly exogenous.

$\Rightarrow$Both LSDV and An .of. Cov. estimators of ρ and β will be biased and inconsistent.

FE Estimators and Lagged Dependent Variables

LSDV Bias

LSDV: Reason for bias:

- Define $z_{it} = \begin{bmatrix} D_{it} & y_{it-1} & x_{it} \end{bmatrix}$ and $\gamma = \begin{bmatrix} \alpha \\ \rho \\ \beta \end{bmatrix}$
- $y_{it} = z_{it}\gamma + \varepsilon_{it}$
- $\mathbb{E}[y|z] = z\gamma + \mathbb{E}[\varepsilon|z] \neq z\gamma$ since $\mathbb{E}[\varepsilon|y] \neq 0$
  - $\mathbb{E}[\varepsilon_{it}|y_{is}] = 0$ for $s < t$
  - $\mathbb{E}[\varepsilon_{it}|y_{is}] \neq 0$ for $s \geq t$
- Failure of orthogonality condition

FE Estimators and Lagged Dependent Variables

Analysis of Covariance Bias

An . of . Cov Reason for bias:

- $[y_{iy} - \overline{y}_i] = [y_{it-1} - \overline{y}_{i-1}]\rho + [x_{it} - \overline{x}_i]\beta + [\varepsilon_{it} - \overline{\varepsilon}_i]$
- $\mathbb{E}[y_{is-1} - \overline{y}_{i-1}][\varepsilon_{it} - \overline{\varepsilon}_i] \neq 0$ since
  - $\mathbb{E}[\overline{y}_{i-1}\varepsilon_{it}] \neq 0$
  - $\mathbb{E}[y_{is-1}\overline{\varepsilon}_i] \neq 0$
- If $T \to \infty$, bias $\to 0$.
- If $N \to \infty$, bias $\nrightarrow 0$

FE Estimators and Lagged Dependent Variables

Solution

Solution: First differencing and IV

- $T > 2, [y_{it} - y_{it-1}] = [y_{it-1} - y_{it-2}]\rho + [x_{it} - x_{i-1}]\beta + \varepsilon_{it} + \varepsilon_{it-1}$
- So, α_i is eliminated
- $y_{it-2}, (y_{it-2} - y_{it-3}), x_{it-1}, (x_{it-2} - x_{it-2})$ are valid IV.
- Orthogonality assumption hold
- lags > 1 of dependent variable have to go further XXX yo find valid IV.

FE Estimators and Lagged Dependent Variables

Other Pitfalls

Other pitfalls with FE estimators:

- Waste of “between” estimation.
- loss of degree of freedom $\rightarrow$ loss of efficiency.
- effect of time-invariant explanatory factors eliminated.
- out of sample predictions.

Between Estimator

$$a_i = a + v_i$$
$$\overline{y}_i = \alpha + \overline{x}_i\beta + v_i + \overline{\varepsilon}_i = \alpha\overline{x}_i\beta + \eta_i$$

Estimation by OLS

- Standard OLS assumption on ε_{it}
- In contrast to within/FE, no need to assume i specific effect fixed.
- $cov(\eta_i, \overline{x}_i) = cov(v_i, \overline{x}_i) = 0$
- Unbiased and consistent but not efficient.

Between Estimator

Notations:

- $\overline{\overline{y}} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} y_{it}$
- $\overline{\overline{x}} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} x_{it}$
- Moment matrices of overall sums of squares and cross-products
  - $S_{xx}^{o} = \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \overline{\overline{x}})(x_{it} - \overline{\overline{x}})'$
  - $S_{xy}^{o} = \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \overline{\overline{x}})(y_{it} - \overline{\overline{y}})'$
- Moment matrices of "Within"
  - $S_{xx}^{w} = \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \overline{x}_i)(x_{it} - \overline{x}_i)'$
  - $S_{xy}^{w} = \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \overline{x}_i)(y_{it} - \overline{y}_i)'$
- Moment matrices of "Between"
  - $S_{xx}^{b} = \sum_{i=1}^{N} (\overline{x} - \overline{\overline{x}})(\overline{x}_i - \overline{\overline{x}})'$
  - $S_{xy}^{b} = \sum_{i=1}^{N} (\overline{x} - \overline{\overline{x}})(\overline{y}_i - \overline{\overline{y}})'$

Comparison of Estimators

- FE: $b^w = [S^w_{xx}]^{-1} S^w_{xy}$
- Between: $b^b = [S^b_{xx}]^{-1}[S^b_{xy}]$
- OLS: $b^{OLS} = [S^o_{xx}]^{-1} S^o_{xy} = [S^w_{xx} + S^b_{xx}]^{-1}[S^w_{xy} + S^b_{xy}]$

Note that

$$S^w_{xy} = S^w_{xx} b^w$$
$$S^b_{xy} = S^b_{xx} b^b$$

$$\Rightarrow b^{OLS} = F^w b^w + F^b b^b \quad \text{where} \quad F^w = [S^w_{xx} + S^b_{xx}]^{-1} S^w_{xx} = I - F^b$$

Note: Not most efficient; biased and inconsistent if individual specific effects correlated with regressors.

Random Effects Estimator

In contrast to FE model, RE model assume individual differences are random disturbances. Advantage is that linear difference is used. Disadvantage is that it assumes there is no correlation between regressors and random individual disturbances (Hausman Test).

Random Effects Estimator

$$y_{it} = \alpha + x_{it}\beta + v_i + \varepsilon_{it} = \alpha + x_{it}\beta + w_{it}$$

OLS estimator is inefficient, so transform model and apply GLS

Transform Model: $$v^{-\frac{1}{2}} y_{it} = v^{-\frac{1}{2}} \alpha + v^{-\frac{1}{2}} x_{it}\beta + v^{-\frac{1}{2}} w_{it}$$

Feasible GLS since v unknown! Use OLS for transformed model

Hausman Test $(\mathbb{E}[v_i|x_{it}] = 0 \equiv H_0) : w = [b^{FE} - b^{RE}]'[Vas[b^{FE}] - Vas[b^{RE}]]^{-1}[b^{FE} - b^{RE}]$ is the test statistic with di-squared and XXX