Panel Data

Zhentao Shi

The Chinese University of Hong Kong

Data

- Cross-sectional datasets collected at different time points
- Group-specific intercept (one way to handle endogeneity)
- Real data: `https://www.kaggle.com/code/frankshi0/penn-world-table`

Panel Data Structure

- The same individuals are observed over time $t = 1, \ldots, T$
- Independent across $i = 1, \ldots, N$

$$\underset{NT\times 1}{Y} = \begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1T} \\ y_{21} \\ \vdots \\ y_{NT} \end{bmatrix}, \quad \underset{NT\times p}{X} \begin{bmatrix} X'_{11} \\ X'_{12} \\ \vdots \\ X'_{1T} \\ X'_{21} \\ \vdots \\ X'_{NT} \end{bmatrix}$$

$$y_{it} = c + X'_{it}\beta + \alpha_i^* + u_{it}, \qquad \varepsilon_{it} = \alpha_i^* + u_{it}$$

Real data: `https://www.kaggle.com/datasets/frankshi0/nber-ces-manufacturing-industry-database/code`

Panel Data Regression

- Temporal observations over $t = 1, \ldots, T$ for the same i is viewed as a **group**. Temporal dependence is allowed within the group.

$$y_{it} = c + X'_{it}\beta + \varepsilon_{it}$$

  with $E(\varepsilon_{it}) = 0$.
- ε_{it} may be correlated with X_{it}.
- **Composite error**

$$\varepsilon_{it} = \alpha_i^* + u_{it}$$

  with $\mathrm{cov}(u_{it}, X_{it}) = 0$

Section 1

Fixed Effect Models

Fixed Effects

- Consistency of OLS counts on $\mathrm{cov}(\varepsilon_{it}, X_{it}) = \mathrm{cov}(\alpha_i^*, X_{it}) + \mathrm{cov}(u_{it}, X_{it}) = 0$.
- Example
  - Production function (Mundlak, 1961)

$$y_{it} = c + X_{it}'\beta + m_i\gamma + u_{it}$$

    where m_i is the "management quality" of a firm (usually unobservable).
  - $m_i\gamma$, which can be correlated with X_{it}, is captured by α_i^*
- α_i^* can be potentially endogenous (correlated with X_{it})

Static Linear Model

- Model

$$
\begin{aligned}
y_{it} &= c + X_{it}'\beta + \varepsilon_{it} \\
&= (c + \alpha_i^*) + X_{it}'\beta + u_{it} \\
&= \alpha_i + X_{it}'\beta + u_{it}
\end{aligned}
$$

  where $\alpha_i = c + \alpha_i^*$ is the new **individual-specific** intercept. Also called the **fixed effect**.
- β is a p-dimensional **homogenous** slope coefficient
- $(N + p)$ parameters $(\alpha_1, \alpha_2, \ldots, \alpha_N; \beta)$

Least Squares Dummy Variable Estimator

- Direct Estimation with N dummy variables

$$y_{it} = \sum_{j=1}^{N} \alpha_j \mathbf{I}(i = j) + X_{it}'\beta + u_{it}$$

$$Y = D\alpha + X\beta + u, \qquad D := I_N \otimes \mathbf{1}_T,$$

where $Y \in \mathbb{R}^{NT}$, $D \in \mathbb{R}^{NT \times N}$, $X \in \mathbb{R}^{NT \times p}$, $\alpha \in \mathbb{R}^{N}$, $\beta \in \mathbb{R}^{p}$, and $u \in \mathbb{R}^{NT}$.

LSDV

The LSDV estimator can be written in closed-form:

$$\begin{bmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{bmatrix} = \left(\begin{bmatrix} D & X \end{bmatrix}' \begin{bmatrix} D & X \end{bmatrix} \right)^{-1} \begin{bmatrix} D & X \end{bmatrix}' Y$$

- The slope coefficient

$$\widehat{\beta} = \left(X' M_D X\right)^{-1} X' M_D Y,$$

  where $M_D := I_{NT} - D\left(D'D\right)^{-1} D'$.
- The fixed effects

$$\widehat{\alpha} = (D'D)^{-1} D' \left(Y - X\widehat{\beta} \right).$$

Within-Group Transformation

- Inner-outer optimization: given any β, the OLS estimator

$$\hat{\alpha}_i = T^{-1} \sum_{t=1}^{T} (y_{it} - X'_{it}\beta) = \bar{y}_i - \bar{X}'_i\beta,$$

  where $\bar{y}_i = T^{-1} \sum_{t=1}^{T} y_{it}$ is the **within group average**, and so is $\bar{X}_i$.
- Substitute $\hat{\alpha}_i$ into $y_{it} = \alpha_i + X'_{it}\beta + u_{it}$ and rearrange:

$$\tilde{y}_{it} = \tilde{X}'_{it}\beta + \tilde{u}_{it}$$

  where $\tilde{y}_{it} = y_{it} - \overline{y}_i$ is the **within-group transformation**, and $\tilde{X}_{it}$ and $\tilde{u}_{it}$ are defined similarly.
- This transformation eliminates the N parameters $(\alpha_i)_{i=1}^{N}$.

Alternative Interpretation of Within-Group Transformation

- Recall the model
$$y_{it} = \alpha_i + X_{it}'\beta + u_{it}.$$
- For each i, average over $t = 1, \ldots, T$:
$$\bar{y}_i = \alpha_i + \bar{X}_i'\beta + \bar{u}_i.$$
- Subtraction:
$$\tilde{y}_{it} = \tilde{X}_{it}'\beta + \tilde{u}_{it}.$$

  eliminates the fixed effects.
- No intercept, by construction.

Data in Blocks

- Stack into long vector/matrix of all data

$$\underset{NT\times 1}{\begin{bmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \vdots \\ \tilde{y}_N \end{bmatrix}} = \underset{NT\times p}{\begin{bmatrix} \tilde{X}_1 \\ \tilde{X}_2 \\ \vdots \\ \tilde{X}_N \end{bmatrix}} \underset{p\times 1}{\beta} + \underset{NT\times 1}{\begin{bmatrix} \tilde{u}_1 \\ \tilde{u}_2 \\ \vdots \\ \tilde{u}_N \end{bmatrix}}.$$

- Compact expression of data:

$$\tilde{Y} = \tilde{X}\beta + \tilde{u}.$$

Within Estimator

- Within estimator (or equivalently the FE estimator):

$$\widehat{\beta}_{FE} = \left(\tilde{X}'\tilde{X}\right)^{-1} \tilde{X}'\tilde{Y}$$

- The fixed effects are estimated as

$$\hat{\alpha}_i = \bar{y}_i - \bar{X}_i'\hat{\beta}_{FE}$$

- Consistency (T fixed, $N \to \infty$) is achieved if strict exogeneity holds.

Strict Exogeneity

- A necessary condition for the consistency of $\widehat{\beta}_{FE}$:

$$E\left[\left(X_{it} - \bar{X}_i\right)\left(u_{it} - \bar{u}_i\right)\right] = 0.$$

- A sufficient condition is **strict exogeneity**:

$$E\left[X_{it} u_{is}\right] = 0 \quad \text{for all } s, t \in \{1, \ldots, T\}.$$

- Interpretation: no correlation of the error term u_{is} with the regressor X_{it} in the past, the present, and the future.

Exogeneity

- Notice that

$$\tilde{X}_{it} = X_{it} - \frac{1}{T}\sum_{s=1}^{T} X_{iS} = \left(1 - \frac{1}{T}\right) X_{it} - \frac{1}{T}\sum_{s \neq t} X_{is},$$

  is a linear combination of $\{X_{i1}, \ldots, X_{iT}\}$.
- **Contemporaneous exogeneity**:
  $E\left(X_{it}u_{it}\right) = 0 \quad$ for all $t \in \{1, \ldots, T\}$.
- **Sequential exogeneity**:
  $E\left(u_{it}X_{is}\right) = 0 \quad$ for all $s \leq t \in \{1, \ldots, T\}$. The error term is uncorrelated with the past and present regressor.
- Neither of the above two kinds of exogeneity produces consistent $\widehat{\beta}_{FE}$.

Variance Estimation for FE

- Under homoskedasticity:

$$\widehat{\sigma}_u^2 = \frac{1}{N\left(T-1\right)} \sum_{i=1}^{N} \sum_{t=1}^{T} \widehat{\tilde{u}}_{it}^2.$$

where $\widehat{\tilde{u}}_{it} = \tilde{y}_{it} - \tilde{X}_{it} \widehat{\beta}_{FE}$

- Asymptotic normality

$$\left(\widehat{\sigma}_u^2 \left(\tilde{X}' \tilde{X}\right)^{-1}\right)^{-1/2} \left(\widehat{\beta}_{FE} - \beta^0\right) \Rightarrow N\left(0, I_K\right).$$

First Difference (FD)

- Alternative way to eliminate FE
- Recall

$$y_{it} = \alpha_i + X_{it}'\beta + u_{it},$$
$$y_{i,t-1} = \alpha_i + X_{i,t-1}'\beta + u_{i,t-1}.$$

- Subtraction:

$$\Delta y_{it} = \Delta X_{it}'\beta + \Delta u_{it},$$

  where $\Delta y_{it} = y_{it} - y_{i,t-1}$ is the first-differenced variable, and ΔX_{it} and Δu_{it} are defined similarly.
- Collect the FD variables into ΔY and ΔX:

$$\Delta Y \quad \text{is of size } N(T-1) \times 1,$$

$$\Delta X \quad \text{is of size } N(T-1) \times p.$$

FD Estimator

- The FD estimator is:

$$\widehat{\beta}_{FD} = \left(\Delta X'\Delta X\right)^{-1} \Delta X'\Delta Y.$$

- A necessary condition for consistency is

$$E\left[\Delta X_{it}\Delta u_{it}\right] = 0,$$

which is slightly weaker than strict exogeneity.

Section 2

Random Effect Models

Random Effects

- Recall the model

$$\begin{aligned} y_{it} &= c + X'_{it}\beta + \varepsilon_{it} \\ &= c + X'_{it}\beta + \alpha_i^* + u_{it}, \end{aligned}$$

where the composite error $\varepsilon_{it} = \alpha_i^* + u_{it}$, with $u_{it} \sim$ iid $(0, \sigma_u^2)$, $\alpha_i^* \sim$ iid $(0, \sigma_\alpha^2)$, and they are uncorrelated with X_{it}.

Efficient Estimation of RE Model

- $E\left[X_{it}'\varepsilon_{it}\right] = 0$.
- OLS is consistent, but inefficient due to violation of homoskedasticity:

$$\underset{T\times T}{S} := E\left[\varepsilon_i\varepsilon_i'\right] = \begin{bmatrix} \sigma_\alpha^2 + \sigma_u^2 & \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma_u^2 & \cdots & \sigma_\alpha^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 + \sigma_u^2 \end{bmatrix}.$$

- Generalized Least Squares (GLS) is the efficient estimator.

Generalized Least Squares

- Rewrite

$$y_{it} = c + X_{it}'\beta + \varepsilon_{it} = W_{it}'\theta + \varepsilon_{it}$$

where $W_{it} := (1, X_{it}')'$ and $\theta = (c, \beta')'$

- The (infeasible) GLS estimator is:

$$\begin{aligned}
\widehat{\theta}_{RE}^{infeasible} &= \left(\sum_{i=1}^{N} W_i' S^{-1} W_i \right)^{-1} \sum_{i=1}^{N} W_i' S^{-1} y_i \\
&= \left(W' \mathbf{S}^{-1} W \right)^{-1} W' \mathbf{S}^{-1} Y
\end{aligned}$$

where $\mathbf{S} = I_T \otimes S$

Feasible GLS

- Step 1: use OLS and obtain $\widehat{\varepsilon}_{it} = y_{it} - W_{it}\widehat{\theta}_{OLS}$
  - Estimate the diagonal term and the off-diagonal term in S as

$$\begin{aligned}\widehat{S}_{diag} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{\varepsilon}_{it}^{2} \\ \widehat{S}_{off} &= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T(T-1)}\sum_{t=1}^{T}\sum_{s\neq t}\widehat{\varepsilon}_{it}\widehat{\varepsilon}_{is}\end{aligned}$$

  respectively, to obtain $\hat{S}$.

- Step 2: The feasible GLS (FGLS) is

$$\widehat{\theta}_{RE} = \left(\sum_{i=1}^{N} W_i'\hat{S}^{-1}W_i\right)^{-1}\sum_{i=1}^{N} W_i'\hat{S}^{-1}y_i$$

Fixed Effects vs. Random Effects

- FE is more general:
  - But does not allow time-invariant X_i.
- RE does not cope with endogenous α_i^*.
- Hausman test is traditionally used to distinguish the two models.
- Data demo: `https://www.kaggle.com/code/jipann/panel-data-estimation-in-python`

Section 3

Dynamic Panel Data Models

Dynamic Panel Data

- The simplest dynamic panel model is

$$y_{it} = \alpha_i + \beta y_{i,t-1} + u_{it},$$

  where $|\beta| < 1$, $u_{it} \sim$ iid $(0, \sigma^2)$ over i and t, and $\mathrm{Cov}(u_{it}, y_{i,t-1}) = 0$.
- Replace X_{it} in the static model by the lagged dependent variable $y_{i,t-1}$ to model the dynamic feedback.
- Other regressors can be added on the right-hand side.

FE Estimator

- How about estimate β by the FE estimator?
  Let $\tilde{Y}_{-1}$ be the demeaned variable of the lagged dependent variable, and then

$$\widehat{\beta}_{FE} = \frac{\tilde{Y}'_{-1}\tilde{Y}}{\tilde{Y}'_{-1}\tilde{Y}_{-1}}$$

- It is easy to see

$$\widehat{\beta}_{FE} - \beta_0 = \frac{\tilde{Y}'_{-1}\tilde{U}}{\tilde{Y}'_{-1}\tilde{Y}_{-1}},$$

  where strict exogeneity is violated.

Nickell Bias

- Serious consequence: the FE estimator is inconsistent under "small T, large N" (Nickell, 1981).
- A numerical demonstration of the Nickell bias.
  `https://www.kaggle.com/code/jipann/nickell-bias`

Another Angle: First Difference

- Recall

$$\Delta y_{it} = \beta \Delta y_{i,t-1} + \Delta u_{it}$$

- It follows

$$\begin{aligned} \widehat{\beta}_{FD} - \beta_0 &= \frac{\sum_{i,t} \Delta y_{i,t-1} \Delta u_{it}}{\sum_{i,t} \Delta y_{i,t-1}^2} \\ &= \frac{\sum_{i,t} \left(y_{i,t-1} - y_{i,t-2}\right)\left(u_{it} - u_{i,t-1}\right)}{\sum_{i,t} \left(y_{i,t-1} - y_{i,t-2}\right)^2} \end{aligned}$$

Inherent Endogeneity in FD

- The expected value of the numerator is

$$
\begin{aligned}
& E\left[\left(y_{i,t-1}-y_{i,t-2}\right)\left(u_{it}-u_{i,t-1}\right)\right] \\
= & E\left[y_{i,t-1}u_{it}\right]-E\left[y_{i,t-1}u_{i,t-1}\right]-E\left[y_{i,t-2}u_{it}\right]+E\left[y_{i,t-2}u_{i,t-1}\right] \\
= & 0-\sigma_u^2-0+0=-\sigma_u^2 \neq 0
\end{aligned}
$$

- $\widehat{\beta}_{FD}$ is inconsistent under finite T.

Remedy: A further Lag as Instrument

- Anderson and Hsiao (1981): $\Delta y_{i,t-2}$ is a valid IV for the regressor $\Delta y_{i,t-1}$
  - $\Delta y_{i,t-2} = y_{i,t-2} - y_{i,t-3}$ is uncorrelated with Δu_{it}.
  - $\Delta y_{i,t-2}$ is correlated with $\Delta y_{i,t-1}$.
- 2SLS:
$$\widehat{\beta}_{IV} = \frac{\sum_{i,t} \Delta y_{i,t-2} \Delta y_{i,t}}{\sum_{it} \Delta y_{i,t-2} \Delta y_{i,t-1}}.$$
- Consistent and asymptotic normal.

More Lags as IV

- Arellano and Bond (1991):
  $(y_{i,t-2}, y_{i,t-3}, y_{i,t-4}, \ldots, y_{i,0})$ are all valid IV.
- Use GMM for estimation.
- The more IVs, the more efficient (in theory).
- Optimal weighting matrix is needed for efficiency.
- The practical issue of "too many instruments."

Summary

- Panel data
- FE estimator
- RE estimator
- Static panel model
- Dynamic panel model

- Rich information
- Big data