EC708 Discussion 6
Linear Panel Data

Yan Liu

Department of Economics
Boston University

March 5, 2021

Outline

Linear Panel Model Setup

$$y_{it} = x_{it}'\beta + \epsilon_{it}, \quad i = 1, \ldots, N,\ t = 1, \ldots, T$$

Level of time:

$$\underset{N\times 1}{y_t} = \underset{N\times k}{X_t}\,\beta + \underset{N\times 1}{\epsilon_t}, \quad t = 1, \ldots, T$$

Level of individual ("medium" notation):

$$\underset{T\times 1}{y_i} = \underset{T\times k}{X_i}\,\beta + \underset{T\times 1}{\epsilon_i}, \quad i = 1, \ldots, N$$

"Large" notation:

$$\underset{NT\times 1}{Y} = \underset{NT\times k}{X}\beta + \underset{NT\times 1}{\epsilon}$$

Different Exogeneity Assumptions

We focus on large-N asymptotics with T fixed. So we view cross section observations as i.i.d. but allow for arbitrary time dependence across t.

1. Contemporaneous exogeneity

$$\mathbb{E}[\epsilon_t | X_t] = 0, \quad t = 1, \dots, T$$

2. Sequential exogeneity

$$\mathbb{E}[\epsilon_t | X_t, X_{t-1}, \dots, X_1] = 0, \quad t = 1, \dots, T$$

3. Strict exogeneity

$$\mathbb{E}[\epsilon_t | X_1, \dots, X_T] = 0, \quad t = 1, \dots, T$$

Pooled OLS

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

Assume that

1. $\mathbb{E}[\epsilon_{it} x_{it}] = 0$;
2. $\text{rank}\left(\sum_{t=1}^{T} \mathbb{E}[x_{it} x_{it}']\right) = k.$

Ignore the panel structure and treat data as a giant cross section:

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

Error Component Model

Error component model:

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

- α_i: individual-specific effect;
- u_{it}: idiosyncratic (i.i.d.) errors.

In vector we can write $\epsilon_i = \mathbf{1}_T \alpha_i + u_t$ where $\mathbf{1}_T$ is a $T \times 1$ vector of 1's.

Depending on whether we assume α_i is correlated with x_{it}:

- Random effects assumes zero correlation between α_i and x_{it};
- fixed effects allows for arbitrary dependence between α_i and x_{it}.

Random Effects Estimator

Assumption RE.1:

- Strict exogeneity: $\mathbb{E}[u_{it}|X_i, \alpha_i] = 0$;
- orthogonality: $\mathbb{E}[\alpha_i|X_i] = \mathbb{E}[\alpha_i] = 0$ (KEY assumption);

Assumption RE.2: Equicorrelated random effects structure

$$\Omega \equiv \mathbb{E}[\epsilon_i \epsilon_i'|X_i] = \begin{pmatrix} \sigma_\alpha^2 + \sigma_u^2 & \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma_u^2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \cdots & \cdots & \sigma_\alpha^2 + \sigma_u^2 \end{pmatrix} = \sigma_\alpha^2 J_T + \sigma_u^2 I_T$$

where I_T is a $T \times T$ identity matrix and $J_T = \mathbf{1}_T \mathbf{1}_T'$.

Assumption RE.3: rank $\left(\mathbb{E}[X_i'\Omega^{-1}X_i]\right) = k$.

$$\hat{\beta}_{RE} = \left(\sum_{t=1}^{N} X_i'\Omega^{-1}X_i\right)^{-1} \sum_{t=1}^{N} X_i'\Omega^{-1}y_i.$$

Random Effects Estimator

"Big" Notation

- Define $P = I_N \otimes J_T/T$ and $Q = I_{NT} - P$. Then

$$\begin{aligned} V \equiv I_N \otimes \Omega &= \sigma_u^2(I_N \otimes I_T) + \sigma_\alpha^2(I_N \otimes J_T) \\ &= \sigma_u^2(P+Q) + T\sigma_\alpha^2 P \\ &= \underbrace{(\sigma_u^2 + T\sigma_\alpha^2)}_{=\sigma_1^2} P + \sigma_u^2 Q. \end{aligned}$$

- It turns out that $V^{-1} = \sigma_1^{-2}P + \sigma_u^{-2}Q$.
  - follows from $(\sigma_1^{-2}P + \sigma_u^{-2}Q)(\sigma_1^2 P + \sigma_u^2 Q) = P + 0 + 0 + Q = I_{NT}$
- We can write the RE estimator as

$$\hat{\beta}_{RE} = (X'(\sigma_1^{-2}P + \sigma_u^{-2}Q)X)^{-1}X'(\sigma_1^{-2}P + \sigma_u^{-2}Q)y.$$

Random Effects Estimator

Inference

Asymptotic normality follows from

$$\sqrt{n}(\hat{\beta}_{RE} - \beta) = \left(\frac{1}{N} \sum_{i=1}^{N} X_i' \Omega^{-1} X_i \right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i' \Omega^{-1} \epsilon_i .$$

- By LLN, $\frac{1}{N} \sum_{i=1}^{N} X_i' \Omega^{-1} X_i \xrightarrow{p} \mathbb{E}[X_i' \Omega^{-1} X_i]$.
- By CLT, $\frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i' \Omega^{-1} \epsilon_i \xrightarrow{d} N(0, \mathbb{E}[X_i' \Omega^{-1} \epsilon_i \epsilon_i' \Omega^{1} X_i])$.

- Under Assumption RE.2, asymptotic variance $\Sigma = \mathbb{E}[X_i' \Omega^{-1} X_i]^{-1}$.
- When $\mathbb{E}[\epsilon_i \epsilon_i' | X_i] \neq \mathbb{E}[\epsilon_i \epsilon_i']$, or $\mathbb{E}[\epsilon_i \epsilon_i']$ does not have RE structure, we need a cluster-robust variance matrix estimator

$$\hat{\Sigma}_n = \left(\frac{1}{N} \sum_{i=1}^{N} X_i' \hat{\Omega}^{-1} X_i \right)^{-1} \frac{1}{N} \sum_{i=1}^{N} X_i' \hat{\Omega}^{-1} \hat{\epsilon}_i \hat{\epsilon}_i' X_i \hat{\Omega}^{-1} X_i \left(\frac{1}{N} \sum_{i=1}^{N} X_i' \hat{\Omega}^{-1} X_i \right)^{-1} .$$

Fixed Effects Estimator

Assumption FE.1:

- Strict exogeneity: $\mathbb{E}[u_{it}|X_i, \alpha_i] = 0$;

Fixed effects transformation:

$$y_{it} - \overline{y}_i = (x_{it} - \overline{x}_i)'\beta + u_{it} - \overline{u}_i$$

Demeaning operator: $Q_T = I_T - J_T/T$. Then $Q = I_N \otimes Q_T$ and

$$Q_T y_i = Q_T X_i \beta + Q_T u_i.$$

Assumption FE.2: $\text{rank}\left(\mathbb{E}[X_i' Q_T X_i]\right) = k$.

$$\hat{\beta}_{FE} = \left(\sum_{i=1}^{N} X_i' Q_T X_i\right)^{-1} \sum_{i=1}^{N} X_i' Q_T y_i.$$

Between and Within Estimators

$$\hat{\beta}_{\text{between}} = (X'PX)^{-1}X'Py, \quad \hat{\beta}_{\text{within}} = (X'QX)^{-1}X'Qy.$$

- Fixed effects estimator is also called the within estimator because it uses time variation within each cross section.
- Between estimator is OLS applied to the time-averaged equation

$$\overline{y}_i = \alpha_i + \overline{x}_i'\beta + \overline{\epsilon}_i$$

which uses only variation between the cross section observations.

- POLS and RE are both linear combinations of between and within estimators:

$$\hat{\beta}_{POLS} = a\hat{\beta}_{\text{between}} + (1-a)\hat{\beta}_{\text{within}},$$
$$\hat{\beta}_{RE} = b\hat{\beta}_{\text{between}} + (1-b)\hat{\beta}_{\text{within}}$$

where $a = (X'X)^{-1}X'PX$ and $b = \left(X'\left(P + \frac{\sigma_1^2}{\sigma_u^2}Q\right)X\right)^{-1}X'PX$.

Relationship between RE and FE Estimators

- As $T \to \infty$ or $\frac{\sigma_u}{\sigma_\alpha} \to 0$, $b \to 0$, and RE coincides with FE.
  - Recall that $\sigma_1^2 = \sigma_u^2 + T\sigma_\alpha^2$.
- It can be shown that

$$\mathrm{Var}(\hat{\beta}_{RE}|X) = (\sigma_1^{-2} X'PX + \sigma_u^{-2} X'QX)^{-1},$$
$$\mathrm{Var}(\hat{\beta}_{FE}|X) = \sigma_u^2 (X'QX)^{-1}.$$

Hence, RE is more efficient than FE.

Hausman Test: FE vs RE

$$H_0 : \mathbb{E}[\alpha_i | X_i] = 0, \quad H_1 : \mathbb{E}[\alpha_i | X_i] \neq 0.$$

- Under H_0: both FE and RE are consistent while RE is more efficient;
- under H_1: only FE is consistent.

Hausman statistic:

$$H = (\hat{\beta}_{FE} - \hat{\beta}_{RE})' \left[\mathrm{Var}(\hat{\beta}_{FE}) - \mathrm{Var}(\hat{\beta}_{RE}) \right]^{-1} (\hat{\beta}_{FE} - \hat{\beta}_{RE})$$

Under H_0, $H \xrightarrow{d} \chi^2_k$. **Caveats:**

- Failure of equicorrelated RE structure leads to a non-standard limiting distribution
- Cannot compare FE and RE coefficients on time-constant variables
- Post model selection size distortion (Discussion 4)

First-Difference Estimator

Assumption FD.1:

- Strict exogeneity: $\mathbb{E}[u_{it}|X_i, \alpha_i] = 0$;

First-differencing transformation:

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

Assumption FD.2: $\text{rank}\left(\sum_{t=2}^{T} \mathbb{E}[\Delta x_{it} \Delta x_{it}']\right) = k$

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

Relationship between FE and FD Estimators

Assumption FE.3: $\mathbb{E}[u_i u_i' | X_i, \alpha_i] = \sigma_u^2 I_T$.

Under Assumption FE.3, FE is more efficient than FD. To see this, define

$$\underset{(T-1)\times T}{D} = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix}$$

Then FD is OLS applied to

$$Dy_i = DX_i\beta + Du_i.$$

Under Assumption FE.3, $\mathbb{E}[(Du_i)(Du_i)' | X_i, \alpha_i] = \sigma_u^2 DD'$ is not spherical. So OLS is not efficient.

Relationship between FE and FD Estimators

A natural thought is to use GLS:

$$\hat{\beta}_{FD,GLS} = \left(\sum_{i=1}^{N} X_i' D' (DD')^{-1} D X_i \right)^{-1} \sum_{i=1}^{N} X_i' D' (DD')^{-1} D y_i.$$

It turns out that $D'(DD')^{-1}D = Q_T$[1] , so $\hat{\beta}_{FD,GLS} = \hat{\beta}_{FE}$.

Forward orthogonal transformation (Arellano and Bover, 1995):

$$(DD')^{-1/2} D v_{it} = \sqrt{\frac{T-t}{T-t+1}} \left[v_{it} - \frac{1}{T-t} (v_{i,t+1} + \cdots + v_{iT}) \right].$$

Alternatively, if $\mathbb{E}[(Du_i)(Du_i)' | X_i, \alpha_i] = \sigma_e^2 I_{T-1}$, FD is efficient.

- Now u_{it} is a random walk, which has substantial serial dependence.

[1]Note that $\underset{T \times T}{\mathcal{H}} = \begin{pmatrix} T^{-1/2} \mathbf{1}_T' \\ (DD')^{-1/2} D \end{pmatrix}$ is such that $\mathcal{H}\mathcal{H}' = J_T/T + D'(DD')^{-1}D = I_T$.

FE in Dynamic Linear Panel AR(1)

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

Within transformation:

$$y_{it} - \overline{y}_i = \rho(y_{i,t-1} - \overline{y}_{i,-1}) + u_{it} - \overline{u}_i$$

where $\overline{y}_{i,-1} = \frac{1}{T}\sum_{t=0}^{T-1} y_{it}$.

- $\overline{y}_{i,-1}$ is correlated with $\overline{u}_i$.
- Results in inconsistency of $\hat{\rho}_{FE}$:

$$\hat{\rho}_{FE} = \rho + \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{i,t-1} - \overline{y}_{i,-1})(u_{it} - \overline{u}_i)}{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{i,t-1} - \overline{y}_{i,-1})^2}$$

- We consider short panel (small T) asymptotics ($N \to \infty$).
- For long panel asymptotics ($N \to \infty, T \to \infty$), refer to Fernández-Val and Weidner (2018).

FE in Dynamic Linear Panel: Nickell (1981) Bias

- Fix T and let $N \to \infty$,

$$\hat{\rho}_{FE} - \rho \xrightarrow{p} -\frac{1+\rho}{T-1}\left\{1 - \frac{1}{T}\frac{1-\rho^T}{1-\rho}\right\} \times \left\{1 - \frac{2\rho}{(1-\rho)(T-1)}\left[1 - \frac{1}{T}\frac{1-\rho^T}{1-\rho}\right]\right\}^{-1}$$

- When $T = 2$,

$$\hat{\rho}_{FE} - \rho \xrightarrow{p} -\frac{1+\rho}{2}.$$

- When T is large (long panel),

$$\underset{N\to\infty}{\text{plim}}\left(\hat{\rho}_{FE} - \rho\right) \approx -\frac{1+\rho}{T-1}.$$

RE and FD in Dynamic Linear Panel

RE estimator:

- $y_{i,t-1}$ also depends on α_i, violating Assumption RE.1.

FD estimator is OLS on

$$y_{it} - y_{i,t-1} = \rho(y_{i,t-1} - y_{i,t-2}) + u_{it} - u_{i,t-1}$$

- $y_{i,t-1} - y_{i,t-2}$ is correlated with $u_{it} - u_{i,t-1}$.

Takeaway: When lagged dependent variable is included as a regressor, FE, RE, and FD fail to account for the endogeneity it brings.

Anderson and Hsiao (1982): First-Differenced IV

- Consider the first-differenced regression:

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

- Assume sequential exogeneity: $\mathbb{E}[u_{it} | y_{i,t-1}, \ldots, y_{i,0}, \alpha_i] = 0$.
- FD is problematic because $\Delta y_{i,t-1}$ is correlated with Δu_{it}.
- Remedy: use $y_{i,t-2}$ or $\Delta y_{i,t-2}$ as an instrument for $\Delta y_{i,t-1}$
    1. IV relevance: $y_{i,t-2} = y_{i,t-1} - \Delta y_{i,t-1}$;
    2. IV validity: $\mathbb{E}[y_{i,t-2} \Delta u_{it}] = \mathbb{E}[\Delta y_{i,t-2} \Delta u_{it}] = 0$.
- Estimator is consistent but inefficient: doesn't exploit all moment conditions.
- IV validity relies on the assumption that u_{it} is serially uncorrelated.

Arellano and Bond (1991)

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

- What are the valid instruments for each period?
  - $t = 2$: no instruments;
  - $t = 3$: $\Delta y_{i2} = y_{i2} - y_{i1}$. IV is y_{i1}.
  - $t = 4$: $\Delta y_{i3} = y_{i3} - y_{i2} = \rho(y_{i2} - y_{i1}) + \Delta u_{i3}$. IVs are y_{i2} and y_{i1}.
  - $t = T$: IVs are $y_{i,T-2}, \ldots, y_{i1}$.
- There are in total $\frac{(T-1)(T-2)}{2}$ IVs and hence moment conditions:

$$\mathbb{E}[(\Delta y_{it} - \rho \Delta y_{i,t-1}) y_{is}] = 0, \quad t = 3, \ldots, T,\ s = 1, \ldots, t-2.$$

- Estimate by two-step GMM.

Arellano and Bond (1991)

Remarks:

- When T is large, using full set of lags as instruments may cause many instruments problem.
- Consistency relies on the assumption that u_{it} is serially uncorrelated.
- Blundell and Bond (1998) point out that the Anderson-Hsiao and Arellano-Bond class of estimators suffer from weak instruments. For example, when $T = 3$, let the first-stage be

$$\Delta y_{i2} = \pi y_{i1} + r_i.$$

Some algebra shows

$$\hat{\pi} \xrightarrow{p} (\rho - 1)\frac{k}{k + \sigma_\alpha^2/\sigma_u^2}, \quad k = \frac{1-\rho}{1+\rho}.$$

$\text{plim}_{N\to\infty}\hat{\pi} \to 0$ if $\rho \to 1$ (persistent dynamics) or $\sigma_\alpha^2/\sigma_u^2 \to \infty$.

Blundell and Bond (1998)

$$y_{it} = \rho y_{i,t-1} + \underbrace{\alpha_i + u_{it}}_{\epsilon_{it}}$$

To reduce weak instruments problem,

- Arellano and Bover (1995) add moments

$$\mathbb{E}[\epsilon_{it}\Delta y_{i,t-1}] = 0, \quad t = 3, \ldots, T$$

- Blundell and Bond (1998): Δy_{i1} is observed, additional moment

$$\mathbb{E}[\epsilon_{i2}\Delta y_{i1}] = 0.$$

Needs restrictions on initial conditions generating y_{i0}.

Blundell and Bond (1998)

- Specify $y_{i0} = \frac{\alpha_i}{1-\rho} + \epsilon_{i0}$.
  - $\alpha_i/(1-\rho)$ is unconditional "mean" of y_{it} under stationarity.
- Then $\mathbb{E}[\epsilon_{i2}\Delta y_{i1}] = 0$ is equivalent to

$$\mathbb{E}[(\alpha_i + u_{i2})(u_{i1} + (\rho - 1)\epsilon_{i0}] = 0.$$

- Necessary conditions: $\mathbb{E}[\epsilon_{i0}\alpha_i] = \mathbb{E}[\epsilon_{i0}u_{i2}] = 0$.

In sum, Blundell and Bond (1998) use the following moment conditions:

$$\mathbb{E}[(\Delta y_{it} - \rho\Delta y_{i,t-1})y_{is}] = 0, \quad t = 3, \ldots, T,\ s = 1, \ldots, t-2,$$
$$\mathbb{E}[\epsilon_{it}\Delta y_{i,t-1}] = 0, \quad t = 2, \ldots, T.$$

Bibliography

Anderson, T. W. and Hsiao, C. (1982), "Formulation and estimation of dynamic models using panel data," *Journal of econometrics*, 18, 47–82.

Arellano, M. and Bond, S. (1991), "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations," *The review of economic studies*, 58, 277–297.

Arellano, M. and Bover, O. (1995), "Another look at the instrumental variable estimation of error-components models," *Journal of econometrics*, 68, 29–51.

Blundell, R. and Bond, S. (1998), "Initial conditions and moment restrictions in dynamic panel data models," *Journal of econometrics*, 87, 115–143.

Fernández-Val, I. and Weidner, M. (2018), "Fixed effects estimation of large-t panel data models," *Annual Review of Economics*, 10, 109–138.

Nickell, S. (1981), "Biases in dynamic models with fixed effects," *Econometrica: Journal of the econometric society*, 1417–1426.