EC708 Discussion 6
Linear Panel Data

Yan Liu

Department of Economics
Boston University

February 25, 2022

Outline

Relationship between RE and FE Estimators

Notation

Consider the panel data model **("small" notation)**:

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

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:

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

Unobserved heterogeneity:

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

Assumption RE.1:

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

Assumption RE.2: Equicorrelated random effects structure

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

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

Relationship between RE and FE Estimators

Error Component Structure

- Demeaning operator: $Q_T = I_T - J_T/T$ and $Q = I_N \otimes Q_T$.
- Define $P = I_{NT} - Q$ and $V = I_N \otimes \Omega_T$. Then

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

- P and Q are symmetric and idempotent. Hence,

$$
\begin{aligned}
& PQ = P(I_{NT} - P) = 0 \\
\Rightarrow\ & (\sigma_1^{-2}P + \sigma_u^{-2}Q)(\sigma_1^2 P + \sigma_u^2 Q) = P + 0 + 0 + Q = I_{NT} \\
\Rightarrow\ & V^{-1} = \sigma_1^{-2}P + \sigma_u^{-2}Q.
\end{aligned}
$$

We can write the RE and FE estimators as

$$
\begin{aligned}
\hat{\beta}_{RE} &= \left(\sum_{i=1}^{N} X_i' \Omega_T^{-1} X_i \right)^{-1} \sum_{i=1}^{N} X_i' \Omega_T^{-1} y_i \\
&= (X'(\sigma_1^{-2} P + \sigma_u^{-2} Q) X)^{-1} X'(\sigma_1^{-2} P + \sigma_u^{-2} Q) y, \\
\hat{\beta}_{FE} &= \left(\sum_{i=1}^{N} X_i' Q_T X_i \right)^{-1} \sum_{i=1}^{N} X_i' Q_T y_i \\
&= (X'QX)^{-1} X'Qy.
\end{aligned}
$$

Relationship between RE and FE Estimators

Between and Within Estimators

- $\hat{\beta}_{FE}$ is also called the within estimator because it uses time variation within each cross-section.
- Similarly, we can define the between estimator which uses variation between the cross-section observations:

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

- $\hat{\beta}_{\text{between}}$ is OLS applied to the time-averaged equation

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

Relationship between RE and FE Estimators

Between and Within Estimators

$\hat{\beta}_{RE}$ and $\hat{\beta}_{POLS}$ are both linear combinations of $\hat{\beta}_{\text{between}}$ and $\hat{\beta}_{\text{within}}$:

$$\begin{aligned}
\hat{\beta}_{RE} &= \underbrace{\frac{1}{\sigma_1^2}(X'(\sigma_1^{-2}P + \sigma_u^{-2}Q)X)^{-1}}_{=A(X'PX)^{-1}} X'Py + \underbrace{\frac{1}{\sigma_u^2}(X'(\sigma_1^{-2}P + \sigma_u^{-2}Q)X)^{-1}}_{=B(X'QX)^{-1}} X'Qy \\
&= A\hat{\beta}_{\text{between}} + B\hat{\beta}_{\text{within}}, \\
\hat{\beta}_{POLS} &= \underbrace{(X'X)^{-1}}_{C(X'PX)^{-1}} X'Py + \underbrace{(X'X)^{-1}}_{D(X'QX)^{-1}} X'Qy \\
&= C\hat{\beta}_{\text{between}} + D\hat{\beta}_{\text{within}}.
\end{aligned}$$

We can calculate $A = I_k - B$ and $C = I_k - D$, where

$$B = \left(X'\left(\frac{\sigma_u^2}{\sigma_1^2}P + Q\right)X\right)^{-1} X'QX, \quad D = (X'X)^{-1}X'QX.$$

- What happens when $T \to \infty$ or $\frac{\sigma_u}{\sigma_\alpha} \to 0$?
  $\frac{\sigma_u^2}{\sigma_1^2} = \frac{(\sigma_u/\sigma_\alpha)^2}{(\sigma_u/\sigma_\alpha)^2+T} \to 0 \Rightarrow B \to I_k \Rightarrow$ RE approaches FE.
- We can calculate $\text{Cov}(\hat{\beta}_{\text{between}}, \hat{\beta}_{\text{within}}|X) = 0$,

$$\text{Var}(\hat{\beta}_{\text{between}}|X) = \sigma_1^2(X'PX)^{-1}, \quad \text{Var}(\hat{\beta}_{\text{within}}|X) = \sigma_u^2(X'QX)^{-1},$$

and thus

$$\begin{aligned} &\text{Var}(\widehat{\beta}_{RE}|X) = \text{Cov}(\widehat{\beta}_{RE}, \widehat{\beta}_{\text{within}}|X) = (\sigma_1^{-2}X'PX + \sigma_u^{-2}X'QX)^{-1} \\ \Rightarrow\ &\text{Var}(\widehat{\beta}_{RE} - \widehat{\beta}_{\text{within}}|X) = \text{Var}(\widehat{\beta}_{\text{within}}|X) - \text{Var}(\widehat{\beta}_{RE}|X). \end{aligned}$$

Hence, RE is more efficient than FE.

Relationship between RE and FE Estimators

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:

$$H_N = (\hat{\beta}_{FE} - \hat{\beta}_{RE})'[\mathrm{Var}(\hat{\beta}_{FE}|X) - \mathrm{Var}(\hat{\beta}_{RE}|X)]^{-1}(\hat{\beta}_{FE} - \hat{\beta}_{RE}).$$

Under H_0, $H_N \xrightarrow{d} \chi^2_k$.

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

The two-stage test statistic is

$$t_N(\beta_0) = t_{RE}(\beta_0)1(H_N < \chi^2_{k,1-\alpha}) + t_{FE}(\beta_0)1(H_N > \chi^2_{k,1-\alpha}).$$

Guggenberger (2010) shows that the asymptotic distribution of $t_N(\beta_0)$ is discontinuous in $\gamma_1 = \mathrm{Corr}(\alpha_i, \bar{x}_i)$:

- When $\sqrt{N}\gamma_1 \to \infty$, t_{FE} is almost always used.
- When $\sqrt{N}\gamma_1 \to h < \infty$, Hausman test does not have enough power. t_{RE} is frequently used, leading to invalid second-stage inference.
- Unfortunately, it is impossible to uniformly consistently estimate h.
    - Partial solution: use least favorable critical values (e.g. Andrews and Guggenberger, 2009)

Outline

Relationship between FE and FD Estimators

First-Difference Estimator

Differencing operator:

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

FD estimator is OLS applied to $Dy_i = DX_i\beta + Du_i$:

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

Relationship between FE and FD Estimators

Assumption FE.1 (strict exogeneity): $\mathbb{E}[u_{it}|X_i, \alpha_i] = 0$.

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

- Under Assumption FE.2, $\mathbb{E}[(Du_i)(Du_i)'|X_i, \alpha_i] = \sigma_u^2 DD'$ is not spherical, so OLS is not efficient.
- A natural thought is to use GLS:

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

Relationship between FE and FD Estimators

- It turns out that $D'(DD')^{-1}D = Q_T$, so $\hat{\beta}_{FD,GLS} = \hat{\beta}_{FE}$.
  - Let $\underset{T\times T}{\mathcal{H}} = \begin{pmatrix} T^{-1/2}\mathbf{1}_T' \\ (DD')^{-1/2}D \end{pmatrix}$. Then $\mathcal{H}\mathcal{H}' = I_T$, so that also $\mathcal{H}'\mathcal{H} = J_T/T + D'(DD')^{-1}D = I_T$.
  - Forward orthogonal transformation (Arellano and Bover, 1995):

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

- Under Assumption FE.2, FE is more efficient than FD.
- 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.

Outline

Consider the lagged dependent variable model:

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

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 $\hat{\rho}_{FE}$:

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

- 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),

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

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 equation:

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

- 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.

Dynamic Linear Panel

Arellano and Bond (1991)

Consider the first-differenced equation:

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

- 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.

Remarks:

- When T is large, using full set of lags as instruments may cause many instruments problem.
- 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 regression 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 $\sigma_\alpha^2/\sigma_u^2 \to \infty$.

Dynamic Linear Panel

Blundell and Bond (1998)

Recall

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

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}.

Dynamic Linear Panel

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.

Andrews, D. W. and Guggenberger, P. (2009), “Hybrid and size-corrected subsampling methods,” *Econometrica*, 77, 721–762.

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.

Guggenberger, P. (2010), “The impact of a Hausman pretest on the size of a hypothesis test: The panel data case,” *Journal of Econometrics*, 156, 337–343.

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