EC708 Discussion 5
Linear Panel

Yan Liu

Department of Economics
Boston University

February 17, 2023

Outline

1. Random-Effects Estimator
2. Selected Questions from PS1

Table of Contents

Random-Effects Estimator

Notation

"small":

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

"medium":

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

"Large":

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

Unobserved heterogeneity has the additive error component specification:

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

Assumption 4.1:

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

Under Assumption 4.1, $E[\epsilon_i|X_i] = 0$. Hence, we can consistently estimate β by the (pooled) OLS.

Assumption 4.2:

- $\{\epsilon_i, i = 1, \ldots, N\}$ is i.i.d.;
- Equicorrelated random effects structure

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

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

Random-Effects Estimator

GLS

For the moment, suppose Ω is known. Then,

$$V \equiv E[\boldsymbol{\epsilon}\boldsymbol{\epsilon}'|\mathbf{X}] = I_N \otimes \Omega.$$

The (infeasible) random-effects (RE) estimator is the GLS estimator

$$\begin{aligned}\hat{\beta}_{RE} &= [\mathbf{X}'V^{-1}\mathbf{X}]^{-1}\mathbf{X}'V^{-1}\mathbf{Y} \\ &= \left(\sum_{i=1}^{N} X_i'\Omega^{-1}X_i\right)^{-1}\sum_{i=1}^{N} X_i'\Omega^{-1}Y_i.\end{aligned}$$

This can be implemented by running OLS on transformed regression model

$$\sigma_u V^{-1/2}\mathbf{Y} = \sigma_u V^{-1/2}\mathbf{X}\beta + \sigma_u V^{-1/2}\boldsymbol{\epsilon}. \tag{1}$$

Random-Effects Estimator

Implementation: Quasi De-meaning

Define

$$P = I_N \otimes J_T/T, \quad Q = I_{NT} - P.$$

What happens if we apply P and Q to $\mathbf{X}$?

- Applying P to $\mathbf{X}$ averages observations across time for each i;
- Applying Q to $\mathbf{X}$ demeans observations by subtracting "within" mean

$$P\mathbf{X} = \begin{bmatrix} \bar{X}_1 \\ \vdots \\ \bar{X}_N \end{bmatrix}, \quad Q\mathbf{X} = \begin{bmatrix} X_1 - \bar{X}_1 \\ \vdots \\ X_N - \bar{X}_N \end{bmatrix}.$$

Random-Effects Estimator

Implementation: Quasi De-meaning

How is V related to P and Q?

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

Random-Effects Estimator

Implementation: Quasi De-meaning

- P and Q are symmetric and idempotent. Hence,

$$PQ = P(I_{NT} - P) = 0.$$

- What is V^{-1}?

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

- What is $\sigma_u V^{-1/2}$?

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

Random-Effects Estimator

Implementation: Quasi De-meaning

Premultiplying $\sigma_u V^{-1/2}$ to $\mathbf{X}$ yields

$$\sigma_u V^{-1/2}\mathbf{X} = [(\sigma_u/\sigma_1)P + Q]\mathbf{X}$$
$$= \begin{bmatrix} X_1 - (1-\sigma_u/\sigma_1)\bar{X}_1 \\ \vdots \\ X_N - (1-\sigma_u/\sigma_1)\bar{X}_N \end{bmatrix}.$$

Let $\theta = 1 - \sigma_u/\sigma_1$. Each row of (1) is

$$(y_{it} - \theta\bar{y}_i) = (x_{it} - \theta\bar{x}_i)'\beta + (\epsilon_{it} - \theta\bar{\epsilon}_t),\ i = 1,\ldots,N, t = 1,\ldots,T.$$

Random-Effects Estimator

Between and Within Estimators

Define

$$\hat{\beta}_{\text{between}} = (\mathbf{X}'P\mathbf{X})^{-1}\mathbf{X}'P\mathbf{Y},$$
$$\hat{\beta}_{\text{within}} = (\mathbf{X}'Q\mathbf{X})^{-1}\mathbf{X}'Q\mathbf{Y}.$$

Namely, $\hat{\beta}_{\text{between}}$ uses variation between the cross-section observations while $\hat{\beta}_{\text{within}}$ uses time variation within each cross-section.

Random-Effects Estimator

Between and Within Estimators

$\hat{\beta}_{RE}$ is a linear combination of $\hat{\beta}_{\text{between}}$ and $\hat{\beta}_{\text{within}}$:

$$\hat{\beta}_{RE} = A\hat{\beta}_{\text{between}} + B\hat{\beta}_{\text{within}}, \text{ where } A + B = I_k.$$

What are A and B? Note that

$$\begin{aligned}
\hat{\beta}_{RE} &= (\mathbf{X}'V^{-1}\mathbf{X})^{-1}\mathbf{X}'V^{-1}\mathbf{Y} \\
&= \sigma_1^{-2}(\mathbf{X}'V^{-1}\mathbf{X})^{-1}\mathbf{X}'P\mathbf{Y} + \sigma_u^{-2}(\mathbf{X}'V^{-1}\mathbf{X})^{-1}\mathbf{X}'Q\mathbf{Y} \\
&= \underbrace{\sigma_1^{-2}(\mathbf{X}'V^{-1}\mathbf{X})^{-1}\mathbf{X}'P\mathbf{X}}_{\equiv A}(\mathbf{X}'P\mathbf{X})^{-1}\mathbf{X}'P\mathbf{Y} \\
&\quad + \underbrace{\sigma_u^{-2}(\mathbf{X}'V^{-1}\mathbf{X})^{-1}\mathbf{X}'Q\mathbf{X}}_{\equiv B}(\mathbf{X}'Q\mathbf{X})^{-1}\mathbf{X}'Q\mathbf{Y}
\end{aligned}$$

Random-Effects Estimator

Efficiency

We can calculate $\mathrm{Cov}(\hat{\beta}_{\text{between}}, \hat{\beta}_{\text{within}}|\mathbf{X}) = 0$,

$$\mathrm{Var}(\hat{\beta}_{\text{between}}|\mathbf{X}) = \sigma_1^2(\mathbf{X}'P\mathbf{X})^{-1}, \quad \mathrm{Var}(\hat{\beta}_{\text{within}}|\mathbf{X}) = \sigma_u^2(\mathbf{X}'Q\mathbf{X})^{-1}.$$

Take another arbitrary linear combination

$$\tilde{\beta} = C\hat{\beta}_{\text{between}} + D\hat{\beta}_{\text{within}}, \text{ where } C + D = I_k.$$

Then,

$$\begin{aligned}
\mathrm{Cov}(\hat{\beta}_{RE}, \tilde{\beta}|\mathbf{X}) &= A\mathrm{Var}(\hat{\beta}_{\text{between}}|\mathbf{X})C' + B\mathrm{Var}(\hat{\beta}_{\text{within}}|\mathbf{X})D' \\
&= (\mathbf{X}'V^{-1}\mathbf{X})^{-1}C' + (\mathbf{X}'V^{-1}\mathbf{X})^{-1}D' \\
&= (\mathbf{X}'V^{-1}\mathbf{X})^{-1} = \mathrm{Var}(\hat{\beta}_{RE}|\mathbf{X}).
\end{aligned}$$

Random-Effects Estimator

Efficiency

Hence,

$$\mathrm{Cov}(\hat{\beta}_{RE}, \tilde{\beta} - \hat{\beta}_{RE}|\mathbf{X}) = 0.$$

It follows that

$$\begin{aligned}
\mathrm{Var}(\tilde{\beta}|\mathbf{X}) &= \mathrm{Var}(\hat{\beta}_{RE} + (\tilde{\beta} - \hat{\beta}_{RE})|\mathbf{X}) \\
&= \mathrm{Var}(\hat{\beta}_{RE}|\mathbf{X}) + \mathrm{Var}(\tilde{\beta} - \hat{\beta}_{RE}|\mathbf{X}).
\end{aligned}$$

This means $\hat{\beta}_{RE}$ gives the best linear combination in terms of variance.

Table of Contents

PS1 Q2

Consider the following DGP

$$Y_t = \alpha + D_t(U_{1t} + \beta W_t) + U_{2t}. \tag{2}$$

Suppose D_t is a binary treatment of interest. (U_{1t}, U_{2t}) is independent of W_t and jointly normally distributed. Let $E[W_t] = \mu_W$, $E[U_{jt}] = \mu_j$, and $\mathrm{Var}(U_{jt}) = \sigma_j^2$, $j = 1, 2$.

PS1 Q2(a)

Question: What is the conditional distribution of the treatment effect $Y_t(1) - Y_t(0)|W_t = w$? What is the average treatment effect (ATE)?

Answer: By (2),

$$\begin{aligned} Y_t(1) - Y_t(0) &= [\alpha + (U_{1t} + \beta W_t) + U_{2t}] - [\alpha + U_{2t}] \\ &= U_{1t} + \beta W_t. \end{aligned}$$

It follows that $Y_t(1) - Y_t(0)|W_t = w$ is normally distributed with mean $\mu_1 + \beta w$ and variance σ_1^2. By the law of iterated expectations, the ATE is

$$\begin{aligned} \mathsf{ATE} &= E[Y_t(1) - Y_t(0)] \\ &= E[E[Y_t(1) - Y_t(0)|W_t]] \\ &= E[\mu_1 + \beta W_t] \\ &= \mu_1 + \beta \mu_W. \end{aligned}$$

PS1 Q2(b)

Question: Suppose you can assign D_t as you like. How would you design an experiment to obtain an unbiased estimator of the ATE?

Answer: I would design a completely randomized experiment:

1. given the sample size N, determine the number of units assigned to the treatment N_t such that $1 \leq N_t \leq N - 1$;
2. randomly select N_t units from the sample to receive the treatment.

PS1 Q2(b) (Continued)

Then, an unbiased estimator of the ATE is

$$\hat{\theta}_{\mathsf{ATE}} = \frac{1}{N_t} \sum_{t:D_t=1} Y_t - \frac{1}{N - N_t} \sum_{t:D_t=0} Y_t.$$

To see this, note that

- $Y_t = Y_t(1)$ if $D_t = 1$;
- $Y_t = Y_t(0)$ if $D_t = 0$.

Hence,

$$\hat{\theta}_{\mathsf{ATE}} = \frac{1}{N} \sum_{t=1}^{N} \Big[\frac{D_t \cdot Y_t(1)}{N_t/N} - \frac{(1 - D_t) \cdot Y_t(0)}{(N - N_t)/N} \Big].$$

PS1 Q2(b) (Continued)

Denote $\mathbf{Y}(d) = (Y_1(d), \ldots, Y_N(d))'$, $d = 0, 1$. We have

$$\begin{aligned} & E[\hat{\theta}_{\mathsf{ATE}} | \mathbf{Y}(0), \mathbf{Y}(1)] \\ &= \frac{1}{N} \sum_{t=1}^{N} \Big[\frac{E[D_t | \mathbf{Y}(0), \mathbf{Y}(1)] \cdot Y_t(1)}{N_t / N} \\ &\qquad\qquad - \frac{(1 - E[D_t | \mathbf{Y}(0), \mathbf{Y}(1)]) \cdot Y_t(0)}{(N - N_t)/N} \Big]. \end{aligned}$$

PS1 Q2(b) (Continued)

Given the setup of a completely randomized experiment,

$$E[D_t|\mathbf{Y}(0), \mathbf{Y}(1)] = P(D_t = 1|\mathbf{Y}(0), \mathbf{Y}(1)) = N_t/N.$$

It follows that

$$E[\hat{\theta}_{\mathsf{ATE}}|\mathbf{Y}(0), \mathbf{Y}(1)] = \frac{1}{N}\sum_{t=1}^{N}(Y_t(1) - Y_t(0)).$$

By the law of iterated expectations,

$$\begin{aligned} E[\hat{\theta}_{\mathsf{ATE}}] &= E[E[\hat{\theta}_{\mathsf{ATE}}|\mathbf{Y}(0), \mathbf{Y}(1)]] \\ &= E\Big[\frac{1}{N}\sum_{t=1}^{N}(Y_t(1) - Y_t(0))\Big] = E[Y_t(1) - Y_t(0)] = \mathsf{ATE}. \end{aligned}$$

PS1 Q2

Now suppose D_t is generated according to

$$D_t = 1\{\gamma Z_t \geq U_{3t}\}, \tag{3}$$

where $\gamma > 0$, and Z_t is a binary IV satisfying $Z_t \perp (U_{1t}, U_{2t}, U_{3t})'$, $Z_t \sim$ Bernoulli(0.5), and $U_t = (U_{1t}, U_{2t}, U_{3t})' \sim N(\mu, \Sigma)$, where $\mu = (\mu_1, 0, 0)'$, and

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \rho\sigma_1 \\ 0 & 1 & 0 \\ \rho\sigma_1 & 0 & 1 \end{bmatrix}. \tag{4}$$

PS1 Q2(c)

Question: Express the local average treatment effect (LATE) $E[Y_t(1) - Y_t(0)|D_t(1) > D_t(0)]$ as a function of the underlying parameters. Discuss how the ATE and LATE differ in this example.

Answer: By (3),

$$D_t(1) = 1\{\gamma \geq U_{3t}\}, \quad D_t(0) = 1\{0 \geq U_{3t}\}.$$

Hence,

$$\begin{aligned} D_t(1) > D_t(0) &\iff D_t(1) = 1 \,\&\, D_t(0) = 0 \\ &\iff 0 < U_{3t} \leq \gamma. \end{aligned}$$

PS1 Q2(c) (Continued)

Therefore, by (2) and the moments of the truncated bivariate normal distribution,

$$
\begin{aligned}
\mathsf{LATE} &= E[Y_t(1) - Y_t(0) | D_t(1) > D_t(0)] \\
&= E[U_{1t} + \beta W_t | U_{3t} \in (0, \gamma]] \\
&= \beta \mu_W + E[U_{1t} | U_{3t} \in (0, \gamma]] \\
&= \mu_1 + \beta \mu_W - \rho \sigma_1 \frac{\phi(\gamma) - \phi(0)}{\Phi(\gamma) - \Phi(0)},
\end{aligned}
$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the standard normal density and CDF, respectively. We can see that the ATE and LATE differ by $-\rho\sigma_1 \frac{\phi(\gamma)-\phi(0)}{\Phi(\gamma)-\Phi(0)}$.