10(c). Strict Exogeneity in Panel Data

ISS5096 || ECI

Jaewon ("Jay-one") Yoo

National Tsing Hua University

Outline

1. Fixed Effects under Strict Exogeneity

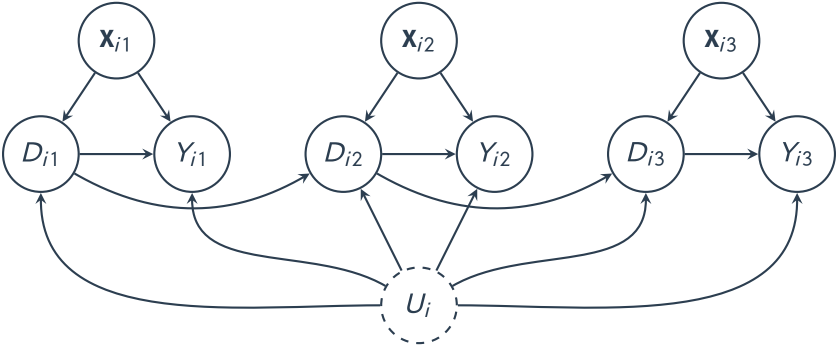
2. Sequential Exogeneity and Dynamic Panels

1/ Fixed Effects under Strict Exogeneity

Basic Idea of Fixed Effects

- "One way" fixed effects generalizes the before/after design.
    - Arbitrary treatment timing, covariates, etc.
    - Units: $i = 1, \ldots, n$
    - Causal ordering with time: covariates $\mathbf{X}_{it}$, treatment D_{it}, outcome Y_{it}
    - History of a variable: $\overline{D}_{it} = (D_{i1}, \ldots, D_{it})$ and $\overline{D}_i \equiv \overline{D}_{iT}$
- Linear fixed effects model: $Y_{it} = \alpha_i + \tau D_{it} + \mathbf{X}'_{it}\beta + \varepsilon_{it}$
    - Key assumption: **strict exogeneity** $\mathbb{E}[\varepsilon_{it} \mid \overline{\mathbf{X}}_i, \overline{D}_i, \alpha_i] = 0$
    - Implies **no feedback** between outcome and treatment ($Y_{it} \nrightarrow D_{i,t+1}$)
    - $\rightsquigarrow$ LDV cannot be a confounder!
    - Imai and Kim (2019, AJPS) give clarification on these identification issues.
- Implicit assumption of **no carryover**? $Y_{it}(d_1, \ldots, d_t) = Y_{it}(d_t)$
    - More a choice of estimand: focuses on **contemporaneous** effect.
    - Treatment history follows observed path through $t-1$:
      $Y_{it}(d_t) = Y_{it}(D_{i1}, \ldots, D_{i,t-1}, d_t)$
    - $\rightsquigarrow$ lags of treatments become part of time-varying confounders.

Strict Exogeneity DAG



Strict exogeneity implied by strict ignorability $Y_{it}(d) \perp\!\!\!\perp \overline{D}_i \mid \overline{\mathbf{X}}_i, U_i$

FE Estimation

- With linear models, two transformations can purge the fixed effects.
- **Within/FE transformation**: $\ddot{Z}_{it} = Z_{it} - T^{-1} \sum_{s=1}^{T} Z_{is}$

$$\ddot{Y}_{it} = \ddot{\mathbf{X}}_{it}' \beta + \tau \ddot{D}_{it} + \ddot{\varepsilon}_{it}$$

  - Time-demeaning Y_{it} purges the time-constant fixed effect.
  - But they retain the same coefficients as the original model.
- **First differences**: $\Delta Z_{it} = Z_{it} - Z_{i,t-1}$

$$\Delta Y_{it} = \Delta \mathbf{X}_{it}' \beta + \tau \Delta D_{it} + \Delta \varepsilon_{it}$$

- Estimation: pooled OLS of either specification, $\widehat{\tau}_{\text{fe}}$, $\widehat{\tau}_{\text{fd}}$
  - Both consistent under strict exogeneity.
  - FE more efficient if original errors, ε_{it}, are serially uncorrelated.
  - FD more efficient if differences, $\Delta \varepsilon_{it}$, are serially uncorrelated.
  - Latter allows for substantial serial dependence in the original errors.

Estimation Notes

- Within estimator can be implemented by adding unit dummy variables.

$$\underset{\alpha,\beta,\tau,\gamma}{\arg\min} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(Y_{it} - \alpha - \mathbf{X}_{it}'\beta - \tau D_{it} - \sum_{k=2}^{n} \gamma_k 1(i=k) \right)^2$$

  - **Least squares dummy variable** estimator reasonable for moderate n.
  - Computationally inefficient for large n (number of dummies grows with n).
- Best practice: cluster variances at the unit level.
  - With CR variance estimators, LSDV "double counts" degrees of freedom.
  - Better to use within estimator in that case.
- Best choice: use canned packages.
  - `fixest` in R, `reghdfe` in Stata.

Non-constant Treatment Effects

- LFE models assume constant treatment effects. What happens if not?
    - OLS typically biased because nonconstant effects induce correlation between treatment and error.
- With no covariates and only treated/control units:

$$\widehat{\tau}_{\text{fe}} \xrightarrow{p} \frac{\mathbb{E}\left[\left(\frac{\sum_t D_{it} Y_{it}}{\sum_t D_{it}} - \frac{\sum_t (1-D_{it}) Y_{it}}{\sum_t (1-D_{it})}\right) S_i^2\right]}{\mathbb{E}[S_i^2]} \neq \tau$$

    - S_i^2 is the within-unit treatment variance.
    - Units with even treatment/control split upweighted.
- Imai, Kim & Wang (2023, AJPS): use a matching to target the ATE.
    - Match treated and control periods within units (also weakens linearity).
    - `PanelMatch` R package.

2/ Sequential Exogeneity and Dynamic Panels

Strict vs. Sequential Exogeneity/Ignorability

- Strict exogeneity/ignorability is **very strong**.
  - Remember: rules out all outcome-treatment feedback.
- Weaker assumption: **sequential ignorability**:

$$Y_{it}(d) \perp\!\!\!\perp D_{it} \mid \overline{\mathbf{X}}_{it}, \overline{D}_{i,t-1}, \alpha_i$$

  - Allow Y_{it} to be related to future $D_{i,t+s}$.
- This implies **sequential exogeneity** of the errors:

$$\mathbb{E}[\varepsilon_{it} \mid \overline{\mathbf{X}}_{it}, \overline{D}_{it}, \alpha_i] = 0.$$

- Estimation to these **dynamic panel models**:
  - Instrumental variables (Arellano and Bond) using lagged difference and levels as instruments (only valid for linear models).
  - Bias correction: estimate the bias and subtract it off (valid for nonlinear models too).

Effect of Lagged Treatments

- Focused on the contemporaneous effect of D_{it}.
- What about treatment histories $Y_{it}(d_{t-1}, d_t)$?
- Very difficult, if not impossible with fixed effects models.
  - Complicated by the effect of treatment on time-varying confounders.
  - Pathways involving $\mathbf{X}_{it}(d_{t-1})$ difficult to identify.
- Possible approach: **propensity score FEs** (Blackwell & Yamauchi, 2021)
  - Include unit dummies in propensity score model.
  - Bias from incidental parameters, but disappears as $T \to \infty$.

Back to the Main Deck!

Contact Information:
jaewon.yoo@iss.nthu.edu.tw
https://j1yoo.github.io/