

Part C: Panel Data Methods

C1: Linear Panel Data Methods Recap

Kirill Borusyak
ARE 213 Applied Econometrics
UC Berkeley, Fall 2023

Panel data methods: Outline

1. Recap of linear panel data methods
2. Canonical DiD and event studies
3. DiD with staggered adoption
4. Synthetic control methods and factor models

C1 outline

- 1 Linear panels: Estimation, inference, efficiency
- 2 Some extensions
- 3 Application

Motivation

- Selection on observables is rarely convincing in cross-sectional data
 - ▶ Self-selection is complex, too many unknown confounders
- To allow for selection on (some) unobservables, leverage repeated observations for the same unit over time — **panel data**
 - ▶ How do outcomes change when treatment changes?
 - ▶ This doesn't resolve the fundamental problem of causal inference but helps controls for confounders that are time-invariant

Linear panel model

- For $i = 1, \dots, I$ and $t = 1, \dots, T$, consider a constant-effects model

$$Y_{it} = \beta' X_{it} + \alpha_i + \varepsilon_{it}$$

where β is of interest; α_i is additive “unobserved heterogeneity”

- Denote $\mathbf{X}_i = (X_{i1}, \dots, X_{iT})$, $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iT})$, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \dots, \varepsilon_{iT})$
- Every (i, t) is observed \implies **balanced panel**; with missing data: unbalanced panel
- **Strict exogeneity**: $\mathbb{E}[\varepsilon_{it} \mid X_{i1}, \dots, X_{iT}, \alpha_i] = 0$ — no time-varying confounders
 - ▶ Precludes $X_{it} \equiv Y_{i,t-1}$ as one of the RHS variables — see “**dynamic panel**” models

Random and fixed effects

- If selection on α_i is allowed, $\mathbb{E}[\alpha_i | \mathbf{X}_i] \neq \text{const}$, α_i is called **fixed effect**
 - ▶ If $\mathbb{E}[\alpha_i | \mathbf{X}_i] = \text{const}$ (no selection on unobservables) $\implies \alpha_i$ is a **random effect**
 - ▶ These labels are not about whether α_i is stochastic (think random sample of $(\alpha_i, \mathbf{X}_i, \varepsilon_i, \mathbf{Y}_i)_{i=1}^N$)
- To estimate β in the FE model, remove α_i in different ways:
 - ▶ “FE regression”: Dummy variable regression = within transformation
 - ▶ First differences
 - ▶ Long differences
- Note: random effect model allows for more methods
 - ▶ “Pooled OLS” regression of Y_{it} on X_{it} across i, t identifies β
 - ▶ “Random effects” Generalized Least Squares estimator

Estimation by dummy variables (FE) regression

- View $\{\alpha_i\}$ as a set of nuisance parameters multiplying dummies for each unit:

$$Y_{it} = \beta' X_{it} + \sum_{j=1}^I \alpha_j W_{ji} + \varepsilon_{it}, \quad W_{ji} \equiv \mathbf{1}[i = j] \quad (*)$$

- ▶ *Note:* don't write e.g. $Y_{it} = \beta' X_{it} + \sum_j \alpha_j + \varepsilon_{it}$,
- By FWL, OLS estimation of $(*)$ is identical to OLS after **within transformation**:

$$\dot{Y}_{it} = \beta' \dot{X}_{it} + \dot{\varepsilon}_{it}, \quad \dot{V}_{it} = V_{it} - \bar{V}_i = V_{it} - \frac{1}{T} \sum_{s=1}^T V_{is}$$

- ▶ *Exercise:* prove this
- ▶ By strict exogeneity, $\mathbb{E} \left[\dot{X}_{it} \dot{\varepsilon}_{it} \right] = \mathbb{E} \left[\left(X_{it} - \frac{1}{T} \sum_{s=1}^T X_{is} \right) \left(\varepsilon_{it} - \frac{1}{T} \sum_{s=1}^T \varepsilon_{is} \right) \right] = 0$
- ▶ This is much faster than using I dummies! Use *reghdfe* in Stata, *fixest* in R

Estimation by differencing

- The FE model also implies the first-difference (FD) specification:

$$\Delta Y_{it} = \beta' \Delta X_{it} + \Delta \varepsilon_{it}, \quad \Delta V_{it} \equiv V_{it} - V_{i,t-1}, \quad t = 2, \dots, T$$

with $\mathbb{E} [\Delta X_{it} \cdot \Delta \varepsilon_{it}] = 0$ by strict exogeneity

- ▶ *Exercise:* with $T = 2$, OLS of estimators FE and FD equations are identical, even in unbalanced panels

- And a long-difference specification: for $h > 1$,

$$Y_{it} - Y_{i,t-h} = \beta' (X_{it} - X_{i,t-h}) + (\varepsilon_{it} - \varepsilon_{i,t-h}), \quad t = h + 1, \dots, T$$

with $\mathbb{E} [(X_{it} - X_{i,t-h}) \cdot (\varepsilon_{it} - \varepsilon_{i,t-h})] = 0$

Asymptotic sequences

- Conventional asymptotic: **short panel**
 - ▶ A growing sample of I units for a fixed number of periods, T
- Alternative asymptotic: **long panel**
 - ▶ The sample grows by increasing both I and T (at the same or different rates)
 - ▶ More appropriate when $I \approx T$

Cluster-robust inference (1)

- Heteroskedasticity-robust SEs rely on

$$\text{Var} \left[\sum_{it} X_{it} \varepsilon_{it} \right] = \mathbb{E} \left[\sum_{it} X_{it} X'_{it} \varepsilon_{it}^2 \right], \quad \text{as } \mathbb{E} [X_{it} X'_{js} \varepsilon_{it} \varepsilon_{js}] = 0 \text{ for } it \neq js$$

- ▶ Fails in panel data because of serial correlation
 - ▶ Serial correlation in ε_{it} is prevalent
 - ▶ In short panels, ε_{it} are serially correlated even if ε_{it} are not
 - ▶ $\Delta \varepsilon_{it}$ is correlated with $\Delta \varepsilon_{i,t-1}$, unless ε_{it} follows a random walk
- Solution: cluster-robust (“clustered”) SEs

Cluster-robust inference (2)

For pooled OLS of Y_{it} on X_{it} without FEs:

$$\hat{\beta} = \left(\sum_i \sum_t X_{it} X'_{it} \right)^{-1} \left(\sum_i \sum_t X_{it} Y_{it} \right)$$

$$\sqrt{I} \cdot (\hat{\beta} - \beta) = \underbrace{\left(\frac{1}{I} \sum_i \left(\sum_t X_{it} X'_{it} \right) \right)^{-1}}_{\xrightarrow{P} \mathbb{E}[\sum_t X_{it} X'_{it}]^{-1}} \cdot \underbrace{\left(\frac{1}{\sqrt{I}} \sum_i \left(\sum_t X_{it} \varepsilon_{it} \right) \right)}_{\xrightarrow{D} \mathcal{N}(0, \text{Var}[\sum_t X_{it} \varepsilon_{it}])}$$

$$I \cdot \widehat{\text{Var}}[\hat{\beta}] = \left(\frac{1}{I} \mathbf{X}' \mathbf{X} \right)^{-1} \cdot \left(\frac{1}{I - \dim(\mathbf{X})} \sum_i \left(\sum_t X_{it} \hat{\varepsilon}_{it} \right) \left(\sum_t X_{it} \hat{\varepsilon}_{it} \right)' \right) \cdot \left(\frac{1}{I} \mathbf{X}' \mathbf{X} \right)^{-1}$$

Cluster-robust inference (3)

- Alternative derivation for the sandwich “meat”:

$$\text{Var} \left[\sum_i \left(\sum_t X_{it} \varepsilon_{it} \right) \right] = \sum_{it} \sum_{js} \mathbb{E} [X_{it} X'_{js} \varepsilon_{it} \varepsilon_{js}] = \sum_i \sum_{t,s=1}^T \mathbb{E} [X_{it} X'_{is} \varepsilon_{it} \varepsilon_{is}]$$

because all terms with $i \neq j$ are zero

- With FEs, same with $(X_{it}, \varepsilon_{it})$ replaced by $(\dot{X}_{it}, \dot{\varepsilon}_{it})$
 - ▶ In FDs, replaced by $(\Delta X_{it}, \Delta \varepsilon_{it})$
- Warning: the approximation requires I to be large. Not enough to have large IT

Choosing between estimators

- Efficiency:
 - ▶ FE estimator is efficient when ε_{it} are serially uncorrelated
 - ▶ FD estimator is efficient when ε_{it} follow a random walk
 - ▶ It's not about persistence in the data (which could come from α_i) but about differential persistence of observations close in time
- Robustness to model misspecification:
 - ▶ E.g. violations of strict exogeneity, measurement error
- Fads: FE estimation is more popular; it *appears* that you've controlled for more
 - ▶ False, and FD are more transparent (especially in more complex situations)
 - ▶ If you use a FE specification, always rewrite it in FD

Outline

- 1 Linear panels: Estimation, inference, efficiency
- 2 Some extensions
- 3 Application

Some extensions

1. Two-way fixed effects
2. Effect dynamics
3. Nonlinear panel models
4. Fixed effects beyond panel data

Two-way fixed effects (TWFE)

- Besides α_i , we may want to include (additive) period effects γ_t to capture shocks that affect all units:

$$Y_{it} = \beta' X_{it} + \alpha_i + \gamma_t + \varepsilon_{it} \quad (\#)$$

- Unlike α_i , period FEs are non-stochastic parameters (on period dummies)
- ($\#$) requires an innocuous normalization on $\{\alpha_i\}$ or $\{\gamma_t\}$
- FE estimator: in balanced panels, OLS from **double-differenced** specification:

$$\ddot{Y}_{it} = \beta' \ddot{X}_{it} + \ddot{\varepsilon}_{it}, \quad \ddot{V}_{it} = (V_{it} - \bar{V}_i) - (\bar{V}_t - \bar{V})$$

- ▶ Follows from FWL, because unit and period dummies are exactly orthogonal

- FD estimator: OLS from $\Delta Y_{it} = \beta' \Delta X_{it} + \Delta \gamma_t + \Delta \varepsilon_{it}$ with period FEs $\Delta \gamma_t$

Modeling effect dynamics

- **Distributed lags** model can be accommodated with no change:

$$Y_{it} = \beta'_0 X_{it} + \beta'_1 X_{i,t-1} + \dots \beta'_L X_{i,t-L} + \alpha_i + \varepsilon_{it}$$

- **Lagged dependent variable** on the RHS: dynamic panel model

$$Y_{it} = \rho Y_{i,t-1} + \beta' X_{it} + \alpha_i + \varepsilon_{it}$$

- ▶ Violates strict exogeneity: ε_{it} can't be mean-independent of $Y_{i,s-1}$ for $s = t + 1$
- ▶ FE estimation is not consistent in short panels: "Nickell bias"
- ▶ Can use Arellano-Bond GMM estimator

Nonlinear panel data models

Nonlinear models with fixed effects are much more complicated. Consider binary choice models:

$$Pr(Y_{it} = 1 \mid X_{it}, \alpha_i) = F(\beta'X_{it} + \alpha_i), \quad F = \text{probit or logistic}$$

- Likelihood estimation of β along with $\{\alpha_i\}$ results in the **incidental parameter problem**: β is inconsistent in short panels
- For logistic regression (but not probit), conditional logit estimation yields consistent $\hat{\beta}$
 - ▶ But not for average partial effects which depend on α_i
- More progress with long panels; see Fernández-Val and Weidner (Annual Review of Economics, 2018)

FEs beyond panel data

There are other types of data with repeated observations:

- Twin studies = repeated observations in the same family
 - ▶ E.g. Ashenfelter and Rouse (1998) estimate returns to schooling for twins
 - ▶ Family FEs control for genetic differences
 - ▶ *Warning*: why does the treatment (e.g. education) vary between twins?
 - ▶ Bound and Solon (1999): first-borns have higher weight, IQ, schooling



Khoa Vu 
@KhoaVuUmn

Your honor, those are my emotional support fixed effects.

FEs beyond panel data (2)

There are other types of data with repeated observations:

- County-level cross-section = repeated observations for the same state
 - ▶ State FEs control for additive state-level unobservables
- In a county-level panel, can include state-by-year FEs:

$$Y_{it} = \beta' X_{it} + \alpha_i + \gamma_{s(i),t} + \varepsilon_{it}$$

- ▶ *Note:* $\sum_{s',t'} \gamma_{s't'} \mathbf{1}[s(i) = s'] \times \mathbf{1}[t = t']$ is correct; $\gamma_{s(i)} \times \delta_t$ is wrong
- ▶ Including $\gamma_{s(i),t}$ demeans Y_{it} and X_{it} by state-year-specific means

FEs beyond panel data (3)

There are other types of data with repeated observations:

- Repeated cross-sections: in each year a new random sample from each state
 - ▶ Can't control for individual heterogeneity
 - ▶ But state FEs control for additive state-level unobservables
 - ▶ Cluster at the state-level if X_{it} only varies by state
- Dyadic data: e.g. how does distance X_{ij} between exporting country i and importing country j affect log trade flow Y_{ij} ? (Gravity equation)

Outline

- 1 Linear panels: Estimation, inference, efficiency
- 2 Some extensions
- 3 Application

Application: Are there good firms?

- Abowd, Kramarz, Margolis (1999): Are there good firms that pay higher wages to the same workers?
 - ▶ How much variation in wages is explained by worker characteristics? By firm characteristics?
 - ▶ Do “better” workers tend to work for “better” firms?
- Use employer-employee matched data for France
 - ▶ A panel of workers i : observe employer ID $j(it)$, experience, wages

AKM model

Model of log-wages Y_{it} :

$$Y_{it} = \beta' X_{it} + \alpha_i + \gamma_{j(it)} + \varepsilon_{it}$$

- X_{it} are time-varying observables, e.g. experience — **not** of interest
- $\{\alpha_i\}$ are worker FEs; $\{\gamma_j\}$ are firm FEs
- ε_{it} captures match-specific wage premium
- Relative firm FEs are identified by **movers**:

$$\Delta Y_{it} = \beta' \Delta X_{it} + (\gamma_{j(it)} - \gamma_{j(i,t-1)}) + \Delta \varepsilon_{it}$$

- ▶ (actual estimation via dummy variable regression, not in first-differences)
- Requires **exogenous mobility** $\mathbb{E}[\varepsilon_{it} \mid X_{it}, \alpha_i, j(it)] = 0$: matching of firms and workers can depend on FEs but not on match quality ε_{it}

Testing exogenous mobility (Card, Heining, Klein 2013)

Do movers from high- γ to low- γ firms lose *less* than movers from low- γ to high- γ gain?

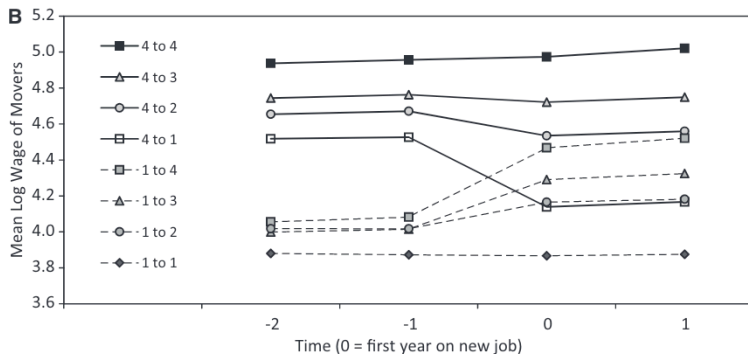


FIGURE V

(use quartiles of average wages paid to other workers; German employer-employee data)

Estimation issues

- Abowd et al. (1999) compute:
 - ▶ Variances of $\hat{\alpha}_i$ and $\hat{\gamma}_{j(it)}$ as worker and firm contributions to wage inequality
 - ▶ Covariance of $\hat{\alpha}_i$ and $\hat{\gamma}_{j(it)}$ as a measure of sorting
- But FEs are not estimated consistently!
 - ▶ For all workers, at most 8 wage observations $\Rightarrow \text{Var} [\hat{\alpha}_i]$ biased \uparrow
 - ▶ For many firms, there are only a few movers $\Rightarrow \text{Var} [\hat{\gamma}_{j(it)}]$ biased \uparrow and $\text{Cov} [\hat{\alpha}_i, \hat{\gamma}_{j(it)}]$ biased \downarrow
- Kline, Saggio, Solvsten (2020) provide a bias correction
 - ▶ Consistent estimate of $\text{Cov} [\alpha_i, \gamma_{j(it)}]$ without consistent estimates of the FEs

Kline et al. findings (for Veneto region in Italy)

TABLE II
VARIANCE DECOMPOSITION^a

	Pooled	Younger Workers	Older Workers
<i>Variance of Firm Effects</i>			
Plug in (PI)	0.0358	0.0368	0.0415
Homoscedasticity Only (HO)	0.0295	0.0270	0.0350
Leave Out (KSS)	0.0240	0.0218	0.0204
<i>Variance of Person Effects</i>			
Plug in (PI)	0.1321	0.0843	0.2180
Homoscedasticity Only (HO)	0.1173	0.0647	0.2046
Leave Out (KSS)	0.1119	0.0596	0.1910
<i>Covariance of Firm, Person Effects</i>			
Plug in (PI)	0.0039	-0.0058	-0.0032
Homoscedasticity Only (HO)	0.0097	0.0030	0.0040
Leave Out (KSS)	0.0147	0.0075	0.0171
<i>Correlation of Firm, Person Effects</i>			
Plug in (PI)	0.0565	-0.1040	-0.0334
Homoscedasticity Only (HO)	0.1649	0.0726	0.0475
Leave Out (KSS)	0.2830	0.2092	0.2744
<i>Coefficient of Determination (R^2)</i>			
Plug in (PI)	0.9546	0.9183	0.9774
Homoscedasticity Only (HO)	0.9029	0.8184	0.9524
Leave Out (KSS)	0.8976	0.8091	0.9489

Extensions

AKM-style models have also been applied in other settings:

- Wages depend on worker FE and city FE (Glaeser and Mare 2001)
- Changes in credit depends on bank FE and client firm FE (in a cross-section; Amiti and Weinstein 2018)