Statistical Modeling and Causal Inference with R

Week 9: Panel Data

Manuel Bosancianu

Max Schaub

November 9, 2020

Hertie School of Governance

Recap

A powerful way to achieve causal identification: leverage multiple observations over time.

The DiD estimator manages to do this with 2 observations in time: prior to and after a treatment was applied.

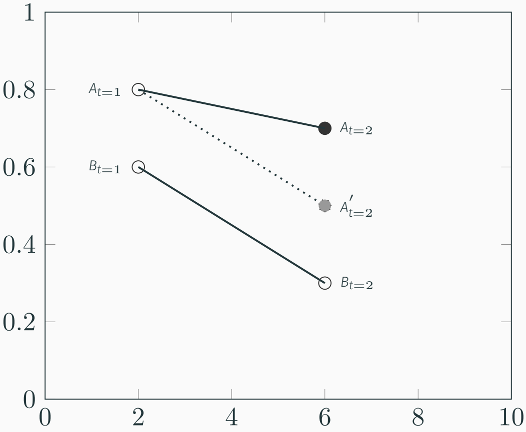
Excellent at ensuring that unobserved time-invariant confounders don't bias estimates (comparisons happen within-units).

Assumption: parallel trends. Had there been no treatment, temporal dynamics for treatment group would have been the same as for control group.

Parallel trends

Effect: $Y^{A}_{t=2} - Y^{A'}_{t=2}$

Not mandatory to have same units over time—works for pooled cross-sections too.



$$\beta = (Y^{A}_{t=2} - Y^{B}_{t=2}) - (Y^{A}_{t=1} - Y^{B}_{t=1})$$

Today's plan

Extend this design to cases with more than 2 time points:

- ✓ Features of 2 "pure" designs: cross-sectional and temporal;
- ✓ Panel data as a mix of CS and T;
- ✓ Estimation: *FE* vs. *RE*
- ✓ Assumptions implicit in strategies

Cross-sectional and temporal designs

Types of data structures

Cross-section: sample of countries, firms, regions, schools, employees ... all measured at same time point, t.

Measurements on $(Y_i; X_i)$ for $i = 1, 2, \ldots, N$ at time t.

Time-series: same unit (country, school, firm, child) measured over multiple time points.

Measurements on $(Y_t; X_t)$ for $t = 1, 2, \ldots, T$ for same unit.

Types of data structures

Panel data: multiple units measured over multiple time points.

Measurements on $(Y_{it}; X_{it})$ for $i = 1, 2, \ldots, N$, and $t = 1, 2, \ldots, T$.

"Pure" CS or TS designs have weaknesses in terms of ability to estimate a causal effect.

Panel data can (theoretically) overcome these weaknesses.

Short didactic example

Card and Krueger (1994) examine whether a minimum price increase influences unemployment.

Treatment: increase in minimum wage in NJ (but not in PA) in Apr. 1992.

Same restaurants over time—Feb. 1992 and Nov.-Dec. 1992.

Static-group comparison

$$Y_{i1}^0 = \theta_i^0 + \delta_1 + \upsilon_{i1}^0$$
$$Y_{i1}^1 = \tau + \theta_i^1 + \delta_1 + \upsilon_{i1}^1$$

θ_i: time-invariant unit-specific effect.

δ_t: time period effect (identical for all units here: δ_1).

$$E[Y_{i1}^1 - Y_{i1}^0] = \tau + E[\theta_i^1 - \theta_i^0] + E[\upsilon_{i1}^1 - \upsilon_{i1}^0] \tag{1}$$

δ_1 disappears—unobserved time-varying factors don't bias the estimate.

Assumptions made

Exogeneity: mean of error is independent of the treatment.

$$E[\upsilon_i^1] = E[\upsilon_i^0] \Rightarrow E[\upsilon_i^1 - \upsilon_i^0] = 0 \tag{2}$$

Random effects: unobserved unit heterogeneity is independent of the treatment.

$$E[\theta_i^1] = E[\theta_i^0] \Rightarrow E[\theta_i^1 - \theta_i^0] = 0 \tag{3}$$

Effect of wage increase: static comparison

If we're willing to make the two assumptions, a static comparison can recover treatment effect.

Model for employment change (Nov. 1992 data)

	DV: No. full-time
(Intercept)	7.56^{***}
	(0.91)
State (NJ)	0.88
	(1.01)
R^2	0.00
Adj. R^2	-0.00
Num. obs.	398

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Heroic random effects assumption—even with controls, hard to make it convincing.

Longitudinal comparison

$$Y_{i0}^{0} = \delta_0^0 + \theta_i + \upsilon_{i0}^{0}$$
$$Y_{i1}^{1} = \tau + \delta_1^1 + \theta_i + \upsilon_{i1}^{1}$$

θ_i: time-invariant unit-specific effect (identical for all units here).

δ_t: time period effect.

$$E[Y_{i1}^{1} - Y_{i0}^{0}] = \tau + E[\delta_1^1 - \delta_0^0] + E[\upsilon_{i1}^{1} - \upsilon_{i1}^{0}] \quad (4)$$

θ_i disappears—unobserved time-invariant unit-specific effect don't bias estimate.

Assumptions made

Exogeneity: $E[\upsilon_{i1}^{1}] = E[\upsilon_{i0}^{0}]$.

Temporal stability: no impact of unobserved time-varying factors.

$$E[\delta_1^1] = E[\delta_0^0] \tag{5}$$

Time-invariant factors (θ_i) don't matter, because the units compared are the same.

Effect of wage increase: temporal comparison

Model for employment change (NJ data)

	DV: No. full-time
(Intercept)	7.72^{***}
	(0.44)
Time (Nov. 1992)	0.72
	(0.62)
R^2	0.00
Adj. R^2	0.00
Num. obs.	647

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Heroic temporal stability assumption.

Panel Data

Benefits

Partly addresses the unit heterogeneity and temporal instability problems.

Relies on a weaker exogeneity assumption.

$$E[Y_{i1}^{1} - Y_{i1}^{0}] - E[Y_{i0}^{1} - Y_{i0}^{0}] = \tau + E[\epsilon_{i1}^{1} - \epsilon_{i1}^{0}] - E[\epsilon_{i0}^{1} - \epsilon_{i0}^{0}] \quad (6)$$

You saw this in previous class: DiD estimator was applied to a specific type of panel data.

Extending the lessons of DiD to instances with $t > 2$ and where, typically, $N \gg T$.

General structure

$$Y_{it} = \beta_0 + \beta_1 D_{it} + \underbrace{\theta_i + \delta_t + \upsilon_{it}}_{\epsilon_{it}} \tag{7}$$

Error term components:

- θ_i: unit fixed effect—captures influence of time-invariant features
- δ_t: time fixed effect—captures influence of time-varying features (affect all units)
- ϵ_{it}: “classical” error term

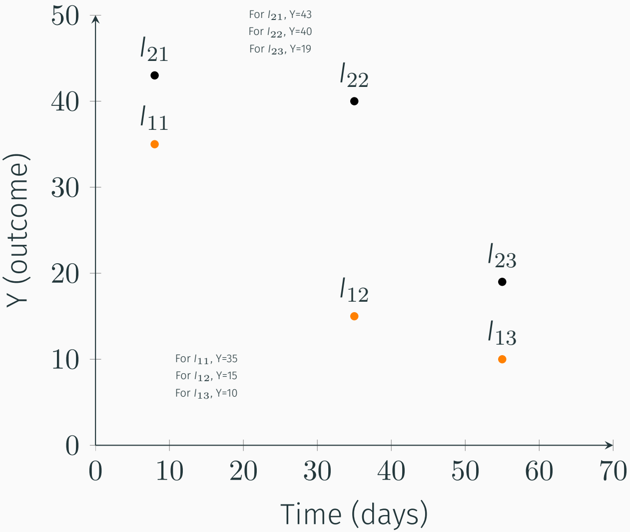
Benefits of panel data I

$$Y_{it} = \beta_0 + \beta_1 D_{it} + \underbrace{\theta_i + \delta_t + \upsilon_{it}}_{\epsilon_{it}} \tag{8}$$

Normally, in this setup, homogeneity assumption is violated: $Cov(D_{it}, \theta_i) \neq 0$.

However, we can partial out θ_i from ϵ_{it} (it's constant in longitudinal data) by "taking out" unit means in the outcome.

Partialling out θ_i



Partialling out θ_i

For group 1, $\overline{Y}$=20. For group 2, $\overline{Y}$=35.

	Y raw	θ_i partialled out
I_{11}	35	15
I_{12}	15	-5
I_{13}	10	-10
I_{21}	43	8
I_{22}	40	5
I_{23}	19	-16

Remaining variance is the relative position of units within a group, e.g. the distance between I_{11} and I_{12} is still 20.

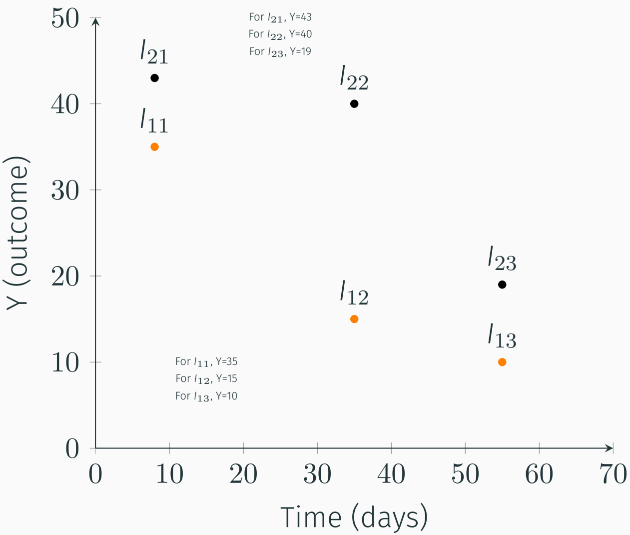
Benefits of panel data II

$$Y_{it} = \beta_0 + \beta_1 D_{it} + \underbrace{\theta_i + \delta_t + \upsilon_{it}}_{\epsilon_{it}} \quad (9)$$

Temporal stability assumption is violated as well: $Cov(D_{it}, \delta_t) \neq 0$.

We can partial out δ_t from ϵ_{it} (it's constant in cross-sectional) by "taking out" period means in the outcome.

Same panel setup



Partialling out δ_t

$\overline{Y}_{period\ 1} = 39$ | $\overline{Y}_{period\ 2} = 27.5$ | $\overline{Y}_{period\ 3} = 14.5$.

	Y raw	δ_t partialled out
I_{11}	35	-4
I_{21}	43	4
I_{12}	15	-12.5
I_{22}	40	12.5
I_{13}	10	-4.5
I_{23}	19	4.5

Remaining variance is the relative position of units within a time period, e.g. the distance between I_{11} and I_{21} is still 8.

DiD redux I

We will cover the mechanics of partialling out θ_i and δ_t in the next section: *Estimation*.

See how similar this is to the ***DiD*** framework!

$$Y_{ct} = \alpha + \beta D_c + \gamma Post_t + \tau(D_c \times Post_t) + \upsilon_{ct} \tag{10}$$

D_c: partials out unit heterogeneity.

$Post_t$: partials out temporal instability.

DiD redux II

$$Employment_{it} = \alpha + \underbrace{\beta State_i}_{\theta_i} + \underbrace{\gamma Time_t}_{\delta_t} + \tau(State_i \times Time_t) + \upsilon_{it} \tag{11}$$

$State_i \times Time_t$ is an indicator for the treatment:

$$State_i \times Time_t = \begin{cases} 1, & \text{if NJ } \textbf{\textit{AND}} \text{ Nov. 1992} \\ 0, & \text{if PA } \textbf{\textit{OR}} \text{ Feb. 1992} \end{cases} \tag{12}$$

Effect of wage increase

Model for employment change

	DV: No. full-time
(Intercept)	10.21^{***}
	(0.94)
State (NJ)	-2.48^{*}
	(1.04)
Time (Nov.)	-2.64^{*}
	(1.33)
State x Time	3.36^{*}
	(1.48)
R^2	0.01
Adj. R^2	0.00
Num. obs.	802

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Covariates (X_{it}) need to be added before jumping to interpretations, but fundamentals are the same.

Estimation

Simplified set-up

Let's assume period effects are not a problem (simplifies notation) for now, as it will make the presentation less cumbersome.

We're only concerned about unit heterogeneity, θ_i.

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \theta_i + \upsilon_{it} \tag{13}$$

Three strategies to estimate:

- ✓ first differences (FD)
- ✓ fixed effects (FE)
- ✓ random effects (RE)

First differences (FD)

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \theta_i + \upsilon_{it} \tag{14}$$

$$Y_{i(t-1)} = \beta_0 + \beta_1 X_{i(t-1)} + \theta_i + \upsilon_{i(t-1)} \tag{15}$$

Though θ_i is a problem, we can eliminate it by differencing.

$$Y_{it} - Y_{i(t-1)} = \beta_1 X_{it} - \beta_1 X_{i(t-1)} + \upsilon_{it} - \upsilon_{i(t-1)} \tag{16}$$

$$\Delta Y_{it} = \beta_1 \Delta X_{it} + \Delta \upsilon_{it} \tag{17}$$

Fixed effects (FE)

Alternative is to "de-mean" variables, by subtracting their mean across time points.

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \theta_i + \upsilon_{it} \tag{18}$$

$$\bar{Y}_i = \beta_0 + \beta_1 \bar{X}_i + \bar{\theta}_i + \bar{\upsilon}_i \tag{19}$$

Keeping in mind that $\bar{\theta}_i = \frac{\sum_{t=1}^{T} \theta_i}{T} = \frac{T * \theta_i}{T} = \theta_i$,

$$Y_{it} - \bar{Y}_i = \beta_1 (X_{it} - \bar{X}_i) + (\upsilon_{it} - \bar{\upsilon}_i) \tag{20}$$

Dummy variables (LSDV)

FE and FD will be mathematically identical *only* in instances with 2 time points.

Related to FE: Least Squares Dummy Variable (LSDV) regression.

Logic: use $i-1$ dummy indicators for units, which capture *all* between-unit heterogeneity.

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \underbrace{\beta_2 U_1 + \beta_3 U_2 + \cdots + \beta_i U_{i-1}}_{\textit{unit dummies}} + \upsilon_{it} \tag{21}$$

LSDV and FE will be mathematically identical.

Fixed effects (FE)

FE (or LSDV) won't allow the use of time-invariant predictors of Y_{it}. They have the same fate as θ_i: de-meaning absorbs them.

FE can be expanded to also incorporate time-period fixed-effects $\Rightarrow$ two-way fixed effects.

Two-way FE models don't allow for purely time-invariant or purely time-varying predictors; only X_{it} work.

Random effects (RE)

One challenge with LSDV is the large number of parameters estimate (all those unit dummies!).

Degrees of freedom are: $n * T - n - k$, with n designating units, T time periods, and k covariates; problematic if T is small.

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \theta_i + \upsilon_{it} \tag{22}$$

Solution: avoid estimating $n - 1$ coefficients by making a simplifying assumption about distribution of θ_i.

Random effects (RE)

Usually, this assumption is that $\theta_i \sim iid(0, \sigma_\theta^2)$.

In this case, we only estimate a variance (of this distribution).

Not so much correcting for heterogeneity bias, but *modeling* it.

However, using the RE estimator requires one strong assumption: $Cov(X_{it}, \theta_i) = 0$, which is encountered rarely in practice.

Assumptions

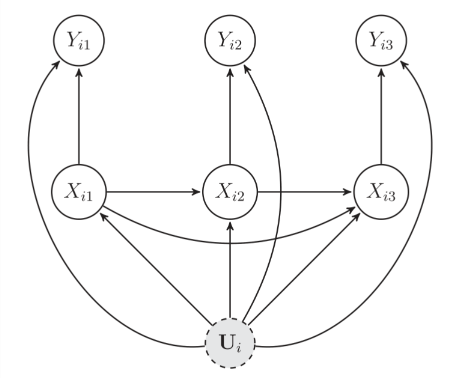
Assumptions embedded in FE

Imai and Kim (2019) refer to the same model, though notation is more complex:

$$\hat{\beta}_{LIN-FE} = argmin \sum_{i=1}^{N} \sum_{t=1}^{T} \{(Y_{it} - \bar{Y}_i) - \beta(X_{it} - \bar{X}_i)\}^2 \tag{23}$$

This specification come at the cost of assuming away the possibility of dynamic temporal relationships.

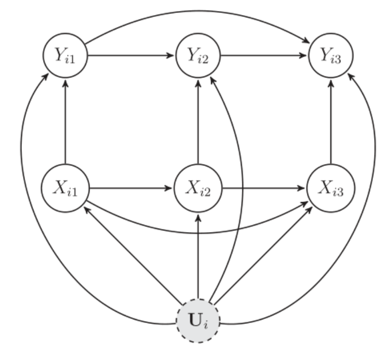
Assumptions of standard FE



1. No unobserved time-varying confounders
2. Past outcomes don't influence current ones
3. Past outcomes don't influence current treatment
4. Past treatment doesn't influence current outcome

Relaxing assumptions I

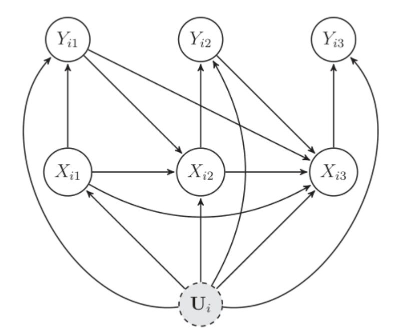
Not much can be done for assumption #1.



Assumption #2: *Outcome*$_{t-1} \nrightarrow$ *Outcome*$_t$. Can be relaxed without biasing τ.

$$Y_{it} = \alpha_i + \beta_1 X_{it} + \beta_2 X_{i(t-1)} + \epsilon_{it}$$

Relaxing assumptions II



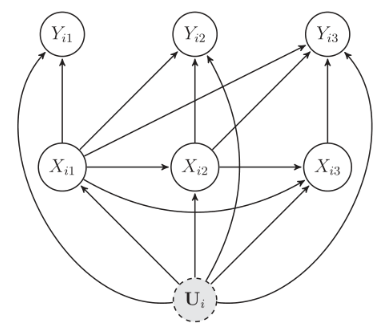
Assumption #3:
Outcome$_{t-1} \nrightarrow$ *Treatment*$_t$.

$$Y_{it} = \alpha_i + \beta X_{it} + \rho Y_{i(t-1)} + \epsilon_{it}$$

Standard OLS isn't good here (can bias downwards other βs) (Achen, 2000).

IV-based approach using X_{i1}, X_{i2} and Y_{i1} as instruments and controlling for U_i and Y_{i2} (Arellano & Bond, 1991).

Relaxing assumptions III



Assumption #4:
Treatment$_{t-1}$ $\nrightarrow$ *Outcome*$_t$.

$$Y_{it} = \alpha_i + \beta_1 X_{it} + \beta_2 X_{i(t-1)} + \epsilon_{it}$$

Also dealt with lagged values of predictors (usually 1-period lag).

Summary

Panel data analysis

Very common type of data, allowing researchers to easily control for confounders.

Offer some leverage over causal inference, given the temporal ordering of observations.

Gives leverage over both within-unit change, and across-unit differences.

Panel data analysis

However, keeping heterogeneity bias and temporal instability in check comes at the expense of dynamic relationships.

Solutions have been found for estimating dynamic and long-run effects, but estimation is not straightforward.

New strategies are still being proposed: `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3555463`.

Remaining challenges: (1) settings where T is long; (2) how to handle "sluggish" variables (see `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=622581`).

Thank you for the kind attention!

Achen, C. H. (2000). Why Lagged Dependent Variables Can Suppress the Explanatory Power of Other Independent Variables. *Annual Meeting of the Political Methodology Section of the American Political Science Association*.

Arellano, M., & 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*(2), 277–297.

Card, D., & Krueger, A. B. (1994). Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey. *The American Economic Review*, *84*(4), 772–793.

Imai, K., & Kim, I. S. (2019). When Should We Use Unit Fixed Effects Regression Models for Causal Inference with Longitudinal Data? *American Journal of Political Science*, *63*(2), 467–490.