New Panel Estimators

Ian McCarthy | Emory University
Workshop on Causal Inference with Panel Data

Table of contents

The Problem of Parallel Trends

Parallel Trends

- Assumption on trends of counterfactual (what if treated never received treatment)
- Central assumption in (essentially) all DD work
- Methods we've discussed are not robust to violations of this assumption

Other Thorny Issues

- Strict exogeneity is a strong assumption
- What if past outcomes affect treatment (standard endogeneity concern)?
- What about treatment turning on, off, and on again?

Explicit counterfactual imputation

- $y_{it}(0) = \beta x_{it} + L_{it} + \varepsilon_{it}$
- $y_{it}(1)$ observed for treated units
- Form $y_{it}(1) - \hat{y}_{it}(0)$ during post-treatment period for ATT estimate

Matching with Panel Data

Panel Matching

- Matching/reweighting based on pre-treatment covariates **and outcomes**
- Kernel/entropy balancing on many moments of covariates, `kbal`
- Trajectory balancing on the path of the pre-treatment variable, `tjbal`

Synthetic Control

The intuition

- Maybe there isn't a good "control" in our analysis
- But maybe could **create** a control with some combination of all possible control groups (donors)
- What is this donor pool? And how do we combine them into a single control?

More formally

- Observed outcome y_{jt}
  - treated group, $j = 1$, so we have y_{1t}
  - all other donor groups, $j = 2, \ldots, J + 1$, we have y_{jt}

  Causal effect:

$$y_{1t} - \sum_{j=2}^{J+1} w_j^* y_{jt},$$

  where w_j^* is a set of optimal weights for each j in the donor pool

In practice

- Weights set to minimize some distance between treatment and control group covariates
- User must decide:
  - Potential donor pool
  - Covariates on which to match
  - Norm to determine weights

Estimable using `synth` in Stata and R

What about inference?

- Re-estimate for each group in the donor pool (as if they were the treated group)
- Comibine results
- Assess whether effect for **true** treatment group is extreme relative to all placebo groups

Benefits of synthetic control

- Parallel pre-trends are essentially guaranteed
- No "extrapoloation" (notorious problem with linear regression)
- Transparency of weights in control group
- Possible to pre-register donor pool and sythnetic control weights

Some caveats

- Doesn't account for reverse causation
- Must have untreated units
- Backdate treatment date under "anticipatory" effects
- Applications remain limited to very few treated units

Multiple treated units

- Difficult estimation due to non-unique weights
- Easy for positive weight assigned to otherwise very different control units (California as control for Georgia)
- Simple solution: synthetic control for each treated unit and aggregate
- Very recent literature working to avoid these issues, `augsynth` in R

Matrix Completion

Simple idea, technically complex

- $y_{it} = \beta x_{it} + L_{it} + \varepsilon_{it}$
- Only observe elements of L_{it} for untreated units
- Need to "complete" the L matrix

But that's too many parameters! So we need some regularization.

In practice

- Include fixed effects explicitly rather than embedded into L
- Implement with `gsynth` in R