Difference-in-Differences Designs

Kosuke Imai
Harvard University

STAT 186/GOV 2002 CAUSAL INFERENCE

Fall 2019
Motivation

- How should we conduct causal inference when repeated measurements are available?
- Two types of variations:
  1. cross-sectional variation within each time period
  2. temporal variation within each unit
- Before-and-after and cross-sectional designs

Can we exploit both variations?
Minimum Wage and Unemployment


- How does the increase in minimum wage affect employment?
- Many economists believe the effect is negative
  - especially for the poor
  - also for the whole economy
- Hard to randomize the minimum wage increase

In 1992, NJ minimum wage increased from $4.25 to $5.05
- Neighboring PA stays at $4.25
- Observe employment in both states before and after increase

NJ and (eastern) PA are similar
- Fast food chains in NJ and PA are similar: price, wages, products, etc.
- They are most likely to be affected by this increase
Difference-in-Differences Design

- Parallel trend assumption

![Graph showing the difference-in-differences design with trend lines for the treatment group (New Jersey) and the control group (Pennsylvania). The graph illustrates the average causal effect estimate, with before and after periods indicated.](image)
Setup:
- Two time periods: time 0 (pre-treatment), time 1 (post-treatment)
- \( G_i \): treatment (\( G_i = 1 \)) or control (\( G_i = 0 \)) group
- \( Z_{it} = tG_i \): treatment assignment indicator for \( t = 0, 1 \)
- Potential outcomes: \( Y_{i0}(0), Y_{i0}(1), Y_{i1}(0), Y_{i1}(1) \)
- Observed outcomes: \( Y_{it} = Y_{it}(Z_{it}) \)

Average treatment effect for the treated:

\[
\tau = \mathbb{E}\{ Y_{i1}(1) - Y_{i1}(0) \mid G_i = 1 \}
\]

Parallel trend assumption:

\[
\mathbb{E}\{ Y_{i1}(0) - Y_{i0}(0) \mid G_i = 1 \} = \mathbb{E}\{ Y_{i1}(0) - Y_{i0}(0) \mid G_i = 0 \}
\]

DiD estimator:

\[
\hat{\tau}_{\text{DiD}} = \{\mathbb{E}(Y_{i1} \mid G_i = 1) - \mathbb{E}(Y_{i0} \mid G_i = 1)\} - \{\mathbb{E}(Y_{i1} \mid G_i = 0) - \mathbb{E}(Y_{i0} \mid G_i = 0)\}
\]

Applicable to repeated cross-section data as well
Two-way fixed effects model:

\[ Y_{it}(z) = \alpha_i + \beta t + \tau z + \epsilon_{it} \]

- \( \mathbb{E}\{Y_{i0}(0)\} = \alpha_i \)
- \( \mathbb{E}\{Y_{i1}(0)\} = \alpha_i + \beta \)
- \( \mathbb{E}\{Y_{i1}(1)\} = \alpha_i + \beta + \tau \)
- \( \mathbb{E}\{Y_{i1}(1) - Y_{i1}(0)\} = \tau \)

Parallel trend assumption:

- \( \mathbb{E}\{Y_{i1}(0) - Y_{i0}(0) \mid G_i = g\} = \beta \)
- Or equivalently \( \mathbb{E}(\epsilon_{i1} - \epsilon_{i0} \mid G_i = g) = 0 \)
- Both \( Z_{it} \) and \( \epsilon_{it} \) can depend on \( \alpha_i \) or unobserved confounders

Least squares estimator equals the nonparametric DiD estimator, i.e., \( \hat{\tau}_{FE} = \hat{\tau}_{DiD} \)

This equivalence does not hold in general beyond the \( 2 \times 2 \) case
Comparison with the Lagged Outcome Model

- Lagged outcome model:
  \[ Y_{i1}(z) = \alpha + \rho Y_{i0} + \tau z + \epsilon_i(z) \]

- Nonparametric identification assumption:
  \[ \{ Y_{i1}(1), Y_{i1}(0) \} \perp \perp Z_{it} \mid Y_{i0} \]
  - can be made conditional on \( X_i \) as well as \( Y_{i0} \)
  - neither stronger nor weaker than the parallel trend assumption
  - same as parallel trend if \( \mathbb{E}( Y_{i0} \mid G_i = 1) = \mathbb{E}( Y_{i0} \mid G_i = 0) \)

- Where does the imbalance in lagged outcome come from?
  - Difference-in-Differences \( \rightsquigarrow \) unobserved time-invariant confounder
  - Lagged outcome directly affects treatment assignment
Relationship between the DiD and Lagged Outcome Estimators

- Least squares estimator:
  \[
  \hat{\tau}_{LD} = \mathbb{E}(Y_{i1} | G_i = 1) - \mathbb{E}(Y_{i1} | G_i = 0) - \hat{\rho}\{(\mathbb{E}(Y_{i0} | G_i = 1) - \mathbb{E}(Y_{i0} | G_i = 0))\}
  \]
  
  If \(\hat{\rho} = 1\), then \(\hat{\tau}_{LD} = \hat{\tau}_{DiD}\)
  
  Assume \(0 \leq \rho < 1\) (stationarity)
  
  Without loss of generality, assume \(\mathbb{E}(Y_{i0} | G_i = 1) \geq \mathbb{E}(Y_{i0} | G_i = 0)\) (monotonicity)
  
  1. If parallel trend holds, \(\mathbb{E}(\hat{\tau}_{LD}) \geq \mathbb{E}(\hat{\tau}_{DiD}) = \tau\)
  2. If ignorability holds, \(\tau = \mathbb{E}(\hat{\tau}_{LD}) \geq \mathbb{E}(\hat{\tau}_{DiD})\)

- Nonparametric estimator (Ding and Li. 2019. Political Anal.):
  
  \[
  \mu_0 = \mathbb{E}\{Y_{i1}(0) | G_i = 1\} = \mathbb{E}\{\mathbb{E}(Y_{i1} | G_i = 0, Y_{i0}) | G_i = 1\}
  \]
  
  (stationarity) \( \partial\mathbb{E}(Y_{i1} | G_i = 0, Y_{i0} = y)/\partial y < 1 \) for all \(y\)
  
  (stochastic monotonicity) \( F_{Y_0}(y | G_i = 1) \leq F_{Y_0}(y | G_i = 0) \) for all \(y\)

Then, the bracketing relationship holds
Adjusting for Baseline Covariates

- Parallel trend assumption conditional on the baseline covariates:
  \[
  \mathbb{E}\{ Y_{i1}(0) - Y_{i0}(0) \mid X_i = x, G_i = 1 \} \\
  = \mathbb{E}\{ Y_{i1}(0) - Y_{i0}(0) \mid X_i = x, G_i = 0 \} \quad \text{for all } x
  \]

- Matching: parallel trend within a pair or a strata

  \[
  \mathbb{E}\{ Y_{i1}(1) - Y_{i1}(0) \mid G_i = 1 \} \\
  = \mathbb{E}\left[ \frac{Y_{i1} - Y_{i0}}{\Pr(G_i = 1)} \cdot \frac{G_i - \Pr(G_i = 1 \mid X_i)}{1 - \Pr(G_i = 1 \mid X_i)} \right]
  \]

- Unconditional parallel trend assumption neither implies nor is implied by conditional parallel trend assumption
Nonlinear Difference-in-Differences

(Athey and Imbens. 2006. *Econometrica*)

- Standard DiD relies upon the linearity assumption
- Not invariant to a nonlinear transformation of outcome (e.g., log)

Temporal change in quantile is identical between the two groups
Formalization of Nonlinear DiD

- Distribution functions: $F_{gt}(y) = \Pr(Y_{it}(0) \leq y \mid G_i = g)$
- Quantile treatment effect for a given $q$:

$$\tau(q) = \tilde{F}_{11}^{-1}(q) - F_{11}^{-1}(q)$$

where $\tilde{F}_{11}(y) = \Pr(Y_{i1}(1) \leq y \mid G_i = 1)$

- We wish to identify $F_{11}(y)$
- Identification assumption:

$$\frac{F_{01}(F_{00}^{-1}(q))}{r_0(q)} = \frac{F_{11}(F_{10}^{-1}(q))}{r_1(q)}$$

for all $q \in [0, 1]$

- Under this assumption,

$$F_{11}(y) = F_{01}[F_{00}^{-1}\{F_{10}(y)\}]$$

- One treated unit $i = 1$ receiving the treatment at time $T$
- Quantity of interest: $Y_{1T} - Y_{1T}(0)$
- Create a synthetic control using past outcomes
- Weighted average:

$$\hat{Y}_{1T}(0) = \sum_{i=2}^{N} \hat{w}_i Y_{iT}$$

where the weights balance past outcomes

$$\hat{w} = \text{argmin}_{w} \sum_{t=1}^{T-1} \left( Y_{1t} - \sum_{i=2}^{N} w_i Y_{it} \right)^2$$

with $\sum_{i=2}^{N} \hat{w}_i = 1$ and $\hat{w}_i \geq 0$

- One could include time-invariant covariates $X_i$
Causal Effect of ETA’s Terrorism

**Figure 1. Per capita GDP for the Basque Country**

The figure illustrates the per capita GDP for the Basque Country from 1955 to 2000. The solid line represents the actual per capita GDP with terrorism, while the dashed line represents the synthetic per capita GDP without terrorism. The graph shows a positive correlation between terrorism activity and the per capita GDP gap, with spikes in terrorist activity followed by increases in the amplitude of the GDP gap.
Placebo Test

Figure 4. A “Placebo Study,” per capita GDP for Catalonia

can do this for all control units and compare them with the treated unit
Figure 4. Per-capita cigarette sales gaps in California and placebo gaps in all 38 control states.

Provide a good fit for per capita cigarette consumption prior to Proposition 99 for the majority of the states in the donor pool. However, Figure 4 indicates also that per capita cigarette sales during the 1970–1988 period cannot be well reproduced for some states by a convex combination of per capita cigarette sales in other states. The state with worst fit in the pre-Proposition 99 period is New Hampshire, with a MSPE of 3437. The large MSPE for New Hampshire does not come as a surprise. Among all the states in the donor pool, New Hampshire is the state with the highest per capita cigarette sales for every year prior to the passage of Proposition 99. Therefore, there is no combination of states in our sample that can reproduce the time series of per capita cigarette sales in New Hampshire prior to 1988. Similar problems arise for other states with extreme values of per capita cigarette sales during the pre-Proposition 99 period.

If the synthetic California had failed to fit per capita cigarette sales for the real California in the years before the passage of Proposition 99, we would have interpreted that much of the post-1988 gap between the real and the synthetic California was also artificially created by lack of fit, rather than by the effect of Proposition 99. Similarly, placebo runs with poor fit prior to the passage of Proposition 99 do not provide information to measure the relative rarity of estimating a large post-Proposition 99 gap for a state that was well fitted prior to Proposition 99. For this reason, we provide several different versions of Figure 4, excluding states beyond a certain level of pre-Proposition 99 MSPE.

Figure 5 excludes states that had a pre-Proposition 99 MSPE of more than 20 times the MSPE of California. This is a very lenient cutoff, discarding only four states with extreme values of pre-Proposition 99 MSPE for which the synthetic method would be clearly ill-advised. In this figure there remain a few lines that still deviate substantially from the zero gap line in the pre-Proposition 99 period. Among the 35 states remaining in the figure, the California gap line is now about the most unusual line, especially from the mid-1990s onward.

Figure 6 is based on a lower cutoff, excluding all states that had a pre-Proposition 99 MSPE of more than five times the MSPE of California. Twenty-nine control states plus California remain in the figure. The California gap line is now clearly the most unusual line for almost the entire post-treatment period.

In Figure 7 we lower the cutoff even further and focus exclusively on those states that we can fit almost as well as California in the period 1970–1988, that is, those states with pre-Proposition 99 MSPE not higher than twice the pre-Proposition 99 MSPE for California. Evaluated against the distribution of the gaps for the 19 remaining control states in Figure 7, the gap for California appears highly unusual. The negative effect in California is now by far the lowest of all. Because this figure includes 19 control states, the probability of estimating a large post-Proposition 99 gap for a state that is well fitted prior to Proposition 99 is much lower than in Figure 5.
Model-based Justification

- The main motivating factor analytic model:

\[ Y_{it}(0) = \gamma_t + \delta_t^T X_i + \xi_t^T U_i + \epsilon_{it} \]

- Generalization of the linear two-way fixed effects model
- Key assumption: there exist weights such that

\[ \sum_{i=2}^{N} w_i X_i = X_1 \quad \text{and} \quad \sum_{i=2}^{N} w_i U_i = U_1 \]

- Another motivating autoregressive model with time-varying covariates:

\[ Y_{it}(0) = \rho_t Y_{i,t-1}(0) + \delta_t^T X_{it} + \epsilon_{it} \]
\[ X_{it} = \lambda_{t-1} Y_{i,t-1}(0) + \Delta_{t-1} X_{i,t-1} + \nu_{it} \]

- Past outcomes can affect current treatment
- No unobserved time-invariant confounders
Generalizing the Difference-in-Differences

🌟 staggered treatment ⇝ What if units go in and out of treatment?


Choose the number of lags $L$ and leads $F$

ATE of Policy Change for the Treated:

\[
\mathbb{E} \left\{ Y_{i,t+F} \left( Z_{it} = 1, Z_{i,t-1} = 0, \{ Z_{i,t-\ell} \}_{\ell=2}^{L} \right) - Y_{i,t+F} \left( Z_{it} = 0, Z_{i,t-1} = 0, \{ Z_{i,t-\ell} \}_{\ell=2}^{L} \right) \mid Z_{it} = 1, Z_{i,t-1} = 0 \right\}
\]

Estimation procedure:

1. Construct a matched set for each treated unit that consists of control units with the identical treatment history up to $L$ time periods
2. Refine covariate balance with a matching/weighting method within a matched set
3. Use the multi-period difference-in-differences estimator:

\[
\frac{1}{N} \sum_{i=1}^{N} \sum_{t=L+1}^{T-F} Z_{it} \left\{ \left( Y_{i,t+F} - Y_{i,t-1} \right) - \sum_{i' \in M_{it}} w_{it}' \left( Y_{i',t+F} - Y_{i',t-1} \right) \right\}
\]
Empirical Application (1)

- ATT with $L = 4$ and $F = 1, 2, 3, 4$
- We consider democratization and authoritarian reversal
- Examine the number of matched control units
- 18 (13) treated observations have no matched control

Democratization

Authoritarian Reversal

Acemoglu et al. (2018)
Scheve & Stasavage (2012)

Number of matched control units
Improved Covariate Balance

Before Matching

Before Refinement

Mahalanobis Distance Matching

Propensity Score Matching

Propensity Score Weighting

Standardized Mean Differences for Democratization

Standardized Mean Differences for Authoritarian Reversal

Years relative to the administration of treatment

Acemoglu et al. (2018)

Scheve & Stasavage (2012)

Kosuke Imai (Harvard University)

Difference-in-Differences Designs

Causal Inference (Fall 2019)
Estimated Causal Effects

Mahalanobis Matching
Up to 5 matches
Up to 10 matches

Propensity Score Matching
Up to 5 matches
Up to 10 matches

Propensity Score Weighting

Estimated Effect of Democratization

Estimated Effect of Authoritarian Reversal

Years relative to the administration of treatment

Kosuke Imai (Harvard University) Difference-in-Differences Designs Causal Inference (Fall 2019) 21 / 22
Concluding Remarks

- Difference-in-differences design:
  - fully exploit the panel data structure
  - cross sectional and before-and-after designs do not
  - parallel trend assumption
  - adjusts for time-invariant unobserved confounders
  - tradeoff between dynamics and unobservables

- Extensions:
  - adjusting for baseline covariates
  - nonlinear difference-in-differences
  - systhetic controls
  - multiperiod difference-in-differences estimator

- Readings: ANGRIST AND PISKE. CHAPTER 5