# CAUSAL INFERENCE WITH PANEL DATA: NEW METHODS AND REMAINING CHALLENGES

Guido Imbens – Stanford University

(based on work with Alexander Almeida, Dmitry Arkhangelsky, Susan Athey, Eva Lestant, Alexia Olaizola, Zhaona Qu, Davide Viviano)

NABE, Oct 28, 2024

# OUTLINE

# 1.1 INTRODUCTION

- Panel with $N$ units and $T$ periods.
- Each unit/period pair $(i, t)$ is characterized by a pair of potential outcomes $(Y_{it}(0), Y_{it}(1))$ and a binary treatment $W_{it} \in \{0, 1\}$
- We observe for each unit the pair $(W_i, Y_i)$ where

$$Y_{it} \equiv Y_{it}(W_{it}) = \begin{cases} Y_{it}(0) & \text{if } W_{it} = 0, \\ Y_{it}(1) & \text{if } W_{it} = 1. \end{cases}$$

- The focus is on the average treatment effect for the treated:

$$\tau \equiv \sum_{i,t} W_{it}(Y_{it}(1) - Y_{it}(0)) \Big/ \sum_{i,t} W_{it}$$

## 1.II INTRODUCTION: EXAMPLES

- What is the effect of an advertising campaign on consumer spending (units are media markets, time periods are days)?
- What is effect of changes in state regulation (units are states, periods are days or weeks)?
- What was effect of Brexit on UK GDP (units are countries, time periods may be quarters or years)?
- Lots of variation in empirical settings in
  - number of units,
  - number of time periods,
  - correlation patterns over time,
  - correlation patterns across units.

## 1.III INTRODUCTION: ESTIMATORS

- Lots of estimators (with variation within classes)
  - Differences in Differences (DID) / Two-Way-Fixed-Effects (TWFE)
  - Synthetic Control Methods
  - Factor Models / Matrix Completion
  - Synthetic Difference In Differences
- Estimators are motivated in different ways, sometimes based on models, sometimes algorithmic
- Even though not all estimators can be used in all settings, they address closely related problems, more than much of the literature suggests.

# 1.IV DID/TWFE

- Difference In Differences / Two Way Fixed Effect Regression:

$$\min_{\alpha,\beta,\tau^{\mathrm{DID}}} \sum_{i,t} \left( Y_{it} - \alpha_i - \beta_t - \tau^{\mathrm{DID}} W_{it} \right)^2$$

$$\text{if } \mathbf{W} = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & \dots & 0 & 1 & 1 & 1 \end{pmatrix}$$

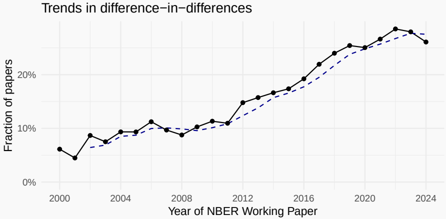
block assignment (common in practice)

$$\text{then} : \hat{\tau}^{\mathrm{DID}} = \overline{Y}^{\mathrm{post,treat}} - \overline{Y}^{\mathrm{pre,treat}} - \left( \overline{Y}^{\mathrm{post,control}} - \overline{Y}^{\mathrm{pre,control}} \right)$$

## 1.V DID/TWFE

- Often in settings where policy decisions are made at aggregate level: evaluation of state level regulations with individual data (Card-Krueger, 1995, restaurant data to evaluate state-level minimum wage regulations).
- Lots of recent DID work: Goodman-Bacon (2021), Callaway & Sant'Anna (2021), Sun & Abraham (2021), surveys: Roth, Sant'Anna, Bilinski & Poe (2021), De Chaisemartin & d'Haultfoeuille (2020) dealing with complications in settings with staggered adoption, allowing for general treatment effect heterogeneity.

# 1.VI INTRODUCTION: DIFFERENCE IN DIFFERENCES




- (graph courtesy of Paul Goldsmith-Pinkham)
- DID/TWFE is
  - Popular
  - Easy to use, transparent.
  - But, ..., not necessarily very good in practice
- Remember: DDDiD (Don't Do Difference in Differences)

# 1.VII INTRODUCTION: THE PROBLEM WITH DID/TWFE

- Very restrictive model for $Y_{it}(0)$

$$Y_{it}(0) = \alpha_i + \beta_t + \varepsilon_{it}$$

  E.g., with states as units, it assumes all systematic differences between California and Texas, or between Delaware and Alaska, are captured by additive component $\alpha_i$.

- Two alternatives:
  - Make potential outcome model more flexible
  - Estimate model locally (put more weight on similar units and similar time periods)

# 1.VIII INTRODUCTION: THE DATA (FOCUS ON CASE WITH SINGLE TREATED UNIT/PERIOD)

$$\mathbf{Y} \equiv \begin{pmatrix} Y_{11} & Y_{12} & Y_{13} & \dots & Y_{1T} \\ Y_{21} & Y_{22} & Y_{23} & \dots & Y_{2T} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Y_{N1} & Y_{N2} & Y_{N3} & \dots & Y_{NT} \end{pmatrix} \qquad \text{(Outcome)}$$

$$\mathbf{W} \equiv \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix} \qquad \text{(Treatment)},$$

- Setting with single treated unit/period allows easier comparison of estimators.
- (Covariates $X_{it}$ not that important conceptually or in practice, but can

# 1.IX INTRODUCTION: OTHER ASSIGNMENT MECHANISMS POSSIBLE

$$\mathbf{W} = \begin{pmatrix} 1 & 0 & 0 & \dots & 1 \\ 0 & 1 & 0 & \dots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 1 & \dots & 0 \end{pmatrix}$$

general assignment

$$\mathbf{W} = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 1 \\ 0 & 0 & 1 & \dots & 1 & 1 \end{pmatrix}$$

staggered adoption
(common in practice)

- Existing estimators (ctd):
  2. Synthetic Control (SC, Abadie et al, 2003, 2010):

     Define SC weights $\omega_i$

$$\omega = \arg \min_{\omega \geq 0, \sum_{i=1}^{N-1} \omega_i = 1} \sum_{t=1}^{T-1} \left( Y_{Nt} - \sum_{j=1}^{N-1} \omega_j Y_{jt} \right)^2$$

     Then:

$$\hat{\tau}^{SC} = Y_{NT} - \sum_{i=1}^{N-1} \omega_i Y_{iT}$$

# 1.XI INTRODUCTION: OTHER ESTIMATORS

- Existing estimators (ctd):
   3. DIFP (Doudchenko-Imbens, 2016, Fernan-Pinto 2021), Modified SC with free intercept

$$\hat{\tau}^{\mathrm{DIFP}} = Y_{NT} - \omega_0 - \sum_{i=1}^{N-1} \omega_i Y_{iT}$$

$$\omega = \arg \min_{\omega_j \geq 0, j \geq 1, \sum_{i=1}^{N-1} \omega_i = 1} \sum_{t=1}^{T-1} \left( Y_{Nt} - \omega_0 - \sum_{j=1}^{N-1} \omega_j Y_{jt} \right)^2$$

# 1.XII INTRODUCTION: OTHER ESTIMATORS

- Existing estimators (ctd):
  3. DIFP (Doudchenko-Imbens, 2016, Fernan-Pinto 2021),

     Can be written as weighted DID/TWFE regression

$$\min_{\alpha,\beta,\tau^{\text{DIFP}}} \sum_{i,t} \omega_i^{\text{SC}} \left(Y_{it} - \alpha_i - \beta_t - \tau^{\text{DIFP}} W_{it}\right)^2$$

     with SC weights.

     More attractive than DID/TWFE because upweights "similar" units.

## 1.XIII INTRODUCTION:

- Existing estimators (ctd):

  4. Horizontal Regression:

$$\lambda = \arg\min_{\lambda} \sum_{i=1}^{N-1} \left( Y_{iT} - \lambda_0 - \sum_{t=1}^{T-1} \lambda_t Y_{it} \right)^2$$

$$\text{Then :} \qquad \hat{\tau}^{\text{HZ}} = Y_{NT} - \sum_{t=1}^{T-1} \lambda_t Y_{Nt}$$

- Comparison:
  - $\hat{\tau}^{\text{DIFP}}$ regresses $Y_{Nt}$ on $Y_{1t}, \ldots, Y_{N-1,t}$ ($N$ regr, $T-1$ obs),
  - $\hat{\tau}^{\text{HZ}}$ regresses $Y_{iT}$ on $Y_{i1}, \ldots, Y_{i,T-1}$ ($T$ regr, $N-1$ obs)

# 1.XIV INTRODUCTION: OTHER ESTIMATORS

- Existing estimators (ctd):
  5. Matrix Completion (MC, Athey, Bayati, Doudchenko, Imbens & Khosravi, 2021):

$$\min \sum_{i,t} \left( Y_{it} - \alpha_i - \beta_t - \tau^{\mathrm{MC}} W_{it} - L_{it} \right)^2 + \lambda \|\mathbf{L}\|_{\mathrm{nn}}$$

  - ★ Nuclear norm for matrix, is sum of singular values.
  - ★ Leads to factor representation for $L_{it} = \sum_{r=1}^{R} \gamma_{ir}\delta_{rt}$.
  - ★ Regularization parameter $\lambda$, chosen through cross-validation, determines rank $R$.
  - ★ Strict generalization of DID with data-driven flexible low rank model for outcomes, should dominate DID in large samples.

## 1.XV INTRODUCTION: OTHER ESTIMATORS

- Existing estimators (ctd):
  6. Synthetic Difference in Differences (SDID, Arkhangelsky, Athey, Hirshberg, Imbens  Wager, 2021)

$$\min_{\alpha,\beta,\tau^{\text{SDID}}} \sum_{i,t} \omega_i \lambda_t \left( Y_{it} - \alpha_i - \beta_t - \tau^{\text{SDID}} W_{it} \right)^2$$

$$\omega = \arg \min_{\omega \geq 0, \sum_{i=1}^{N-1} \omega_i = 1} \sum_{t=1}^{T-1} \left( Y_{Nt} - \sum_{j=1}^{N-1} \omega_j Y_{jt} \right)^2 \quad \text{(like SC)}$$

$$\lambda = \arg \min_{\lambda \geq 0, \sum_{t=1}^{T-1} \lambda_t = 1} \sum_{i=1}^{N-1} \left( Y_{iT} - \sum_{t=1}^{T-1} \lambda_j Y_{jt} \right)^2$$

## 1.XVI INTRODUCTION: OTHER ESTIMATORS

- Existing estimators (ctd):
  7. Generalized Mundlak Estimator (Arkhangelsky & Imbens, 2024)
     - ★ Mundlak (1978) shows that DID/TWFE is identical to

$$\min_{\gamma,\delta,\tau} \sum_{i,t} \left(Y_{it} - \gamma \overline{W}_{i\cdot} - \delta \overline{W}_{\cdot t} - \tau W_{it}\right)^2$$

     - ★ One could adjust more flexibly for $\overline{W}_{i\cdot}$ and $\overline{W}_{\cdot t}$ rather than just linearly
     - ★ Makes a difference with general assignment patterns, but not with single treated unit/period.

## 2.I TWO ISSUES WITH EXISTING ESTIMATORS

1. The estimators treat the matrices  and  as row and column exchangeable (Aldous, 1981). The implications are that
   - If one were to swap the columns corresponding to 2019 and 1983, that would not change the estimates $\hat{\tau}$ for DID, SC, HZ, MC, SDID, DIFP.
   - If we are trying to estimate $Y_{N,2020}$, having pre-treatment observations for $t = 1980, \ldots, 1999$ is just as valuable as observations for $t = 2001, \ldots, 2019$.
   - Time ordering is irrelevant.
   - This is highly implausible.
   - These estimators ignore very relevant information.

# 2.II TWO ISSUES WITH EXISTING ESTIMATORS

2 SC type estimators rely on regression for estimating weights

$$\omega = \arg \min_{\omega \geq 0, \sum_{i=1}^{N-1} \omega_i = 1} \sum_{t=1}^{T-1} \left( Y_{Nt} - \sum_{j=1}^{N-1} \omega_j Y_{jt} \right)^2$$

- This does not easily generalize to staggered adoption, and to other assignment patterns: it would require regression with many missing covariates.

## 2.III ISSUES WITH INFERENCE

- Multiple proposals for doing inference
  - Regression-based methods (for DID)
  - Placebo methods
  - Conformal inference
- Not clear how to choose between them.
- Not clear what appropriate repeated sampling thought experiment is.
- focus in literature on validity, largely ignoring power

# 3.1 NEW ESTIMATOR: DOUBLY WEIGHTED CAUSAL PANEL (DWCP)

Minimize over $\alpha_i$, $\beta_t$, $\mathbf{L}$, given tuning parameters $(\lambda^U, \lambda_T, \lambda^{nn})$

$$\sum_{i,t} \exp\left(-\lambda^T d^T(T,t) - \lambda^U d^U(N,i)\right)\left(Y_{it} - \alpha_i - \beta_t - L_{it} - \tau_{NT}^{\text{DWCP}} W_{it}\right)^2 + \lambda^{nn} \|\mathbf{L}\|,$$

- distance metrics:

$$d^U(N,i) = \sum_{t=1}^{T-1} (Y_{it} - Y_{Nt})^2 \qquad d^T(T,t) = |T - t|$$

- Is related to DID, SC, MC, SDID estimators

# 3.II NEW ESTIMATOR: DOUBLY WEIGHTED CAUSAL PANEL (DWCP)

Minimize over $\alpha_i$, $\beta_t$, $\mathbf{L}$, given tuning parameters $(\lambda^U, \lambda_T, \lambda^{nn})$

$$\sum_{i,t} \exp\left(-\lambda^T d^T(T,t) - \lambda^U d^U(N,i)\right) \left(Y_{it} - \alpha_i - \beta_t - L_{it} - \tau_{NT}^{DWCP} W_{it}\right)^2 + \lambda^{nn} \|\mathbf{L}\|,$$

distance metrics

$$d^U(N,i) = \sum_{t=1}^{T} (Y_{it} - Y_{Nt})^2 \qquad d^T(T,t) = |T - t|$$

- Like DID, but with weights and factor component

# 3.III NEW ESTIMATOR: DOUBLY WEIGHTED CAUSAL PANEL (DWCP)

Minimize over $\alpha_i$, $\beta_t$, $\mathbf{L}$, given tuning parameters $(\lambda^U, \lambda_T, \lambda^{nn})$

$$\sum_{i,t} \exp\left(-\lambda^T d^T(T,t) - \lambda^U d^U(N,i)\right) \left(Y_{it} - \alpha_i - \beta_t - L_{it} - \tau_{NT}^{\mathrm{DWCP}} W_{it}\right)^2 + \lambda^{nn} \|\mathbf{L}\|,$$

distance metrics

$$d^U(N,i) = \sum_{t=1}^{T} (Y_{it} - Y_{Nt})^2 \qquad d^T(T,t) = |T - t|$$

- Like SC, but with unit fixed effects and factor component, and different weights

# 3.IV NEW ESTIMATOR: DOUBLY WEIGHTED CAUSAL PANEL (DWCP)

Minimize over $\alpha_i$, $\beta_t$, $\mathbf{L}$, given tuning parameters $(\lambda^U, \lambda_T, \lambda^{nn})$

$$\sum_{i,t} \exp\left(-\lambda^T d^T(T,t) - \lambda^U d^U(N,i)\right)\left(Y_{it} - \alpha_i - \beta_t - L_{it} - \tau_{NT}^{\mathrm{DWCP}} W_{it}\right)^2 + \lambda^{nn}\|\mathbf{L}\|,$$

distance metrics

$$d^U(N,i) = \sum_{t=1}^{T}(Y_{it} - Y_{Nt})^2 \qquad d^T(T,t) = |T - t|$$

- Like matrix completation estimator, but with weights

# 3.V NEW ESTIMATOR: DOUBLY WEIGHTED CAUSAL PANEL (DWCP)

Minimize over $\alpha_i$, $\beta_t$, $\mathbf{L}$, given tuning parameters $(\lambda^U, \lambda_T, \lambda^{nn})$

$$\sum_{i,t} \exp\left(-\lambda^T d^T(T,t) - \lambda^U d^U(N,i)\right)\left(Y_{it} - \alpha_i - \beta_t - L_{it} - \tau_{NT}^{\text{DWCP}} W_{it}\right)^2 + \lambda^{nn} \|\mathbf{L}\|,$$

distance metrics

$$d^U(N,i) = \sum_{t=1}^{T}(Y_{it} - Y_{Nt})^2 \qquad d^T(T,t) = |T-t|$$

- Like SDID estimator, but with different weights and factor component

# 3.VI NEW ESTIMATOR: KEY FEATURE 1

Minimize over $\alpha_i$, $\beta_t$, $\mathbf{L}$, given tuning parameters $(\lambda^U, \lambda_T, \lambda^{\text{nn}})$

$$\sum_{i,t} \exp\left(-\lambda^T d^T(T,t) - \lambda^U d^U(N,i)\right)\left(Y_{it} - \alpha_i - \beta_t - L_{it} - \tau_{NT}^{\text{DWCP}} W_{it}\right)^2 + \lambda^{\text{nn}} \|\mathbf{L}\|,$$

distance metrics

$$d^U(N,i) = \sum_{t=1}^{T} (Y_{it} - Y_{Nt})^2 \qquad d^T(T,t) = |T - t|$$

- Unit weights are distance based, so can deal with general assignment patterns

# 3.VII NEW ESTIMATOR: KEY FEATURE 2

Minimize over $\alpha_i$, $\beta_t$, $\mathbf{L}$, given tuning parameters $(\lambda^U, \lambda_T, \lambda^{nn})$

$$\sum_{i,t} \exp\left(-\lambda^T d^T(T,t) - \lambda^U d^U(N,i)\right)\left(Y_{it} - \alpha_i - \beta_t - L_{it} - \tau_{NT}^{\mathrm{DWCP}} W_{it}\right)^2 + \lambda^{nn} \|\mathbf{L}\|,$$

distance metrics

$$d^U(N,i) = \sum_{t=1}^{T} (Y_{it} - Y_{Nt})^2 \qquad d^T(T,t) = |T - t|$$

- time weights decline with distance in time.

# 3.V OTHER NEW ESTIMATORS

- DWCP is just one example of estimator that satisfies desiderata:
  - Allows time ordering to matter
  - Can be applied to general assignment patterns.
  - Can incorporate covariates
- Possible generalizations:
  - Include distance measure based on covariates

$$d^C(i, j) = (X_i - X_j)^\top \Sigma_X^{-1} (X_i - X_j)$$

  - Include distance measure based on assignment pattern

$$d^A(i, j) = \sum_{t=1}^{T} (W_{it} - W_{jt})^2$$

# 3.III SOME SIMULATIONS: DESIGN

- Simulation Designs from earlier SDID paper (Athey et al, 2021), report results for all 9 combinations of outcomes/treatments they consider.
- Estimated flexible model on real data set ($N$ = 50, $T$ = 50, CPS panels, Penn world tables) to get relatively realistic DGPs. Many simulation designs in the literature make no effort to ensure realistic DGPs
  - Estimate rank 4 model on full panel.
  - Extract two way fixed effects and remainder
  - Estimate autoregressive process on residuals.
  - Estimate assignment process for real regulations on factors so assignment is not uncorrelated with outcomes, but is uncorrelated with residuals.

# 3.IV SIMULS: RMSE, BEST IS ORANGE WORST IS BLUE

| outcome | treatment | DWCP | SDID | SC | DID | MC | DIFP |
|---|---|---|---|---|---|---|---|
| CPS logwage | min wage | 1.00 | 1.09 | 1.13 | 1.46 | 1.07 | 1.12 |
| CPS urate | min wage | 1.02 | 1.12 | 1.06 | 1.00 | 1.11 | 1.06 |
| CPS hours | min wage | 1.00 | 1.08 | 1.07 | 1.20 | 1.06 | 1.07 |
| CPS log wage | gun law | 1.00 | 1.11 | 1.13 | 1.33 | 1.06 | 1.14 |
| CPS log wage | abortion | 1.00 | 1.10 | 1.11 | 1.46 | 1.05 | 1.14 |
| CPS log wage | random | 1.00 | 1.12 | 1.13 | 1.14 | 1.05 | 1.14 |
| PENN GDP | democracy | 1.00 | 1.10 | 2.13 | 6.19 | 1.27 | 1.60 |
| PENN GDP | education | 1.00 | 1.08 | 2.44 | 4.74 | 1.29 | 1.50 |
| PENN GDP | random | 1.00 | 1.08 | 2.67 | 6.92 | 1.22 | 2.51 |

- DWCP estimator almost uniformly better
- Directly generalizes to general assignment mechanisms
- DID very poor for some DGPs

# 4.1 VARIANCE ESTIMATION

- What assumptions do we need to estimate the variance?
- Model

$$Y_{it}(0) = \mu_{it} + \varepsilon_{it}, \quad \mathbb{E}[\varepsilon_{it}] = 0, \quad \mathbb{V}(\varepsilon_{it}) = \sigma^2_{it}$$

- Then, if we can estimate $\mu_{it}$:

$$\mathbb{V}(\hat{\tau}) \approx \sigma^2_{NT}$$

## 4.II EXISTING VARIANCE ESTIMATORS

- How do we estimate the variance $\sigma^2_{NT}$?
- $\hat{\mathbb{V}}^{\text{marg}}$: Based on robust variance for regression estimators,

$$Y_{it} = \alpha_i + \beta_t + \tau W_{it} + \varepsilon_{it} = X_{it}^\top \theta + \varepsilon_{it}$$

$$\hat{\mathbb{V}}^{\text{marg}} = \left( \sum_{i,t} X_{it} X_{it}^\top \right)^{-1} \left( \sum_{i,t} \hat{\varepsilon}_{it}^2 X_{it} X_{it}^\top \right) \left( \sum_{i,t} X_{it} X_{it}^\top \right)^{-1}$$

Approximately variance for estimated treatment effect $\hat{\tau}$:

$$\hat{\sigma}_{NT}^{2,\text{marg}} \equiv \frac{1}{NT-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (1 - W_{it}) \hat{\varepsilon}_{it}^2$$

## 4.III EXISTING VARIANCE ESTIMATORS

- $\mathbb{V}^{\text{unit}}$: Unit-placebo method: pretend each of the other $N-1$ units is treated in period $T$, estimate treatment effect for that unit, average squared error for those pseudo treatment effects.
  - Iterate through units $i = 1, \ldots, N-1$. Estimate the treatment effect using the outcome data $Y_{j,s}$, $j \neq N$, and $(j, s) \neq (i, T)$. The error is $\hat{\varepsilon}_{iT} = Y_{iT} - \hat{Y}_{iT}$.
  - Average the squared errors,

$$\hat{\sigma}^{2,\text{unit}}_{NT} \equiv \frac{1}{N-1} \sum_{i=1}^{N-1} \hat{\varepsilon}^2_{iT}$$

- Abadie et al (2010), Doudchenko and Imbens (2016)

# 4.IV EXISTING VARIANCE ESTIMATORS FOR PANEL DATA

- $\mathbb{V}^{\text{time}}$: Time-placebo method: pretend unit $N$ is treated in the other $T-1$ time periods, estimate effect, average squared errors.
  - Iterate through $t = 1, \ldots, T-1$. Estimate the treatment effect using the outcome data $Y_{j,s}$, $s \neq T$, and $(j,s) \neq (N,t)$. The error is $\hat{\varepsilon}_{Nt} = Y_{Nt} - \hat{Y}_{Nt}$. Average the squared errors,

$$\hat{\sigma}^{2,\text{time}}_{NT} \equiv \frac{1}{T-1} \sum_{t=1}^{T-1} \hat{\varepsilon}^2_{Nt}$$

- Doudchenko and Imbens (2016), Chernozhukov et al (2021)

# 4.IV EXISTING VARIANCE ESTIMATORS

- Marginal variance method

$$\hat{\sigma}_{NT}^{2,\text{marg}} \equiv \frac{1}{NT-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (1 - W_{it}) \hat{\varepsilon}_{it}^{2}$$

- Unit-placebo method

$$\hat{\sigma}_{NT}^{2,\text{unit}} \equiv \frac{1}{N-1} \sum_{i=1}^{N-1} \hat{\varepsilon}_{iT}^{2}$$

- Time-placebo method

$$\hat{\sigma}_{NT}^{2,\text{time}} \equiv \frac{1}{T-1} \sum_{t=1}^{T-1} \hat{\varepsilon}_{Nt}^{2}$$

# 4.V HOW DO THE EXISTING METHODS PERFORM?

- Take as outcome average log earnings from CPS for each of 50 states, for 40 years.
- For each of $N \times T = 2,000$ pairs $(i, t)$ calculate the estimated pseudo treatment effect $\hat{\tau}_{it}$.
- Calculate for each of $2,000$ pairs $(i, t)$ each of the three standard errors, $\sqrt{\hat{V}_{it}^{\text{marg}}}$, $\sqrt{\hat{V}_{it}^{\text{unit}}}$, $\sqrt{\hat{V}_{it}^{\text{time}}}$

# 4.VI AVERAGES, STANDARD DEVIATIONS, AND CORRELATIONS OF STANDARD ERRORS

| | Mean | SD | Time | Unit | Marg | Cond. |
|---|---|---|---|---|---|---|
| Time | 10.17 | 7.02 | 1.00 | -0.08 | -0.61 | 0.88 |
| Unit | 11.76 | 3.79 | -0.08 | 1.00 | -0.11 | 0.20 |
| Marg | 12.35 | 0.01 | -0.61 | -0.11 | 1.00 | -0.67 |
| Cond. | 12.63 | 11.01 | 0.88 | 0.20 | -0.67 | 1.00 |

(a) California Smoking

| | Mean | SD | Time | Unit | Marg | Cond. |
|---|---|---|---|---|---|---|
| Time | 0.06 | 0.02 | 1.00 | -0.05 | -0.40 | 0.79 |
| Unit | 0.06 | 0.01 | -0.05 | 1.00 | -0.13 | 0.34 |
| Marg | 0.06 | 0.00 | -0.40 | -0.13 | 1.00 | -0.44 |
| Cond. | 0.06 | 0.03 | 0.79 | 0.34 | -0.44 | 1.00 |

(b) CPS Log Wage

# 4.VII QUESTIONS

- The three variance estimators (marg, unit-placebo, time-placebo) have approximately the same mean, and mean is equal to the variance of the estimator
  $\Longrightarrow$ all three are approximately valid.
- The standard deviation of these variances is quite different.
- There is substantial variation in the three variance estimators (and negative correlations)
  $\Longrightarrow$ they can be quite different for any given case.
- What should we use in practice?

# 4.VIII TOY EXAMPLE

- Toy setting with single treated unit/period, unit $N$, period $T$,

$$Y_{NT}(1) = 0 \quad \forall i, t$$

- Simple Model:

$$Y_{it}(0) = \varepsilon_{it}$$

- Estimator

$$\hat{Y}_{NT}(0) = 0 \qquad \hat{\tau} = Y_{NT}(1) - \hat{Y}_{NT}(0) = 0$$

- Error:

$$\tau - \hat{\tau} = Y_{NT}(0) - \hat{Y}_{NT} = \varepsilon_{NT}$$

# 4.IX ROW AND COLUMN EXCHANGEABILITY (RCE)

- Suppose the matrix $\varepsilon$ is row and column and exchangeable (RCE). *(this may not be realistic in proper panel setting, and we may need to first adjust for some autocorrelation structure)*
- RCE $\Longrightarrow$ there is a random $N \times T$ matrix $\varepsilon^*$ with typical element

$$\varepsilon^*_{it} = f(\kappa, \nu_i, \xi_t, \eta_{it}),$$

with all $(\kappa, \nu_i, \xi_t, \eta_{it})$ jointly independent, and $\varepsilon \overset{d}{\sim} \varepsilon^*$.

- Assume (given model for expected values $\mu_{it}$):

$$\mathbb{E}[f(\kappa, \nu_i, \xi_t, \eta_{it}) | \kappa, \nu_i, \xi_t] = 0$$

## 4.X FOUR VARIANCES

- Marginal variance:

$$\sigma^{2,\text{marg}}(\kappa) \equiv \mathbb{V}\left( f(\kappa, \nu_i, \xi_t, \eta_{it}) \middle| \, \kappa \right),$$

- Conditional variance given unit:

$$\sigma^{2,\text{unit}}(\kappa, \nu_i) \equiv \mathbb{V}\left( f(\kappa, \nu_i, \xi_t, \eta_{it}) \middle| \, \kappa, \nu_i \right)$$

- Conditional variance given unit:

$$\sigma^{2,\text{time}}(\kappa, \xi_t) \equiv \mathbb{V}\left( f(\kappa, \nu_i, \xi_t, \eta_{it}) \middle| \, \kappa, \xi_t \right)$$

- Conditional variance given unit and time period:

$$\sigma^{2,\text{cond}}(\kappa, \nu_i, \xi_t) \equiv \mathbb{V}\left( f(\kappa, \nu_i, \xi_t, \eta_{it}) \middle| \, \kappa, \nu_i, \xi_t \right)$$

## 4.XI FOUR VARIANCES

- The expected values of all variances, cond. on $\kappa$, are all equal (because the conditional means are all zero, $\mathbb{E}[f(\kappa, \nu_i, \xi_t, \eta_{it})|\kappa, \nu_i, \xi_t] = 0$)

$$\mathbb{E}\left[\sigma^{2,\text{cond}}(\kappa, \nu_i, \xi_t)\middle|\,\kappa\right] = \sigma^{2,\text{marg}}(\kappa)$$

$$\mathbb{E}\left[\sigma^{2,\text{unit}}(\kappa, \xi_t)\middle|\,\kappa\right] = \sigma^{2,\text{marg}}(\kappa)$$

$$\mathbb{E}\left[\sigma^{2,\text{time}}(\kappa, \xi_t)\middle|\,\kappa\right] = \sigma^{2,\text{marg}}(\kappa)$$

## 4.XII THREE VARIANCE ESTIMATORS

- Which variance would be preferable?
- There is an ordering in terms of the variance of these variances:

$$\mathbb{V}(\sigma^{2,\text{marg}}(\kappa)) \leq \left\{ \begin{array}{c} \mathbb{V}(\sigma^{2,\text{unit}}(\kappa, \xi_t)) \\ \\ \mathbb{V}(\sigma^{2,\text{time}}(\kappa, \xi_t)) \end{array} \right\} \leq \mathbb{V}(\sigma^{2,\text{cond}}(\kappa, \nu_i, \xi_t))$$

- No general ordering between $\mathbb{V}(\sigma^{2,\text{unit}}(\kappa, \xi_t))$ and $\mathbb{V}(\sigma^{2,\text{time}}(\kappa, \xi_t))$.

# 4.XII THREE VARIANCE ESTIMATORS

- The conditionality principle (Cox 1958 example)
  - We flip a coin to decide whether to use measurement instrument $X = A$ or measurement instrument $X = B$.
  - The measurement satisfies $Y|X \sim \mathcal{N}(\theta, \sigma_X^2)$,

$$\sigma_X^2 = \begin{cases} 1 & \text{if } X = A \\ 100 & \text{if } X = B \end{cases}$$

  - Marginal variance: $\mathbb{V}^{\text{marg}} = 50.5 = \mathbb{V}(Y)$
  - Conditional variance: $\mathbb{V}^{\text{cond}} = \sigma_X^2 = \mathbb{V}(Y|X)$
- Both valid, but we should use $\mathbb{V}^{\text{cond}}$

# 4.XIII THREE VARIANCE ESTIMATORS

- Leans into row exchangeability:

$$\tilde{\sigma}^{2,\text{unit}}_{NT} = \frac{1}{N-1}\sum_{i=1}^{N-1} \varepsilon^2_{iT} \approx \sigma^{2,\text{unit}}(\kappa, \xi_t) \qquad (\text{unit} - \text{placebo}).$$

- Leans into column exchangeability

$$\tilde{\sigma}^{2,\text{time}}_{NT} = \frac{1}{T-1}\sum_{t=1}^{T-1} \varepsilon^2_{Nt} \approx \sigma^{2,\text{time}}(\kappa, \xi_t) \qquad (\text{time} - \text{placebo}).$$

- Leans into both

$$\tilde{\sigma}^{2,\text{marg}}_{NT} = \frac{1}{NT-1}\sum_{t=1}^{T}\sum_{i=1}^{N} \mathbf{1}_{(i,t)\neq(N,T)}\,\varepsilon^2_{it} \approx \sigma^{2,\text{marg}}(\kappa) \qquad (\text{marginal}).$$

## 4.XIV IMPLICATIONS

- $\tilde{\sigma}^{2,\mathrm{time}}_{NT}$ or $\tilde{\sigma}^{2,\mathrm{unit}}_{NT}$ are preferable to $\tilde{\sigma}^{2,\mathrm{marg}}_{NT}$.
- Choice between $\tilde{\sigma}^{2,\mathrm{time}}_{NT}$ and $\tilde{\sigma}^{2,\mathrm{unit}}_{NT}$ not clear.
- $\sigma^{2,\mathrm{cond}}_{NT}$ would be preferable to all, but how to estimate it?

- Can we do something that gets closer to $\sigma^{2,\mathrm{cond}}(\kappa, \nu_i, \xi_t)$?
  - Not immediately: only a single unit/period pair with $(\nu_i, \xi_t)$.
  - Need a model for variance $\sigma^2_{it}$.

# 4.XV A NEW VARIANCE ESTIMATOR

- (Simple) model for conditional variance

$$\log\left(\sigma^{2,\text{cond}}(\kappa, \nu_i, \xi_t)\right) = \kappa + \nu_i + \xi_t.$$

- Estimate $\kappa$, $\nu_i$, and $\xi_t$ using regression

$$\log\left(\hat{\varepsilon}_{it}^2\right) = \kappa + \nu_i + \xi_t + \zeta_{it}$$

- Estimate conditional variance

$$\hat{\sigma}^{2,\text{cond}} = \exp\left(1.2704 + \hat{\kappa} + \hat{\nu}_i + \hat{\xi}_t\right)$$

- aside: The adjustment factor exp(1.2704) adjusts for the fact that if $\varepsilon \sim \mathcal{N}(0, \sigma^2)$, then

$$\mathbb{E}\left[\ln\left(\varepsilon^2\right)\right] = \ln(\sigma^2) + \int_{-\infty}^{\infty} \ln(z^2) \frac{1}{\sqrt{2\pi}} \exp\left(-z^2/2\right) dz = \ln(\sigma^2) + \frac{\Gamma(1/2)}{\sqrt{\pi}} (\ln(2) + \psi(1/2)) \approx \ln(\sigma^2) + \exp(-1.2704).$$

# 4.XV A NEW VARIANCE ESTIMATOR

$$\text{estimate}: \ln(\hat{\varepsilon}_{it}^2) = \alpha_i + \beta_t \qquad \tilde{\alpha}_i = \ln\left(\frac{\exp(\alpha_i)}{\frac{1}{N}\sum_{j=1}^N \exp(\alpha_j)}\right)$$

so that the average of $\exp(\alpha_i)$ is equal to one, and similar for $\tilde{\beta}_t$.

Calculate the dof by first calculating the var of the $\exp(\alpha_i)$:

$$\hat{V}(\exp(\tilde{\alpha}_i)) = \frac{1}{N}\sum_{i=1}^{N}\left(\exp(\tilde{\alpha}_i) - \frac{1}{N}\sum_{j=1}^{N}\exp(\tilde{\alpha}_j)\right)^2$$

We want to think of the $\exp(\tilde{\alpha}_i)$ as $\sim \chi(M_\alpha)/M_\alpha$. Hence

$$V(\exp(\tilde{\alpha})) = \frac{2M_\alpha}{M_\alpha^2} = \frac{2}{M_\alpha}, \qquad M_\alpha = \frac{2}{\hat{V}(\exp(\tilde{\alpha}_i))}$$

# 4.XVI AVERAGES, STANDARD DEVIATIONS, AND CORRELATIONS OF STANDARD ERRORS

| | | | | Unit Placebo | | | Time Placebo | | | Marginal | | | Conditional | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Data Set | S.D. | $M_\alpha$ | $M_\beta$ | mean | (s.d.) | cov | mean | (s.d.) | cov | mean | (s.d.) | cov | mean | (s.d.) | cov |
| cal_Y | 32.77 | 0.39 | 6.87 | 0.36 | 0.12 | 0.94 | 0.31 | 0.21 | 0.95 | 0.38 | 0.00 | 0.96 | 0.39 | 0.34 | 0.95 |
| cps_hours | 1.34 | 4.26 | 13.37 | 0.64 | 0.09 | 0.94 | 0.63 | 0.14 | 0.95 | 0.64 | 0.00 | 0.95 | 0.66 | 0.24 | 0.94 |
| cps_lwage | 0.44 | 5.18 | 7.61 | 0.13 | 0.03 | 0.93 | 0.13 | 0.04 | 0.92 | 0.13 | 0.00 | 0.94 | 0.14 | 0.07 | 0.91 |
| cps_urate | 0.02 | 7.11 | 8.48 | 0.67 | 0.15 | 0.94 | 0.67 | 0.14 | 0.94 | 0.69 | 0.00 | 0.95 | 0.69 | 0.25 | 0.94 |
| germ | 8.95 | 1.10 | 2.95 | 0.19 | 0.08 | 0.91 | 0.16 | 0.12 | 0.89 | 0.20 | 0.00 | 0.92 | 0.23 | 0.21 | 0.92 |
| mariel | 0.20 | 9.44 | 3.69 | 0.63 | 0.23 | 0.93 | 0.64 | 0.20 | 0.93 | 0.67 | 0.00 | 0.95 | 0.65 | 0.29 | 0.94 |
| guns | 9.71 | 0.19 | 1.91 | 0.35 | 0.17 | 0.93 | 0.21 | 0.33 | 0.93 | 0.39 | 0.00 | 0.96 | 0.25 | 0.34 | 0.90 |

Real panel simulation studies - TWFE

- All four standard errors are the same on average, and do fine for coverage, but they vary greatly in terms of standard deviation.
- we should prefer the one with larger standard deviation

## 4.XVII SOME SIMULATIONS

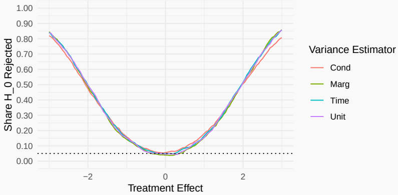
- Some Simulations (homoskedastic case)

$$\varepsilon_{it} \sim \mathcal{N}(0, \sigma_{it}^2) \qquad \ln(\sigma_{it}^2) = \lambda_{\text{unit}} z_i + \lambda_{\text{time}} v_t,$$

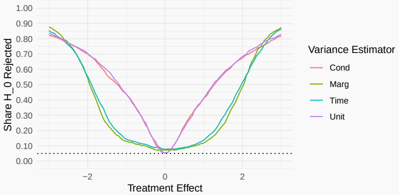
$$z_i \sim \mathcal{X}(1), \quad v_t \sim \mathcal{X}(1),$$

(a) homoskedastic case: $\lambda_{\text{unit}} = 0, \ \lambda_{\text{time}} = 0$
(b) unit-heteroskedastic case: $\lambda_{\text{unit}} = 1, \ \lambda_{\text{time}} = 0$
(c) time-heteroskedastic case: $\lambda_{\text{unit}} = 0, \ \lambda_{\text{time}} = 1$
(d) fully heteroskedastic case: $\lambda_{\text{unit}} = 1, \ \lambda_{\text{time}} = 1$

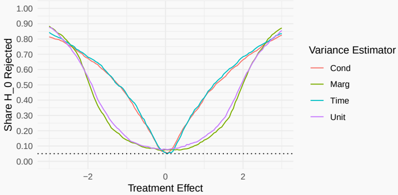
# 4.XVIII FOUR POWER CURVES WITH SIMULATED DATA



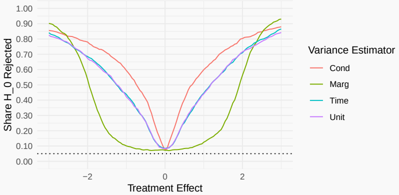

(a) $\lambda_{unit} = 0, \quad \lambda_{time} = 0$




(b) $\lambda_{unit} = 0, \quad \lambda_{time} = 1$



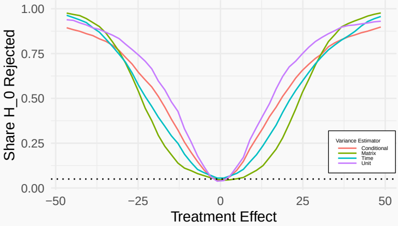

(c) $\lambda_{unit} = 1, \quad \lambda_{time} = 0$



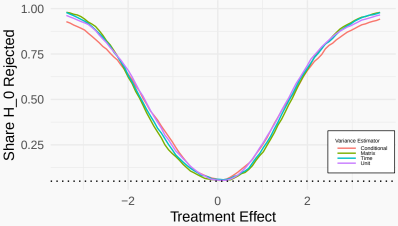

(d) $\lambda_{unit} = 1, \quad \lambda_{time} = 1$

Power curves for simulation studies with heterogeneous variance structure estimated using a two-way-fixed-effects estimator for imputing missing values.

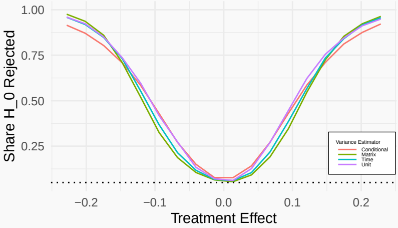
# 4.XIX FOUR POWER CURVES WITH ACTUAL DATA



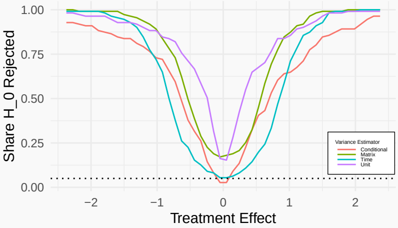

(a) California Prop. 99




(b) CPS - hours




(c) CPS - lwage




(d) Penn World Tables

# 4.XX VARIANCES

- Variance Estimation is tricky in panel data settings
- Validity of variance estimators is not sole consideration, power also matters.
- Full robustness to misspecification is not always possible.
- The choice between unit and time placebo variance estimators is not clear
- There are other approaches beyond unit and time placebo variance estimators based on modelling the heteroskedasticity.

# 5. CONCLUSION

- More to be done on estimation:
  - row-column-exchangeability is not a good assumption, but underlies many estimators.
  - Combination of outcome modelling and local estimation, similar to cross-section approaches may be effective.
- More to be done on variance estimation
  - Power and validity are both important
  - Focus on choice between unit-placebo and time-placebo is misplaced.

# REFERENCES


- Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program." *Journal _of the American statistical Association* 105, no. 490 (2010): 493-505.
- Aldous, D.J., 1981. Representations for partially exchangeable arrays of random variables. *Journal of Multivariate Analysis*, 11(4), pp.581-598.
- Almuzara, Martín. Heterogeneity in transitory income risk. Working Paper, 2020.
- Arkhangelsky, Dmitry, and David Hirshberg. "Large-sample properties of the synthetic control method under selection on unobservables." arXiv preprint arXiv:2311.13575 2 (2023).

# REFERENCES (CTD)


- Arkhangelsky, Dmitry, and Guido Imbens. Causal models for longitudinal and panel data: A survey. *Econometrics Journal* 2024.
- Arkhangelsky, Dmitry, and Guido W. Imbens. "Fixed effects and the generalized Mundlak estimator." *Review of Economic Studies* 91, no. 5 (2024): 2545-2571.
- Arkhangelsky, Dmitry, Guido W. Imbens, Lihua Lei, and Xiaoman Luo. "Design-Robust Two-Way-Fixed-Effects Regression for Panel Data." *Quantitative Economics*, (2024).
- Arkhangelsky, Dmitry, Susan Athey, David A. Hirshberg, Guido W. Imbens, and Stefan Wager. "Synthetic difference-in-differences." *American Economic Review* 111, no. 12 (2021): 4088-4118.

# REFERENCES (CTD)


- Arkhangelsky, Dmitry, and Guido Imbens. "Causal models for longitudinal and panel data: A survey." *The Econometrics Journal* (2024).
- Athey, Susan, Mohsen Bayati, Nikolay Doudchenko, Guido Imbens, and Khashayar Khosravi. "Matrix completion methods for causal panel data models."*Journal of the American Statistical Association* 116, no. 536 (2021).
- Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. "How much should we trust differences-in-differences estimates?." *The Quarterly journal of economics* 119, no. 1 (2004): 249-275.
- Bottmer, Lea, Guido W. Imbens, Jann Spiess, and Merrill Warnick. "A design-based perspective on synthetic control methods." *Journal of Business & Economic Statistics* 42, no. 2 (2024): 762-773.

# REFERENCES (CTD)


- Callaway, Brantly, and Pedro HC Sant'Anna. "Difference-in-differences with multiple time periods." *Journal of econometrics* 225, no. 2 (2021).
- De Chaisemartin, Clément, and Xavier d'Haultfoeuille. "Two-way fixed effects estimators with heterogeneous treatment effects." *American economic review* 110, no. 9 (2020): 2964-2996.
- Chernozhukov, Victor, Kaspar Wüthrich, and Yinchu Zhu. "An exact and robust conformal inference method for counterfactual and synthetic controls." *Journal of the American Statistical Association* 116, no. 536 (2021): 1849-1864.
- Doudchenko, Nikolay, and Guido W. Imbens. Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. No. w22791. National Bureau of Economic Research, 2016.

# REFERENCES (CTD)


- Ferman, Bruno, and Cristine Pinto. "Synthetic controls with imperfect pretreatment fit." *Quantitative Economics* 12, no. 4 (2021): 1197-1221.
- Goldsmith-Pinkham, Paul. "Tracking the Credibility Revolution across Fields." arXiv preprint arXiv:2405.20604 (2024).
- Goodman-Bacon, Andrew. "Difference-in-differences with variation in treatment timing." *Journal of econometrics* 225, no. 2 (2021): 254-277.
- Imbens, Guido, Nathan Kallus, and Xiaojie Mao. "Controlling for unmeasured confounding in panel data using minimal bridge functions: From two-way fixed effects to factor models." arXiv preprint arXiv:2108.03849 (2021).

# REFERENCES (CTD)


- Mundlak, Yair. "On the pooling of time series and cross section data." Econometrica (1978): 69-85.
- Roth, Jonathan, Pedro HC Sant'Anna, Alyssa Bilinski, and John Poe. "What's trending in difference-in-differences? A synthesis of the recent econometrics literature." *Journal of Econometrics* 235, no. 2 (2023): 2218-2244.
- Sekhon, Jasjeet S., and Yotam Shem-Tov. "Inference on a new class of sample average treatment effects." *Journal of the American Statistical Association* 116, no. 534 (2021): 798-804.

# REFERENCES (CTD)


- Splawa-Neyman, Jerzy, Dorota M. Dabrowska, and Terrence P. Speed. "On the application of probability theory to agricultural experiments. Essay on principles. Section 9." *Statistical Science* (1990): 465-472.
- Sun, Liyang, and Sarah Abraham. "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." *Journal of econometrics* 225, no. 2 (2021): 175-199.
- White, Halbert. "Maximum likelihood estimation of misspecified models." *Econometrica* (1982): 1-25.