Causal Panel Data Models

Guido Imbens

Stanford University)

ICML

Wednesday, July 20th, 2022, 12.30-1.30pm

Based on work with Dmitry Arkhangelsky, Susan Athey, Mohsen Bayati, Lea Bottmer, Raj Chetty, Nick Doudchenko, Skip Hirshberg, Nathan Kallus, Hyunseung Kang, Khas Khosravi, Luihua Lei, Xiaoman Luo, Xiaojie Mao,Jann Spiess, Stefan Wager, Yuhao Wang, Merrill Warnick, Ruoxuan Xiong

Arkhangelsky, Dmitry, Susan Athey, David A. Hirshberg, Guido W. Imbens, and Stefan Wager. Synthetic difference in differences, 2019. AAHIW

Arkhangelsky, Dmitry, and Guido Imbens. The role of the propensity score in fixed effect models. No. w24814. National Bureau of Economic Research, 2018. AI1

Arkhangelsky, D., & Imbens, G. W. (2021). Double-robust identification for causal panel data models (No. w28364). National Bureau of Economic Research, AI2

Arkhangelsky, Dimtry, Guido Imbens, Lihua Lei, and Xiaoman Luo, (2021) Double Robust Two-way Fixed Effect Regression For Panel Data. AILL

Athey, Susan, Mohsen Bayati, Mohsen, Nick Doudchenko, Guido Imbens, and Khashayar Khosravi, (2018). Matrix completion methods for causal panel data models, forthcoming *Journal of the American Statistical Association*. ABDIK

Athey, Susan, Raj Chetty, and Guido Imbens. "Combining experimental and observational data to estimate treatment effects on long term outcomes." arXiv preprint arXiv:2006.09676 (2020), ACI.

Athey, Susan, Raj Chetty, Guido Imbens, and Hyunseung Kang. "Estimating treatment effects using multiple surrogates: The role of the surrogate score and the surrogate index." arXiv preprint arXiv:1603.09326 (2016), ACIK.

Athey, Susan, and Guido W. Imbens. Design-based analysis in difference-in-differences settings with staggered adoption, forthcoming *Journal of Econometrics*. AI3

Bottmer, Lea, Guido Imbens, Jann Spiess, and Merrill Warnick. "A Design-Based Perspective on Synthetic Control Methods." arXiv preprint arXiv:2101.09398 (2021). BISW

Doudchenko, Nikolay, and Guido W. Imbens. Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. No. w22791. NBER 2016. DI

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), IKM.

Imbens, Guido, Nathan Kallus, Xiaojie Mao, and Yuhao Wang. "Long-term causal inference under persistent confounding via data combination." arXiv preprint arXiv:2202.07234 (2022), IKMW.

Xiong, Ruoxuan, Susan Athey, Mohsen Bayati, and Guido W. Imbens. "Optimal experimental design for staggered rollouts." Available at SSRN 3483934 (2019). XABI

1. Introduction

We are interested in the causal effect of a binary treatment $w \in \{0, 1\}$ on a scalar outcome Y.

■ Observe multiple units (individuals, states, countries, firms) repeatedly over time (days, weeks, years), Y_{it}, $i = 1, \ldots, N$, $t = 1, \ldots, T$, **panel/longitudinal data**.

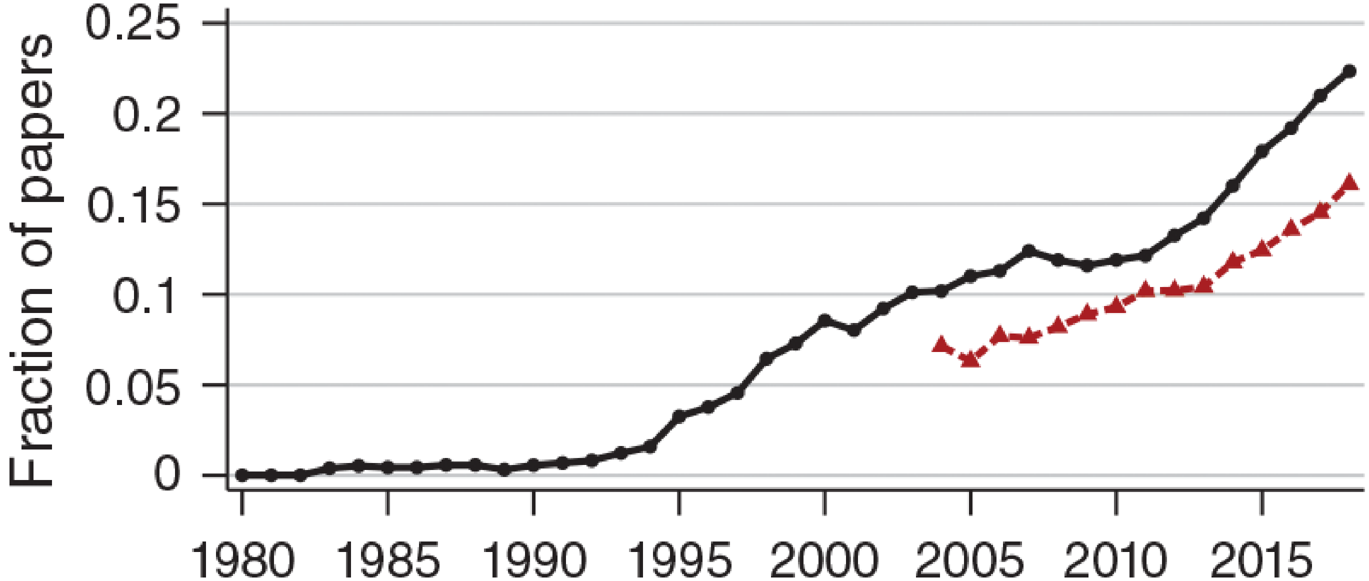
■ Possibly $N >> T$, $N << T$, $N \approx T$, N and T may be large or modest (so regularization may be a problem).

■ Binary treatment W_{it} may vary over time and across units.

Huge amount of empirical work in this area.

See documentation of prevalence of this setting in Bertrand, Duflo & Mullainathan (2004), De Chaisemartin d'Haultfoeuille (2020), Currie, Kleven & Zwiers (2020).

Panel A. Difference-in-differences



2.A General Set Up: we observe (in addition to covariates – not essential, ignored here):

$$\mathbf{Y}_{N\times T} = \begin{pmatrix} Y_{11} & Y_{12} & Y_{13} & \dots & Y_{1T} \\ Y_{21} & Y_{22} & Y_{23} & \dots & Y_{2T} \\ Y_{31} & Y_{32} & Y_{33} & \dots & Y_{3T} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Y_{N1} & Y_{N2} & Y_{N3} & \dots & Y_{NT} \end{pmatrix} \quad \text{(realized outcome).}$$

$$\mathbf{W}_{N\times T} = \begin{pmatrix} 1 & 1 & 0 & \dots & 1 \\ 0 & 0 & 1 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 1 & \dots & 0 \end{pmatrix} \quad \text{(binary treatment).}$$

- rows of $\mathbf{Y}$ and $\mathbf{W}$ correspond to units (individuals, states, firms, countries), $i = 1, \dots, N$, *ex ante* exchangeable,
- columns of $\mathbf{Y}$ and $\mathbf{W}$ correspond to time periods (days, weeks, years), $t = 1, \dots, T$, ordered.

For the time being, no dynamic treatment effects (see Bojinov, Rambachan, Shepard, 2020). Dynamics create interesting complications.

- Outcomes only depend on contemporaneous treatment, **not** on future or past treatments.

- **Implication:** Two potential outcomes for each unit/time-period: $Y_{it}(0)$, $Y_{it}(1)$.

- **no restrictions** on temporal correlation in $Y_{it}(0)$, outcomes in absence of treatment.

- **Observe** realized outcome

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

In terms of potential outcome matrices $\mathbf{Y}(0)$ and $\mathbf{Y}(1)$:

$$\mathbf{Y}(0) = \begin{pmatrix} ? & ? & \checkmark & \dots & ? \\ \checkmark & \checkmark & ? & \dots & \checkmark \\ ? & \checkmark & ? & \dots & \checkmark \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ ? & \checkmark & ? & \dots & \checkmark \end{pmatrix} \quad \mathbf{Y}(1) = \begin{pmatrix} \checkmark & \checkmark & ? & \dots & \checkmark \\ ? & ? & \checkmark & \dots & ? \\ \checkmark & ? & \checkmark & \dots & ? \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \checkmark & ? & \checkmark & \dots & ? \end{pmatrix}.$$

$Y_{it}(0)$ observed iff $W_{it} = 0$, $Y_{it}(1)$ observed iff $W_{it} = 1$.

In order to estimate the target, the **average treatment effect for the treated**,

$$\tau = \frac{\sum_{i,t} W_{it} \{Y_{it}(1) - Y_{it}(0)\}}{\sum_{it} W_{it}} = \frac{\sum_{i,t} W_{it} \{Y_{it} - Y_{it}(0)\}}{\sum_{it} W_{it}},$$

(or other average, *e.g.*, overall average effect) we need to **impute** the missing potential outcomes in $\mathbf{Y}(0)$ (and possibly in $\mathbf{Y}(1)$ for other estimands).

Examples

Example I: The Mariel Boatlift Study 44 controls, 1 treated, 6 pretreatment periods, 7 posttreatment periods

Example II: California Smoking Regulation 49 controls, 1 treated, 17 pretreatment periods, 13 posttreatment periods

Example III: German Re-unification 16 controls, 1 treated, 30 pretreatment periods, 14 posttreatment periods

Example IV: Minimum Wage Change 16 controls, 321 treated, 78 pretreatment periods, 1 posttreatment periods

2.B Common Shapes of Data Matrices: (relative size of N and T, important for relative merits of various methods)

$$\mathbf{Y}^{\text{square}} = \begin{pmatrix} ? & ? & \checkmark & \dots & ? \\ \checkmark & \checkmark & ? & \dots & \checkmark \\ ? & \checkmark & ? & \dots & \checkmark \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ ? & \checkmark & ? & \dots & \checkmark \end{pmatrix} \quad (N \approx T)$$

$$\mathbf{Y}^{\text{tall}} = \begin{pmatrix} ? & ? & \checkmark \\ \checkmark & \checkmark & ? \\ ? & \checkmark & \checkmark \\ \checkmark & \checkmark & ? \\ \checkmark & \checkmark & \checkmark \\ ? & \checkmark & ? \\ ? & \checkmark & ? \\ \vdots & \vdots & \vdots \\ ? & \checkmark & \checkmark \end{pmatrix} \quad (N >> T)$$

$$\mathbf{Y}^{\text{fat}} = \begin{pmatrix} ? & ? & \checkmark & \checkmark & ? & ? & \checkmark & \checkmark & \dots & ? \\ \checkmark & \checkmark & \checkmark & ? & ? & \checkmark & \checkmark & ? & \dots & \checkmark \\ ? & \checkmark & ? & \checkmark & \checkmark & ? & ? & \checkmark & \dots & \checkmark \\ \checkmark & ? & ? & \checkmark & \checkmark & ? & \checkmark & ? & \dots & \checkmark \end{pmatrix} \quad (N << T)$$

2.C Common Patterns of Treatment (issue raised in recent literature)

$$\mathbf{Y}^{\text{gen}} = \begin{pmatrix} ? & ? & \checkmark & \checkmark & \dots & ? \\ \checkmark & \checkmark & ? & \checkmark & \dots & \checkmark \\ ? & \checkmark & ? & \checkmark & \dots & \checkmark \\ ? & \checkmark & \checkmark & ? & \dots & \checkmark \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ ? & \checkmark & ? & \checkmark & \dots & \checkmark \end{pmatrix}$$

(general)

$$\mathbf{Y}^{\text{stag}} = \begin{pmatrix} \checkmark & \checkmark & \checkmark & \checkmark & \dots & \checkmark \\ \checkmark & \checkmark & \checkmark & \checkmark & \dots & ? \\ \checkmark & \checkmark & \checkmark & \checkmark & \dots & ? \\ \checkmark & \checkmark & \checkmark & ? & \dots & ? \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \checkmark & ? & ? & ? & \dots & ? \end{pmatrix}$$

(staggered adoption)

$$\mathbf{Y}^{\text{single}} = \begin{pmatrix} \checkmark & \checkmark & \checkmark & \checkmark & \dots & \checkmark \\ \checkmark & \checkmark & \checkmark & \checkmark & \dots & \checkmark \\ \checkmark & \checkmark & \checkmark & \checkmark & \dots & \checkmark \\ \checkmark & \checkmark & \checkmark & \checkmark & \dots & \checkmark \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \checkmark & \checkmark & \checkmark & ? & \dots & ? \end{pmatrix}$$

(single unit)

$$\mathbf{Y}^{\text{last}} = \begin{pmatrix} \checkmark & \checkmark & \checkmark & \checkmark & \dots & \checkmark \\ \checkmark & \checkmark & \checkmark & \checkmark & \dots & \checkmark \\ \checkmark & \checkmark & \checkmark & \checkmark & \dots & \checkmark \\ \checkmark & \checkmark & \checkmark & \checkmark & \dots & ? \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \checkmark & \checkmark & \checkmark & \checkmark & \dots & ? \end{pmatrix}$$

(last period)

$$\mathbf{Y}^{\text{block}} = \begin{pmatrix} \checkmark & \checkmark & \checkmark & \checkmark & \dots & \checkmark & \checkmark \\ \checkmark & \checkmark & \checkmark & \checkmark & \dots & \checkmark & \checkmark \\ \checkmark & \checkmark & \checkmark & \checkmark & \dots & \checkmark & \checkmark \\ \checkmark & \checkmark & \checkmark & ? & \dots & ? & ? \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \checkmark & \checkmark & \checkmark & ? & \dots & ? & ? \end{pmatrix}$$

(block)

$$\mathbf{Y}^{\text{single/last}} = \begin{pmatrix} \checkmark & \checkmark & \checkmark & \checkmark & \dots & \checkmark \\ \checkmark & \checkmark & \checkmark & \checkmark & \dots & \checkmark \\ \checkmark & \checkmark & \checkmark & \checkmark & \dots & \checkmark \\ \checkmark & \checkmark & \checkmark & \checkmark & \dots & \checkmark \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \checkmark & \checkmark & \checkmark & \checkmark & \dots & ? \end{pmatrix}$$

(single unit/ period)

2.D Special Cases I (unconfoundedness):

- $\mathbf{Y}^{\text{tall,last}}$, treatment only in last period, tall matrix ($N >> T$), many treated units, many control units. **Huge literature following Rosenbaum and Rubin (1983).** Key econ applications: Lalonde, 1986; Dehejia & Wahba, 1999.

- Methods: matching, inverse propensity score weighting, double robust methods. (for surveys, see Imbens, 2004; Rubin, 2006)

- Key Assumption **unconfoundedness**:

$$W_{iT} \perp\!\!\!\perp (Y_{iT}(0), Y_{iT}(1)) \;\Big|\; Y_{i1}, \ldots, Y_{iT-1}$$

Matching Estimator: for each treated unit i with $W_{i,T} = 1$, find control unit $j(i)$ with $W_{j(i),T} = 0$ such that $Y_{it} \approx Y_{j(i),t}$ for all pre-treatment periods $t = 1, \ldots, T-1$

$$\hat{\tau} = \frac{1}{N_T} \sum_{i:W_{iT}=1} \Big(Y_{iT} - Y_{j(i),T} \Big)$$

Regression Estimator: Estimate regression function

$$\mu_0(y_1, \ldots, y_{T-1}) = E[Y_{iT} | W_{iT} = 0, Y_{it} = y_t, \ldots, Y_{i,T-1} = y_{T-1}]$$

flexibly (eg using random forests, neural nets) and then estimate

$$\hat{\tau} = \frac{1}{N_T} \sum_{i:W_{iT}=1} \Big(Y_{iT} - \hat{\mu}_0(Y_{i1}, \ldots, Y_{i,T-1}) \Big)$$

or **double robust estimator**.

Simple linear regression version (**"horizontal regression"**):

$$\hat{\beta} = \arg\min \sum_{i:W_{iT}=0} \left(Y_{iT} - \beta_0 - \sum_{t=1}^{T-1} \beta_t Y_{it} \right)^2$$

$$\hat{\tau} = \frac{1}{N_T} \sum_{i:W_{iT}=1} \left(Y_{iT} - \hat{\beta}_0 - \sum_{t=1}^{T-1} \hat{\beta}_t Y_{it} \right)$$

- Note: N_C units in objective function, T regressors.

- Would need **regularization** if N_C is small relative to T.

- Suppose only a single treated unit, say unit N, then estimated treatment effect is linear combination of period T outcomes **and** linear combination of unit N outcomes:

$$\hat{\tau} = Y_{N,T} - \lambda_0^{\text{unconf}} - \sum_{i=1}^{N-1} \lambda_i^{\text{unconf}} Y_{iT} = Y_{N,T} - \delta_0^{\text{unconf}} - \sum_{t=1}^{T-1} \delta_t^{\text{unconf}} Y_{Nt}$$

Special Cases II (synthetic control):

■ $\mathbf{Y}^{\text{fat,single}}$, single (or few) treated unit, fat matrix ($T > N$), multiple treated periods, multiple pre-treatment periods. **Huge literature since Abadie and Gardeazabal (2003).**

Methods: synthetic control and modifications (Abadie, Diamond and Hainmueller, 2010, 2015; AAHIW; DI; Ben-Michael, Feller, Rothstein, 2018)

Simple linear regression version ("vertical regression"):

$$\hat{\gamma} = \arg\min \sum_{t:W_{Nt}=0} \left(Y_{Nt} - \gamma_0 - \sum_{i=1}^{N-1} \gamma_i Y_{i,t} \right)^2$$

- Note: T_{pre} units in objective function, N regressors.

- Abadie, Diamond and Hainmueller (2010) impose restrictions on γ: $\gamma_0 = 0$, $\gamma_i \geq 0$, $\sum_{i=1}^{N-1} \gamma_i = 1$. DI suggest relaxing those and adding regularization.

Given $\hat{\gamma}$:

$$\hat{\tau} = \frac{1}{T_1} \sum_{i:W_{Nt}=1} \left(Y_{Nt} - \hat{\gamma}_0 - \sum_{i=1}^{N-1} \hat{\gamma}_t Y_{it} \right)$$

■ Suppose only a single treated period, period T, then estimated treatment effect is linear combination of period T outcomes **and** linear combination of unit N outcomes:

$$\hat{\tau} = Y_{N,T} - \lambda_0^{\text{sc}} - \sum_{i=1}^{N-1} \lambda_i^{\text{sc}} Y_{iT} = Y_{N,T} - \delta_0^{\text{sc}} - \sum_{t=1}^{T-1} \delta_t^{\text{sc}} Y_{Nt}$$

■ The coefficients $\lambda_i^{\text{unconf}}$ and λ_i^{sc} are chosen very differently. (same for δ_t^{unconf} and δ_t^{sc})

■ How should we choose between **unconfoundedness** versus **synthetic control** approaches?

■ Unconfoundedness and SC exploit different correlation patterns in matrix.

Special Cases III (two-way-fixed-effect (twfe) estimation):

■ $\mathbf{Y}^{\text{square,block}}$, square matrix ($T \approx N$), block of treated units and periods.

Methods: two-way-fixed-effect regression (**TWFE**), (*e.g.*, Bertrand, Duflo, Mullainathan, 2004)

$$(\hat{\mu}, \hat{\alpha}, \hat{\beta}, \hat{\tau}) = \arg\min \sum_{i=1}^{N} \sum_{t=1}^{T} (Y_{it} - \mu - \alpha_i - \beta_t - \tau W_{it})^2$$

3. General Comments / Themes of this Talk

1. Lots of combinations of shapes and assignment patterns in $\mathbf{W}$ not covered well, and deliniations between cases not clear.

2. TWFE does not work well with heterogenous treatment effects outside of block assignment case. (Callaway & St'Anna, 2020; Goodman-Bacon 2017; Sun & Abraham, 2020; De Chaisemartin & d'Haultfoeuille, 2020; Borusyak & Jaravel, 2017).

3. **Estimation** with both N and T modest is challenging because regularization is difficult (AAHIW, AI3)

4. **Inference** with both N and T modest is challenging (AAHIW, AI3; Rambachan & Roth, 2020; Chernozhukov, Wuthrich & Zhu, 2019)

5. There may be **both** correlation patterns over time, similar for all units (exploited by horizontal/unconfoundeness regression) **and** correlation patterns across units, stable over time (exploited by vertical/SC regression). Horizontal or vertical regression **cannot** exploit both.

6. TWFE can be an attractive model for baseline (control) outcome when N and T are modest, but may be **too restrictive when N and T are large**.

4. Today: Four Main themes

1. More **flexible outcome model** than TWFE, *e.g.,* factor models, matrix completion, ABDIK, Amjad, Shah & Shen (2018)

2. Estimate TWFE **locally**. (weight towards treated units/periods), AAHIW, Ben-Michael, Feller & Rothstein (2018)

3. Design-based approaches with explicit **models for assignment mechanism** instead of, or in addition to, models for outcomes **(double robust estimators)**. AI1, AI2, AI3, AILL, BISW

4. Combining Short Term Experimental and Long Term Observational Data. Experiments may contain information on short term outcomes with high degree of internal validity. Observational data may contain detailed information on long term outcomes, with low degree of internal validity. ACI, ACIK, IKM, IKMW.

4.A. Factor models: more flexible model than TWFE:

$$\mathbf{Y}(0) = \mathbf{L} + \mathbf{E} \quad \mathbf{E} \perp\!\!\!\perp \mathbf{L}, \quad \mathbf{L} \text{ low rank matrix}$$

Regularization for $\mathbf{L}$ is important. Fixing rank is too restrictive. Use singular value decomposition:

$$\mathbf{L}_{N\times T} = \mathbf{S}_{N\times N}\mathbf{\Sigma}_{N\times T}\mathbf{R}_{T\times T}$$

singular values $\sigma_i(\mathbf{L})$ are eigenvalues of $\mathbf{\Sigma}$.

Estimator based on matrix completion (Candes & Plan, 2010)

$$\min_{\mathbf{L},\alpha,\beta} \frac{1}{\sum_{i,t}(1-W_{it})} \sum_{(i,t):W_{it}=0} (Y_{it} - \alpha_i - \beta_t - L_{it})^2 + \lambda_L \|\mathbf{L}\|_*$$

$$\|\mathbf{L}\|_* = \sum_{j=1}^{\min(N,T)} \sigma_i(\mathbf{L}) \quad \text{(nuclear norm, like LASSO)}$$

leads to low rank for $\hat{\mathbf{L}}$.

■ **Note**: regularize only $\mathbf{L}$ but not fixed effects.

Illustration: Stock Market Return Data

We use daily returns for 2453 stocks over 10 years (3082 days). We create sub-samples by looking at the first T daily returns of N randomly sampled stocks for pairs of (N, T) such that $N \times T = 4900$, ranging from fat to thin: $(N, T) = (10, 490), \ldots, (70, 70), \ldots, (490, 10)$.

$$\mathbf{Y} = \begin{pmatrix} \checkmark & \checkmark & \checkmark & \checkmark & \checkmark & \checkmark & \checkmark & \checkmark \\ \checkmark & \checkmark & \checkmark & \checkmark & \checkmark & \checkmark & \checkmark & \checkmark \\ \checkmark & \checkmark & \checkmark & \checkmark & \checkmark & \checkmark & ? & ? \\ \checkmark & \checkmark & \checkmark & \checkmark & \checkmark & \checkmark & ? & ? \end{pmatrix}$$

(fat) favorable for **vertical/SC** regression

$$\mathbf{Y} = \begin{pmatrix} \checkmark & \checkmark & \checkmark & \checkmark & \checkmark & \checkmark & \checkmark \\ \checkmark & \checkmark & \checkmark & \checkmark & \checkmark & \checkmark & \checkmark \\ \checkmark & \checkmark & \checkmark & \checkmark & \checkmark & \checkmark & \checkmark \\ \checkmark & \checkmark & \checkmark & \checkmark & \checkmark & \checkmark & \checkmark \\ \checkmark & \checkmark & \checkmark & \checkmark & ? & ? & ? \\ \checkmark & \checkmark & \checkmark & \checkmark & ? & ? & ? \\ \checkmark & \checkmark & \checkmark & \checkmark & ? & ? & ? \end{pmatrix}$$

(square) favorable for low rank regression

$$\mathbf{Y} = \begin{pmatrix} \checkmark & \checkmark & \checkmark & \checkmark \\ \checkmark & \checkmark & \checkmark & \checkmark \\ \checkmark & \checkmark & \checkmark & \checkmark \\ \checkmark & \checkmark & \checkmark & \checkmark \\ \checkmark & \checkmark & \checkmark & \checkmark \\ \checkmark & \checkmark & \checkmark & \checkmark \\ \checkmark & \checkmark & ? & ? \\ \checkmark & \checkmark & ? & ? \end{pmatrix}$$

(thin) favorable for **horizontal/unconfoundedness** regression

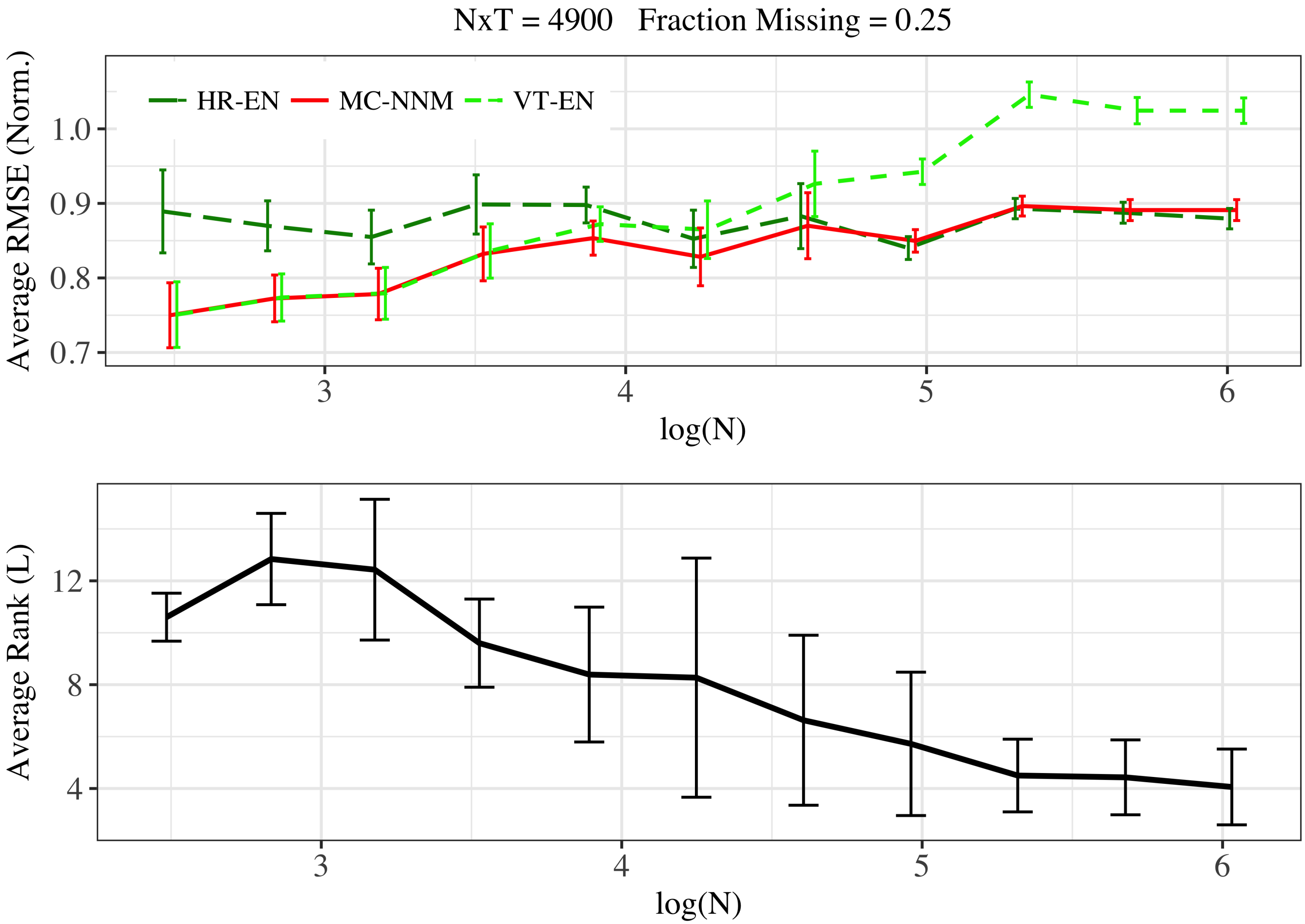
We compare **HR-EN** (horizontal, unconfoundedness regression) and **VT-EN** (vertical, synthetic control regression), both with elastic net regularization, and **MC-NNM** (matrix completion with nuclear norm regularization)

Results

◇ Vertical regression (VT-EN) does much better than horizontal regression (HR-EN) when $T_0 >> N_0$ (VT-EN is regression with T_0 observations and N_0 regressors).

◇ Horizontal regression (HR-EN) does much better than vertical regression (VT-EN) when $N_0 >> T_0$ (HR-EN is regression with N_0 observations and T_0 regressors).

◇ **MC-NNM does as well as or better than HR-EN and VT-EN in all cases, adapts to shape of matrix**


NxT = 4900 Fraction Missing = 0.25
HR-EN
MC-NNM
VT-EN
Average RMSE (Norm.)
1.0
0.9
0.8
0.7
3
4
5
6
log(N)
Average Rank (L)
12
8
4
3
4
5
6
log(N)

4.B. Local TWFE Estimation: Observation I

Synthetic Control is weighted linear regression without unit fixed effects:

$$(\widehat{\tau}^{\text{SC}}, \widehat{\beta}) = \arg\min_{\tau,\beta} \sum_{i=1}^{N} \sum_{t=1}^{T} (Y_{it} - \beta_t - \tau W_{it})^2 \times \omega_i^{\text{SC}}$$

- regression with time fixed effects and SC weights.
- Why no unit fixed effects α_i in the synthetic control regression specification?
- Why no time weights?

Observation II

TWFE is unweighted regression with unit and time fixed effects:

$$\hat{\tau}^{\text{TWFE}} = \arg\min_{\tau,\beta,\alpha} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(Y_{it} - \alpha_i - \beta_t - \tau W_{it}\right)^2$$

- regression with time fixed effects and unit fixed effects, no weights.

- Why no unit weights ω_i in the TWFE regression specification?

- Why no time weights?

Synthetic Difference In Differences (AAHIW)

$$\hat{\tau}^{\text{SDID}} = \arg\min_{\tau,\gamma,\alpha} \sum_{i=1}^{N} \sum_{t=1}^{T} (Y_{it} - \alpha_i - \beta_t - \tau W_{it})^2 \times \omega_i^{\text{SC}} \times \lambda_t^{\text{SC}}$$

Regression with unit and time fixed effects, and with unit and time weights.

■ this is a TWFE model estimated **"locally,"** dropping distant units and distant time periods.

SDID has

◇ unit fixed effects (like TWFE, unlike SC)

◇ unit weights (like SC, unlike TWFE)

◇ time weights (unlike SC, unlike TWFE)

Time weights satisfy:

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

subject to

$$\lambda_t \geq 0, \quad \sum_{t=1}^{T-1} \lambda_t = 1.$$

(or down-weight observations from distant past.)

CPS-motivated Simulations: based on wages for women in the March outgoing rotation groups in the Current Population Survey (CPS) for the years 1979 to 2019 (as in Bertrand, Duflo, Mullainathan, 2004).

We simulate outcome data using the model

$$\mathbf{Y} = \mathbf{L} + \mathbf{F} + \tau\mathbf{W} + \mathbf{E}$$

where the rows ε_i of $\mathbf{E}$ have a multivariate Gaussian distribution $\varepsilon_i \sim \mathcal{N}(0, \mathbf{\Sigma})$,
we choose
$\mathbf{L}$ (rank four matrix),
$\mathbf{F}$ (fixed effect matrix with $F_{it} = \alpha_i + \beta_t$)
and Σ to fit the CPS data.

First, we fit

$$\mathbf{M} = \text{arg} \min_{\mathbf{M}:\text{rank}(\mathbf{M})=4} \sum_{i,t} (Y_{it} - M_{it})^2,$$

We then estimate $\mathbf{\Sigma}$ by fitting an AR(2) model to the residuals of $Y_{it} - M_{it}$. We decompose the systematic component $\mathbf{M}$ into an additive (fixed effects) term $\mathbf{F}$ and an factor component $\mathbf{L}$, with

$$F_{it} = \alpha_i + \beta_t$$

$$\alpha_i = \frac{1}{T} \sum_{s=1}^{T} M_{is} - \frac{1}{NT} \sum_{it} M_{it} \quad \beta_t = \frac{1}{N} \sum_{j=1}^{N} M_{jt} - \frac{1}{NT} \sum_{it} M_{it}$$

$$L_{it} = M_{it} - F_{it}$$

■ This decomposition of $\mathbf{M}$ into an additive two-way fixed effect component $\mathbf{F}$ and an interactive component $\mathbf{L}$ enables us to study the sensitivity of different estimators to the presence of different types of systematic effects.

For the treatment we choose the treated units so that the assignment mechanism is correlated with the systematic components $\mathbf{L}$ and $\mathbf{F}$.

We set $W_{it} = D_i \mathbf{1}_{t>T_0}$, where D_i is a binary exposure indicator generated as

$$D_i \Big| \varepsilon_i, \alpha_i, \mathbf{L}_i \sim \text{Bernoulli}\left(\frac{\exp(\phi_\alpha \alpha_i + \phi_L \mathbf{L}_i)}{1 + \exp(\phi_\alpha \alpha_i + \phi_L \mathbf{L}_i)}\right)$$

We choose ϕ as the coefficient estimates from a logistic regression of an observed binary characteristic of the state D_i on $\mathbf{L}_i$ and α_i. We consider three different choices for D_i, relating to minimum wage laws, abortion rights, and gun control laws.

	$\frac{\|\mathbf{F}\|_F}{\sqrt{nT}}$	$\frac{\|\mathbf{L}\|_F}{\sqrt{nT}}$	$\sqrt{\|\|\Sigma\|\|}$	AR(2)	RMSE SDID	RMSE SC	RMSE DID	RMSE MC	Bias SDID	Bias SC	Bias DID	Bias MC
Baseline	0.990	0.100	0.080	(.01,-.06)	**0.027**	0.046	0.049	0.035	**0.008**	0.032	0.022	0.016
Restricted Outcome Model												
No Corr	0.990	0.100	0.080	(.00,.00)	**0.028**	0.046	0.049	0.035	0.008	0.031	0.021	0.015
No low rank	0.990	0.000	0.080	(.00,.00)	0.016	0.019	**0.015**	**0.015**	-0.001	0.007	-0.001	-0.001
No fixed effects	0.000	0.100	0.080	(.00,.00)	0.028	**0.022**	0.049	0.035	0.007	0.006	0.021	0.015
Only Noise	0.000	0.000	0.080	(.00,.00)	0.016	**0.013**	0.015	0.015	-0.001	-0.000	-0.001	-0.001
No Noise	0.990	0.100	0.000	(.00,.00)	**0.003**	0.026	0.048	0.004	0.002	0.010	0.022	0.000
Assignment Process												
Gun Law	0.990	0.100	0.080	(.01,-.06)	**0.026**	**0.026**	0.046	0.036	0.009	-0.006	0.015	0.016
Abortion	0.990	0.100	0.080	(.01,-.06)	**0.024**	0.039	0.045	0.031	0.003	0.028	0.008	0.005
Random	0.990	0.100	0.080	(.01,-.06)	**0.023**	0.026	0.044	0.031	-0.001	-0.004	-0.003	-0.003
Outcome Variable												
Hours	0.790	0.400	0.460	(.06,.00)	0.189	0.205	0.201	**0.182**	0.110	-0.098	0.087	0.102
U-rate	0.750	0.440	0.490	(-.02,-.01)	**0.172**	0.187	0.330	0.232	0.070	0.116	0.286	0.176
Assignment Block Size												
$T_{\text{post}} = 1$	0.990	0.100	0.080	(.01,-.06)	**0.047**	0.055	0.068	0.048	0.012	0.023	0.037	0.019
$N_{\text{T}} = 1$	0.990	0.100	0.080	(.01,-.06)	**0.069**	0.074	0.138	0.090	0.002	0.018	0.018	0.008
$T_{\text{post}} = N_{\text{T}} = 1$	0.990	0.100	0.080	(.01,-.06)	0.118	0.125	0.167	**0.113**	0.004	0.013	0.024	0.004

- SDID does better than SC (because of presence of fixed effects $\mathbf{F}$) and better than DID (because of presence of low rank component $\mathbf{L}$) and better than MC-NNM (because of double robustness).

- Improvement of SDID over SC is weaker with random assignment.

4.C Experimental Design Questions, BISW, XABI

- Suppose we have a randomized experiment.

  ◇ Should we estimate the causal effect as the difference in means of treated and means of controls, or using SC type methods?

  ◇ Does SC have guarantees under randomization?

- SC methods for analysis can have much lower variance than difference in means.

- SC is generally biased, but bias can be removed by including linear restriction on weights.

Simulation Experiment Based on CPS. Randomly choose single state for (pseudo) treatment in single (last) year. Outcome is average log wage by state and year. $N = 50$ states, $T = 40$ years.

	Diff in Means	SC	Modified Unbiased SC
Bias	0	-0.007	0
rmse	0.105	0.051	0.048
Ave se	0.105	0.051	0.048

- If we design an experiment where we can choose which units are treated, given data on past outcomes, how should we choose the units treated, and how should we analyze the data.

- Choosing units carefully can be very useful depending on the target:

  - ◇ average effect for treated units only (choose units for treatment in center of convex hull),

  - ◇ average effect for all units (choose units for treatment that span the space)

- If we can design an experiment where we can choose which units are treated, and when they are treated, given outcomes on other units, how should we choose units and periods to be treated?

Start with few units treated, and slowly increase number of treated units. (XABI)

4.D Combining Short Term Experiments and Long Term Observational Data, ACIK, ACI, IKM, IKMW

How can we systematically combine experimental and observational data to answer questions that neither can directly answer?

• Methods for doing so will make experiments more valuable by extending the value beyond the narrow questions they were intended for.

• Methods for doing so will make observational studies more credible by grounding them using experimental data.

Two Settings:

■ I. Surrogacy

◇ Observational sample contains information on early and long term outcomes (but not on treatment).

◇ Experimental data contains information on treatment and early outcomes (but not on long term outcomes).

ACIK

■ II. Observational Data with Limited Internal Validity:

◇ Observational sample contains information on treatment, early and late outcomes.

◇ Experimental data contains information on treatment and early outcomes (but not on long term outcomes).

ACI, IKM, IKMW

6. Conclusion

◇ Lots of new developments in causal panel data models.

◇ Improved understanding of two-way-fixed-effect models.

◇ Design approaches to estimation and experimental design.

◇ Double robust approaches.

◇ Lots more to be done.

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* 105.490 (2010): 493-505.

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. ”Comparative politics and the synthetic control method.” *American Journal of Political Science* 59.2 (2015): 495-510.

Altonji, Joseph G., and Rosa L. Matzkin. ”Cross section and panel data estimators for nonseparable models with endogenous regressors.” *Econometrica* 73, no. 4 (2005): 1053-1102.

Amjad, Muhammad, Devavrat Shah, and Dennis Shen. ”Robust synthetic control.” *The Journal of Machine Learning Research* 19, no. 1 (2018): 802-852.

Arkhangelsky, Dmitry, Susan Athey, David A. Hirshberg, Guido W. Imbens, and Stefan Wager. Synthetic difference in differences, 2019.

Arkhangelsky, Dmitry, and Guido Imbens. The role of the propensity score in fixed effect models. No. w24814. National Bureau of Economic Research, 2018.

Arkhangelsky, D., & Imbens, G. W. (2021). Double-robust identification for causal panel data models (No. w28364). National Bureau of Economic Research.

Athey, Susan, Mohsen Bayati, Mohsen, Nick Doudchenko, Guido Imbens, and Khashayar Khosravi, (2018). Matrix completion methods for causal panel data models, forthcoming *Journal of the American Statistical Association*.

Athey, Susan, Raj Chetty, and Guido Imbens. ”Combining experimental and observational data to estimate treatment effects on long term outcomes.” arXiv preprint arXiv:2006.09676 (2020), ACI.

Athey, Susan, Raj Chetty, Guido Imbens, and Hyunseung Kang. "Estimating treatment effects using multiple surrogates: The role of the surrogate score and the surrogate index." arXiv preprint arXiv:1603.09326 (2016), ACIK.

Athey, Susan, and Guido W. Imbens. Design-based analysis in difference-in-differences settings with staggered adoption, forthcoming *Journal of Econometrics*

Athey, Susan, and Scott Stern. An empirical framework for testing theories about complimentarity in organizational design. No. w6600. National Bureau of Economic Research, 1998.

Baker, Andrew, David F. Larcker, and Charles CY Wang. "How Much Should We Trust Staggered Difference-In-Differences Estimates?." Available at SSRN 3794018 (2021).

Ben-Michael, Eli, Avi Feller, and Jesse Rothstein. "The Augmented Synthetic Control Method." forthcoming *Journal of the American Statistical Association*.

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.

Bojinov, Iavor, Ashesh Rambachan, and Neil Shephard. "Panel experiments and dynamic causal effects: A finite population perspective." arXiv preprint arXiv:2003.09915 (2020).

Borusyak, Kirill, and Xavier Jaravel. "Revisiting event study designs." Available at SSRN 2826228 (2017).

Bottmer, Lea, Guido Imbens, Jann Spiess, and Merrill Warnick. "A Design-Based Perspective on Synthetic Control Methods." arXiv preprint arXiv:2101.09398 (2021).

Callaway, Brantly, and Pedro HC Sant'Anna. "Difference-in-differences with multiple time periods." *Journal of Econometrics* (2020).

Candes, Emmanuel J., and Yaniv Plan. "Matrix completion with noise." *Proceedings of the IEEE* 98, no. 6 (2010): 925-936.

De Chaisemartin, Clément, and Xavier d'Haultfoeuille. "Two-way fixed effects estimators with heterogeneous treatment effects." *American Economic Review* 110, no. 9 (2020): 2964-96.

Currie, Janet, Henrik Kleven, and Esmée Zwiers. "Technology and big data are changing economics: Mining text to track methods." *AEA Papers and Proceedings*, vol. 110, pp. 42-48. 2020.

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.

Freyaldenhoven, Simon, Christian Hansen, and Jesse M. Shapiro. "Pre-event trends in the panel event-study design." *American Economic Review* 109, no. 9 (2019): 3307-38.

Goodman-Bacon, Andrew. Difference-in-differences with variation in treatment timing. No. w25018. National Bureau of Economic Research, 2018.

Imbens, Guido W. "The role of the propensity score in estimating dose-response functions." *Biometrika* 87, no. 3 (2000): 706-710.

Imbens, Guido W. "Nonparametric estimation of average treatment effects under exogeneity: A review." *Review of Economics and statistics* 86, no. 1 (2004): 4-29.

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), IKM.

Imbens, Guido, Nathan Kallus, Xiaojie Mao, and Yuhao Wang. "Long-term causal inference under persistent confounding via data combination." arXiv preprint arXiv:2202.07234 (2022), IKMW.

Imbens, Guido W., and Donald B. Rubin. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press, 2015.

Rambachan, Ashesh, and Jonathan Roth. "An honest approach to parallel trends." Unpublished manuscript, Harvard University. (2019).

Rosenbaum, Paul R., and Donald B. Rubin. "The central role of the propensity score in observational studies for causal effects." *Biometrika* 70, no. 1 (1983): 41-55.

Roth, Jonathan, and Pedro HC Sant'Anna. "When Is Parallel Trends Sensitive to Functional Form?." arXiv preprint arXiv:2010.04814 (2020).

Rubin, Donald B. *Matched sampling for causal effects*. Cambridge University Press, 2006.

Sun, Liyang, and Sarah Abraham. "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." *Journal of Econometrics* (2020).

R Packages available from Abadie-Diamond-Hainmueller,
Athey-Bayati-Doudchenko-Imbens-Khosravi,
Arkhangelsky-Athey-Hirshberg-Imbens-Wager.