Doubly Robust Proximal Synthetic Controls

Hongxiang (David) Qiu

Department of Statistics, the Wharton School, University of Pennsylvania

Table of Contents

Motivation of synthetic controls: causal inference with panel data

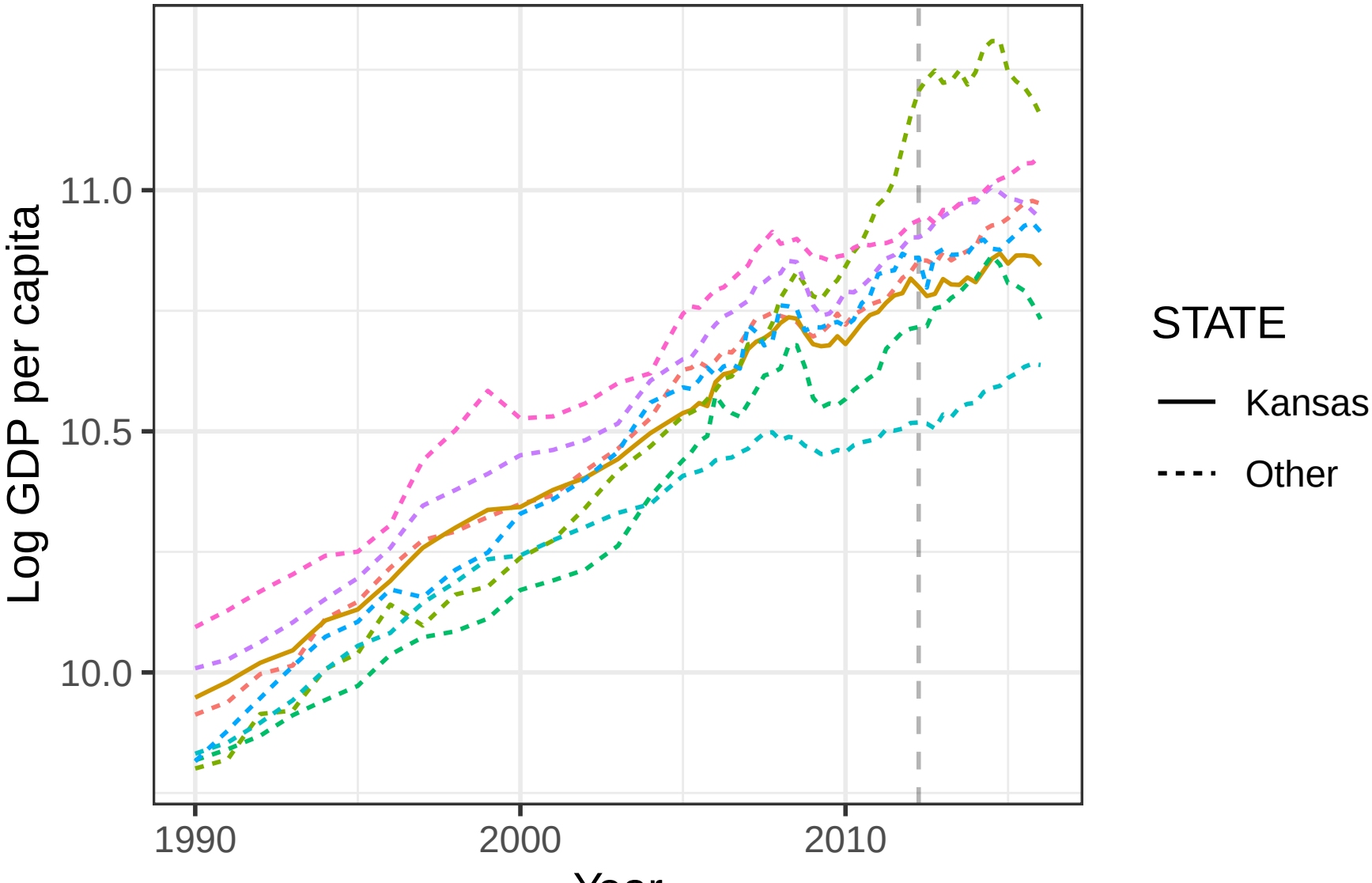
A common scenario:

- Intervention on a single unit (e.g., country, state, hospital, etc.)
- Observe time series data of the treated unit and some untreated units
- How to estimate the causal effect of this intervention?

Example:

- An aggressive tax cut in the state of Kansas U.S. in Q1 2012
- What is the effect of this intervention on GDP in Kansas?

Motivation of synthetic controls: causal inference with panel data



Motivation of synthetic controls: causal inference with panel data

Notable challenges compared to iid setting:

- Lack of randomization in treatment assignment
  - among units
  - across time periods
- Serial correlation
  - within units
  - potentially across units

Idea behind classical synthetic controls

Some notations:

- Total number of time periods: T
- Intervention time: T_0
- Unit index: treated$= 0$; control$= 1, \ldots, N$
- Outcome of unit i at time t: $Y_{t,i}$
- Counterfactual outcome of treated unit corresponding to treatment and control: $Y_{t,0}(1)$ and $Y_{t,0}(0)$
- Causal estimand (ATT): $\phi^*(t) := \mathbb{E}[Y_{t,0}(1) - Y_{t,0}(0)]$ at $t > T_0$
- Main challenge: learn about $Y_{t,0}(0)$ for $t > T_0$

Idea behind Abadie's classical synthetic controls

Intuition:

- Impute $Y_{t,0}(0)$ with control units' contemporary outcomes $Y_{t,i}$
- Consider this linear latent factor model [Abadie and Gardeazabal, 2003, Abadie et al., 2010, 2015]

$$Y_{t,0}(0) = U_t^\top \alpha_0 + \epsilon_{t,0}$$
$$Y_{t,i} = U_t^\top \alpha_i + \epsilon_{t,i}$$

  U_t: latent time-varying factor (confounder)

  α_i: unit-specific coefficient

  $\epsilon_{t,i}$: exogenous zero-mean random noise

- Under this model, $\mathbb{E}_\epsilon[Y_{t,0}(0)] = \sum_{i=1}^N w_i \mathbb{E}_\epsilon[Y_{t,i}]$ for weights w_i such that $\alpha_0 = \sum_{i=1}^N w_i \alpha_i$.

Abadie's synthetic controls

- Find the weights by fitting treated unit's pre-treatment trajectory:

$$\hat{w} = \underset{w}{\operatorname{argmin}} \sum_{t=1}^{T_0} \left(Y_{t,0} - \underbrace{\sum_{i=1}^{N} w_i Y_{t,i}}_{\text{synthetic cnotrol}} \right)^2$$

  (originally with constraint $w_i \geq 0, \sum_{i=1}^{N} w_i = 1$)

- Estimate the ATT $\phi^*(t)$ with $Y_{t,0} - \sum_{i=1}^{N} \hat{w}_i Y_{t,i}$ $(t > T_0)$

Proximal synthetic controls

- Many other ways to form a synthetic control have been proposed, but most still assume a linear model.
- A notable exception: based on proximal causal inference, Shi et al. [2021] proposed a method allowing for nonlinear models

What is proximal causal inference in the iid setting?

Proximal synthetic controls

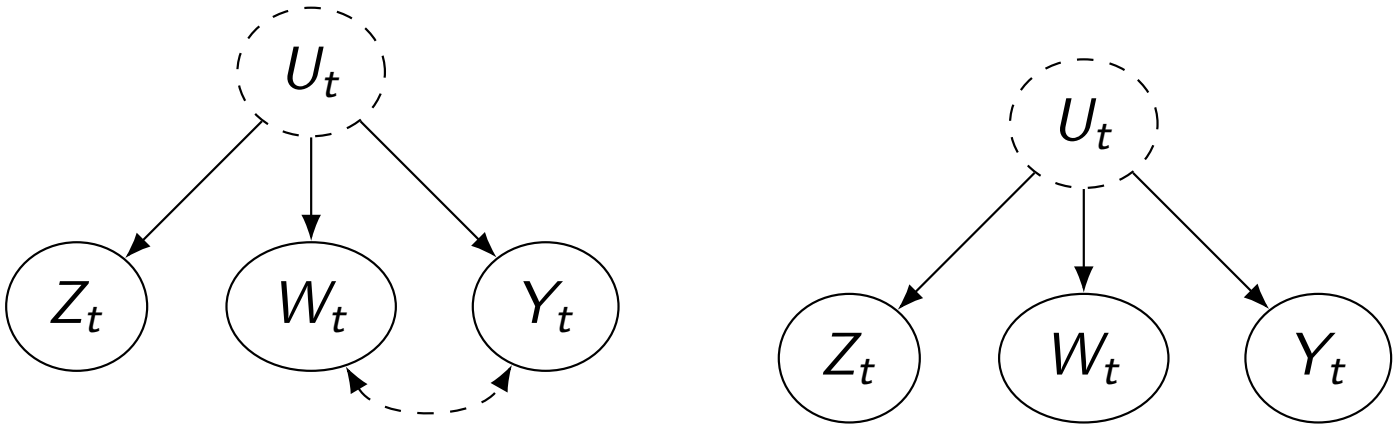
- Many other ways to form a synthetic control have been proposed, but most still assume a linear model.
- A notable exception: based on proximal causal inference, Shi et al. [2021] proposed a method allowing for nonlinear models

What is proximal causal inference in the iid setting?

- Some degree of unmeasured confounding allowed
- Provided two proxies of unmeasured confounder are observed
- One proxy can be related to treatment; the other can be related to outcome
- How are these related to synthetic controls?

Proximal synthetic controls

- Split control units into two groups: donors (outcomes denoted by W) and non-donor control units (outcomes denoted by Z)[1]
- W defines set of proxies to model $Y(0)$
- Z defines set of proxies to identify representation of $Y(0)$ based on W
- Key assumption 1: $Z_t \perp\!\!\!\perp (Y_t, W_t) \mid U_t$ (implied by linear latent factor model)



[1] From now on, I use Y to denote treated unit's outcome

Proximal synthetic controls

- Key assumption 2: there exists an *outcome confounding bridge function* h^* such that $\mathbb{E}[Y_t(0) \mid U_t] = \mathbb{E}[h^*(W_t) \mid U_t]$.
- h^* is linear if we assume a linear latent factor model
- Shi et al. [2021] showed that
  1. $\phi^*(t) := \mathbb{E}[Y_t(1) - Y_t(0)] = \mathbb{E}[Y_t - h^*(W_t)]$ for $t > T_0$;
  2. h^* satisfies $\mathbb{E}[Y_t - h^*(W_t) \mid Z_t] = 0$ for $t \leq T_0$.[2]
- Estimation based on generalized method of moments (GMM).
- Key contribution: h^* can be flexibly modeled and need not be linear.
- However, h^* must be correctly specified.

[2] h^* is the unique solution under a completeness condition

Table of Contents

Intuition: connect to "usual ATT"

Consider this (over) simplification to the setting of iid "individuals":

- Regard each time t (not unit i!!!) as the index for "individuals"
- At time t, regard control units' outcomes as covariates/proxies for "individual" t
- $A_t := \mathbb{1}(t > T_0)$ is treatment indicator for "individual" t
- Suppose that individuals are iid (so $\phi^*(t) = \phi^*$ is constant)

Intuition: connect to "usual ATT"

Consider this (over) simplification to the setting of iid "individuals":

- Regard each time t (not unit i!!!) as the index for "individuals"
- At time t, regard control units' outcomes as covariates/proxies for "individual" t
- $A_t := \mathbb{1}(t > T_0)$ is treatment indicator for "individual" t
- Suppose that individuals are iid (so $\phi^*(t) = \phi^*$ is constant)

Under these simplifications, $\phi^*(t)$ is the "usual ATT" in iid settings.

Intuition: connect to "usual ATT"

Cui et al. [2020] showed that the influence function of the "usual ATT" is

$$\frac{A_t Y_t}{\Pr(A_t = 1)} - (1 - A_t) q^*(Z_t) \frac{Y_t - h^*(W_t)}{\Pr(A_t = 1)} - A_t \frac{h^*(W_t) - \phi^*}{\Pr(A_t = 1)}.$$

- h^* defined as in Shi et al. [2021]
- q^* is a *treatment confounding bridge function* that captures the weight for treatment assignment:

$$\mathbb{E}[q^*(Z_t) \mid U_t, A_t = 0] = \frac{\Pr(A_t = 1 \mid U_t)}{\Pr(A_t = 0 \mid U_t)}.$$

Gaps between iid setting and panel data setting

- Data are not iid.
- $A_t = \mathbb{1}(t > T_0)$ is not random, so $\Pr(A_t = 1)$ and their definition of q^* are not meaningful.

I will use t_- (t_+) to denote a general pre-(post-)treatment time

Our solution

We need some assumptions similar to iid

- $(Y_t(0), W_t) \mid U_t$ is identically distributed for all t (implied by linear latent factor model).
- U_{t_+} is identically distributed for all t_+.[3]

[3] Can be relaxed

Our solution

We need some assumptions similar to iid

- $(Y_t(0), W_t) \mid U_t$ is identically distributed for all t (implied by linear latent factor model).
- U_{t_+} is identically distributed for all t_+.[3]

We need to define q^* while avoiding introducing A_t as a random variable:

- Assume that there exists q^* that captures a likelihood ratio:

$$\mathbb{E}[q^*(Z_{t_-}) \mid U_{t_-} = u] = \frac{\mathrm{d}P_{U_{t_+}}}{\mathrm{d}P_{U_{t_-}}}(u).$$

[3] Can be relaxed

Novel identification results

Theorem (Weighting identification)

$$\phi^*(t_+) = \mathbb{E}[Y_{t_+} - q^*(Z_{t_-})Y_{t_-}]$$

and q^* *satisfies*

$$\mathbb{E}[q^*(Z_{t_-}) \mid W_{t_-} = w] = \frac{\mathrm{d}P_{W_{t_+}}}{\mathrm{d}P_{W_{t_-}}}(w).$$

An implicit implication: distribution of W_{t_+} should be dominated by W_{t_-}.

Novel identification results

Theorem (Doubly robust identification)

$$\phi^*(t_+) = \mathbb{E}[Y_{t_+} - q(Z_{t_-})(Y_{t_-} - h(W_{t_-})) - h(W_{t_+})]$$

if $h = h^*$ *or* $q = q^*$.

Only need to correctly specify one of h^* and q^*.

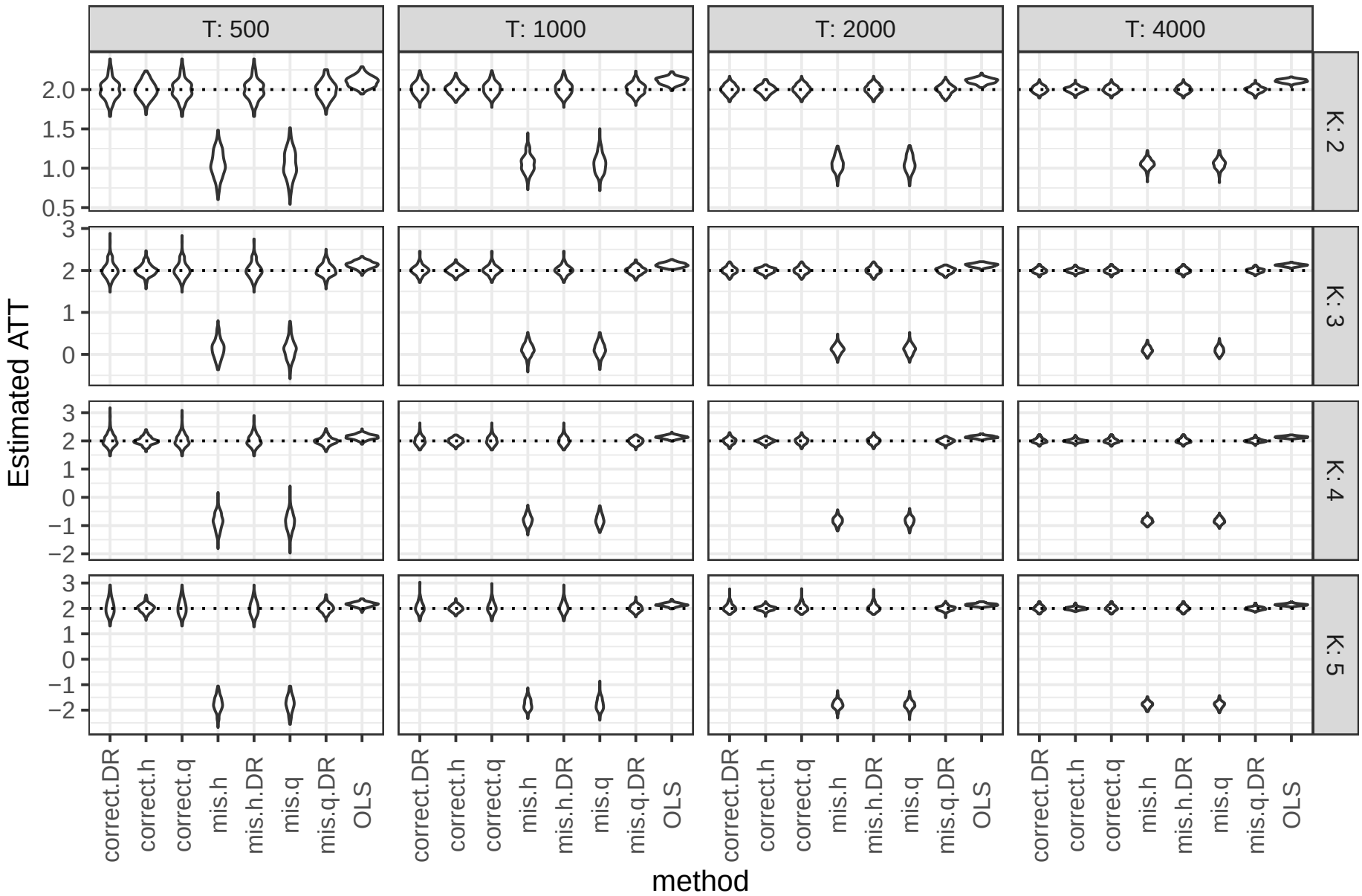
Doubly robust estimation and inference based on generalized method of moments (GMM).

Simulation

Methods compared:

- OLS + Proximal synthetic control methods based on h^* only, q^* only, and both h^* and q^*
- Consider cases where
  - both h^* and q^* are correctly specified
  - h^* or q^* is misspecified

Simulation: sampling distribution



Kansas data analysis

- Quarterly data of 50 U.S. sates from 1990–2016 (105 quarters)
- Remove time trend: fit a quadratic curve of time to control states' outcomes and take residuals for all states
- Time trend removal is important to make covariate shift assumption plausible
- Choice of donors W: we run Abadie's original synthetic control method and choose states with large weights: North Dakota, South Carolina, Texas, Washington
- Linear model for h^*
- Log-linear model for q^*: to restrict model complexity, only a subset of non-donor control states are included in the model for q^* (chosen based on similarity to Kansas):
  1. Iowa
  2. Iowa, South Dakota
  3. Iowa, South Dakota, Oklahoma

Kansas data analysis

Point estimate (95% confidence interval)

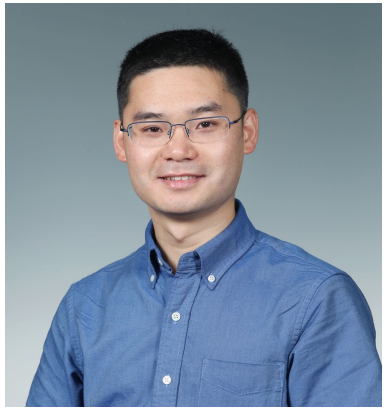
Method	tax cut (Q1 2012)	placebo (Q1 2008)
`Abadie's SC`	-0.048	0.029
`OLS`	-0.069 (-0.087, -0.050)	0.026 (2.6×10^{-6}, 0.052)
`DR`	-0.077 (-0.126, -0.028)	0.004 (-0.068, 0.077)
`DR2`	-0.095 (-0.147, -0.043)	-0.005 (-0.039, 0.030)
`DR3`	-0.103 (-0.228, -0.021)	-0.007 (-0.059, 0.046)
`Outcome bridge`	-0.104 (-0.150, -0.058)	0.012 (-0.069, 0.093)
`Treatment bridge`	-0.031 (-0.087, 0.024)	-0.028 (-0.063, 0.008)
`Treatment bridge2`	-0.017 (-0.032, -0.002)	-0.042 (-0.056, -0.0027)
`Treatment bridge3`	-0.016 (-0.029, -0.003)	-0.048 (-0.097, 0.001)

Conclusion

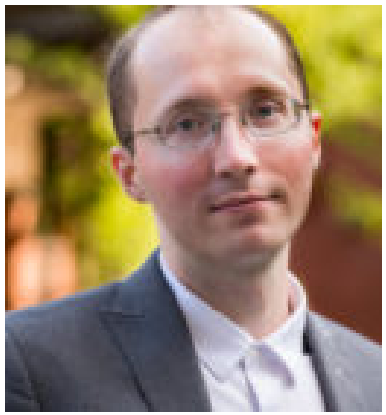
Using ideas from proximal causal inference, we have developed *doubly robust* methods to estimate ATT in synthetic control settings.

Collaborators

Xu Shi



Wang Miao



Edgar Dobriban

Eric Tchetgen Tchetgen

arXiv preprint: `https://arxiv.org/abs/2210.02014`

References

A. Abadie and J. Gardeazabal. The economic costs of conflict: A case study of the Basque country. *American Economic Review*, 93(1):113–132, 2003. ISSN 00028282. doi: 10.1257/000282803321455188.

A. Abadie, A. Diamond, and A. 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):493–505, 2010. ISSN 01621459. doi: 10.1198/jasa.2009.ap08746.

A. Abadie, A. Diamond, and J. Hainmueller. Comparative Politics and the Synthetic Control Method. *American Journal of Political Science*, 59(2):495–510, 2015. ISSN 15405907. doi: 10.1111/ajps.12116.

Y. Cui, X. Shi, and W. Miao. Semiparametric proximal causal inference. *arXiv preprint arXiv:2011.08411v1*, 2020.

B. Ferman and C. Pinto. Synthetic controls with imperfect pretreatment fit. *Quantitative Economics*, 12(4):1197–1221, 2021. ISSN 1759-7323. doi: 10.3982/qe1596.

X. Shi, W. Miao, M. Hu, and E. Tchetgen Tchetgen. Theory for identification and Inference with Synthetic Controls: A Proximal Causal Inference Framework. *arXiv preprint arXiv:2108.13935v3*, 2021.

Discussion

Relaxing stationarity:

- We can drop stationarity assumption on U_{t_+} and consider an ATT averaged over post-treatment time periods: $\sum_{t_+=T_0+1}^{T} \phi^*(t_+)\ell(t_+)$ for given importance time weight $\ell(t_+)$
- Similar GMM estimator, but conservative standard error (because of non-centered moment equation at every t)

Covariates:

- Our methods can incorporate covariates into h^* and q^* models, similarly to proximal causal inference in iid setting
- Alternatively, they can be included in proxies W or Z.