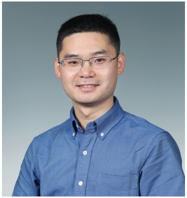
Doubly Robust Proximal Synthetic Controls

Hongxiang (David) Qiu

Department of Statistics, the Wharton School, University of Pennsylvania

Collaborators

Xu Shi



Wang Miao

Edgar Dobriban

Eric Tchetgen Tchetgen

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

Table of Contents

Motivation: causal inference with panel data

A common scenario:

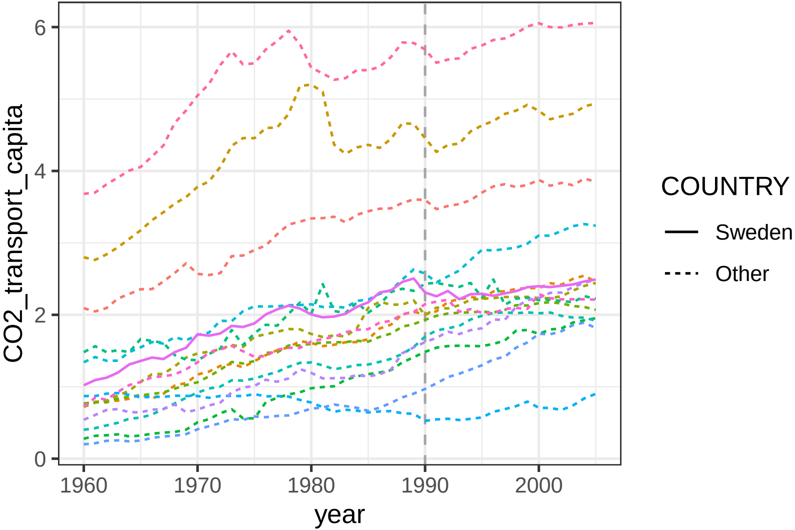
- Intervention on a single unit (e.g., country, state, etc.)
- Observe time series data of treated unit and a few untreated units
- How to estimate the causal effect of this intervention?

Example:

- A carbon tax and a value-added tax on transport fuel were issued in Sweden in 1990
- What is the effect of this (composite) intervention on per-capita CO_2 emission from transportation in Sweden?

Motivation: causal inference with panel data

Example data:



Motivation: causal inference with panel data

Notable challenges compared to "usual causal inference" with iid data:

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

Motivation: causal inference with panel data

Notable challenges compared to “usual causal inference” with iid data:

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

Somewhat strong assumptions appear necessary in panel data setting.

Idea behind classical synthetic controls

Some notations:

- Total number of time periods: T
- Intervention time: T_0
- Unit index: treated$= 0$; control$= 1, \dots, N$
- 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 $t > T_0$

Idea behind classical synthetic controls

Some notations:

- Total number of time periods: T
- Intervention time: T_0
- Unit index: treated$= 0$; control$= 1, \dots, N$
- 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 $t > T_0$

To estimate the ATT $\phi^*(t)$,

- $Y_{t,0}(1) = Y_{t,0}$ is observed
- how to learn about $Y_{t,0}(0)$?

Idea behind classical synthetic controls

Intuition:

- Might 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

- Use a weighted average/linear combination of control units to serve as a *synthetic control*
- 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

- Abadie's proposal essentially requires no random noise $\epsilon_{t,i}$ (otherwise, regression with measurement error in covariates)
- 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

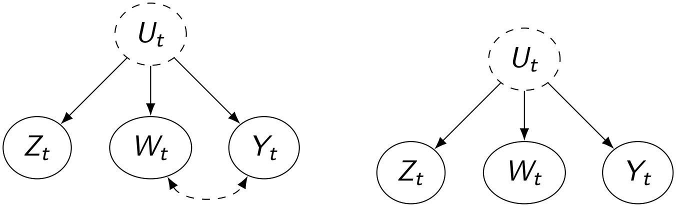
- Abadie's proposal essentially requires no random noise $\epsilon_{t,i}$ (otherwise, regression with measurement error in covariates)
- Many other ways to form a synthetic control have been proposed, but most still assume a linear model.
- A notable exception: based on proximal causal inference, Shi 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

- We split control units into two groups: donors (denoted by W) and non-donor control units (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
- W and Z are IVs for U; Z is an IV for W
- Key assumption 1: $Z_t \perp\!\!\!\perp (Y_t, W_t) \mid U_t$



[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 $\mathbb{E}[Y_t(0) \mid U_t] = \mathbb{E}[h^*(W_t) \mid U_t]$.
- 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) with a parametric model of h^*.
- 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 an 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 an 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. It can be identified via weighting or the influence function.

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:

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

This influence function is doubly robust.

Gaps between iid setting and panel data setting

- Data are not iid.
- A_t 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.
- 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.
- 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^* such that

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

- By Bayes Theorem, confounding=covariate shift.

[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* $q = q^*$.

Therefore, if we specify parametric models for h^* and q^*, only one needs to be correct.

Table of Contents

Doubly robust estimation with GMM

- Parametric models h_α for h^*, q_β for q^*, and $\phi_\lambda(t)$ for $\phi^*(t)$
- α, β, λ are model parameters to be estimated
- Arbitrary user-specified functions g_h and g_q
- Dimensions of $g_h(z)$ and $g_q(w)$ are higher than α and β, resp.

Doubly robust estimation with GMM

Define moment function

$$G_t : \theta \mapsto \begin{pmatrix} \mathbb{1}(t \leq T_0) \{[Y_t - h_\alpha(W_t)] g_h(Z_t)\} \\ \mathbb{1}(t > T_0) \{\psi - g_q(W_t)\} \\ \mathbb{1}(t \leq T_0) \{q_\beta(Z_t) g_q(W_t) - \psi\} \\ \mathbb{1}(t > T_0) \{\phi_\lambda(t) - [Y_t - h_\alpha(W_t)] + \psi_-\} \\ \mathbb{1}(t \leq T_0) \{\psi_- - q_\beta(Z_t)(Y_t - h_\alpha(W_t))\} \end{pmatrix}.$$

Equation for estimating h^*

Equations for estimating q^*

Equations for estimating $\phi^*(t)$

Doubly robust estimation with GMM

Why define G_t this way?

- A key condition of GMM is that $\mathbb{E}[G_t(\theta^*)] = 0$ for truth θ^* and all t

$$\mathbb{E}[[Y_{t_-} - h^*(W_{t_-})]g_h(Z_{t_-})] = 0$$
$$\mathbb{E}[g_q(W_{t_+})] = \psi^* = \mathbb{E}[q^*(Z_{t_-})g_q(W_{t_-})]$$
$$-\phi^*(t_+) + \mathbb{E}[Y_{t_+} - h^*(W_{t_+})] = \psi_-^* = \mathbb{E}[q^*(Z_{t_-})(Y_{t_-} - h^*(W_{t_-}))]$$

- We split one equation involving expectation in pre- and post-treatment time periods into separate equations so that $\mathbb{E}[G_t(\theta^*)] = 0$ *for all* t
- The condition of centered moment is especially important to obtain a correct standard error

Doubly robust estimation with GMM

GMM estimator:

$$\operatorname*{argmin}_{\theta} \left\{ \frac{1}{T} \sum_{t=1}^{T} G_t(\theta) \right\}^{\top} \Omega_T \left\{ \frac{1}{T} \sum_{t=1}^{T} G_t(\theta) \right\}$$

Ω_T: user-specified symmetric positive definite matrix (e.g., identity)

Theorem

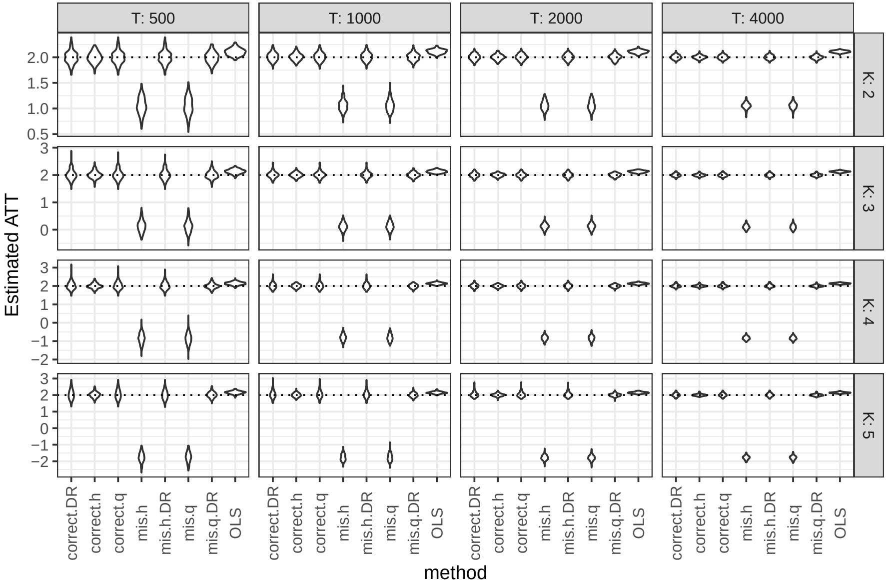
Under conditions, the GMM estimator is root-n consistent for the ATT and asymptotically normal as $T \to \infty$, if h^* or q^* is correctly specified.

Simulation

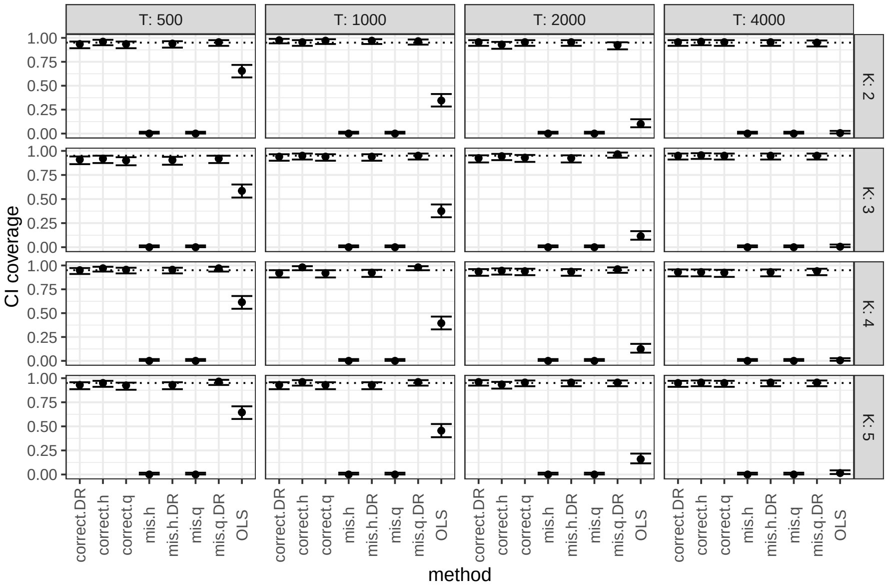
Methods compared:

- 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



Simulation: CI coverage



Sweden data analysis

- Yearly data of 15 countries from 1960–2005 (46 years)
- Remove time trend: fit a quadratic curve of time to control countries' outcomes and take residuals for all countries
- 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 countries with large weights: Belgium, Denmark, Greece and New Zealand
- Linear model for h^*
- Log-linear model for q^*: to restrict model complexity, only a subset of other control countries are included in the model for q^* (chosen based on geographical distance from Sweden):
  1. Iceland
  2. Iceland, France
  3. Iceland, France, Switzerland

Sweden data analysis

Method	Sweden tax	placebo at 1980
Abadie's SC	-0.286	0.008
OLS	-0.209 (-0.312, -0.107)	-0.009 (-0.046, 0.029)
DR	-0.321 (-0.451, -0.192)	-0.013 (-0.116, 0.090)
DR2	-0.302 (-0.418, -0.186)	-0.015 (-0.101, 0.072)
DR3	-0.314 (-0.476, -0.153)	-0.011 (-0.242, 0.219)
Outcome bridge	-0.346 (-0.479, -0.214)	0.001 (-0.086, 0.087)
Treatment bridge	-0.120 (-0.189, -0.052)	-0.002 (-0.004, -0.000)
Treatment bridge2	-0.143 (-0.275, -0.011)	0.011 (0.008, 0.013)
Treatment bridge3	-0.145 (-0.246, -0.044)	0.017 (-0.250, 0.283)

Discussion

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

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 $\mathbb{E}[G_t(\theta^*)] \neq 0$ for every t but $\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}[G_t(\theta^*)] = 0$)

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.

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.

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.