Dynamic Causal Inference & Synthetic Control

Benjamin O. Harrison

9 February 2025

Econometrics/Reading Session

Potential Outcome Framework: Time series

EMORY UNIVERSITY

Recap

What about Dynamic Systems

Setup

Application in Macroeconomics

The Synthetic Control Methods

EMORY UNIVERSITY

Introduction

Setup

Implementation

Properties of SCM

Empirical Implementation

Inference

Conclusion

Part I

Potential Outcome Framework: Time series

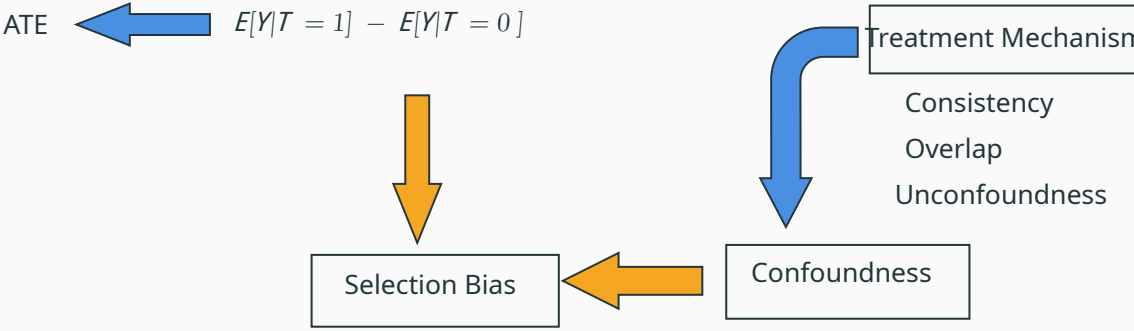
Recap

Quick Recap

- The core of inferring Causality lies in uncovering ”Counterfactuals”.

$$Y(T = 1) - Y(T = 0)$$

- **Key:** The central idea behind quantifying causal effects lies in the Treatment Mechanism. Holland, 1986



What about Dynamic Systems

Dynamic Causal Effects

- In dynamic settings, quantifying causality is more complex.
  - Both treatment and outcome evolve over time.
  - Dynamic feedback between treatment and outcome.
- Traditionally, causal relationships in macro literature relied on structural models.
  - Restrictions based on economic theory are imposed to identify causal effects.
  - Much of the literature relies on parametric linear models.
  - Widely used framework includes SVARs and Local Projections
- The strong functional form assumptions may implicitly impose substantial restrictions on the underlying dynamic causal effects.

Setup

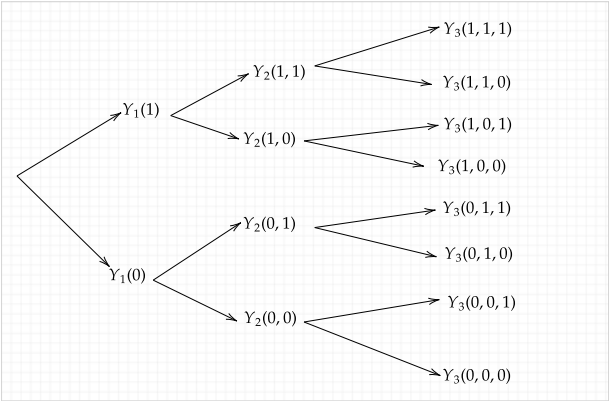
Potential Outcome Time series Framework

- Consider a single unit.
- At each time $t = 1, \cdots T$, the unit is exposed to a scalar treatment $W_t \in \mathcal{W} = \{0, 1\}$.
- Define the stochastic process $\{W_t\}_{t=1}^{T}$ as the **Assignment Process**, where each $W_t \in \mathcal{W}$.
- Potential Outcome Process: For any deterministic $\omega_{1:T} \in \mathcal{W}^T$, the stochastic process

$$Y_{1:T}(\omega_{1:T}) := (Y_1(\omega_{1:T}), Y_2(\omega_{1:T}), \ldots, Y_T(\omega_{1:T}))^\top$$

- Each time t potential outcome $Y_t(\omega_{1:T}) : \mathcal{W}^T \to \mathbb{R}$
- This definition shows that the potential outcome can depend on future treatments.
- Think of time t potential outcome process as the cross-sectional framework in which the outcome of unit i depends on the treatment of everyone.

Illustration



Assumption (Non-Anticipating {No Interference})

For any deterministic $\{\omega_t\}_{t\geq 1}, \{\omega'_t\}_{t\geq 1}$ *with* $\omega_t, \omega'_t \in \mathcal{W}$*;*

$$Y_t\left(\omega_{1:t}, \{\omega_s\}_{s\geq t+1}\right) = Y_t\left(\omega_{1:t}, \{\omega'_s\}_{s\geq t+1}\right) \quad \forall t \geq 1 \quad a.s$$

Assumption (Output {Consistency})

The output process $\{W_t, Y_t\}_{t\geq 1} = \{W_t, Y_t(W_{1:t})\}_{t\geq 1}$*. The* $\{Y_t\}_{t\geq 1}$ *is called the outcome process.*
In short, $Y_{1:T} = (Y_1(W_1), Y_2(W_{1:2}), \cdots Y_T(W_{1:T}))$ *is the random process whose realization is observed.*

Assumption (Sequential Probabilistic assignment{Overlap})

$0 < P(W_t = \omega \mid \mathcal{F}_{t-1}) < 1$ *with probability one for all* $w \in \mathcal{W}$*.*
$\mathcal{F}_t$ *is determined by* $\left\{W_t, \left\{Y_t(\omega_{1:t}), \omega_{1:t} \in \mathcal{W}^t\right\}\right\}_{t\geq 1}$*.*

Direct Potential Outcome System

Definition

Any $\left\{W_t, \left\{Y_t\left(\omega_{1:t}\right) : \omega_{1:t} \in \mathcal{W}^t\right\}\right\}_{t\geq 1}$ satisfying Assumptions 1-3 is a direct potential outcome system.

- **Non-Anticipating on Assignment Process:** Assume that W_t does not explicitly influence future assignment W_s , for $s > t$.
- W_t should directly affect potential outcome process Y_t but not indirectly through future assignment.
- Dynamic causal effect at time t: $Y_t(\omega_{1:t}) - Y_t(\omega'_{1:t})$
- Notation: Define potential outcome for time $t + h$ as

$$Y_{t+h}(\omega) := Y_{t+h}\left(W_{1:t-1}, \omega, W_{t+1:t+h}\right)$$

- For $t \geq 1, h \geq 0$, and any fixed $\omega, \omega' \in \mathcal{W}$, the time- t, h-period ahead:
  - Impulse causal effect : $Y_{t+h}(\omega) - Y_{t+h}\left(\omega'\right)$
  - Average Treatment Effect: $\mathbb{E}\left[Y_{t+h}(\omega) - Y_{t+h}\left(\omega'\right)\right]$

Application in Macroeconomics

Impulse Response Function

- Quantifying the effect of Monetary Policy on Economy.
- In these literature, the quantity of interest is Impulse response function.

$$IRF = \mathbb{E}\left[Y_{t+h} \mid W_t = \omega\right] - \mathbb{E}\left[Y_{t+h} \mid W_t = \omega'\right]$$

- Traditionally, the IRF has been estimated by modeling the entire dynamic system (VAR's).
- Recently, Local Projections estimate IRF using by running simple OLS.
- Two Questions poised by Rambachan and Shephard, 2021
  - When is IRF causal?
  - How can we estimate IRF non-parametrically?
- Spoiler Alert: The IRF is causal only under randomization of the Treatment.

Estimation

- The IRF has a causal interpretation if the shock or intervention is exogenous.
- In our framework, this holds when

$$W_t \perp \left(W_{1:t-1}, W_{t+1:t+h}, \left\{Y_{t+h}\left(w_{1:t+h}\right) : w_{1:t+h} \in \mathcal{W}^{t+h}\right\}\right)$$

- This is equivalent to the Ignorability assumption or CIA.
- Based of this assumption:
  - Angrist et al., 2018 employed a semiparametric estimates of monetary policy effects using Inverse probability weighting estimators.
  - Recently, Ballinari and Wehrli, 2024 employs Double debiased machine learning to estimate the IRF.

Remarks

- The Direct Potential Outcome system framework places no functional form restrictions.
- No common time series assumptions such as invertibility.
- The VAR and LPs methodologies are special case of this framework.

Part II

The Synthetic Control Methods

Introduction

Motivation

- Quantify the effects of policy intervention that takes place at an Aggregate level (States, country, Firms).
- Study the effect of terrorist conflicts in Basque country on the Basque economy. Abadie and Gardeazabal, 2003
- The intervention affects all units. (Overlap assumption is violated.)
- Comparative case studies: Compare the evolution of an aggregate outcome for the treated unit and control unit. Card, 1993
- The main limitation:
  - Degree of ambiguity in the choice of comparison group.
  - The inferential techniques of these comparison case studies only quantifies sampling uncertainty.
  - No inference on the uncertainty about the control group being a good counterfactual for the treated unit.
- Do not Capture heterogeneous effect across different subpopulation. Gunsilius, 2023

Synthetic Control

- Synthetic control (SC) proposed **abadie2023**; Abadie et al. (2010) reduces the ambiguity and subjectivity by quantifying how affinity.
- In practice, it is difficult to find a single unit that best matches the most relevant characteristics of the treated unit.
- Uses a combination of control units (donor pool) to approximate the counterfactual for treated units.
- Captures the relative contribution of each control unit to the counterfactual.

Setup

Famework

- Suppose that we observe $J + 1$ units in periods $1, 2, \cdots T$.
- Unit one is exposed to the intervention of interest (that is,"treated") during periods $T_0 + 1, \cdots T$.
- The remaining J units are an untreated reservoir of potential controls (a "donor pool").
- Let Y_{it}^I be the outcome that would be observed for unit i at time t if unit i is exposed to the intervention in periods $T_0 + 1$ to T.
- Let Y_{it}^N be the outcome that would be observed for unit i at time t in the absence of the intervention.
- We aim to estimate the effect of the intervention on the treated unit,

$$\tau_{1t} = Y_{1t}^I - Y_{1t}^N = Y_{1t} - Y_{1t}^N$$

for $t > T_0$, and Y_{1t} is the outcome for unit one at time t.

Assumption

- *No Anticipation Assumption* *i.e.* $Y_{it}^{I} = Y_{it}^{N}$ *for* $t \in \{1, \cdots T_0\}$
- *Sequential Treatment:* *The treatment does not turn off i.e. it continues to affect the treated unit in all* $T_0 + 1, \cdots, T$
- *No Interference Assumption*

Implementation

Implementation

- A SC is a weighted average of units from a donor pool.
- Let $\boldsymbol{W} = (w_2, \ldots, w_{J+1})'$ s.t. $w_j \geq 0$ for $j = 2, \ldots, J+1$ and $w_2 + \cdots + w_{J+1} = 1$.
- The goal is to obtain $\left(w_2^*, \ldots, w_{J+1}^*\right)$ such that:

$$\begin{array}{ll} \sum_{j=2}^{J+1} w_j^* Y_{j1} = Y_{11}, & \sum_{j=2}^{J+1} w_j^* Y_{j2} = Y_{12}, \\ \sum_{j=2}^{J+1} w_j^* Y_{jT_0} = Y_{1T_0}, & \text{and} \quad \sum_{j=2}^{J+1} w_j^* \mathbf{Z}_j = \mathbf{Z}_1 \end{array}$$

- Let $\mathbf{X}_1 = \left(\mathbf{Z}_1', \bar{Y}_1^{\mathbf{K}_1}, \ldots, \bar{Y}_1^{\mathbf{K}_M}\right)'$ be a $(k \times 1)$ vector of pre-intervention characteristics for the treated unit. Similarly, let $\mathbf{X}_0$ is $\left(\mathbf{Z}_j', \bar{Y}_j^{\mathbf{K}_1}, \ldots, \bar{Y}_j^{\mathbf{K}_M}\right)'$ be a $(k \times J)$ matrix which contains the same variables for the unaffected units.
- The vector $\boldsymbol{W}^* = \left(w_2^*, \ldots, w_{J+1}^*\right)'$ is chosen to minimize $\|\boldsymbol{X}_1 - \boldsymbol{X}_0 \boldsymbol{W}\|$, subject to our weight constraints.
- $\|\mathbf{X}_1 - \mathbf{X}_0 \mathbf{W}\|_{\mathbf{V}} = \sqrt{(\mathbf{X}_1 - \mathbf{X}_0 \mathbf{W})' \mathbf{V} (\mathbf{X}_1 - \mathbf{X}_0 \mathbf{W})}$

Implementation (cont.)

- V represents a symmetric positive semi-definite matrix chosen by the researcher or by data-driven methods (out-of-sample validation).
- For a post-intervention period t (with $t > T_0$) the SC estimator is:

$$\widehat{\tau}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt}$$

Properties of SCM

Motivating Model

- Recall that $\tau_{1t} = Y_{1t} - Y_{1t}^N$
- The SC estimator is a finite sample estimator i.e. so asymptotic properties may not apply in this context.
- To evaluate the validity of the SC estimator, we compute the $Bias(\tau_{1t})$.
- suppose we model the counterfactual with a factor model :

$$Y_{it}^N = \theta_t Z_{it} + \lambda_t \mu_i + \varepsilon_{it}$$

where $\boldsymbol{Z}_i$ are observed features, $\boldsymbol{\lambda}_t$ are unobserved factors, and ε_{it} is a unit-level transitory shock, modeled as random noise.

- This generalizes the DiD method as it allows for unobserved confounders to vary over time.
- As pointed out earlier, SCM seeks for $\{w\}_{j=2}^{J+1}$ s.t

$$\sum_{j=2}^{J+1} w_j^* Y_{jt} = Y_{1t} \quad \forall t = 1, .., T_0$$

- If this match holds, Bias = 0.

- This match holds exactly only if $(Y_{11}, \ldots, Y_{1T_0}, \mathbf{Z}_1')$ belongs to the convex hull of $\{(Y_{21}, \ldots, Y_{2T_0}, \mathbf{Z}_2'), \ldots, (Y_{J+11}, \ldots, Y_{J+1T_0}, \mathbf{Z}_{J+1}')\}$.

Proposition

Suppose that $E\left|\varepsilon_{jt}\right|^p < \infty$ *for some* $p > 2$. *Let* $\sigma_{jt}^2 = E\left|\varepsilon_{jt}\right|^2$, $\sigma_j^2 = (1/T_0)\sum_{t=1}^{T_0}\sigma_{jt}^2$, $\bar{\sigma}^2 = \max_{j=2,\ldots J+1}\sigma_j^2$, *and* $\bar{\sigma} = \sqrt{\bar{\sigma}^2}$. *Similarly, let* $m_{p,jt} = E\left|\varepsilon_{jt}\right|^p$, $m_{p,j} = (1/T_0)\sum_{t=1}^{T_0} m_{p,jt}$, *and* $\bar{m}_p = \max_{j=2,\ldots J+1} m_{p,j}$. *We obtain that, for* $t > T_0$,

$$|E[\widehat{\tau}_{1t} - \tau_{1t}]| \leq C(p)^{1/p}\left(\frac{\bar{\lambda}^2 F}{\underline{\xi}}\right) J^{1/p} \max\left\{\frac{\bar{m}_p^{1/p}}{T_0^{1-1/p}}, \frac{\bar{\sigma}}{T_0^{1/2}}\right\}$$

- The bias of the estimator can be bounded by a function that goes to zero as the number of pretreatment periods increases.

Important Remarks of SC

- Proposition 1 demonstrates that the credibility of the synthetic control depends extent to which it can fit the trajectory of Y_{1t} for an extended pre-intervention period.
- There is no ex-ante guarantee that SC will be a good fit.
- This happens when $(Y_{11}, \ldots, Y_{1T_0}, \mathbf{Z}_1')$ falls far from the convex hull of $\{(Y_{21}, \ldots, Y_{2T_0}, \mathbf{Z}_2'), \ldots, (Y_{J+11}, \ldots, Y_{J+1T_0}, \mathbf{Z}_{J+1}')\}$.
- If the fit is poor, Abadie et al., 2010 recommends against the use of Synthetic control .
- Settings with small T_0, large J, and large noise create substantial risk of over-fitting.
- We can reduce interpolation bias by restricting the donor pool to regions with similar characteristics to the treated.

Empirical Implementation

Tobacco Control Program

- In 1988, California passed proposition 99, anti-tobacco law.
  - Introduced a 25-cent per pack state excise tax.
  - Banned vending machines and single cigarette sales.
  - Revenue-funded health and anti-tobacco programs.
- We investigate the impact of Proposition 99 on smoking prevalence using synthetic control methodology.

Data and Sample

- Annual state-level data from 1970 - 2000.
- Proposition was implemented in January 1989, $\implies T_0 = 19$ years .
- Create a Synthetic California counterfactual, using the remaining 49 states as donor Pool during the preintervention periods.
- Donor Pool is restricted to 38 states, as some states did implement anti-tobacco policies.
- The Outcome of interest is Annual per capita cigarette sales.
- Included Predictors such as ”Average Retail Price of Cigarette”, ”Per-capita state income”, and ”Per-capita beer consumption”.
- Choose $\bar{Y}_i^K = (CigSales_{1975}, CigSales_{1980}, CigSales_{1988})$

Traditional Comparison group

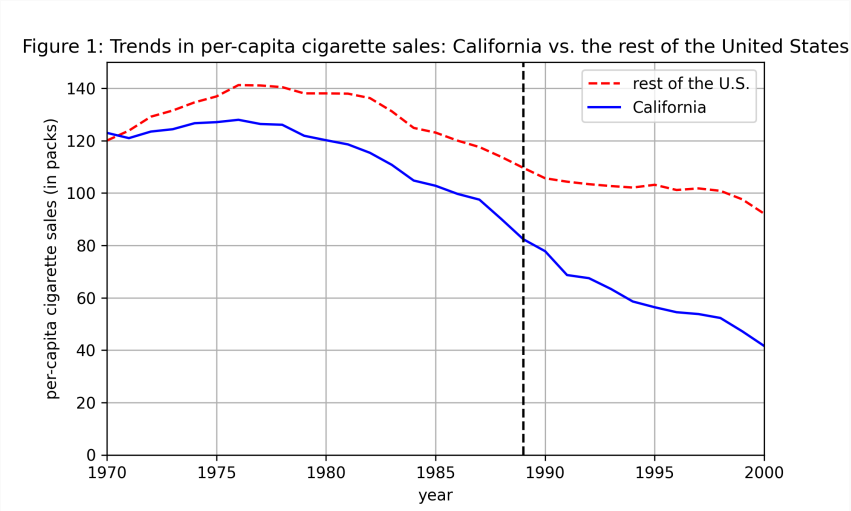


Figure 2: Average of all units in donor pool as a Comparison group

Synthetic Control

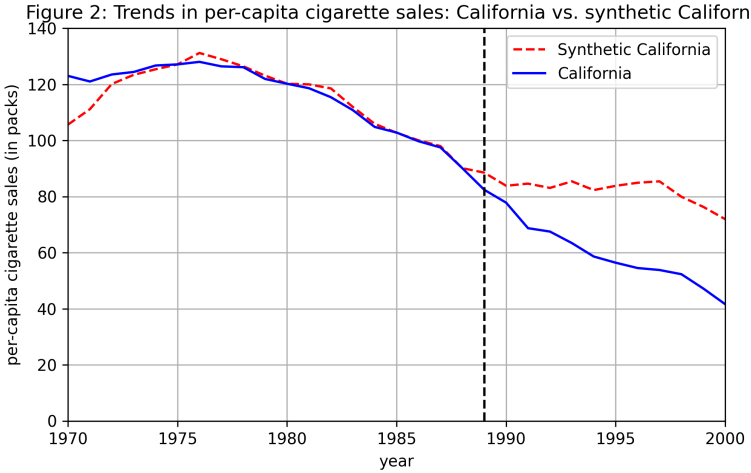


Figure 3: Trends in cigarette sales of the Synthetic california and Carlifonia

Treatment effect

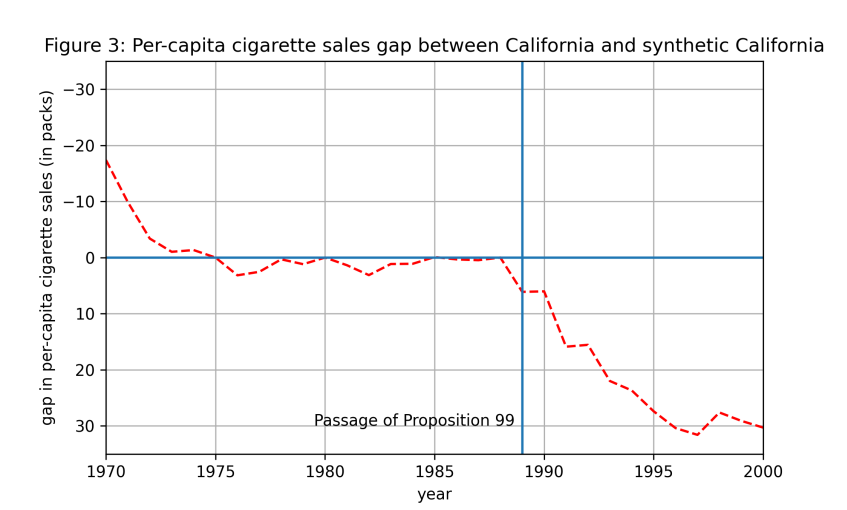


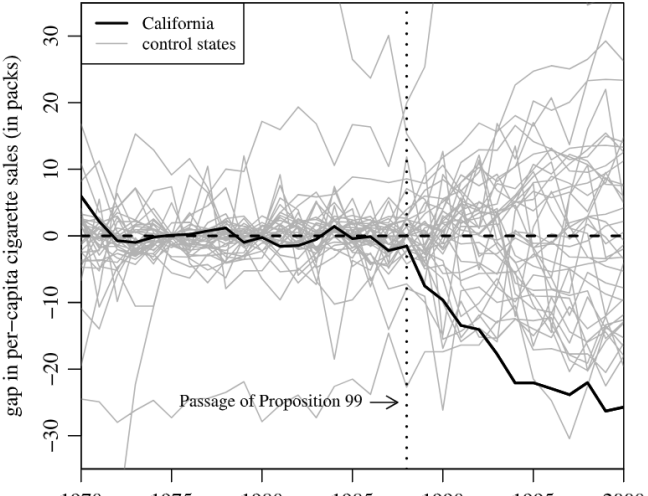
Figure 4: Impact of Proposition 99 on cigarette sales

Inference

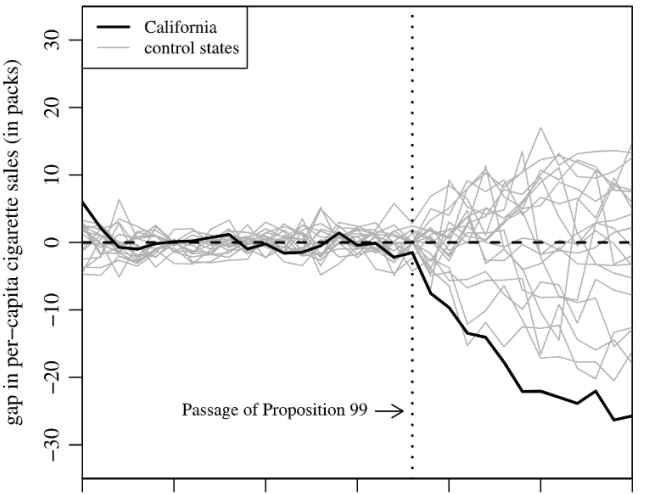
Inference

- To evaluate the significance of our estimates, we pose the question of whether our results could be driven entirely by chance.
- Large sample inferential techniques is not well suited for this methodology.
- Abadie et al. (2010) propose a mode of inference for the synthetic control framework that is based on permutation methods.
- A permutation distribution can be obtained by iteratively reassigning the treatment to the units in the donor pool and estimating placebo effects in each iteration.
- The effect of the treatment on the unit affected by the intervention is deemed to be significant when its magnitude is extreme relative to the permutation distribution.

Placebo Test



Placebo Test



Conclusion

Comment

- Can the Synthetic control methods be extended to identifying IRF?
- Recall that the dynamic causal effect requires a comparison between the observed treatment path and the comparison treatment path.
- Recent literature by Gonçalves et al., 2024 provides a new definition of IRF which takes an average over all possible comparison of the treatment paths.
- Perhaps a Synthetic control method can be extended to provide a synthetic version of the treatment path as a comparison to the observed path??

Thanks!

benjamin.harrison@emory.edu
benhars.com
@benhars

References

- Abadie, A., Diamond, A., & Hainmueller, J. (2010). **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.
- Abadie, A., & Gardeazabal, J. (2003). **The economic costs of conflict: A case study of the basque country.** *American economic review*, *93*(1), 113–132.
- Angrist, J. D., Jordà, Ò., & Kuersteiner, G. M. (2018). **Semiparametric estimates of monetary policy effects: String theory revisited.** *Journal of Business & Economic Statistics*, *36*(3), 371–387.
- Ballinari, D., & Wehrli, A. (2024). **Semiparametric inference for impulse response functions using double/debiased machine learning.** *arXiv preprint arXiv:2411.10009*.

- Card, D. (1993). **Minimum wages and employment: A case study of the fast food industry in new jersey and pennsylvania.**
- Gonçalves, S., Herrera, A. M., Kilian, L., & Pesavento, E. (2024). **State-dependent local projections.** *Journal of Econometrics*, 105702.
- Gunsilius, F. F. (2023). **Distributional synthetic controls.** *Econometrica*, *91*(3), 1105–1117.
- Holland, P. W. (1986). **Statistics and causal inference.** *Journal of the American statistical Association*, *81*(396), 945–960.
- Rambachan, A., & Shephard, N. (2021). **When do common time series estimands have nonparametric causal meaning.** *Manuscript, Harvard University*.