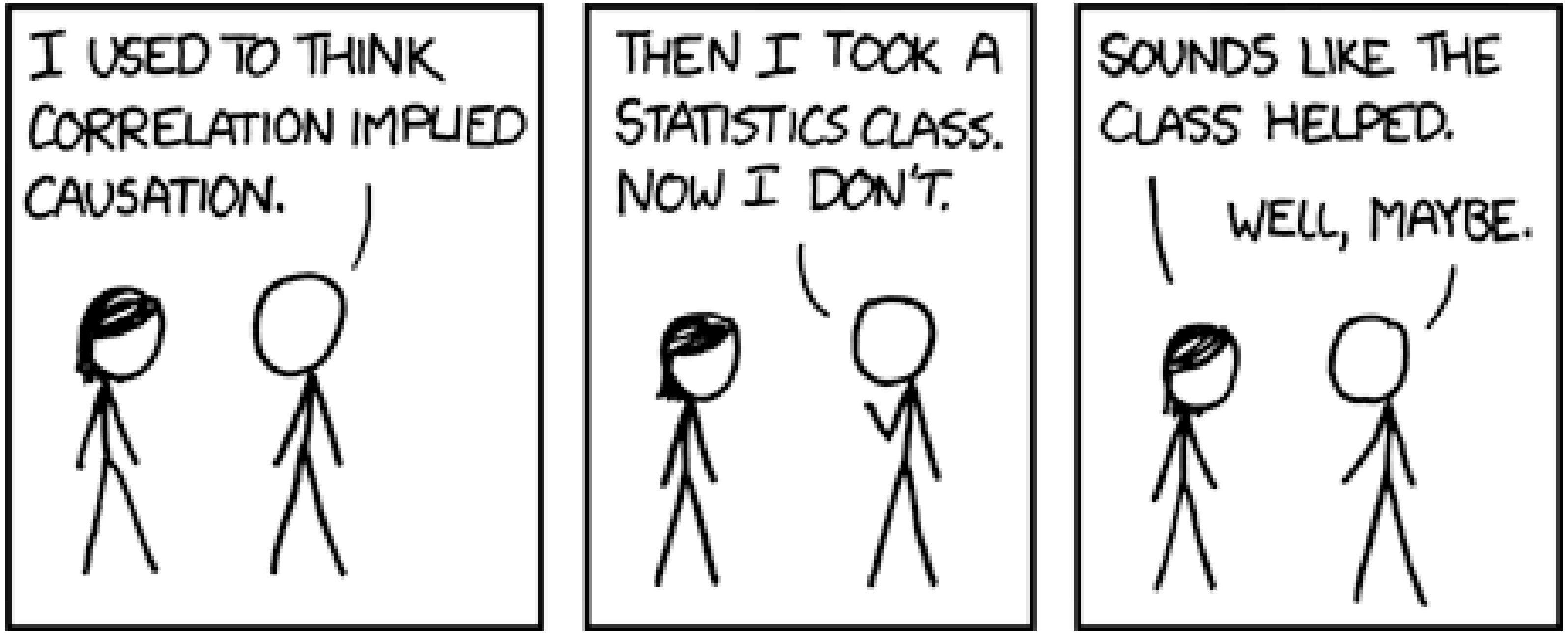
Causal Inference in Marketing: Learning from Quasi-experiments

Kathy Li

UT Austin, McCombs School of Business

What is Causal Inference?



Inferring the effect of one thing on another

Did X cause Y?

Golden standard of causal inference are randomized experiments

But we can’t always run an experiment

Enter quasi-experiments!

Avery et al. 2012; Pauwels and Neslin 2015; Soysal and Krishnamurthi 2015 Wang and Goldfarb 2017

Puranam, Narayan and Kadiyali 2017

Kim, Lee and Gupta 2020
Seiler, Tuchman and Yao 2021

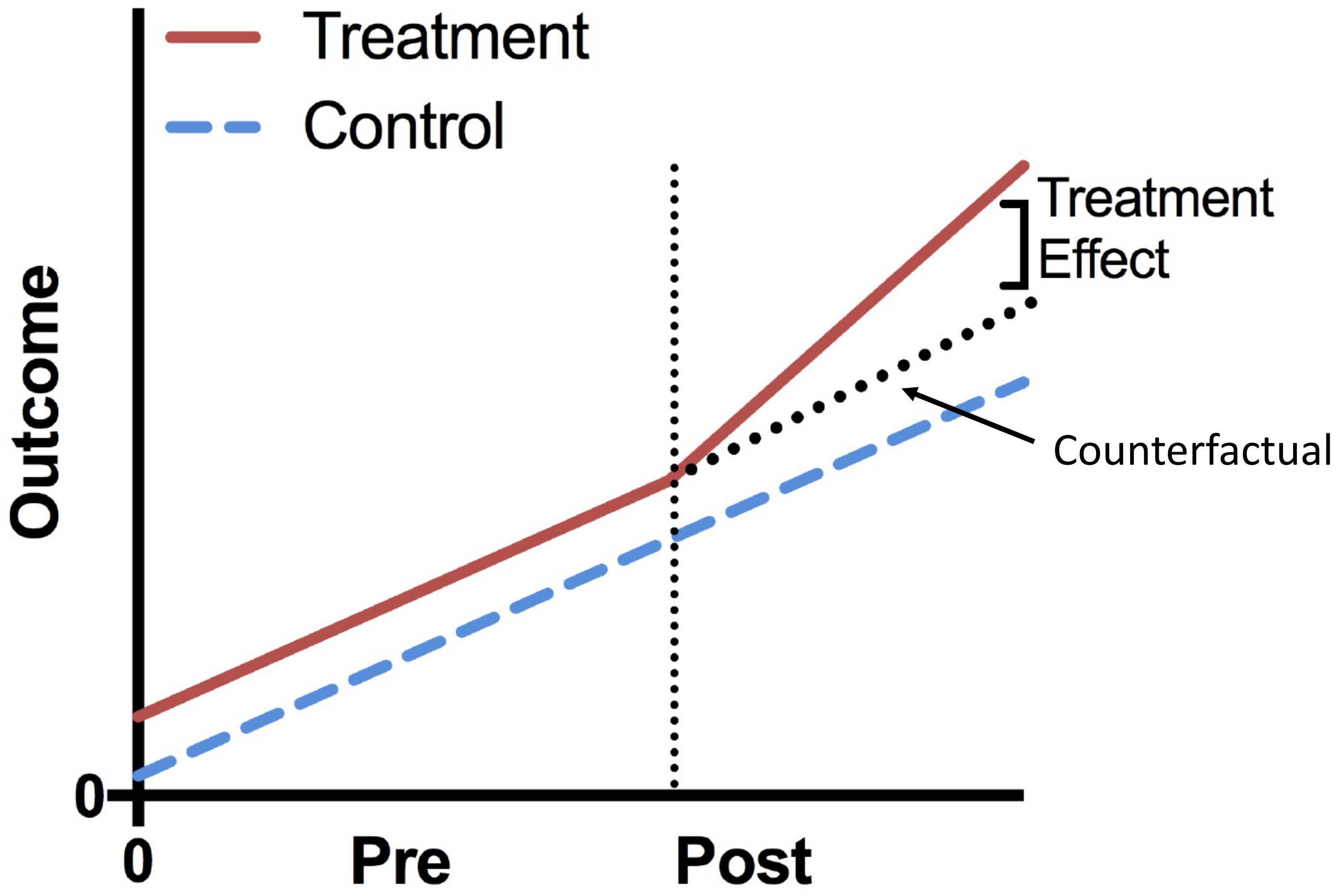
(Quasi-experimental)
Marketing and policy interventions are prevalent

- Data Setting: Pre/Post Treatment/Control
- Impact of
  - Regulations/legislations/policy interventions (in one or few states) on sales (e.g., Abadie et al. 2010)
  - Online paywall on sales (Pattabhiramiah et al. 2019)
  - Payment disclosure on prescriptions (Guo et al. 2020)
  - Advertising on word of mouth (Tirunillai and Tellis 2017)
  - Media services on piracy search (Lu et al. 2021)
  - Soda tax on sales (Kim et al. 2020)
  - Minimum wage change on service perceptions (Puranam et al. 2021)
  - Food labeling on consumer behavior (Araya et al. 2022)

Challenge: Randomized control design is infeasible

Solution: Quasi-experimental methods

Popular Method: Differences-in-Differences



$$Y_{it} = \beta_0 + \beta_1 TREAT_i + \beta_2 POST_t + \beta_3 (TREAT_i)(POST_t) + u_{it}$$

Parallel Trends Assumption

Treatment unit would have been parallel to the control unit in the absence of the treatment

What if it is violated?

Synthetic Control Methods to the rescue!

Synthetic Control Method

(Abadie and Gardeazabal 2003; Abadie, Diamond and Hainmueller 2010)

"Arguably the most important innovation in the evaluation literature in the last fifteen years"

– Susan Athey and Guido Imbens

Main idea

- Whereas DID essentially require equal weights on control units, SC allows the weights on control units to vary (and allow zero weights)
- Additional flexibility to match controls to treatment
- Use weighted average of control units to create a synthetic version of treatment unit

Synthetic Control Method Overview

Using pre-treatment data, select weights **w** to minimize

$$\sum_{t=1}^{T_1} \left[y_{treat,t} - y'_{control,t} w\right]^2$$

s.t. $\sum_{j=1}^{N_{co}} w_j = 1$ and $w_j \geq 0$, $j = 1, \ldots, N_{co}$

Note: need long enough time series

Relaxing restrictions -> more flexible methods

Actual Sales

$$ATT = \frac{1}{T_2} \sum_{T_1+1}^{T} (y_{treat,t} - \hat{y}^0_{treat,t})$$

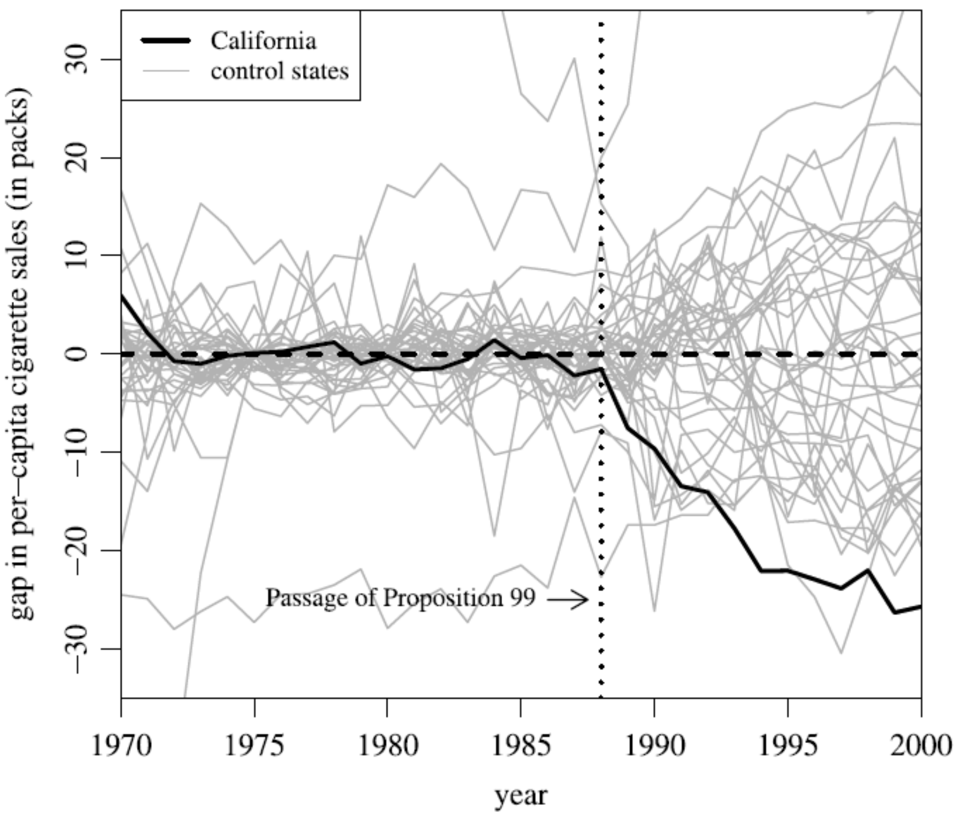
Counterfactual

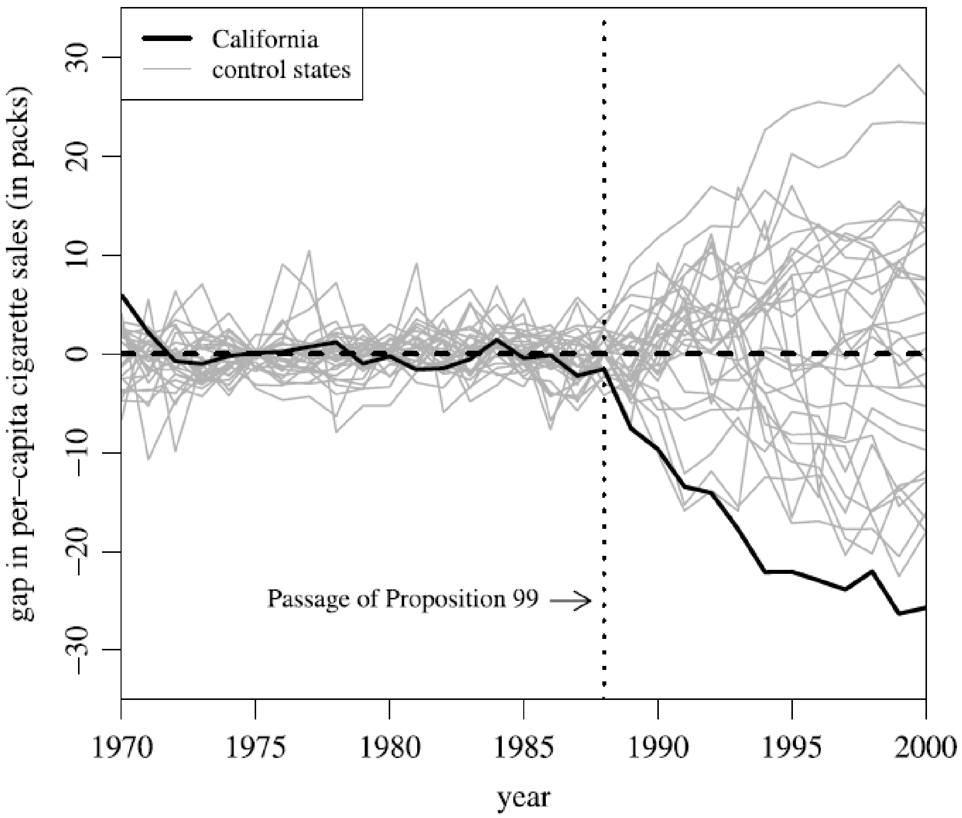
$$\hat{y}^0_{treat,t} = \hat{w}_1 y_{control1,t} + \ldots + \hat{w}_{N\,co} y_{controlN_{co},t}$$

How do researchers quantify uncertainty using synthetic control method?

Previously, researchers used placebo tests

- Strong symmetry assumption treatment and controls have similar data variation (variance)
- No confidence bounds, hypothesis testing or (general) p-values

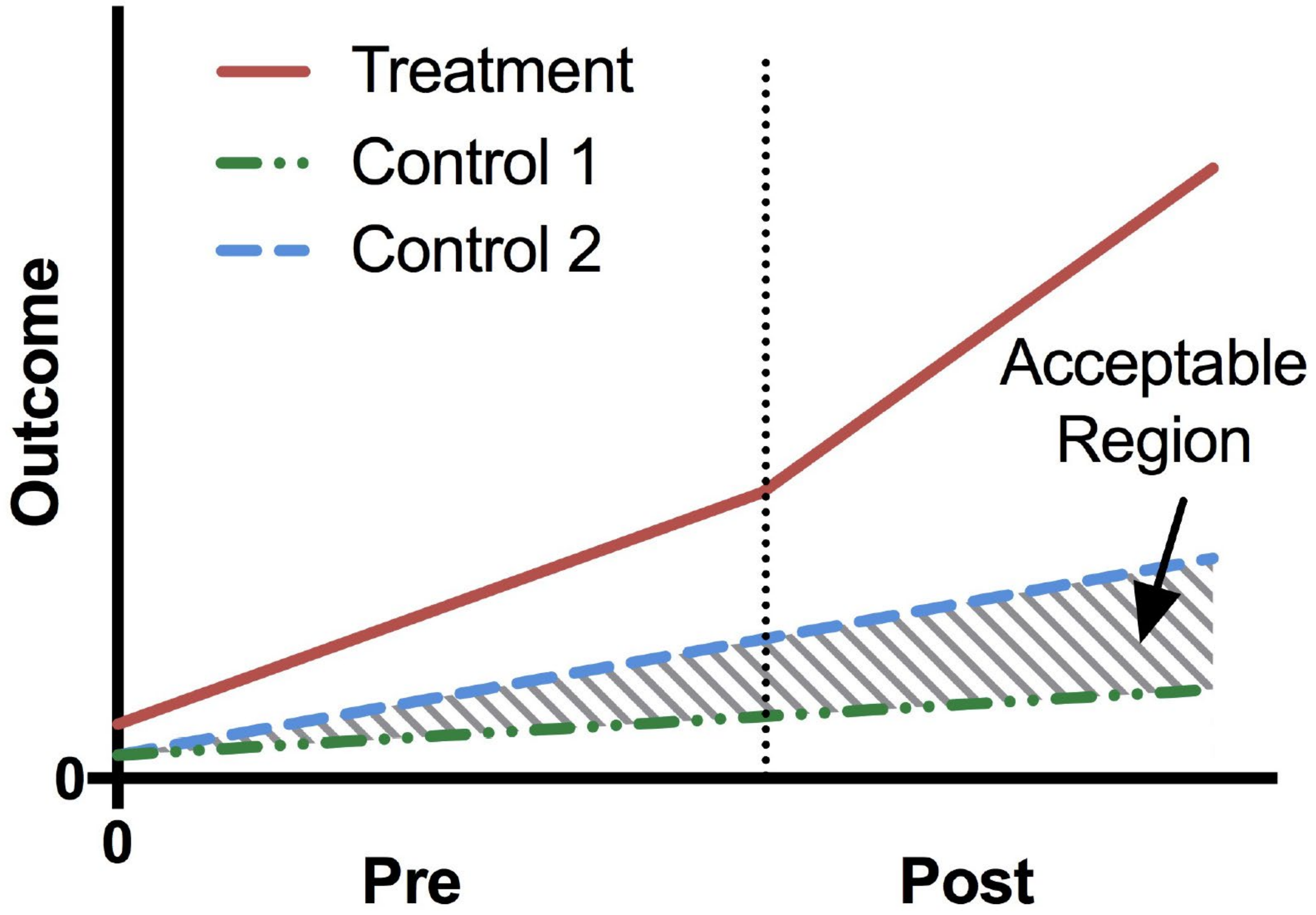




How to quantify uncertainty when using synthetic control method? (Li 2020, *JASA*)

- Previously, researchers had to rely on placebo tests that have strong symmetry assumptions
- Li (2020) develops formal inference theory for the synthetic control method and
- Proves that subsampling procedure can be used to calculate confidence intervals, conduct hypothesis testing and obtain p-values

What if we need even more flexible methods?



One such flexible method:
HCW (Hsiao, Ching and Wan 2012) or OLS method (see Li and Bell 2017 for inference)

Problem: Overfitting

Solution: Regularization

1) Bayesian:
   - Kim, Lee and Gupta 2020 - Bayesian shrinkage priors to solve the sparsity problem and get Bayesian inference as a benefit of that approach
   - Pang, Liu and Xu 2022 – Bayesian factor model

2) Frequentist:
   - Ben-Michael et al 2018 (ridge), Carvalho et al 2018 (lasso), Doudchenko and Imbens 2016 (elastic net)
   - Factor model (will discuss next)

Factor Model Approach Overview

(Gobillon and Magnac 2016, Chan and Kwok 2016, Xu 2017)

- Similar in spirit but mechanically different to synthetic control
  1) Project the control units onto a lower dimensional factor space (largest eigenvectors of control data)
  2) *Regress the treatment unit's outcome on the latent factors to recover the factor loadings (weights) without any constraints to predict the treated counterfactuals*

- Uses all the control units' information to estimate common factors without discarding information
- Can handle a large number of control units
- Can handle treatment outside range of the control units
- Implicit regularization leads to better out of sample predictions
- Easily applied to multiple treatment units because dimension reduction to obtain factors only has to be done once

- Also called interactive fixed effects model, generalized synthetic control

*Note: needs long time series

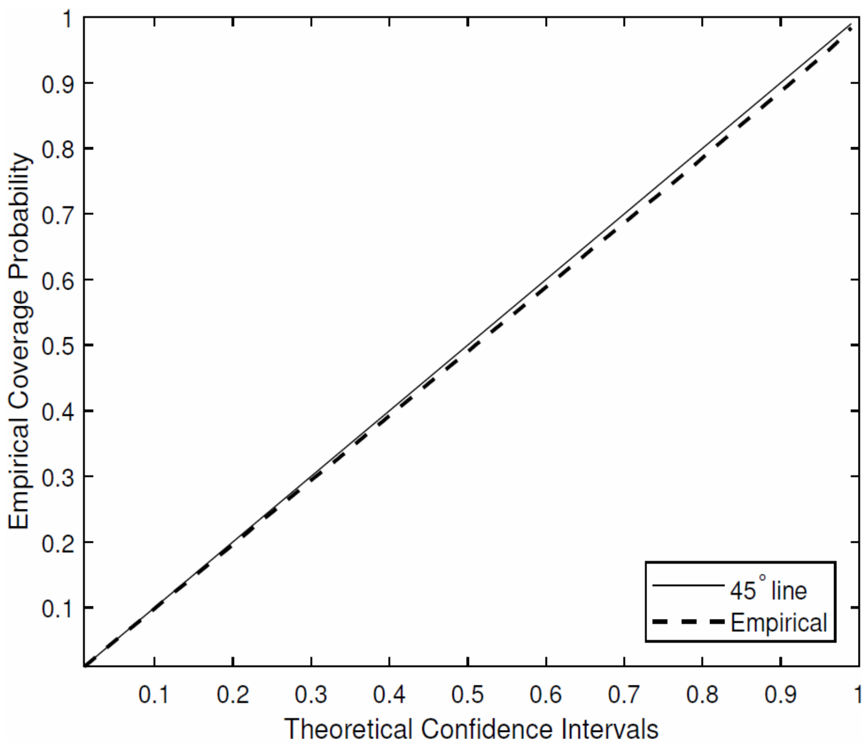
Inference for Factor Model Approach

(Li and Sonnier 2023, *JMR)*

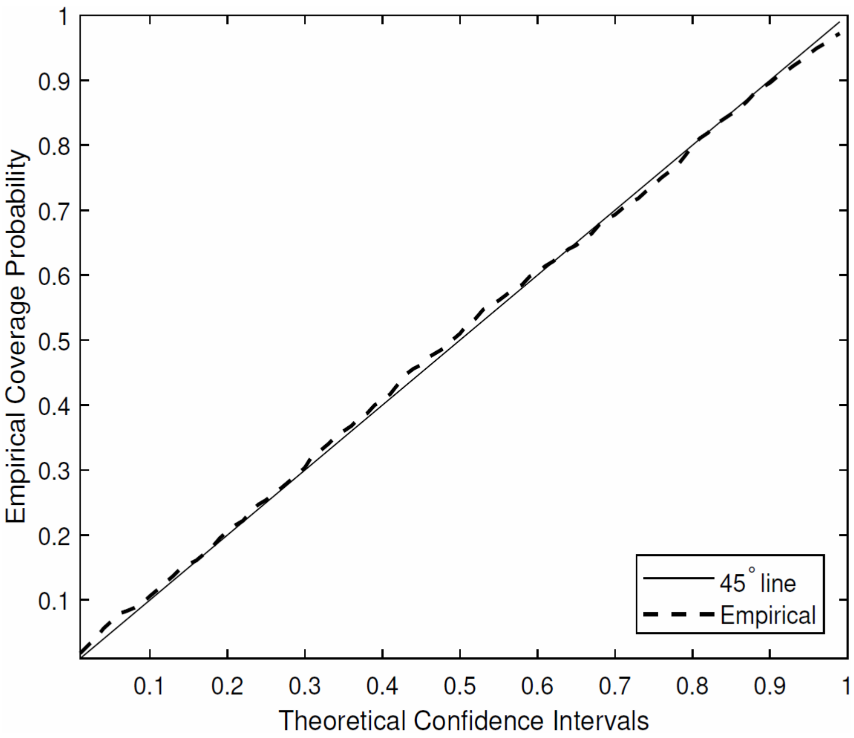
- Develop the inference theory for factor model approach to estimate treatment effects
  - Quantify uncertainty through hypothesis testing and confidence intervals
  - Computationally more efficient than bootstrap and placebo/permutation procedures
  - Allows treatment and control unit outcomes to have different distributions (different error variance is a special case)
- Especially important as factor model approach increasingly popular in marketing
  (Guo et al 2020, Nair et al 2021, Pattabhiramaiah et al 2019, Lovett et al 2019)
- Our inference method recovers coverage probabilities whereas Xu's (2017) bootstrap procedure often leads biased confidence intervals
  (because Xu 2017 bootstrap draws treatment error by resampling from control unit's estimated errors)

Equal Error Variance Coverage Probabilities

Our inference method



Bootstrap procedure

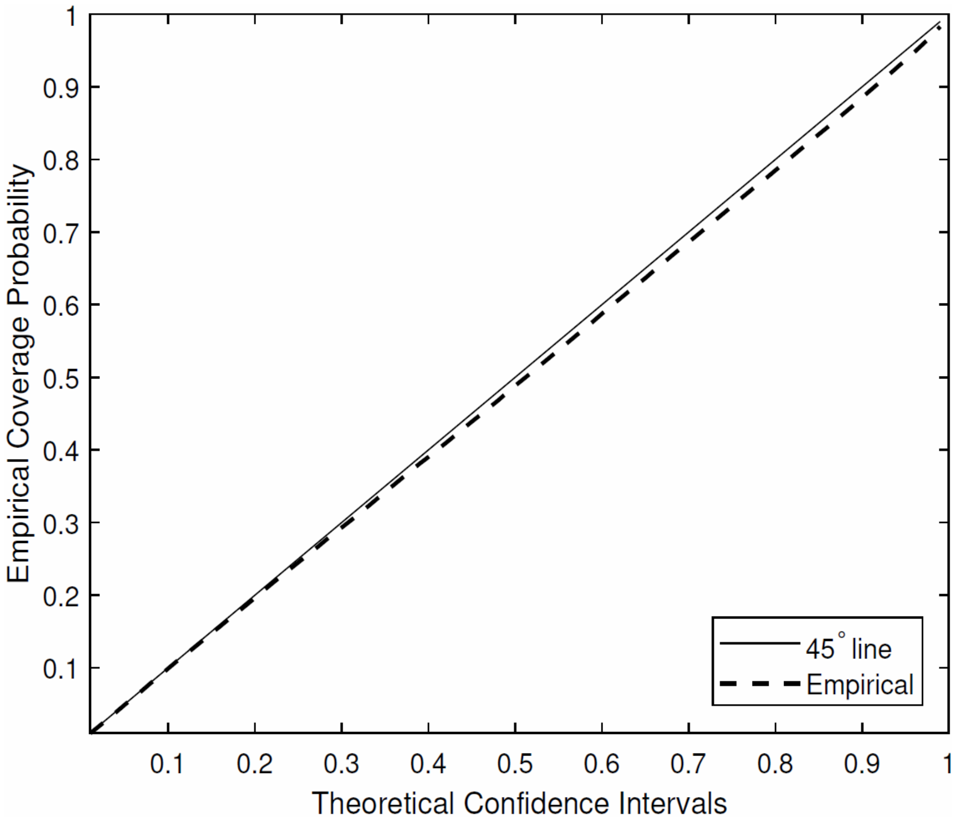


Takeaway: Our inference method and extant bootstrap procedure recovers the coverage probabilities (dashed line overlaps 45 degree line) when treatment and control units have equal error variance

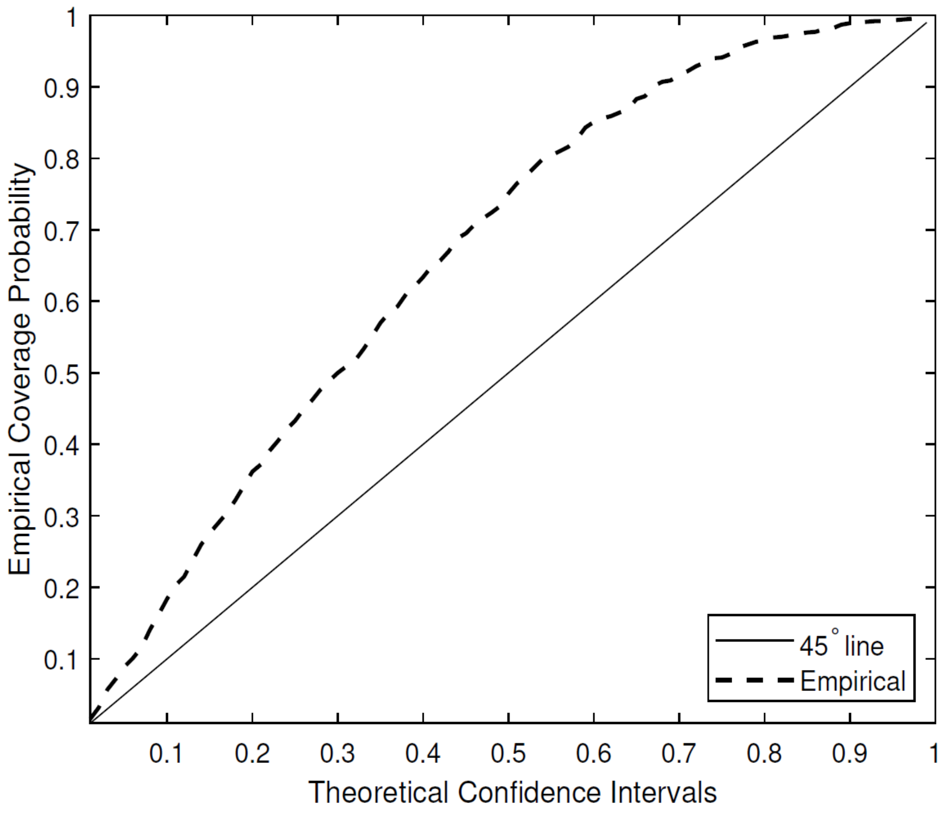
Unequal Error Variance Coverage Probabilities

Treatment's error variance smaller than control error variance

Our inference method



Bootstrap procedure

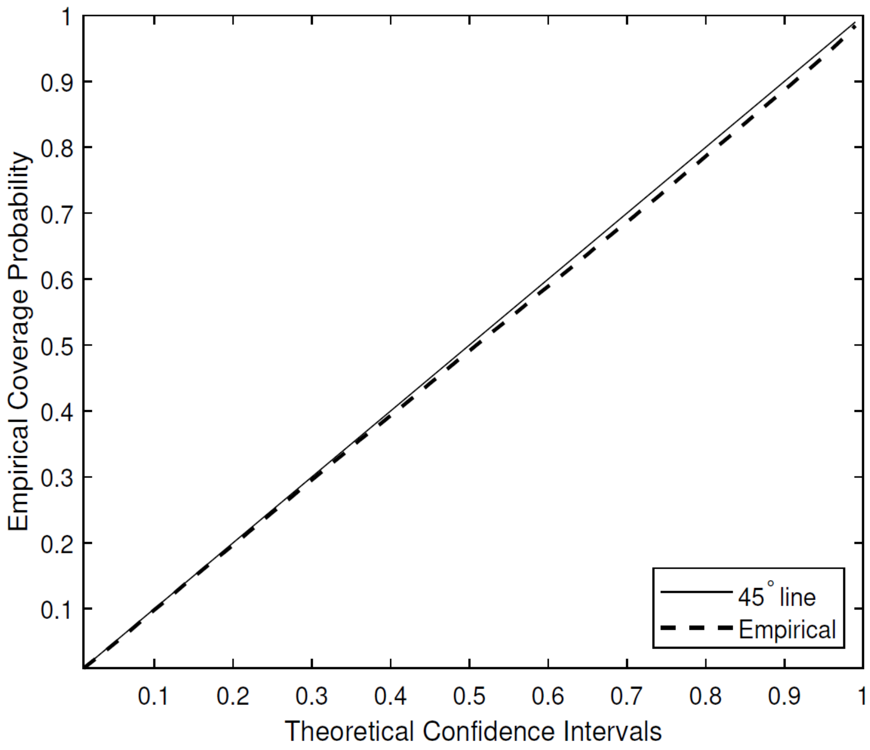


Takeaway: Our inference method recovers the coverage probabilities while extant bootstrap procedure cannot and results in over coverage when treatment error variance is smaller than control units' error variance

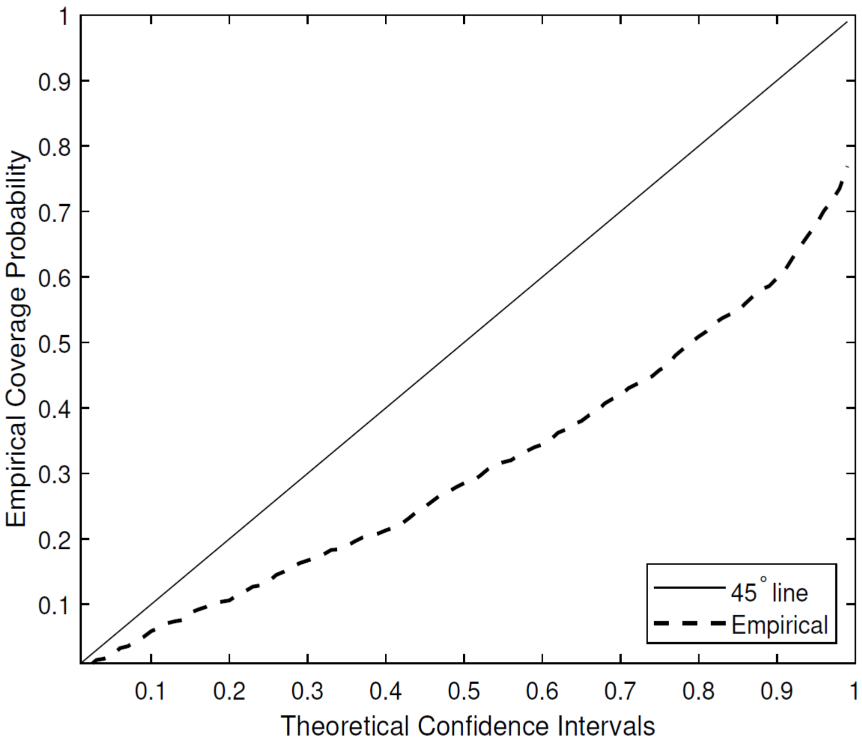
Unequal Error Variance Coverage Probabilities

Treatment's error variance greater than control error variance

Our inference method



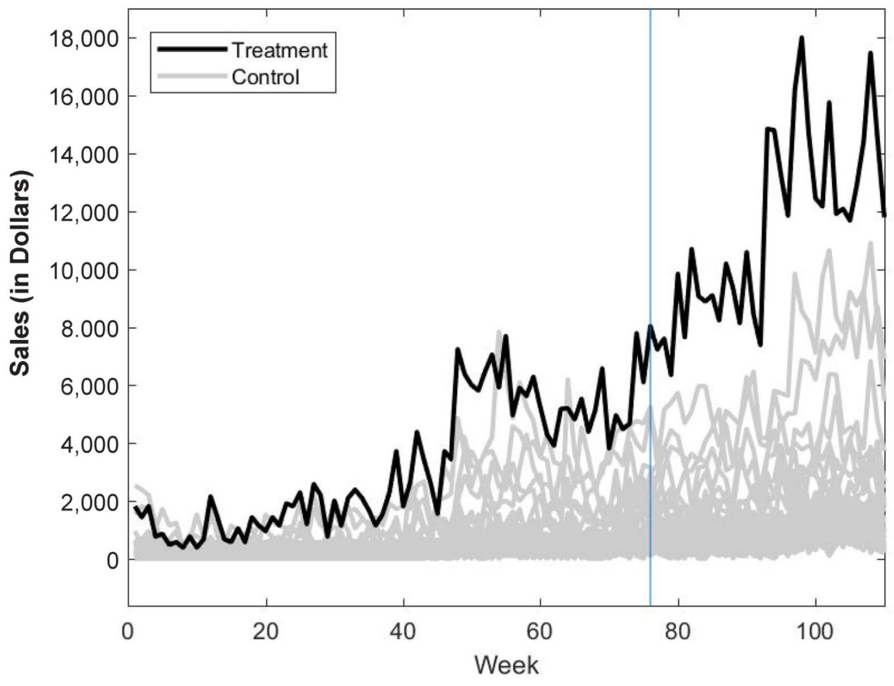
Bootstrap procedure

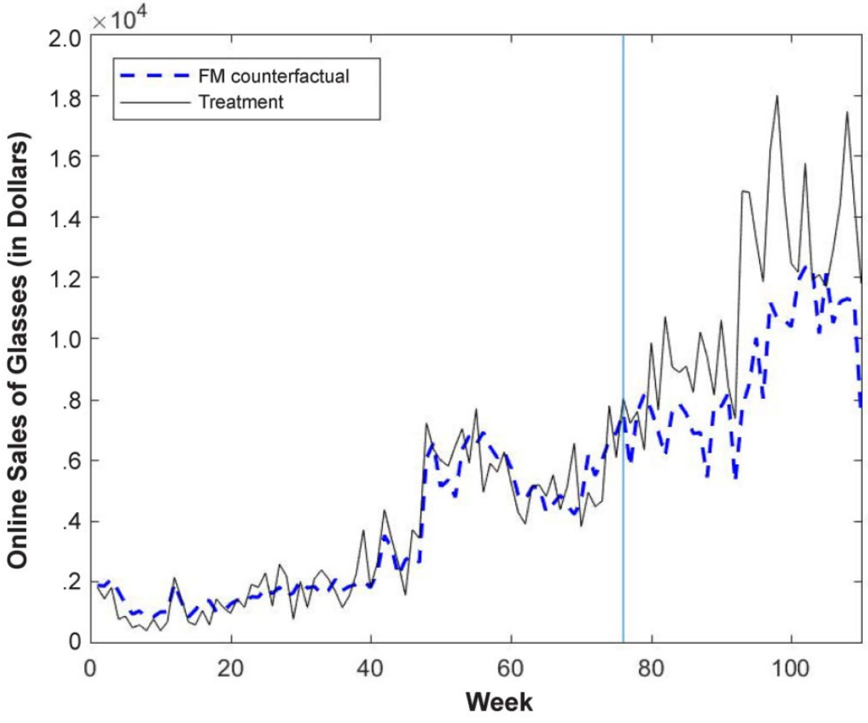


Takeaway: Our inference method recovers the coverage probabilities while extant bootstrap procedure cannot and results in under coverage when treatment error variance is greater than control units' error variance

Showroom Opening using Factor Model

Raw data shows treatment outside range of controls

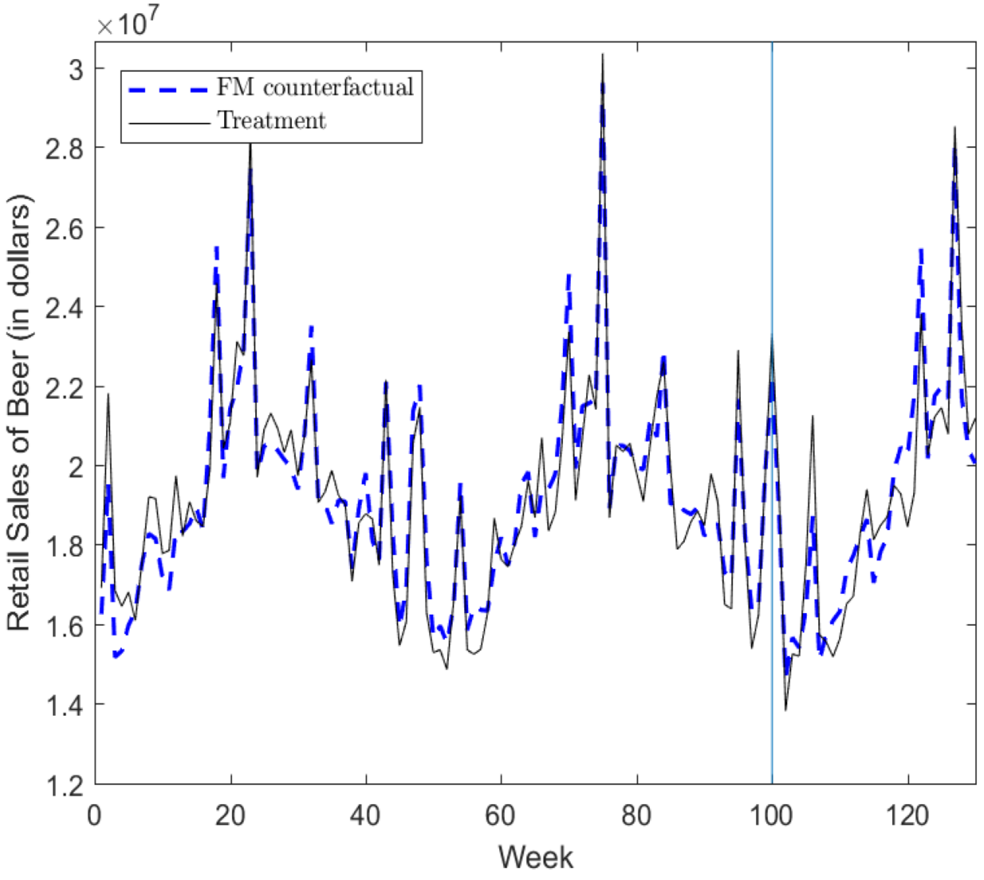




Method	ATT	ATT%	95% Confidence Interval	Ratio
Factor model	2,446	27.2%	[1,842, 3,050]	1.00
Xu bootstrap	2,446	27.2%	[2,146, 2,734]	$\boxed{.486}$

Xu bootstrap CI width are about half of CI width using our inference
likely due to under-coverage

Recreational Marijuana Legalization on Beer in California using Factor Model



Adj R squared 0.907

Retail sale of recreational marijuana in California decreases weekly beer sales by $88,400 or 0.464%, which is not statistically significant

Method	ATT	ATT%	95% Confidence Interval	Ratio
Factor model	−.8840	−.464%	[−4.074, 2.306]	1.00
Xu bootstrap	−.8840	−.464%	[−2.290, .824]	$\boxed{.488}$

Xu bootstrap CI width are about half of CI width using our inference
likely due to under-coverage

So far, we have talked about synthetic control method and the more flexible factor model method and their inference

Recall that both need long time series

What if we don’t have enough time periods?

Treatment outside range of control units

Treatment within range of control units

Augmented Difference-in-Differences

Li and Van den Bulte 2023
Marketing Science

use scaled average of control units

Forward Difference-in-Differences

Li 2023
Marketing Science

select a relevant subset of control units using forward selection algorithm and then apply DID

Augmented DID Method

(Li and Van den Bulte 2023, *Marketing Science*)

- Main idea: Use scaled average of the control units to construct the counterfactual

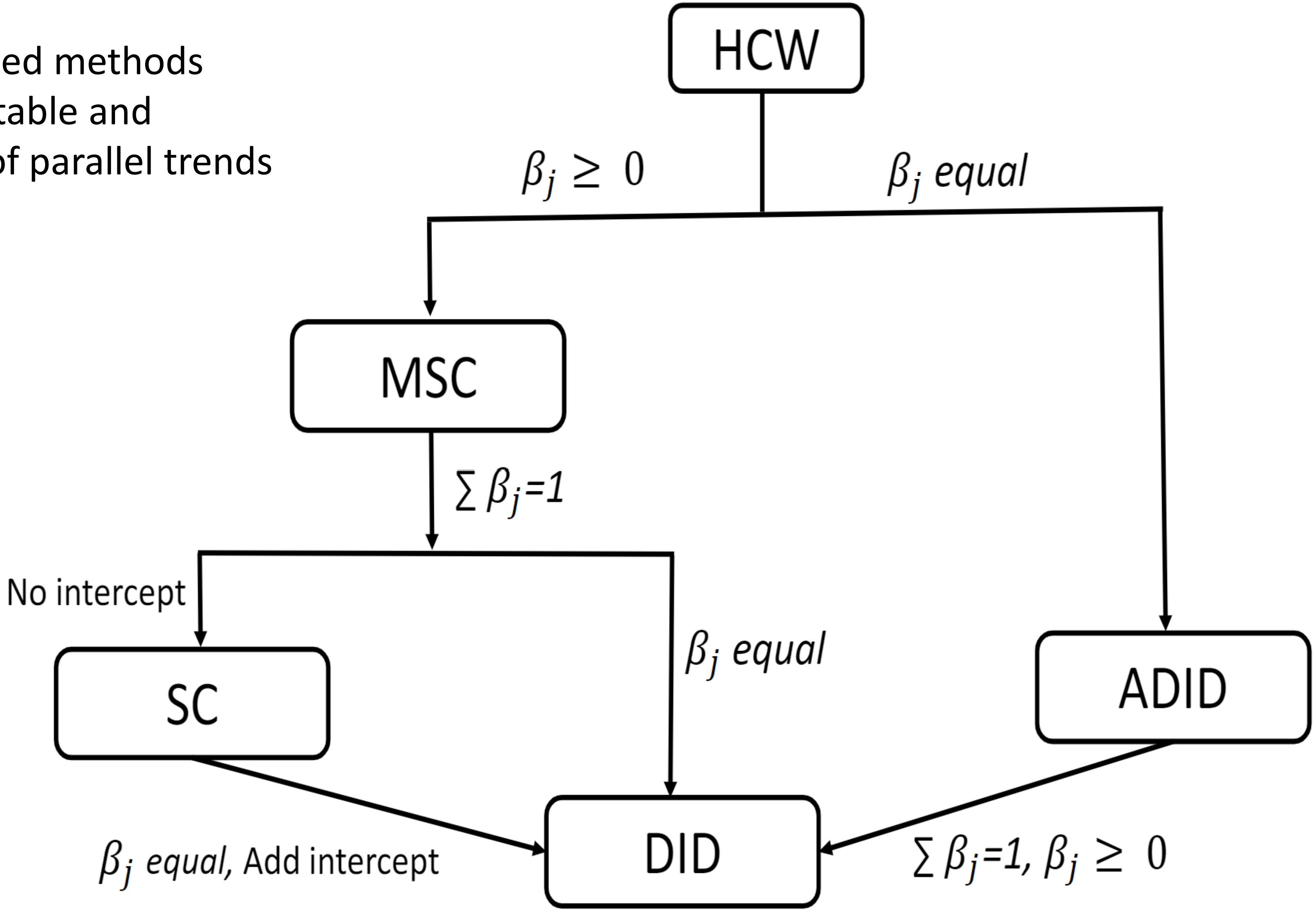
$$y^{0}_{tr,t} = \delta_1 + \delta_2 \bar{y}_{co,t} + e_{t,}$$

- Advantages:
  - Simple, easy-to-implement method
  - More flexible than DID so it can better handle heterogeneity between treatment and control units
  - Easy-to-compute confidence intervals / inference

- Using analytical proofs, simulations, and nine empirical applications, we document the attractive properties of ADID and provide guidance on what method(s) to use when.

See Section 2 of Augmented DID paper

- overview of related methods
- discussion of testable and untestable part of parallel trends assumption

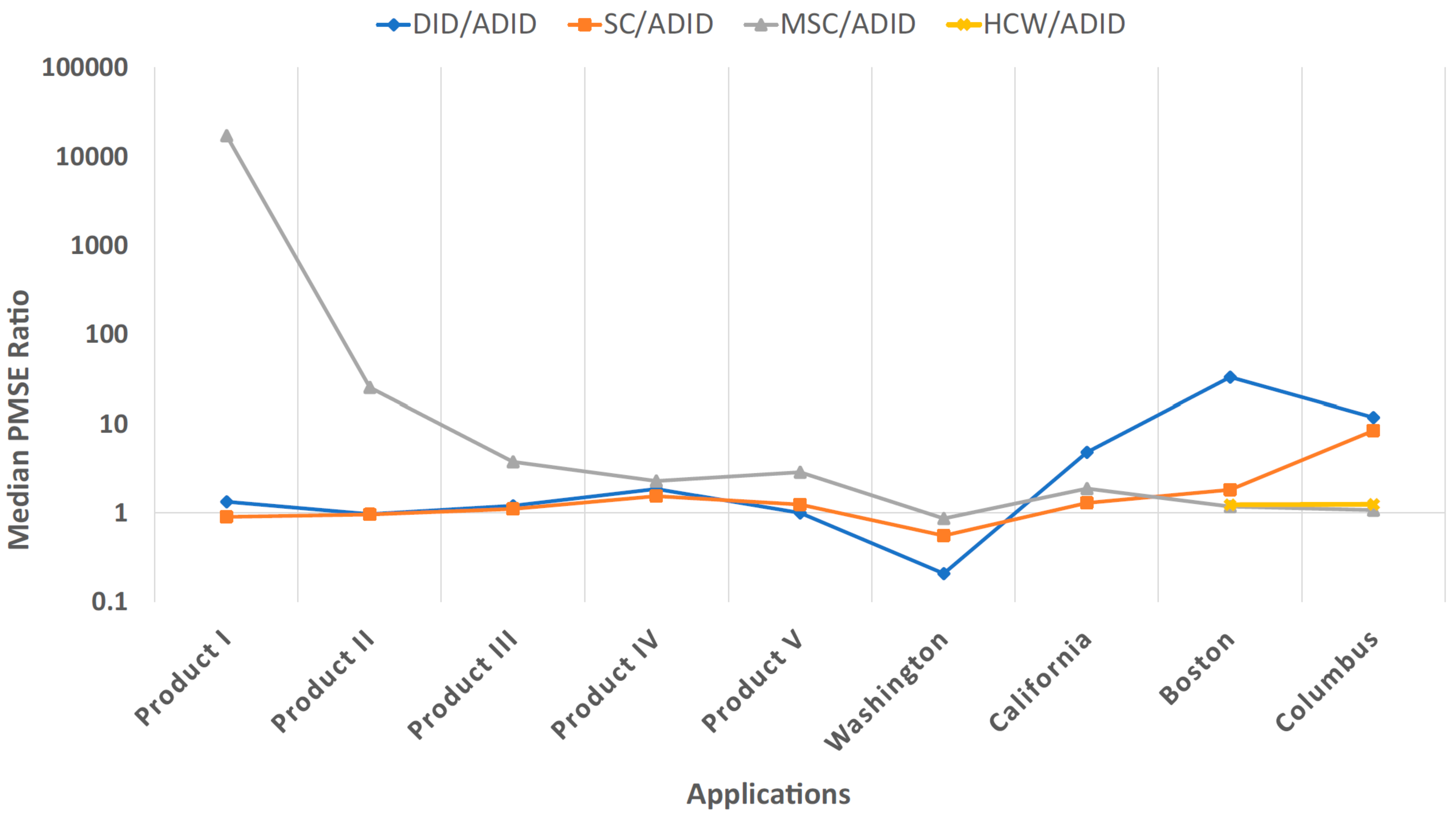


See Section 2 of Augmented DID paper

- overview of related methods
- discussion of testable and untestable part of parallel trends assumption

Identifying parallel trends assumption by method

Method	Identifying parallel trends assumption
DID	Treated unit's outcome would have been parallel to the average control units' outcomes in the absence of treatment
ADID	Treated unit's outcome would have been parallel to a slope-adjusted average of the control units' outcomes in the absence of treatment
SC	Treated unit's outcome would have been the weighted average of control units' outcome in the absence of treatment
MSC	Treated unit's outcome would have been parallel to a linear combination with nonnegative weights of the control units' outcome in the absence of treatment
HCW	Treated unit's outcome would have been parallel to a linear combination of control units' outcome in the absence of treatment



DID/ADID
SC/ADID
MSC/ADID
HCW/ADID
100000
10000
1000
100
10
1
0.1
Median PMSE Ratio
Product I
Product II
Product III
Product IV
Product V
Washington
California
Boston
Columbus
Applications

Forward Difference-in-Differences

(Li 2023, *Marketing Science*)

Forward Difference-in-Differences
(Li 2023, *Marketing Science)*

- Main idea: forward selection algorithm* to select a relevant subset of control units and then apply DID

- Advantages
  - Can accommodate any number of control units and any number of pre and post-treatment time periods
  - Solves the overfitting problem
  - Develop widely applicable normal inference theory (key advantage over other alternative flexible methods - SC, MSC, HCW)
  - Prove consistency

*Shi and Huang (2023) proposes Forward HCW method, which uses forward selection algorithm and then applies HCW method. However, Forward HCW needs large enough pre and post-treatment time periods, may still suffer from overfitting, and there is no inference theory for non-stationary data

Notation

Let $y^1_{tr,t}$ and $y^0_{tr,t}$ denote the outcome of treated unit in period *t* with and without treatment, respectively

$$ATT = \frac{1}{T_2} \sum_{t=T_1+1}^{T} (y_{tr,t} - y^0_{tr,t})$$

$$y^0_{tr,t} - \bar{y}_{N_{co},t} = \alpha + u_t, \qquad t = 1, ..., T.$$

Use pre-treatment data and OLS:

$$y_{tr,t} - \bar{y}_{N_{co},t} = \alpha + u_t, \qquad t = 1, ..., T_1$$

$$\hat{\alpha}_{N_{co}} = \frac{1}{T_1} \sum_{t=1}^{T_1} (y_{tr,t} - \bar{y}_{N_{co},t})$$

Estimated fitted in sample curve/counterfactual:

$$\hat{y}^0_{tr,t} = \hat{\alpha}_{N_{co}} + \bar{y}_{N_{co},t}, \qquad t = 1, ..., T$$

$$\widehat{ATT} = \frac{1}{T_2} \sum_{t=T_1+1}^{T} (y_{tr,t} - \hat{y}^0_{tr,t})$$

Forward DID Algorithm

- Choose the best subset (highest R^2) of size one control unit
- Then choose best subset of size two that includes first control unit
- Continue to choose greater subsets that has one more control unit than previous subset

- Out of the N_{co} models, choose the model with the highest R^2

Note: No overfitting since all submodels estimate one parameter

Store Opening on Sales

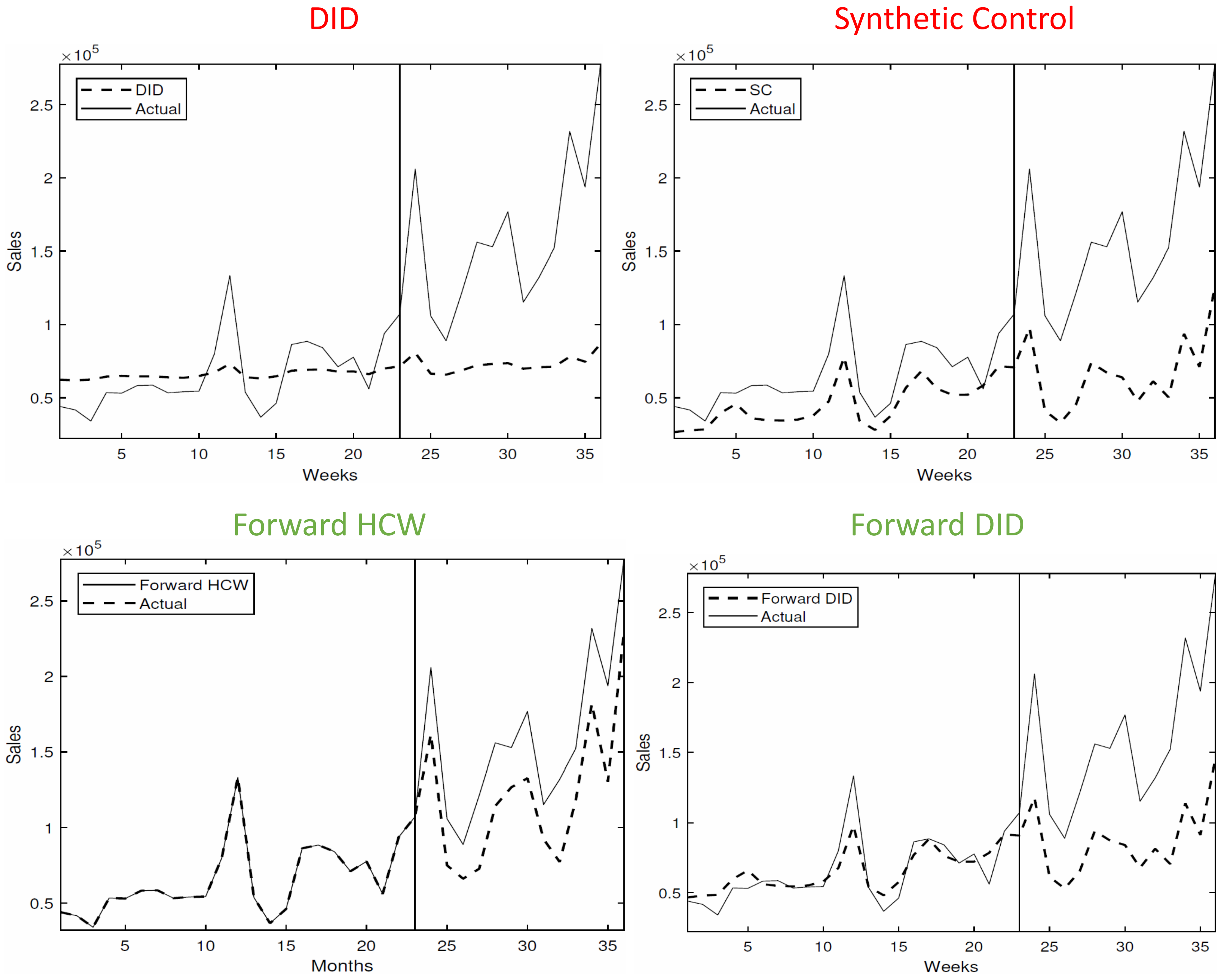
- What is the causal effect of opening a store on retailer's sales?

- January 2014 to December 2016

- Store openings
  - Atlanta, GA (December 2014)
  - San Diego, CA (January 2015)
  - San Jose, CA (January 2015)

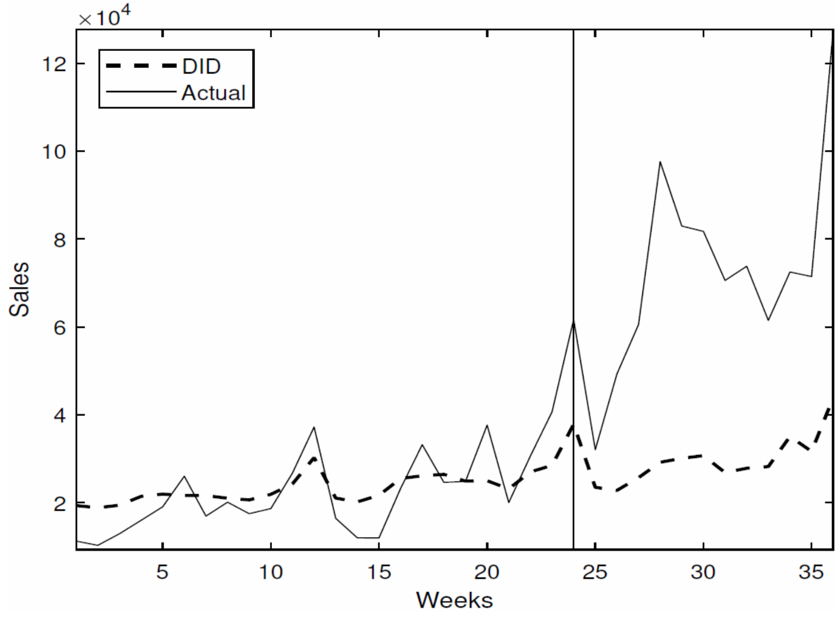
- Apply Forward DID compare to DID, SC and Forward HCW

Store Opening in Atlanta

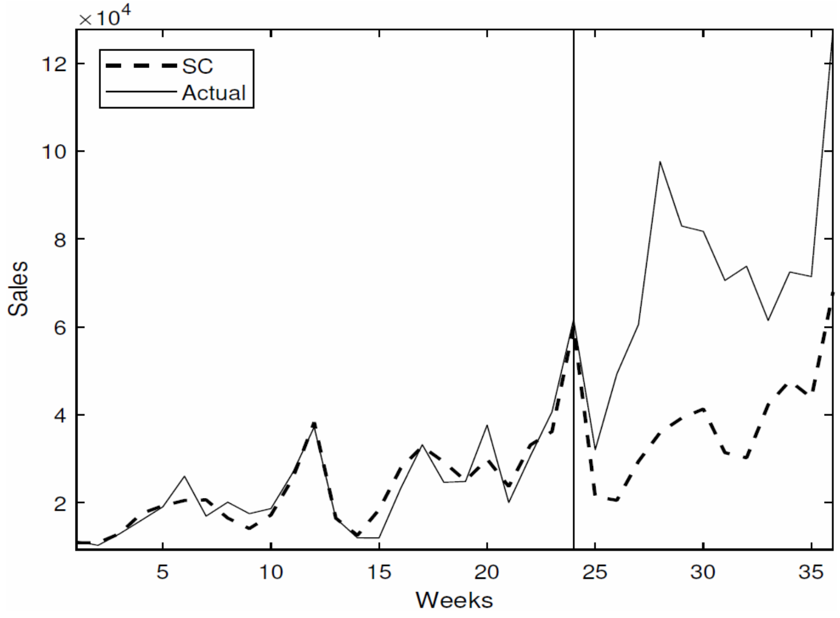


Store Opening in San Diego

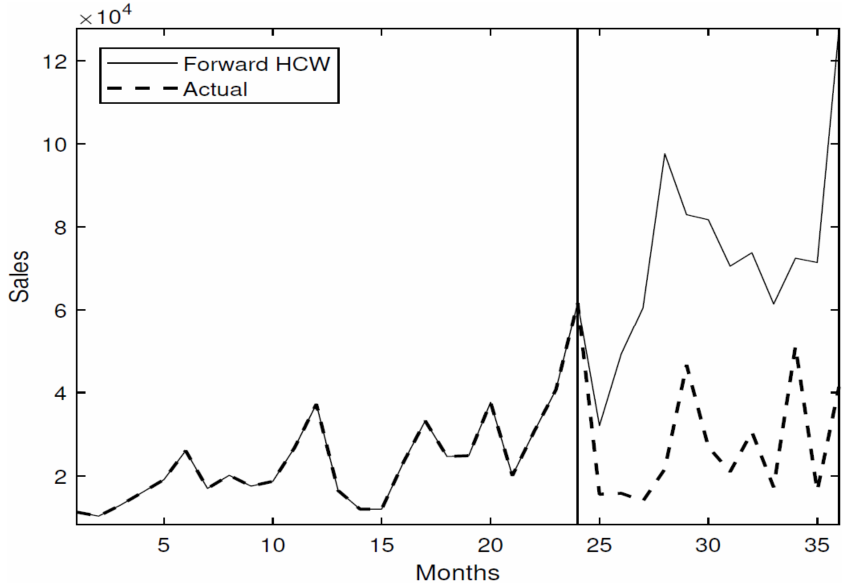
DID



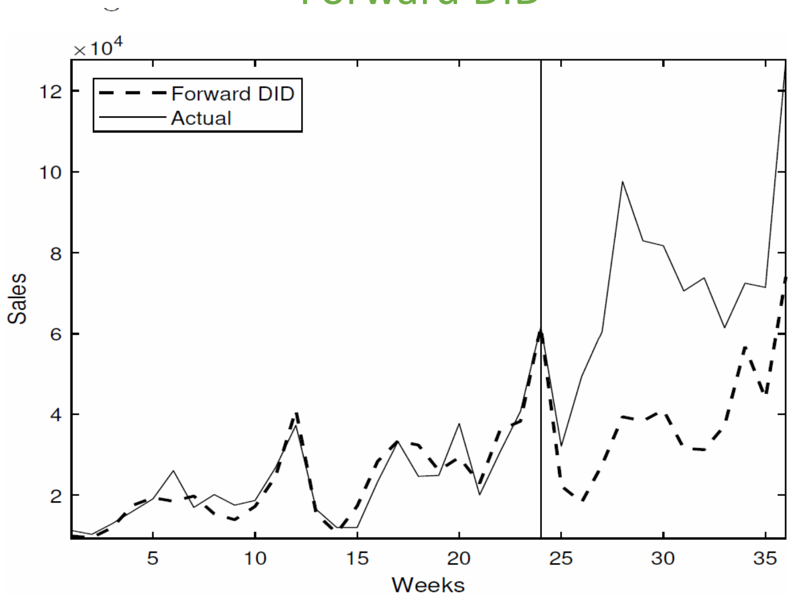
Synthetic Control



Forward HCW

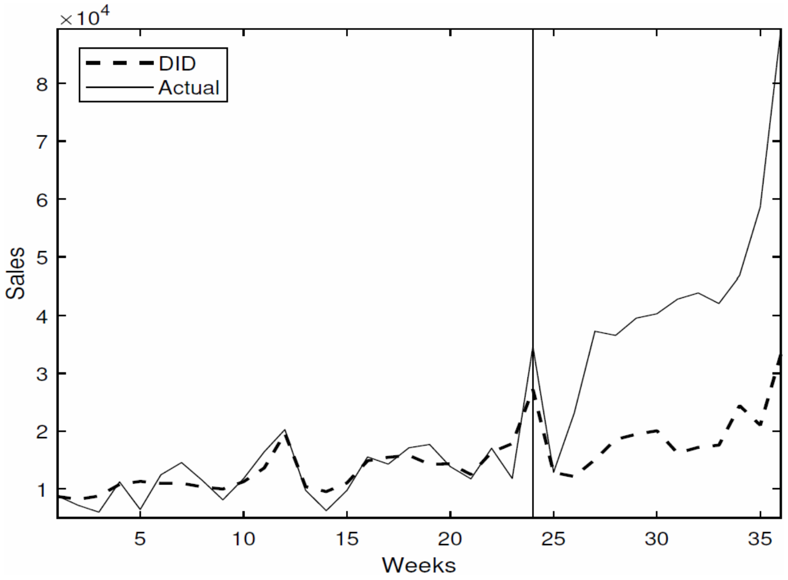


Forward DID

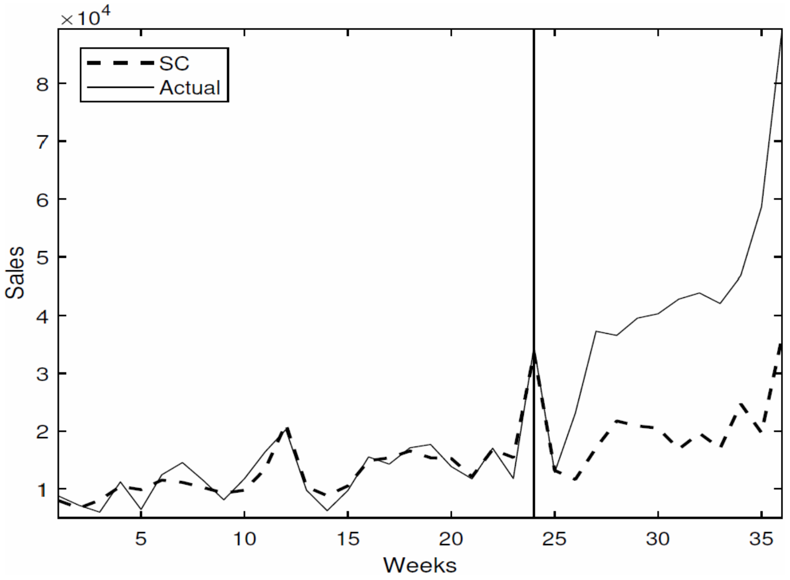


Store Opening in San Jose

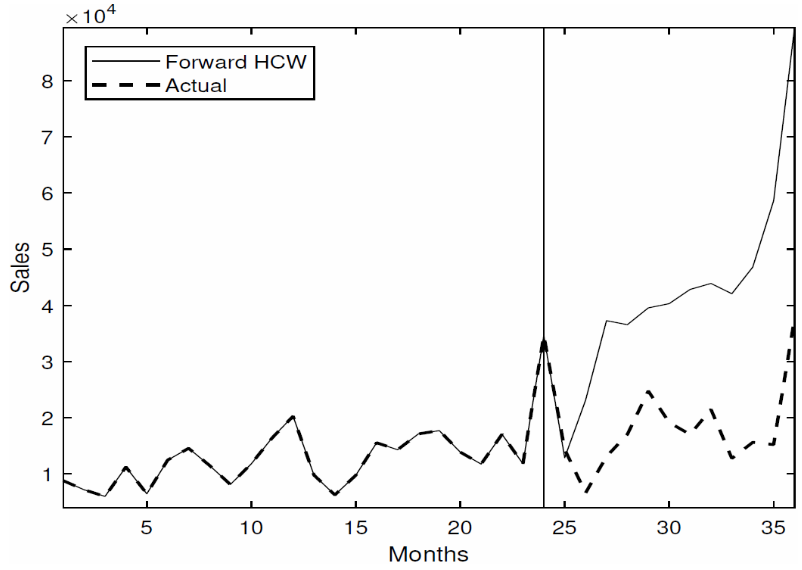
DID



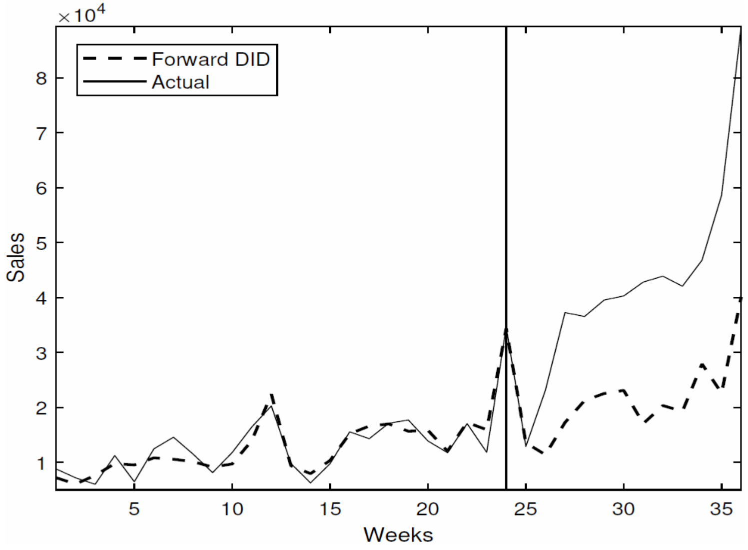
Synthetic Control



Forward HCW



Forward DID



Causal Effect (ATT) and Confidence Intervals

	ATT				ATT%				R^2			
City	DID	SC	FHCW	FDID	DID	SC	FHCW	FDID	DID	SC	FHCW	FDID
Atlanta	89,144	95,314	45,208	75,143	121	142	38.6	86.1	0.229	0.048	0.992	0.762
San Diego	43,840	35,861	47,136	35,055	148	95.4	179	91.3	0.547	0.919	0.999	0.890
San Jose	23,782	22,814	24,874	21,409	125	114	132	113	0.784	0.907	1.000	0.895

Same story as figures

	95% CI		CI Width Ratio
City	DID	FDID	DID/FDID
Atlanta	[74,882,103,402]	[67,212, 83,074]	1.80
San Diego	[38,405, 49,274]	[33,399, 38,426]	2.16
San Jose	[21,875, 25,690]	[21,331, 23,998]	1.43

Forward DID has narrower confidence intervals

We need out-of-sample comparisons by method

Out-of-sample PMSE using Backdating

City	Atlanta			San Diego			San Jose		
T_0/Ratio	$\frac{DID}{FDID}$	$\frac{SC}{FDID}$	$\frac{FHCW}{FDID}$	$\frac{DID}{FDID}$	$\frac{SC}{FDID}$	$\frac{FHCW}{FDID}$	$\frac{DID}{FDID}$	$\frac{SC}{FDID}$	$\frac{FHCW}{FDID}$
12	2.28	2.36	0.461	3.14	1.03	3.99	0.358	0.817	2.12
14	3.08	7.85	5.26	2.59	0.977	3.12	0.705	1.36	3.54
16	4.41	9.51	2.79	3.16	1.41	9.50	0.468	1.47	3.25
18	2.26	7.57	3.90	2.59	1.39	9.11	0.428	1.17	4.82
20	1.60	5.50	1.60	2.54	0.798	3.66	0.453	0.740	1.94

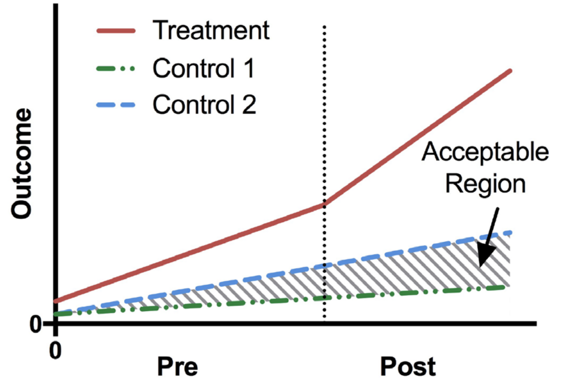
All ratios except one greater than one so Forward DID performs better out of sample

Recap: Augmented DID and Forward DID are especially useful when…

- DID is not applicable so we need more flexible estimators
- We have short (although still applies to long) pre and post treatment time periods
- Note: Unlike SC (which uses a weighted average), both of these methods use a simple average of control units, which facilitates straightforward inference

Boundary condition:

Forward DID should not be used when it identifying parallel trends assumption is violated (e.g. treatment is outside the range of the control units – can check graphically in data before analysis)



Difference-in-Differences

Synthetic Control

More flexible estimators

Less flexible
(more restrictive)

More flexible
(less restrictive)

Use the most restrictive method that is still correct (i.e. parallel trends identifying assumption holds)

Two Step Synthetic Control, Li and Shankar 2023
Management Science

1. Formal test of (pre)parallel trends assumption
2. Framework to recommend appropriate method that balances bias and variance

How to arrive at more flexible estimators from synthetic control method?

1) Change the weights or add an intercept
   Doudchenko and Imbens 2016, Li 2020
   Hsiao Ching and Wan 2012
   Augmented DID Li and Van den Bulte 2023, *Mkt Sci*
   Synthetic DID (introduce time weights) Arkhangelsky et al 2021 *AER*
2) Dimension reduction via factor model aka generalized synthetic control
   Xu 2017, Inference - Li and Sonnier 2023, *JMR*
3) Choose a relevant subset of control units
   Shi and Huang 2022, Li 2023
4) Bayesian
   Kim, Lee and Gupta 2020 *JMR*, Pang, Liu and Xu 2022

Data characteristics of research problem guide the choice of method

- number of pre/post time periods
- number of treatment and control units
- treatment unit within or outside range of control units

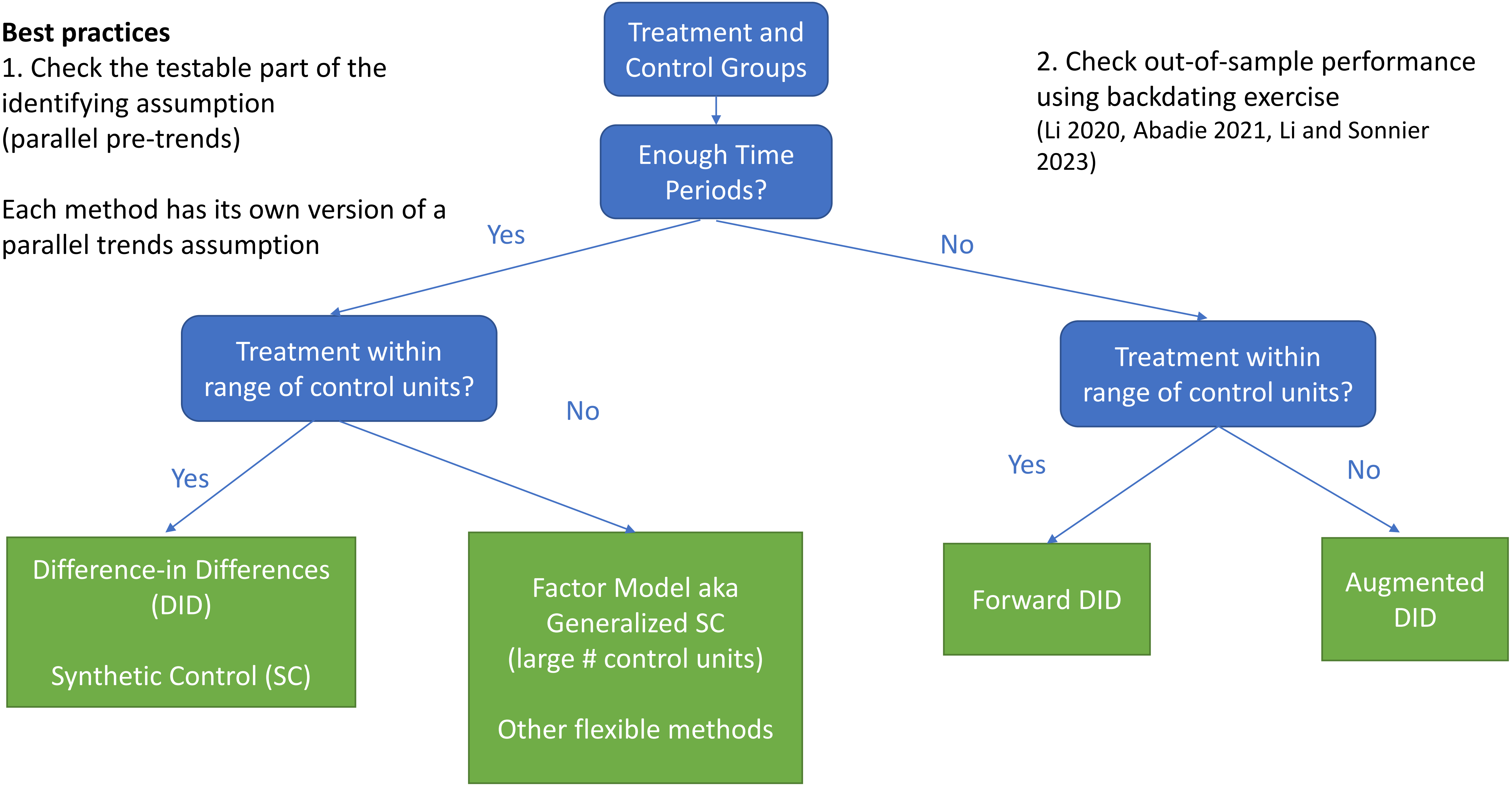
Goal: Use the simplest (i.e. most restrictive) method that has parallel pre-trends without overfitting

Best practices

1. Check the testable part of the identifying assumption (parallel pre-trends)

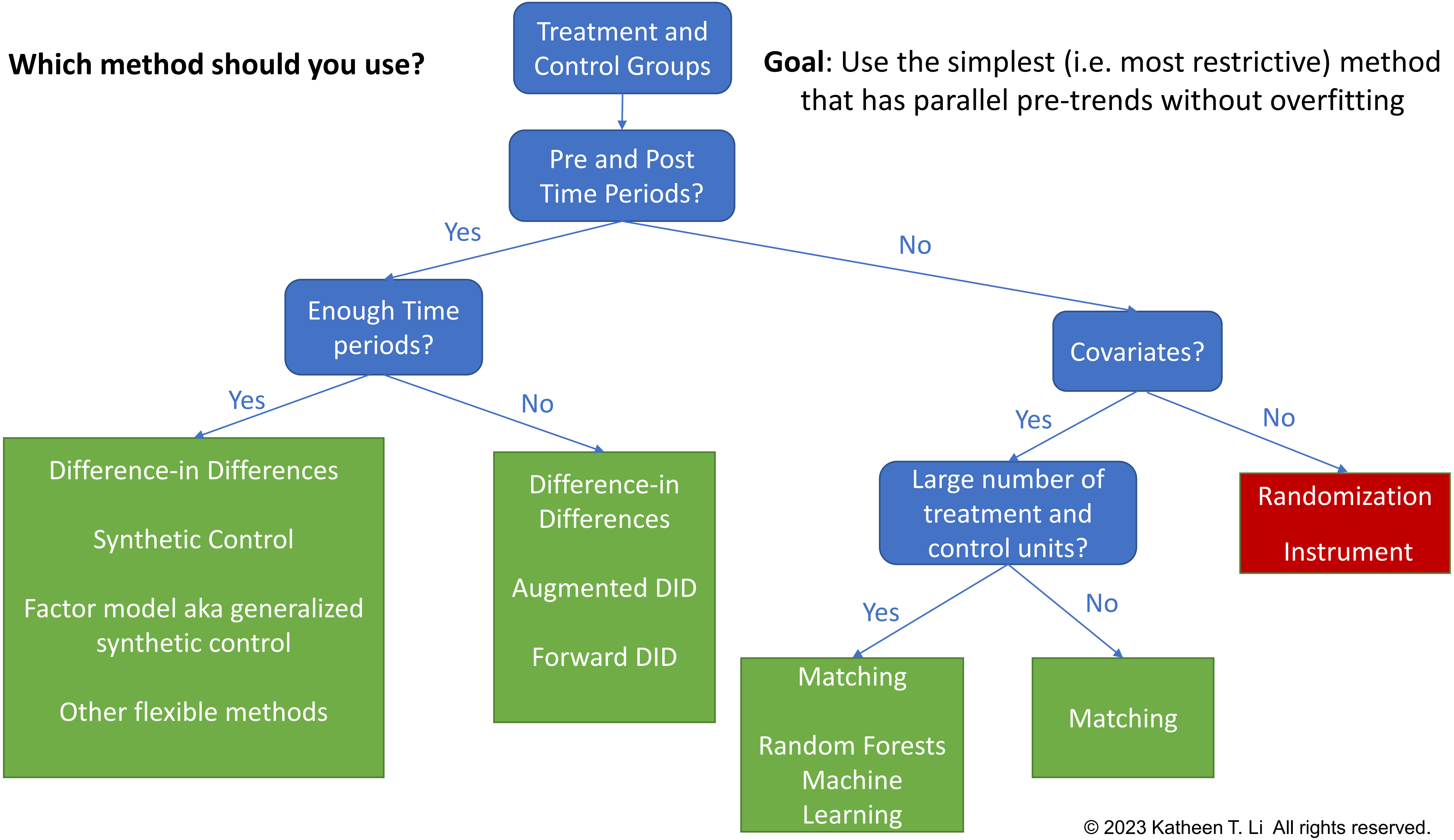
Each method has its own version of a parallel trends assumption

2. Check out-of-sample performance using backdating exercise (Li 2020, Abadie 2021, Li and Sonnier 2023)



Which method should you use?

Goal: Use the simplest (i.e. most restrictive) method that has parallel pre-trends without overfitting



Conclusion

- Causal inference is fundamental to science and knowledge generation
- DID is most popular, widely used quasi-experimental method
- Synthetic control method started the renaissance in more flexible estimators (good for small to moderate # treatment, large enough time periods)
- When to use which method is guided by data characteristics / satisfying corresponding identifying assumption
- Expanding the in causal inference toolkit available to marketing scholars (factor model, Bayesian, Augmented DID, Forward DID) allows researchers to answer previously unanswerable questions!

Thank you!

kathleen.li@mccombs.utexas.edu

- Abadie, Alberto (2021) "Using synthetic controls: Feasibility, data requirements, and methodological aspects." *Journal of Economic Literature* 59(2):391-425.
- Abadie, Alberto, Alexis Diamond, and Jens Hainmueller (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 (2010): 493-505.
- Abadie, Alberto, and Javier Gardeazabal (2003) "The economic costs of conflict: A case study of the Basque Country." *American Economic Review* 93(1):113-132.
- Ben-Michael, Eli, Avi Feller, and Jesse Rothstein (2021) "The augmented synthetic control method." *Journal of the American Statistical Association* 116(536):1789-1803.
- Carvalho, Carlos, Ricardo Masini, and Marcelo C. Medeiros (2018) "ArCo: An artificial counterfactual approach for high-dimensional panel time-series data." *Journal of Econometrics* 207(2): 352-380.
- Chan, Marc and Simon Kwok (2016) "Policy Evaluation with Interactive Fixed Effects," working paper, School of Economics, University of Sydney.
- Doudchenko, Nikolay, and Guido W. Imbens (2016) "*Balancing, regression, difference-in-differences and synthetic control methods: A synthesis*. No. w22791. National Bureau of Economic Research"
- Gobillon, Laurent, and Thierry Magnac (2016) "Regional policy evaluation: Interactive fixed effects and synthetic controls." *Review of Economics and Statistics* 98(3):535-551.
- Hsiao, Cheng, H. Steve Ching, and Shui Ki Wan (2012) "A panel data approach for program evaluation: measuring the benefits of political and economic integration of Hong Kong with mainland China." *Journal of Applied Econometrics* 27(5): 705-740.
- Kim, Sungjin, Clarence Lee, and Sachin Gupta (2020) "Bayesian synthetic control methods." *Journal of Marketing Research* 57(5):831-852. https://doi.org/10.1177/002224372093623
- Pang, Xun, Licheng Liu, and Yiqing Xu (2022) "A Bayesian alternative to synthetic control for comparative case studies." *Political Analysis* 30(2):269-288.
- Shi, Zhentao and Huang, Jingyi (2023) "Forward-selected panel data approach for program evaluation." *Journal of Econometrics.* 234(2):512-535.
- Xu, Yiqing (2017) "Generalized synthetic control method: Causal inference with interactive fixed effects models." *Political Analysis* 25(1): 57-76.

- Li, Kathleen T. and Christophe Van den Bulte (2023), "Augmented Difference-in-Differences" *Marketing Science* https://doi.org/10.1287/mksc.2022.1406
- Li, Kathleen T. and Garrett Sonnier (2023), "Statistical Inference for the Factor Model Approach to Estimate Causal Effects in Quasi-Experimental Settings," *Journal of Marketing Research* https://doi.org/10.1177/00222437221137533
- Li, Kathleen T. (2020), "Statistical Inference for Average Treatment Effects Estimated by Synthetic Control Methods," *Journal of the American Statistical Association*, 115(532): 2068-2083. https://doi.org/10.1080/01621459.2019.1686986
- Li, Kathleen T. and David R. Bell (2017), "Estimation of Average Treatment Effects with Panel Data: Asymptotic Theory and Implementation," *Journal of Econometrics*, 197(1): 65-75. https://doi.org/10.1016/j.jeconom.2016.01.011
- Li, Kathleen T. (2023) "A Simple Forward Differences-in-Differences Method" *Marketing Science https://doi.org/10.1287/mksc.2022.0212*
- Li, Kathleen T. and Venkatesh Shankar (2023), "A Two-Step Synthetic Control Approach for Estimating Causal Effects of Marketing Events," *Management Science* https://doi.org/10.1287/mnsc.2023.4878