Randomization Tests in Observational Studies with Staggered Adoption of Treatment

Panos Toulis
panos.toulis@chicagobooth.edu

Econometrics and Statistics
University of Chicago, Booth School of Business

Joint work with Azeem Shaikh (U Chicago, Economics)

Conference on Research on Economic Theory and Econometrics, Naxos, 2023

Introduction

We consider the problem of inference in observational studies in which units adopt treatment at varying times.

Problems of this type relate to panel data and widely used methods, such as difference-in-differences (Snow, 1855; Card and Krueger, 1993) and *synthetic controls* (Abadie and Gardeazabal, 2003; Abadie et al., 2010; Abadie, 2019).

The majority of these methods rely on specification of the outcome model; see Abadie and Cattaneo (2018).

We propose a method that exploits only variation in treatment adoption (i.e., design-based). The method relies on randomization tests and is robust.

Contributions

- In contrast to earlier literature, our method exploits the availability of multiple units that adopt treatment at different times.
- We make a novel connection between such settings and survival analysis in statistics.
- We develop a randomization test for the null hypothesis of no treatment effect for all units and time periods. We prove that our test is valid asymptotically when the treatment model satisfies the *proportional hazards* condition.
- Our test can be robust to misspecification of the treatment model under certain symmetry assumptions on the test statistic (reminiscent of “doubly robust” methods).

Setup

- Unit indexed by $i = 1, \ldots, n$ and time indexed by $t \in [0, \infty)$. Observations are censored at t_{max}.
- T_i = time at which unit i adopts treatment (at most once).
  [T_i not observed if $T_i > t_{\text{max}}$.]
- X_{it} = covariates of unit i observed at times $t = 1, \ldots, t_{\text{max}}$.
- $Y_{it}(r)$ = potential outcome of unit i at times $t = 1, \ldots, t_{\text{max}}$ under $T_i = r$.
  [X,Y observed at "lower frequency" than T; see obs. frequencies ▷]

Observed outcomes therefore satisfy:

$$Y_{it} = Y_{it}(T_i). \tag{1}$$

Shorthand notation: $\boldsymbol{X} = \{X_{it} : i, t\}$, $\boldsymbol{Y} = \{Y_{it} : i, t\}$ and $\boldsymbol{T} = \{T_i : i\}$.

Also, $\mathbb{Y} = \{Y_{it}(r) : r, i, t\}$ denotes the full schedule of potential outcomes.

Goal

We want to test the “no effect” hypothesis:

$$H_0 : Y_{it}(r) = Y_{it}(r'), \text{for all } r, r' \in [0, \infty), i = 1, \ldots, N,\ t = 1, \ldots, t_{\text{max}}.$$

Goal

We want to test the "no effect" hypothesis:

$$H_0 : Y_{it}(r) = Y_{it}(r'), \text{for all } r, r' \in [0, \infty), i = 1, \ldots, N,\ t = 1, \ldots, t_{\text{max}}.$$

We make an unconfoundedness assumption on $\mathbb{Y}$:

$$\mathbb{Y} \perp\!\!\!\perp \boldsymbol{T} \mid \boldsymbol{X}. \tag{A1}$$

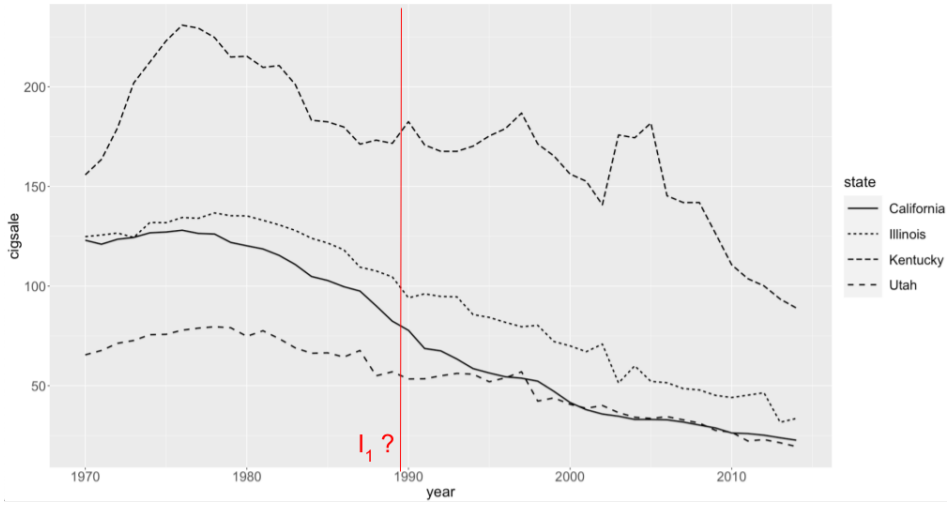
Under H_0 and (A1), focus shifts to $p(\boldsymbol{T} \mid \boldsymbol{X})$ — the "propensity score".

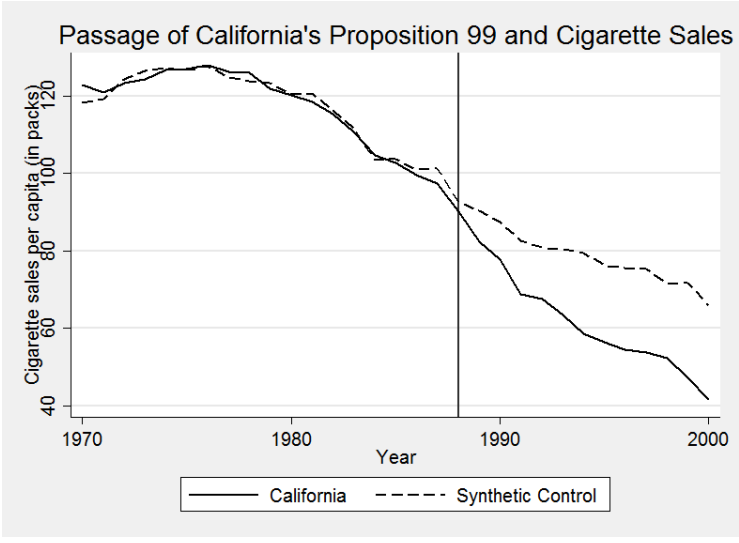
Knowledge of $p(.)$ immediately leads to a valid randomization test based on the p-value:

$$\text{pval} = \mathsf{E}\big\{ S_n(\boldsymbol{T}'; \boldsymbol{Y}, \boldsymbol{X}) \geq S^{\text{obs}} \mid \boldsymbol{Y}, \boldsymbol{X} \big\},\ \boldsymbol{T}' \sim p(\boldsymbol{T}' \mid \boldsymbol{X}),$$

where $S^{\text{obs}} = S_n(\boldsymbol{T}; \boldsymbol{Y}, \boldsymbol{X})$ is the observed value of the test statistic.

Illustration — Tobacco legislation of Abadie et al. (2010)



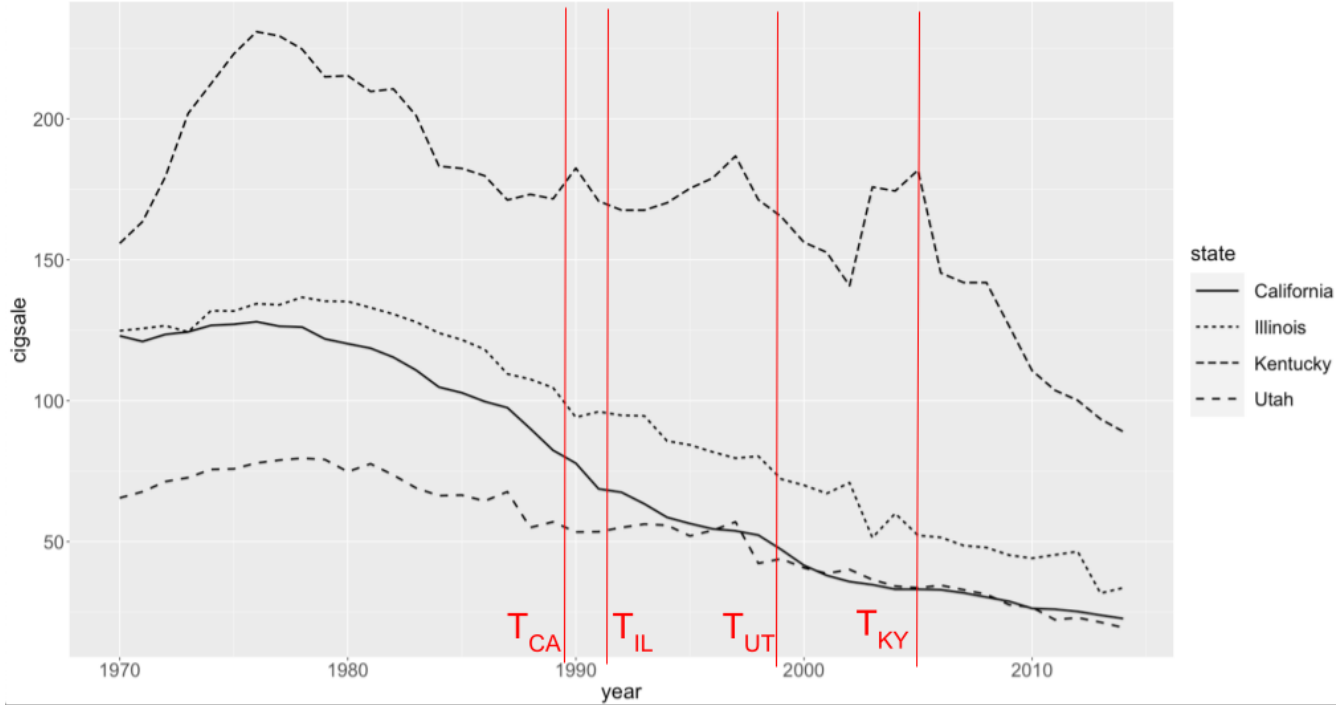


The "synthetic control" method of Abadie et al (2010) constructs synthetic data predicting the outcomes for the treated *had the unit been in control*.

Inference was based on randomizing the identity of the **first** treated unit (I_1).

This assumes that $p(\boldsymbol{T} \mid \boldsymbol{X})$ is uniform, which is unrealistic in many settings.

Illustration — Tobacco legislation of Abadie et al. (2010)



In general, the randomization test needs to shuffle the red lines along the time axis according to a *plausible* $p(\boldsymbol{T} \mid \boldsymbol{X})$.

So, what can we say about $p(\boldsymbol{T} \mid \boldsymbol{X})$? (generally unknown)

Model for $p(\boldsymbol{T} \mid \boldsymbol{X})$ — Connection to survival analysis

- Natural connection to event data models and survival analysis.
  [$T_i < t_{\text{max}}$ = event, T_i = event time].
- One of the most widely used event models relies on proportional hazards.
- This is a *design-based* approach to inference: Exploit only variation in $\boldsymbol{T}$.

Proportional hazards

Assume:

(a) $T_1, \ldots, T_n$ are independent conditional on $\boldsymbol{X}$.

(b*) T_i depends only on $\boldsymbol{X}_i = \{X_{it} : t = 1, \ldots, t_{\text{max}}\}$. [despite different observation frequencies].

(c) $T_i \mid \boldsymbol{X}_i$ has density such that

$$\lim_{\delta \to 0} P\{T_i \in [t, t+\delta] \mid T_i \geq t, \boldsymbol{X}_i\} = \lambda(t) \exp(X_{it}' \beta). \tag{2}$$

Here, $\beta \in \mathbf{R}^d$ are fixed parameters; $\lambda(t)$ is the *baseline hazard* rate. This rate may be unknown.

The key restriction of (2) is that the hazard rate of an event at t is decomposed between a temporal term and covariate term. May fail if these two terms actually interact.

Eq. (2) leads to the celebrated Cox proportional hazards model.

Cox proportional hazards model

Under PH, we obtain the partial likelihood:

$$L(\beta) = \prod_{i=1}^{n} \left(\frac{\exp(X'_{iT_i}\beta)}{\sum_{j \in \mathcal{R}_i} \exp(X'_{jT_i}\beta)} \right)^{1-\delta_i}. \tag{3}$$

- $\delta_i = \mathbf{1}(T_i > t_{\text{max}})$ indicates censorship.
- $\mathcal{R}_i = \{j : T_j \geq T_i\}$ is the *risk set* at T_i (units "competing" with i for an event at T_i).

We will assume that the maximand of (3) leads to a consistent estimator of β:

$$\hat{\beta}_n \xrightarrow{p} \beta. \tag{A2}$$

Sufficient conditions can be found in many standard textbooks; e.g., (Andersen and Gill, 1982).

Our approach

As Abadie et al (2010), we condition on $T_{(1)} = \min_i T_i$, the first treatment event. Let $I_1 = \arg\min_i T_i$ be the first treated unit.

This is useful because under H_0 and (A1), we have:

$$\boldsymbol{Y} \perp\!\!\!\perp I_1 \mid T_{(1)}, \boldsymbol{X}.$$

Moreover,

$$P(I_1 = i \mid T_{(1)}, \boldsymbol{X}) = \frac{\exp(X'_{iT_{(1)}}\beta)}{\sum_{j=1}^{n} \exp(X'_{jT_{(1)}}\beta)} \triangleq \omega_i(\beta).$$

We can then randomize the identity of the first treated unit conditional on $T_{(1)}$:

$$\text{pval} = \sum_{i=1}^{n} \omega_i(\beta) \mathbf{1}\{S_n(i, T_{(1)}; \boldsymbol{Y}, \boldsymbol{X}) \geq S^{\text{obs}} \mid T_{(1)}, \boldsymbol{Y}, \boldsymbol{X}\}$$

leads to an (infeasible) exact test.

Concrete procedure

(1) Using $(\boldsymbol{X}, \boldsymbol{Y})$, fit Cox proportional hazards model and get $\hat{\beta}_n$.

(2) Calculate weights $\omega_i(\hat{\beta}_n)$.

(4) Define a test statistic to depend on $\boldsymbol{T}$ only through $I_1, T_{(1)}$.

(3) Calculate the p-value:

$$\text{pval} = \sum_i \omega_i(\hat{\beta}_n) \mathbf{1}\{S_n(i, T_{(1)}; \boldsymbol{Y}, \boldsymbol{X}) \geq S^{\text{obs}} \mid T_{(1)}, \boldsymbol{Y}, \boldsymbol{X}\}.$$

Theorem

Under PH and (A1) and (A2), the above procedure implies an asymptotically valid test for H_0.

Concrete procedure

(1) Using $(\boldsymbol{X}, \boldsymbol{Y})$, fit Cox proportional hazards model and get $\hat{\beta}_n$.

(2) Calculate weights $\omega_i(\hat{\beta}_n)$.

(4) Define a test statistic to depend on $\boldsymbol{T}$ only through $I_1, T_{(1)}$.

(3) Calculate the p-value:

$$\text{pval} = \sum_i \omega_i(\hat{\beta}_n) \mathbf{1}\{S_n(i, T_{(1)}; \boldsymbol{Y}, \boldsymbol{X}) \geq S^{\text{obs}} \mid T_{(1)}, \boldsymbol{Y}, \boldsymbol{X}\}.$$

Theorem

Under PH and (A1) and (A2), the above procedure implies an asymptotically valid test for H_0.

Notes:

- First known result on validity of randomization tests under estimated propensity score.
- The PH assumption is testable; e.g., (Xue and Schifano, 2017).
- When $\omega_i = 1/n$, we obtain a "uniform test", which is equivalent to the procedure employed in (Abadie et al., 2010) for synthetic controls.
- Downside: We "throw away" information from other treatments. Could we condition on multiple treatment events ▷ ?

Simulation

Treatment model:

$$X_i \sim U(-10, 10),\ X_{it} = X_i,$$
$$T_i | X_i \sim \text{Exp}(\lambda_i),\ \lambda_i = \exp(X_i \beta),\ \beta = 1. \tag{4}$$

Outcome model:

$$\tilde{Y}_{it}(0) = \rho \tilde{Y}_{i,t-1}(0) + \delta \sqrt{t} + \gamma X_i + \epsilon_{it}$$
$$\tilde{Y}_{it}(1) = \tau + \tilde{Y}_{it}(0).$$

The full schedule of potential outcomes, $\mathbb{Y} = \{Y_{it}(r) : r, i, t\}$ depends only on whether treatment has been adopted at t, such that

$$Y_{it}(r) = \tilde{Y}_{it}(0) + \mathbf{1}(r \leq t)(\tilde{Y}_{it}(1) - \tilde{Y}_{i,t}(0)). \tag{5}$$

- We set $\epsilon_{it} \sim N(0, \sigma^2)$. Assumptions (A1) and (A2) satisfied.
- PH assumption satisfied.

For simplicity, we use the "diff-in-diff" test statistic:

$$S_n = \text{Avg}_{t \geq T_{(1)}} [Y_{I_1 t} - \text{Avg}_{j \neq I_1} (Y_{jt})] - \text{Avg}_{t < T_{(1)}} (Y_{I_1 t} - \text{Avg}_{j \neq I_1} (Y_{jt})].$$

Under the particular model assumptions (and $H_0 : \tau = 0$):

$$S_n \propto \gamma(\bar{X} - X_{I_1}) + \eta_{I_1} - g(\boldsymbol{\eta}), \tag{6}$$

where $\boldsymbol{\eta} \in \mathbf{R}^n$ depends only on ϵ_{it}, and is exchangeable normal r.v.

Eq. (6) shows how $p(I_1 \mid T_{(1)}, \boldsymbol{X})$ affects the randomization distribution of S_n, as long as $\gamma \neq 0$.

We consider three different randomization tests:

(1) "uniform" test: We use $p(I_1 | \ldots) \propto$ const. [misspecified]

(2) "infeasible" test: We use correct $p(I_1 | \ldots)$ based on true Cox model.

(3) "feasible" test: We use estimated $p(I_1 | \ldots)$ from true Cox model.

n	γ	uniform	feasible	infeasible
25	0.00	4.96	5.11	5.00
	0.50	14.26	4.62	4.97
	1.00	17.06	4.48	5.02
	2.00	18.32	4.46	5.06
	5.00	19.40	4.46	5.07
50	0.00	5.10	5.03	5.11
	0.50	14.30	4.82	4.98
	1.00	17.07	4.77	4.95
	2.00	19.22	4.58	4.91
	5.00	20.73	4.77	5.04
100	0.00	4.88	5.07	5.00
	0.50	14.01	4.85	4.99
	1.00	17.12	4.79	5.06
	2.00	19.71	4.85	4.95
	5.00	21.00	4.79	5.11

Table: Rejection rates (%) of 'uniform', 'feasible' and 'infeasible' tests under H_0.

Power ▷ Choice of test statistic ▷

Misspecification

Theorem

Let $P_{n,i}$ = P(reject H_0 | i is treated). Suppose there exists a sequence $j_n \in \{1, \ldots, n\}$ such that, under H_0,

$$\delta_{j_n}(\boldsymbol{T}, \boldsymbol{X}) = \max_i |P_{n,i} - P_{n,j_n}| = o(1). \tag{7}$$

Then, our test is valid asymptotically even when PH does not hold.

Sufficient conditions for (7) include a form of exchangeability of the test statistic wrt to i under H_0 (see paper).

So, our test is valid (asymptotically) when

(a) the Cox treatment model is well-specified, or

(b) when the treatment model is misspecified but the test statistic can be appropriately constructed (e.g., through knowledge of the outcome model).

Reminiscent of *doubly robust* methods.

Simulation under misspecification

We use a different specification for the treatment (known as "accelerated failure"):

$$T_i = \exp(-m(X_i) + \epsilon_i) \ , \tag{8}$$

where $m(x) = 1 - \text{logistic}(2k_1(x+8)) + \text{logistic}(2k_2(x-8))$, $\epsilon_i \sim N(0, 0.4^2)$.

We can control misspecification through (k_1, k_2):

- If (k_1, k_2) = (0, 0), then PH holds. [moreover, T_i are i.i.d.]
- If (k_1, k_2) = (large, small) or (small, large), PH does not hold but misspecification is not too severe.
- If (k_1, k_2) = (large, large), then misspecification is severe.

			$n = 25$		$n = 50$	
k_1	k_2	γ	uniform	feasible (Cox)	uniform	feasible
0.00	0.00	0.00	5.10	5.08	5.01	4.98
0.00	0.00	2.00	4.99	4.77	5.01	4.99
0.00	0.00	5.00	4.93	4.40	4.94	4.69
2.00	0.00	0.00	5.03	3.28	5.10	3.36
2.00	0.00	2.00	8.32	4.57	10.25	4.93
2.00	0.00	5.00	13.98	6.85	17.63	7.98
1.00	1.00	0.00	5.04	4.51	4.98	4.58
1.00	1.00	2.00	8.42	7.57	9.98	9.62
1.00	1.00	5.00	14.51	13.50	17.62	17.61
0.00	2.00	0.00	4.99	3.40	5.04	3.36
0.00	2.00	2.00	8.40	4.48	10.28	4.81
0.00	2.00	5.00	13.63	6.78	17.57	7.93
2.00	2.00	0.00	4.98	4.45	4.94	4.57
2.00	2.00	2.00	8.59	7.91	10.48	10.18
2.00	2.00	5.00	14.62	13.89	17.81	17.79

Table: Rejection rates (%) of ‘uniform’, ‘feasible’ and ‘infeasible’ tests under H_0, i.e., when $\tau = 0$, and under misspecification.

“Robustified” test statistic

Recall that the diff-in-diff test statistic was equivalent to:

$$S_n = \gamma(\bar{X} - X_{I_1}) + \eta_{I_1} - g(\boldsymbol{\eta}),$$

So, we can use a covariate-adjusted test statistic:

$$S_n^{\mathsf{R}} = S_n - \gamma(\bar{X} - X_{I_1}).$$

Thus, $S_n^{\mathsf{R}} = \eta_{I_1} - g(\boldsymbol{\eta})$. Since $\boldsymbol{\eta}$ is exchangeable, knowing the correct distribution of I_1 is not necessary for a valid test.

Of course, this test is infeasible because γ is unknown. We therefore also consider:

$$S_n^{\mathsf{R,f}} = S_n - \hat{\gamma}_n(\bar{X} - X_{I_1}).$$

				$S_n^{\mathsf{R,f}}$		S_n^{R}	
n	k_1	k_2	γ	uniform	feasible (Cox)	uniform	feasible
50	0.00	0.00	0.00	4.93	4.92	4.99	5.00
	0.00	0.00	2.00	5.02	4.98	5.04	5.01
	0.00	0.00	5.00	4.96	4.99	5.04	5.02
	2.00	0.00	0.00	5.14	3.25	5.02	3.32
	2.00	0.00	2.00	4.99	3.30	4.90	3.27
	2.00	0.00	5.00	4.98	3.30	4.97	3.18
	1.00	1.00	0.00	5.26	4.88	5.04	4.62
	1.00	1.00	2.00	5.04	4.64	5.02	4.72
	1.00	1.00	5.00	5.08	4.68	4.97	4.64
	0.00	2.00	0.00	5.19	3.27	4.93	3.25
	0.00	2.00	2.00	5.07	3.43	5.03	3.33
	0.00	2.00	5.00	4.98	3.32	4.88	3.18
	2.00	2.00	0.00	5.11	4.71	4.99	4.52
	2.00	2.00	2.00	4.98	4.66	5.12	4.68
	2.00	2.00	5.00	5.14	4.68	4.90	4.52

Table: Rejection rates (%) of 'uniform', 'feasible' and 'infeasible' tests under H_0, i.e., when $\tau = 0$, and with test statistics defined in (9) and (9) under misspecification of the treatment adoption model.

Application: Tobacco data

Here, we re-analyze the tobacco legislation example in (Abadie et al., 2010).

We index time by $t \in \{\text{"}01/1971\text{"}, \text{"}02/1971\text{"}, \ldots, \text{"}12/2014\text{"}\}$, where we have adopted the "month/year" format and identify 1 with "$01/1971$" and t_{max} with "$12/2014$".

$Y_{i,t}$ = the number of cigarette packets sold in state i at time t.

T_i = time at which state i adopts tobacco tax legislation.$*$

X_{it} = $\{$ log-income/capita, average price levels, %youth population, %unemployment level, %Democratic legislature $\}$.

$*$ Orzechowski and Walker (2014) provide a comprehensive record of tobacco tax increases across states during this time period. We define treatment to be the first time taxes on cigarette packets are increased by at least 50% of current packet value.

Every state except Missouri adopts tobacco legislation in our sample period.

For robustness, we consider multiple alternative specifications.

See Definitions of treatment ▷

Treatment adoption probabilities

State	Prob.	State	Prob.	State	Prob.
Nevada	0.1037	Arkansas	0.0250	Virginia	0.0086
Connecticut	0.1018	Pennsylvania	0.0225	Alabama	0.0081
Rhode Island	0.0681	Louisiana	0.0201	West Virginia	0.0066
North Dakota	0.0617	Ohio	0.0200	Oklahoma	0.0037
Maine	0.0605	Delaware	0.0186	South Carolina	0.0033
Illinois	0.0580	Minnesota	0.0171	South Dakota	0.0027
Wisconsin	0.0517	Tennessee	0.0163	Vermont	0.0026
Texas	0.0491	Montana	0.0151	Utah	0.0023
Nebraska	0.0460	Idaho	0.0139	Iowa	0.0018
California	**0.0440**	Indiana	0.0134	North Carolina	0.0016
New Hampshire	0.0360	Kansas	0.0124	Missouri	0.0014
Wyoming	0.0291	Georgia	0.0124	Kentucky	0.0013
New Mexico	0.0279	Colorado	0.0111	Mississippi	0.0006

Table: Estimated conditional distribution of the identity of the state that first adopted tobacco legislation, i.e., the distribution of $I_1|T_{(1)}, X^{(n)}$.

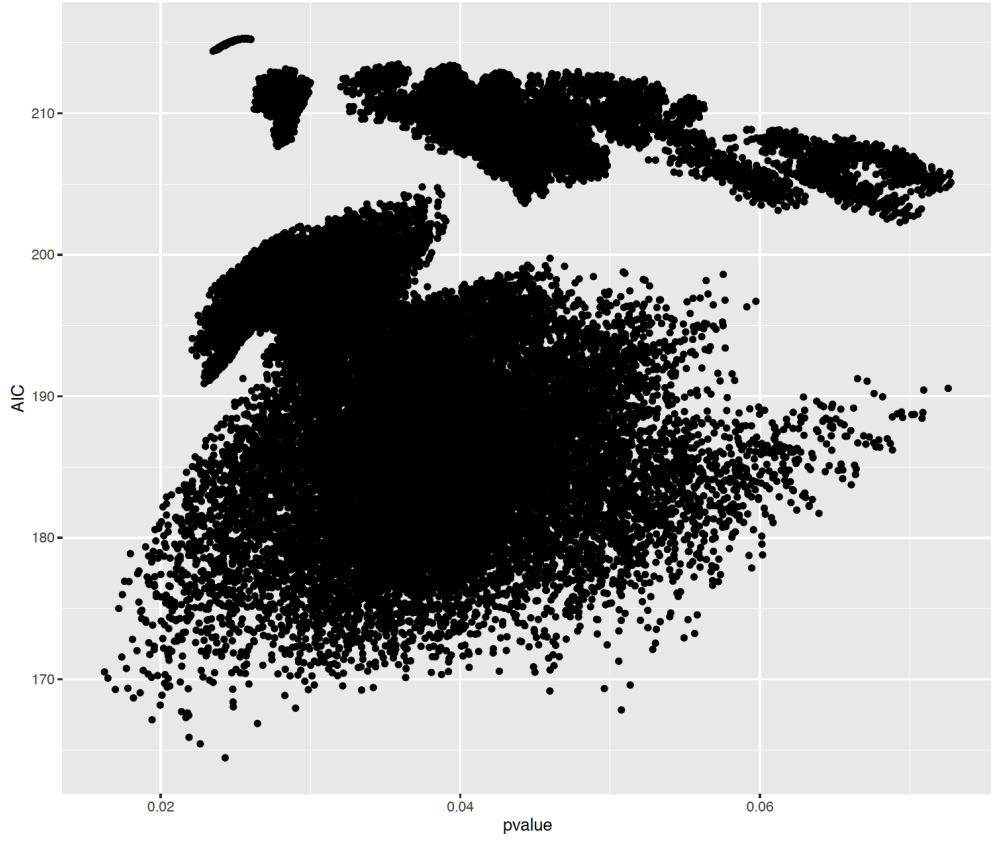
p-value = 0.044. Noticeably larger than 0.026 reported by Abadie et al. (2010), but still significant at 5%.

Robustness — Multiple specifications

Overall, we find 9 states with two possible T_i specifications each.

With 6 covariates, there are $2^6 = 64$ possible Cox models to consider.

In total, there are $2^9 \times 2^6 = 32,768$ possible specifications. As a robustness check, we report the (p-value, AIC) for each one of these specifications.



References

Shaikh, A. and Toulis, P. (2021) Randomization tests in observational studies with staggered adoption of treatment. *Journal of the American Statistical Association, 116(536), pp. 1835-1848.*

Thank you!

References

Abadie, A. (2019). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*.

Abadie, A. and Cattaneo, M. D. (2018). Econometric methods for program evaluation. *Annual Review of Economics*, **10** 465–503.

Abadie, A., Diamond, A. and 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** 493–505.

Abadie, A. and Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque Country. *American Economic Review*, **93** 113–132.

Abadie, A. and L'Hour, J. (2017). A penalized synthetic control estimator for disaggregated data. *Work. Pap., Mass. Inst. Technol., Cambridge, MA*.

Amjad, M., Shah, D. and Shen, D. (2018). Robust synthetic control. *The Journal of Machine Learning Research*, **19** 802–852.

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *The annals of statistics* 1100–1120.

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W. and Wager, S. (2019). Synthetic difference-in-differences. Tech. rep., National Bureau of Economic Research.

Athey, S., Bayati, M., Doudchenko, N., Imbens, G. and Khosravi, K. (2018). Matrix completion methods for causal panel data models. Tech. rep., National Bureau of Economic Research.

Ben-Michael, E., Feller, A. and Rothstein, J. (2019). Synthetic controls and weighted event studies with staggered adoption. *arXiv preprint arXiv:1912.03290*.

Card, D. and Krueger, A. B. (1993). Minimum wages and employment: A case study of the fast food industry in New Jersey and Pennsylvania. Tech. rep., National Bureau of Economic Research.

Cattaneo, M. D., Feng, Y. and Titiunik, R. (2019). Prediction intervals for synthetic control methods. Tech. rep.

Chernozhukov, V., Wuthrich, K. and Zhu, Y. (2018). Practical and robust t-test based inference for synthetic control and related methods.

Doudchenko, N. and Imbens, G. W. (2016). Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. Tech. rep., National Bureau of Economic Research.

Ferman, B. and Pinto, C. (2019). Synthetic controls with imperfect pre-treatment fit. *arXiv preprint arXiv:1911.08521*.

Firpo, S. and Possebom, V. (2018). Synthetic control method: Inference, sensitivity analysis and confidence sets. *Journal of Causal Inference*, **6**.

Hsiao, C., Steve Ching, H. and Ki Wan, S. (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** 705–740.

Lehmann, E. L. and Romano, J. P. (2006). *Testing Statistical Hypotheses*. Springer Science & Business Media.

Orzechowski, W. and Walker, R. (2014). The tax burden on tobacco. *Historical Compilation*, **39**.

Snow, J. (1855). *On the mode of communication of Cholera*. John Churchill.

Xue, Y. and Schifano, E. D. (2017). Diagnostics for the cox model. *Communications for statistical Applications and Methods*, **24** 583–604.

Observation times and Observation frequency

- In general, X_{it}, Y_{it} and T_i can be observed at different "observation frequencies".
- Motivated by our application and real data limitations, we assume that
  - X_{it}, Y_{it} have the same observation frequency, denoted as $t = 1, \ldots, t_{\text{max}}$.
  - T_i are observed at a higher frequency — e.g., month/year vs. only year for X_{it}, Y_{it}.
  - There are no ties, $T_i \neq T_j$ w.p.1.
- (Notation): Thus, "X_{iT_i}" is the value of X at time T_i modulo the observation frequency of X [e.g., log-GDP of state i at the year implied by T_i.]

Back ▷

It is natural to consider tests that condition not just on $T_{(1)}$, but $T_{(1)}$ and $T_{(2)}$, where $T_{(2)}$ is the second-order statistic of $T_1, \ldots, T_n$.

It is possible to see that $P\{I_1 = i, I_2 = j | T_{(1)}, T_{(2)}, X^{(n)}\}$ depends not only on $(X_{i,t} : i \in \mathbb{N}, t \in \{T_{(1)}, T_{(2)}\})$, but also on the integral of $\lambda(t) \exp(X'_{i,t}\beta)$ over $T_{(1)} \leq t \leq T_{(2)}$.

For this reason, we do not pursue such tests further in this paper.

Back ▷

n	γ	$\tau = 0.25$ feasible	$\tau = 0.25$ infeasible	$\tau = 0.5$ feasible	$\tau = 0.5$ infeasible
25	0.00	19.18	19.32	24.05	24.17
	0.50	15.01	14.89	20.35	20.34
	1.00	11.48	11.91	16.38	16.77
	2.00	8.68	9.20	12.43	12.67
	5.00	6.00	6.56	7.96	8.56
50	0.00	32.05	31.99	38.80	38.47
	0.50	26.98	26.93	34.90	35.28
	1.00	21.96	22.07	30.73	30.91
	2.00	16.71	16.59	24.50	24.55
	5.00	10.10	10.43	15.81	16.01
100	0.00	48.00	48.34	58.94	58.89
	0.50	42.09	42.48	54.90	54.90
	1.00	36.11	36.25	49.49	49.58
	2.00	27.63	27.70	41.48	41.51
	5.00	16.80	17.12	27.88	28.28

Table: Rejection rates (%) of 'feasible' and 'infeasible' tests of (**??**) under the alternative hypothesis, i.e., when $\tau = 0.25$ or $\tau = 0.50$.

Back ▷

Choice of S_n

While our theory applies to any choice of test statistic S_n, some choices of test statistics may be preferable in terms of power.

Abadie et al. (2010) suggest, for example, a test statistic of the form

$$\frac{\sum_{t \geq T_{(1)}} (Y_{I_1,t} - \hat{Y}_{I_1,t})^2}{\sum_{t < T_{(1)}} (Y_{I_1,t} - \hat{Y}_{I_1,t})^2}, \tag{9}$$

where $\hat{Y}_{I_1,t}$ is the "synthetic control". See Section 2 of Cattaneo et al. (2019) for a succinct summary of that choice and variations by Abadie et al. (2010), Hsiao et al. (2012), Doudchenko and Imbens (2016), Chernozhukov et al. (2018), Ferman and Pinto (2019), Arkhangelsky et al. (2019), and Abadie and L'Hour (2017), Amjad et al. (2018), Athey et al. (2018) and Ben-Michael et al. (2019), for more extensions.

Other options for S_n include diff-in-diff and t-test statistic. See (Firpo and Possebom, 2018) for numerical experiments in power.

Back ▷

	State	Specification A	Specification B
1	Alabama	05/2004	05/2004
2	**Arkansas**	06/2003	02/1993
3	California	01/1989	01/1989
4	Colorado	01/2005	01/2005
5	Connecticut	04/1989	04/1989
6	**Delaware**	01/1991	08/2003
7	Georgia	07/2003	07/2003
8	**Idaho**	07/1994	06/2003
9	Illinois	07/1989	07/1989
10	Indiana	07/2002	07/2002
11	Iowa	04/2007	04/2007
12	Kansas	07/2002	07/2002
13	Kentucky	06/2005	06/2005
14	Louisiana	08/2002	08/2002
15	**Maine**	11/1997	07/1991
16	**Minnesota**	06/1991	08/2005
17	Mississippi	05/2009	05/2009
18	Missouri	12/2014	12/2014
19	Montana	05/2003	05/2003
20	Nebraska	10/2002	10/2002
21	Nevada	07/1989	07/1989
22	New Hampshire	02/1990	02/1990
23	New Mexico	07/2003	07/2003
24	North Carolina	09/2005	09/2005
25	North Dakota	05/1989	05/1989
26	Ohio	07/2002	07/2002
27	Oklahoma	01/2005	01/2005
28	Pennsylvania	08/1991	08/1991
29	**Rhode Island**	07/1997	07/1993
30	South Carolina	07/2010	07/2010
31	**South Dakota**	03/2003	07/1995
32	Tennessee	07/2002	07/2002
33	Texas	07/1990	07/1990
34	**Utah**	07/1991	07/1997
35	Vermont	07/1995	07/1995
36	Virginia	09/2004	09/2004
37	West Virginia	05/2003	05/2003
38	Wisconsin	05/1992	05/1992
39	**Wyoming**	07/2003	07/1989

Back ▷