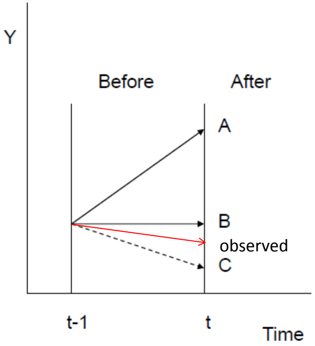
EC4305 Applied Econometrics
Differences-in-Differences

Xinming Du

National University of Singapore

Differences-in-Differences and Causality

Another Counterfeit Counterfactual



- Compare the average outcome before and after:

$$E[Y_{it}|D_{it} = 1] - E[Y_{it-1}|D_{it} = 1]$$

- Assumption:

E(Yit-1 | Dit = 0) = E(Yit-1 | Dit = 1) for causal framework

$$E[Y_{it-1}|D_{it} = 1] = E[Y_{it}|D_{it} = 0]$$

- Counterfactuals $E[Y_{it}|D_{it} = 0]$ could be A, B, or C.

because assuming E(Yit/t-1 | Dit) dont change is bad
time effects happen

- Problem: In addition to policy intervention, there could be many unobserved time varying factors that affect the outcome at time t.

Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania

(Card and Kruger, 1994)

Minimum Wage and Employment

- A classic and controversial question in labour economics: what is and how large is the effect of minimum wages on employment?
- In a competitive labour market, increases in the minimum wage will move us up a downward-sloping labour demand curve $\implies$ employment will fall
- Card and Krueger (1994) analyse the effect of a minimum wage increase in New Jersey using a differences-in-differences (DID) methodology.
- On April 1, 1992, New Jersey increased the state minimum wage from \$4.25 to \$5.05. Pennsylvania's minimum wage stayed at \$4.25.

Backgrounds

- They collected data on employment at fast food restaurants in New Jersey (treatment) and eastern Pennsylvania (control) in February 1992 (before) and again in November 1992 (after).
- These restaurants (Burger King, Wendy's, and so on) are big minimum-wage employers.

Differences-in-Differences Strategy

- DID is a version of fixed effects estimation. Denote
  - Y_{1ist}: Empl. at restaurant i , state s, time t with a *high* w^{min}
  - Y_{0ist}: Empl. at restaurant i , state s, time t with a *low* w^{min}
- Assumption:

$$E[Y_{0ist}|s,t] = \gamma_s + \lambda_t$$

  That is, in the absence of a minimum wage change, employment is determined by the sum of a time-invariant state effect γ_s and a year effect λ_t that is common across states.
- Let D_{st} be a dummy for high-minimum wage states and periods.

Differences-in-Differences Strategy

- Assume $E[Y_{1ist} - Y_{0ist}|s,t] = \delta$ is the treatment effect. Then observed employment Y_{ist} can be written as:

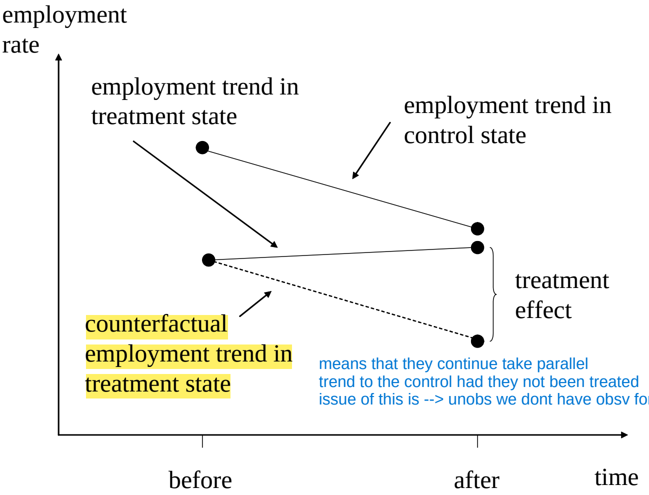
$$Y_{ist} = \gamma_s + \lambda_t + \delta D_{st} + \varepsilon_{ist}$$

delta effect if Dst =1

where $E(\varepsilon_{ist}|s,t) = 0$.

- The key identifying assumption here is that employment trends would be the same in both states in the absence of treatment. Treatment in the only reason that induces a deviation from this common trend. Although the treatment and control states can differ, this difference in captured by the state fixed effect.

Differences-in-Differences Strategy



means that they continue take parallel
trend to the control had they not been treated
issue of this is --> unobs we dont have obsv for this

Differences-in-Differences Strategy

- Employment of New Jersey in February is:

$$E[Y_{ist}|s = NJ, t = Feb] = \gamma_{NJ} + \lambda_{Feb}$$

- Employment of New Jersey in November is:

$$E[Y_{ist}|s = NJ, t = Nov] = \delta + \gamma_{NJ} + \lambda_{Nov}$$

state FE cancels out for same state over time

- The difference between February and November in NJ is:

$$E[Y_{ist}|s = NJ, t = Nov] - E[Y_{ist}|s = NJ, t = Feb] = \delta + \lambda_{Nov} - \lambda_{Feb}$$

Differences-in-Differences Strategy

- Employment of Pennsylvania in February is:

$$E[Y_{ist}|s = PA, t = Feb] = \gamma_{PA} + \lambda_{Feb}$$

- Employment of Pennsylvania in November is:

$$E[Y_{ist}|s = PA, t = Nov] = \gamma_{PA} + \lambda_{Nov}$$

- The difference between February and November in PA is:

$$E[Y_{ist}|s = PA, t = Nov] - E[Y_{ist}|s = PA, t = Feb] = \lambda_{Nov} - \lambda_{Feb}$$

Differences-in-Differences Strategy

- The DID strategy amounts to comparing the change in employment in NJ to the change in employment in PA.
- The population differences-in-differences are:

$$\begin{aligned}\delta =& \{E[Y_{ist}|s = NJ, t = Nov] - E[Y_{ist}|s = NJ, t = Feb]\} \\ & - \{E[Y_{ist}|s = PA, t = Nov] - E[Y_{ist}|s = PA, t = Feb]\}\end{aligned}$$

- δ is estimated using the sample analog of the population means:

$$\begin{aligned}\hat{\delta} =& \{Avg[Y_{ist}|s = NJ, t = Nov] - Avg[Y_{ist}|s = NJ, t = Feb]\} \\ & - \{Avg[Y_{ist}|s = PA, t = Nov] - Avg[Y_{ist}|s = PA, t = Feb]\}\end{aligned}$$

Differences-in-Differences: Table

Table: Average employment per store before and after the New Jersey minimum wage increase

Variable		PA (i)	NJ (ii)	Difference, NJ-PA (iii)
1.	FTE employment before,	23.33	20.44	-2.89
	all available observations	(1.35)	(0.51)	(1.44)
2.	FTE employment after,	21.17	21.03	-0.14
	all available observations	(0.94)	(0.52)	(1.07)
3.	Change in mean FTE	-2.16	0.59	2.76
	employment	(1.25)	(0.54)	(1.36)

- Surprisingly, employment rose in NJ **relative to** PA after the minimum wage change.

Differences-in-Differences: Regression

- We can estimate the DID estimator in a regression framework. Advantages:
  - It is easy to calculate standard errors.
  - We can control for other variables which may reduce the residual variance (lead to smaller standard errors).
  - It is easy to include multiple periods.
  - We can study treatments with different treatment intensity. (e.g. varying increases in the minimum wage for different states).

Differences-in-Differences: Regression

- The typical regression model that we estimate is:

  $$Outcome_{igt} = \beta_1 + \beta_2 Treat_g + \beta_3 Post_t + \beta_4 (Treat_g \times Post_t) + \varepsilon_{igt}$$

  where Treatment = a dummy equals to one if the observation is in the treatment group; Post = post treatment dummy.

- In the Card and Krueger (1994), the equivalent regression model is:

  $$Y_{ist} = \alpha + \gamma NJ_s + \lambda Post_t + \delta(NJ_s \times Post_t) + \varepsilon_{ist}$$

- NJ: a dummy which equal to 1 if the observation is from NJ.
- $Post_t$: a dummy which equal to 1 if the observation is from Nov.

Differences-in-Differences: Regression

- This equation takes the following values:
  - PA Pre: α
  - PA Post: $\alpha + \lambda$
  - NJ Pre: $\alpha + \gamma$
  - NJ Post: $\alpha + \gamma + \lambda + \delta$
- DID estimate:

$$\delta = \text{(NJ Post - NJ Pre)-(PA Post - PA Pre)}$$

A Potential Explanation for the Empirical Findings

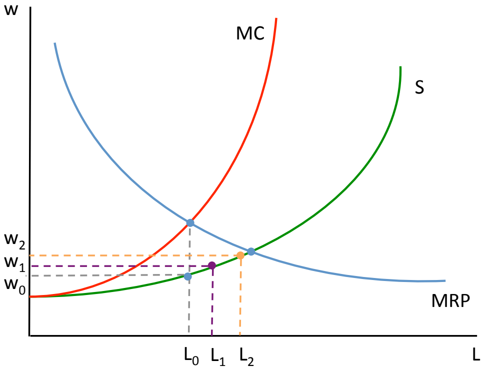
- The study finds the opposite of what we might expect if a higher minimum wage pushes businesses up the labor demand curve. How to explain this seemingly counter-intuitive result?
- A monopsonistic model provides a potential explanation for the observed employment effect of minimum wage.
- Profit maximization problem for a monopsonist employer:

$$max\ \pi(L) = R(L) - w(L)L$$

FOC implies that

for monopsonist employer w'(L) >0 thats why they can pay more

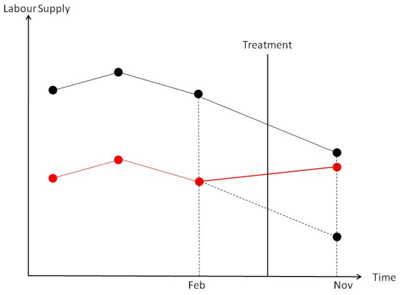
$$\underbrace{R'(L)}_{MRP} = \underbrace{w'(L)L + w(L)}_{MC}$$



- In the absence of minimum wage, a monopsonist employer chooses the employment level (L_0) such that equates the marginal revenue product (MRP) and marginal cost (MC). The equilibrium wage is w_0.
- A minimum wage $w_1 > w_0$ raises employment level to L_1. An increase in minimum wage from w_1 to w_2 further pushes the employment level up to L_2.

Key Assumption I: Common (Parallel) Trends

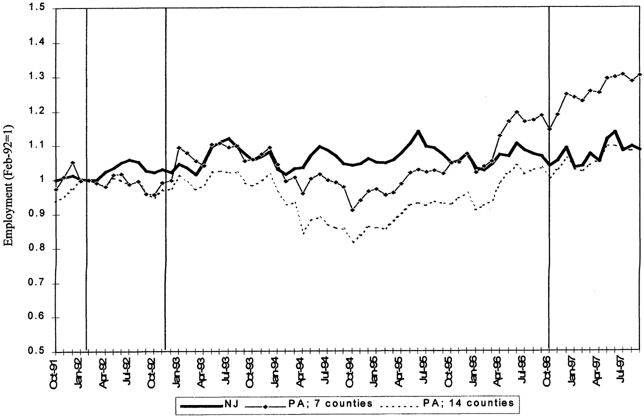
- The key assumption for any DD strategy is that the outcome in treatment and control group would follow the same time trend in the absence of the treatment.
- Common trend assumption is difficult to verify but one often uses pre-treatment data to show that the trends are the same.



Key Assumption I: Common (Parallel) Trends

- Card and Krueger (1994) argues that the seasonal patterns of employment are similar in NJ and eastern PA.
- In an update of their original minimum wage study, Card and Krueger (2000) obtained administrative payroll data for restaurants in NJ and PA for a number of years.
  - A slight decline in employment from February to November 1992 in PA, and little change in NJ over the same period.
  - However, swings of employment often seem to differ substantially in the two states in other periods.
- PA may not provide a very good measure of counterfactual employment rates in NJ in the absence of a policy change, and vice versa.

Figure: Employment in New Jersey and Pennsylvania fast food restaurants



Note: Vertical lines indicate dates of the original Card and Krueger (1994) survey and the October 1996 federal minimum wage increase.

Key Assumption I: Common (Parallel) Trends

- Including leads into the DD model is an easy way to analyze pre-trends.

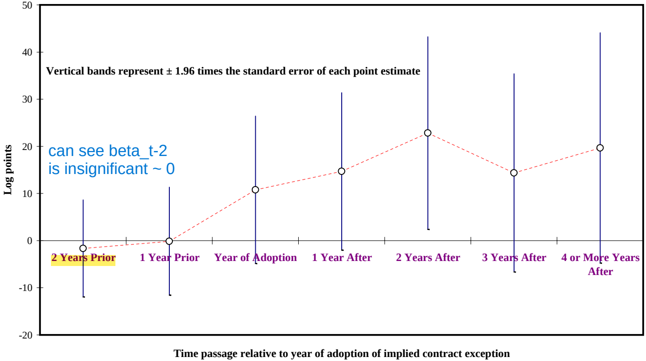
$$Y_{ist} = \gamma_s + \lambda_t + \sum_{\tau=1}^{m} \beta^{-\tau} D_{s,t}^{-\tau} + \sum_{\tau=1}^{q} \beta^{\tau} D_{s,t}^{\tau} + X_{ist}\delta + \varepsilon_{ist}$$

  - $D_{s,t}^{-\tau} = 1$ if $s = NJ$ and $t = t - \tau$, τ periods before treatment time
  - m lags $(\beta^{-1}, \beta^{-2}, ..., \beta^{-m})$ or pre-treatment effects, or anticipatory effects
  - q leads $(\beta^{1}, \beta^{2}, ..., \beta^{q})$ or post-treatment effects.
- Leads can be included to analyze whether the treatment effect changes over time after treatment.
- Using lags, we can test whether effects happen before treatment, namely whether the pre-trends are paralleled.

Key Assumption I: Common (Parallel) Trends

- Autor (2003) includes both leads and lags in a DD model analyzing the effect of increased employment protection on the firm's use of temporary help workers.
- As a rule, U.S. labor law allows "employment at will", which means that workers can be fired for just cause or no cause, at the employer's whim.
- Some states courts have made some exceptions to this employment at will rule and have thus increased employment protection.
- Different states have passed these exceptions at different points in time.

Figure: Estimated impact of state court's adoption of an implied-contract exception to the employment-at-will doctrine on use of temporary workers



Note: The dependent variable is the log of state temporary help employment.

Key Assumption I: Common (Parallel) Trends

- The lags are very close to 0. $\implies$ no evidence for anticipatory effects good.
- The leads show that the effect increases during the first years of the treatment and then remains relatively constant.
- Sometimes, there exists differential pre-trends between the treatment and control groups. To address the problem, we could add group-specific time trends as controls. For example:

$$\begin{aligned} Y_{ist} = & \alpha + \gamma NJ_s + \lambda Post_t + \delta(NJ_s \times Post_t) \\ & + \theta T + \phi(T \times NJ_s) + \varepsilon_{ist} \end{aligned}$$

T is a linear time variable, θ captures the linear time trend for PA while $\theta + \phi$ captures the linear time trend for NJ.

Key Assumption II: No Omitted Variables That Correlate with Treatment Status

- Even if pre-trends are the same one still has to worry about other unobserved policy/economic shocks that occurred at the same time.
- For example, labor protection rule might become more stringent in PA during the same period, which might lower employment.
- DID method attributes any differences in trends between the treatment and control groups, that occur at the same time as the intervention, to that intervention.
- If there are other factors that affect the difference in trends between the two groups, then the estimation will be biased.

Key Assumption II: No Omitted Variables That Correlate with Treatment Status

- It is inherently difficult to test the existence or rule out the unobserved policy/economic shocks. However, there are ways to alleviate/address the concern:

(1) Placebo (Falsification) test:

- Exploit a population you know was NOT affected. E.g., Changes in minimum wage should have little effect on the employment of college graduates.
- Use an outcome variable which is NOT affected by the intervention, but could be affected by the potential unobserved policy shocks.

want to see during period of the intervention interest, that the variable was not affected by other unobsv shocks when it could be affected by it.

(2) IV strategy: exogenous variation of treatment status

Endogenous Intervention

- The DID estimation is appropriate when the interventions are as good as random, conditional on time and group fixed effects.
- Much of the debate around the validity of a DID estimate typically revolves around the possible endogeneity of the interventions themselves. For example,
  - NJ could afford raising minimum wage because it foresaw a strong demand in the labor market.
  - PA remained its minimum because it foresaw an increase in labor supply of less-skilled workers.
- In these scenarios, DID estimator will be biased upward. Why?

Other Potential Pitfalls

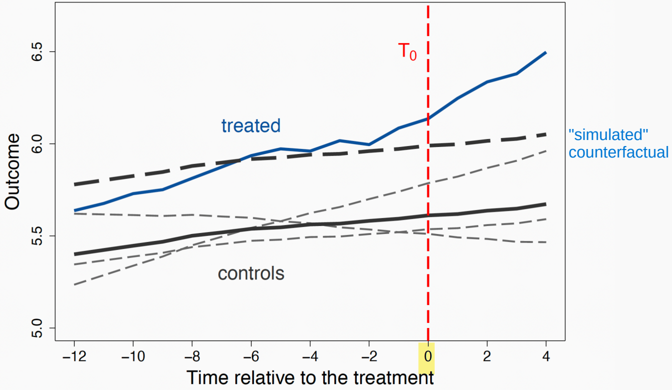
- Suppose you are interested the effects of the generosity of public assistance on labor supply. You design an empirical strategy based on state/time comparisons.
- One potential pitfall in this context arises when the *composition* of the implicit treatment and control groups changes as a result of treatment. The poor people who would in any case have weak labor force attachment might move to states with more generous welfare benefits. "drop out", you cannot control this
- In a DD research design, this sort of program-induced migration tends to make generous welfare programs look worse for labor supply than they really are. How to address the problem?

Synthetic Control

Synthetic Control

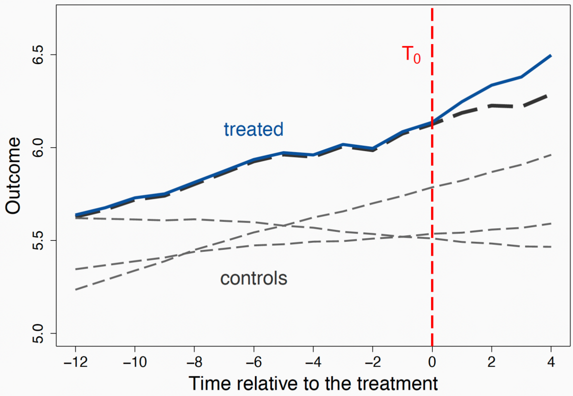
- $J + 1$ units in periods 1,2,3,...,T.
- One treated "1", J controls
- Region "1" is exposed to the intervention after period T_0
- We aim to estimate the treatment effect of the intervention on Region "1"
- Synthetic controls use a weighted average of untreated units' outcomes to impute a counterfactual outcome for the treated united

DID



T_0
treated
"simulated"
counterfactual
controls
Outcome
5.0
5.5
6.0
6.5
−12
−10
−8
−6
−4
−2
0
2
4
Time relative to the treatment

Synthetic Control



Y_1t(0) for T > T0
simulate counterfact

Synthetic Control: Basic Idea

- We observe Y_{it} for $i = 1, 2, ..., J + 1$ and $t = 1, 2, ..., T$. Unit 1 gets treated in T_0
- We use $Y_{1it} = Y_{it}(1)$, $Y_{0it} = Y_{it}(0)$ to denote potential outcomes.
- $\rightarrow$ We observe:
  - $Y_{1t}(0)$ for $t = 1, 2, ..., T_0 - 1$
  - $Y_{1t}(1)$ for $t = T_0, T_0 + 1, ..., T$
  - $Y_{jt}(0)$ for $\forall j,\ t = 1, 2, ., , , T_0 - 1$
  - $Y_{jt}(0)$ for $\forall j,\ t = T_0, T_0 + 1, ..., T$

-> for treat unit, before T0 periods is no treat and after treat in T0 periods, it is a treated PO

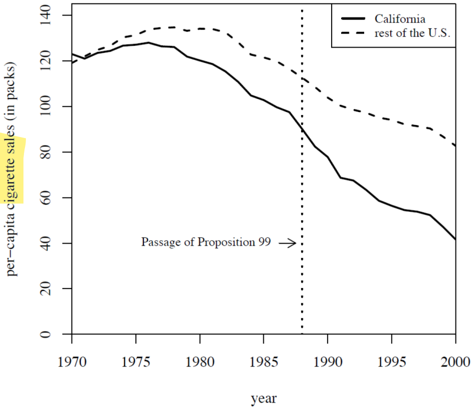
-> for all J control units, before T0 periods and all T0 after periods, they are no treatment

- Idea: infer missing $Y_{1t}(0)$ for $t = T_0, T_0 + 1, ..., T$ from $Y_{jt}(0)$
- Estimator: put each j a weight w_j, $\sum_{j=1}^{J} w_j \times Y_{jt}$ is the counterfactual $Y_{1t}(0)$ for $t = T_0, T_0 + 1$
- Treatment effect in each period is $Y_{1t}(1) - \sum_{j=1}^{J} w_j \times Y_{jt}$

the actual observed treatment outcome of treated unit in post minus the "simulated post outcome had it not been treated"

Example: Cigarette Consumption Tax

- California passed a law in 1989 that raised the cigarette excise tax by 25 cents and implemented a large-scale anti-tobacco media campaign

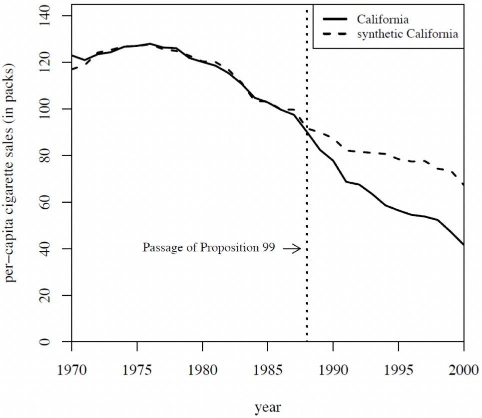


- Clearly a simple diff-in-diff is not going to cut it here

why? pre not parallel

Synthetic California

- Applying standard synthetic control weights renders a pre-trend we could previously only dream about



- What's going on under 'synthetic California'?

Synthetic California

State	Weight	State	Weight
Alabama	0	Montana	0.199
Alaska	-	Nebraska	0
Arizona	-	Nevada	0.234
Arkansas	0	New Hampshire	0
Colorado	0.164	New Jersey	-
Connecticut	0.069	New Mexico	0
Delaware	0	New York	-
District of Columbia	-	North Carolina	0
Florida	-	North Dakota	0
Georgia	0	Ohio	0
Hawaii	-	Oklahoma	0
Idaho	0	Oregon	-
Illinois	0	Pennsylvania	0
Indiana	0	Rhode Island	0
Iowa	0	South Carolina	0
Kansas	0	South Dakota	0
Kentucky	0	Tennessee	0
Louisiana	0	Texas	0
Maine	0	Utah	0.334
Maryland	-	Vermont	0
Massachusetts	-	Virginia	0
Michigan	-	Washington	-
Minnesota	0	West Virginia	0
Mississippi	0	Wisconsin	0
Missouri	0	Wyoming	0

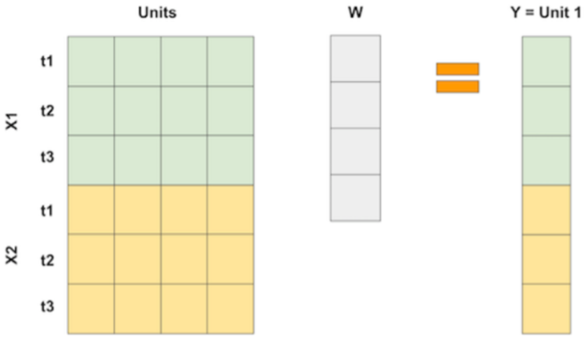
Implementation: Step 1

- Choose potential control units
- Donor pool: untreated reservoir of potential controls.
- Donor pool is not treated and not affected by spillovers.

Implementation: Step 2

- Calculate weight w_j for $j = 1, 2, .., J$
- Goal: make 1 and synthetic 1 based on js before T_0 similar
- Algorithm: optimization
- Input: k characteristics in the pre-period.
  - For each unit, we observe Y_{j1},..., Y_{jT_0-1} and X^1_{j1},..., $X^1_{jT_0-1}$, X^2_{j1},..., $X^2_{jT_0-1}$, ..., X^m_{j1},..., $X^m_{jT_0-1}$
  - If there are m covariates, $k = (m+1) \times T_0 - 1$
  - X_1 is a $k \times 1$ vector, for treated unit 1
  - X_0 is a $k \times J$ vector, for donor pool unit j
- Parameters: two sets of weights:
  1) W is a $J \times 1$ vector
     $(w_1, w_2, ..., w_J)'$: each control unit in the donor pool gets a weight
  2) V is a $k \times k$ symmetric and positive matrix.
     $(v_1, v_2, ..., v_k)$: each characteristics gets a weight

k Characteristics Stack All Pre-Periods



Implementation: Step 2 (con't)

- Minimization problem

$$||X_1 - X_0 W|| = \sqrt{(X_1 - X_0 W)' V (X_1 - X_0 W)}$$

- This gives us optimal pairs ($W1$, $V1$), ($W2$, $V2$),...

- How to choose the best pair? To fit the outcome variable, namely minimize:

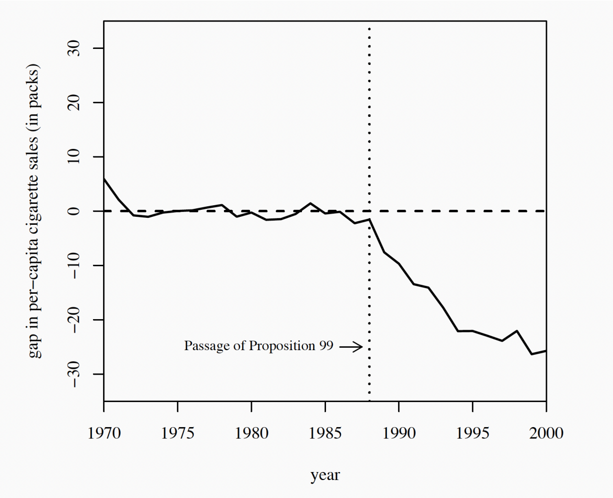
$$\sum_{t=1}^{T_0} (Y_{1t} - \sum_{j=1}^{J} w_j(V) Y_{jt})^2$$

Implementation: Step 3

- Apply the weight in outcomes in control unit in the post period.
- Calculate real treated outcome minus calculated weighted outcome

$$\delta_{1t} = Y_{1t} - \sum_{j=1}^{J} w_j Y_{jt}, \quad \forall t \in \{T_0, T_0 + 1, ..., T\}$$

Point Estimates In Each Year



Standard Error

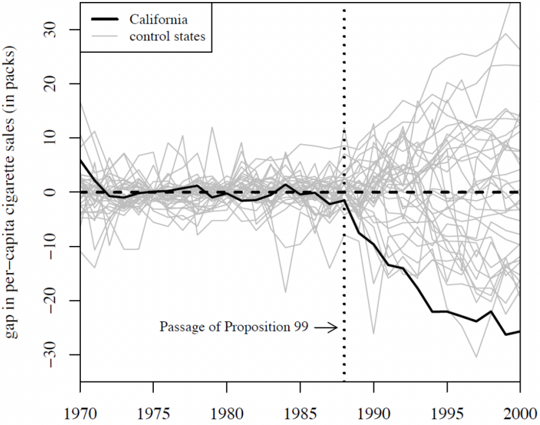
Permutation test:

1) Run placebo synthetic control on all units in the donor pool
2) Compute the treatment effect for each placebo
3) Compare placebos to the treatment effect in the main result
4) Compute empirical p-values:

$$p = \frac{1 + b}{1 + J}$$

- b: number of placebo estimates larger in absolute value than the main estimate
- J: number of placebo estimates

Inference



California
control states
gap in per−capita cigarette sales (in packs)
30
20
10
0
−10
−20
−30
Passage of Proposition 99
1970
1975
1980
1985
1990
1995
2000