17.802 – Lecture

Synthetic Control Methods

Yiqing Xu

MIT

xyq@mit.edu

- The fundamental problem of causal inference
- A statistical solution makes use of the population

$$\text{e.g. } T = E[Y_1] - E[Y_0]$$

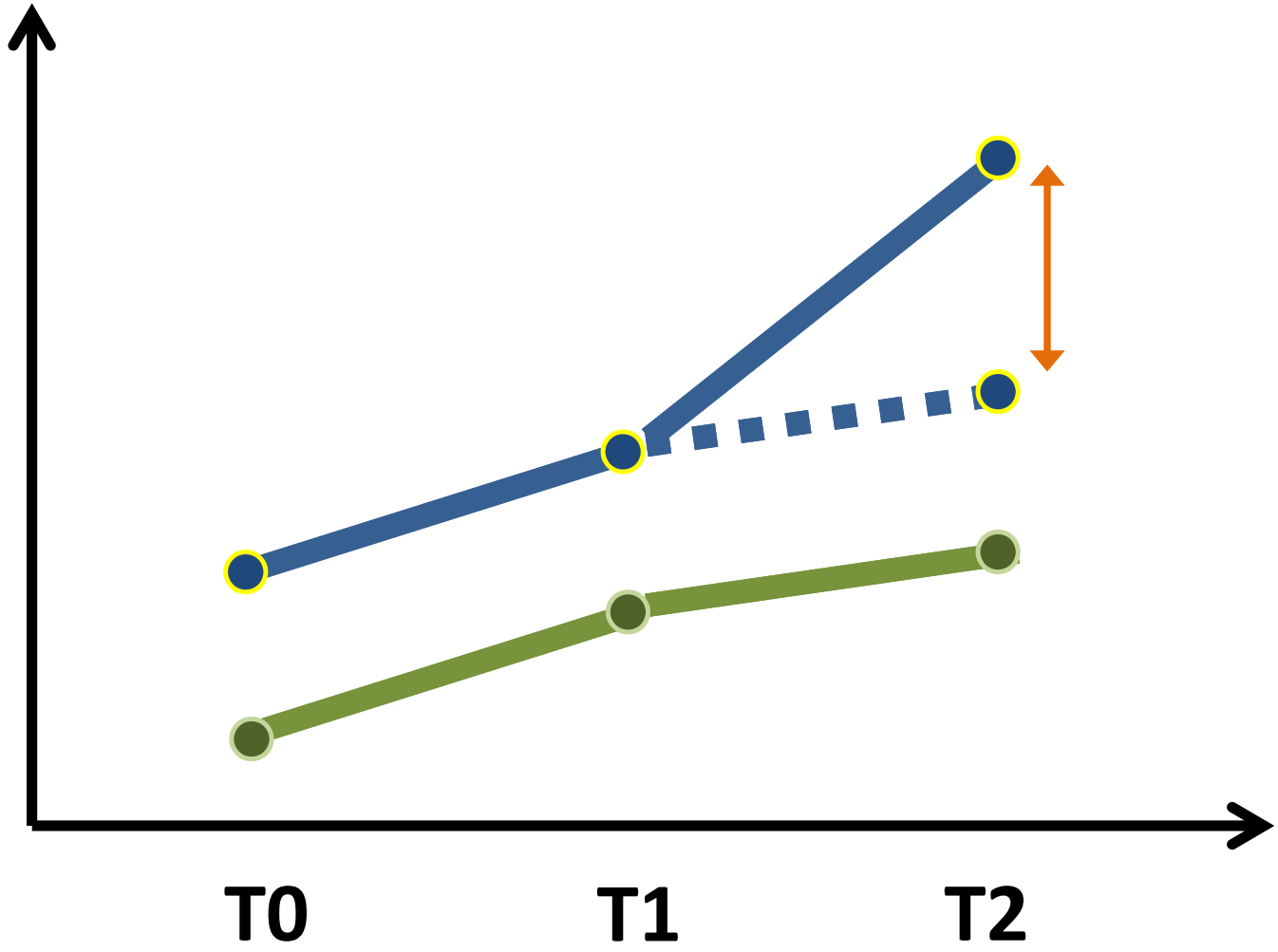
- A scientific solution exploits homogeneity or invariance assumptions

e.g. A rock is a rock.

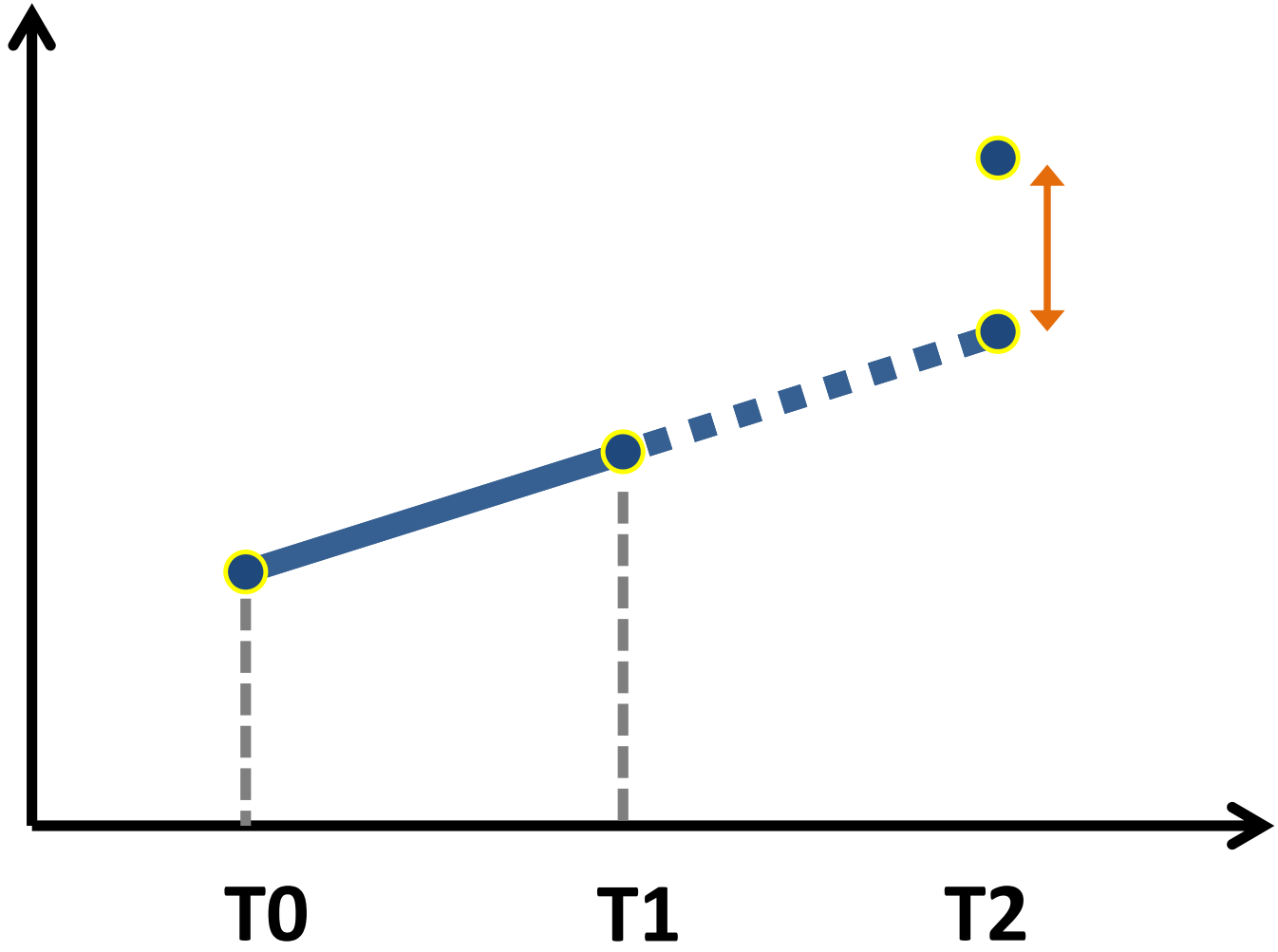
e.g. The long-run growth rate of the US economy is 2.5%.

- Panel data allow us to construct the counterfactuals of the treated units in the post-treatment period using information from both the control group and treatment group in the pre-treatment period

Causal Inference with Panel Data



Solution 1: Time Series Analysis



America leads Russia, but will the gap narrow?

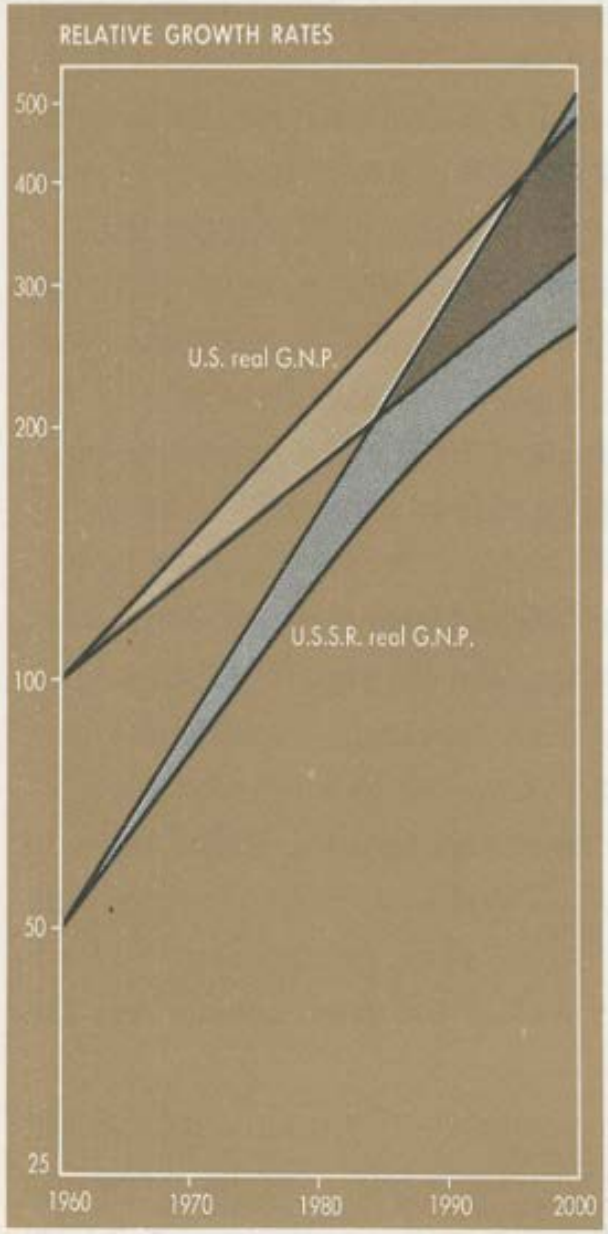
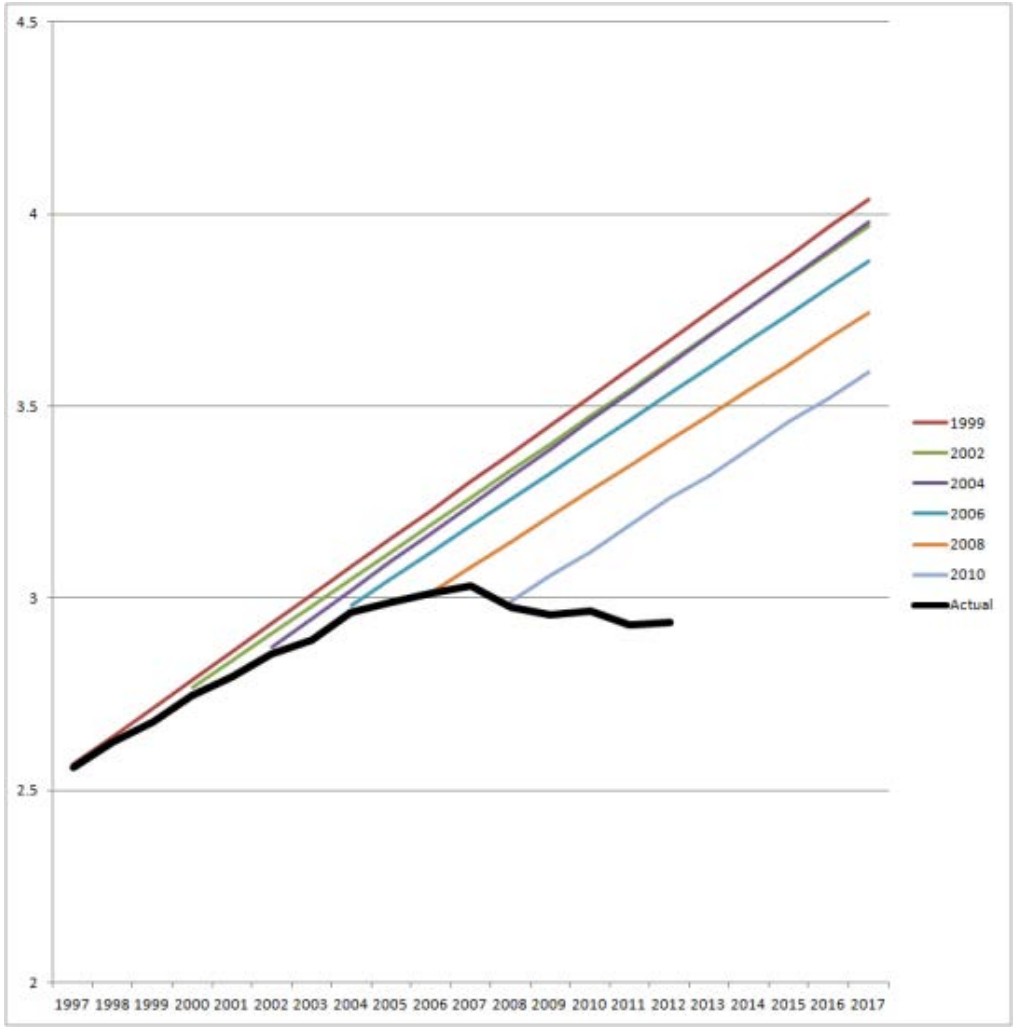
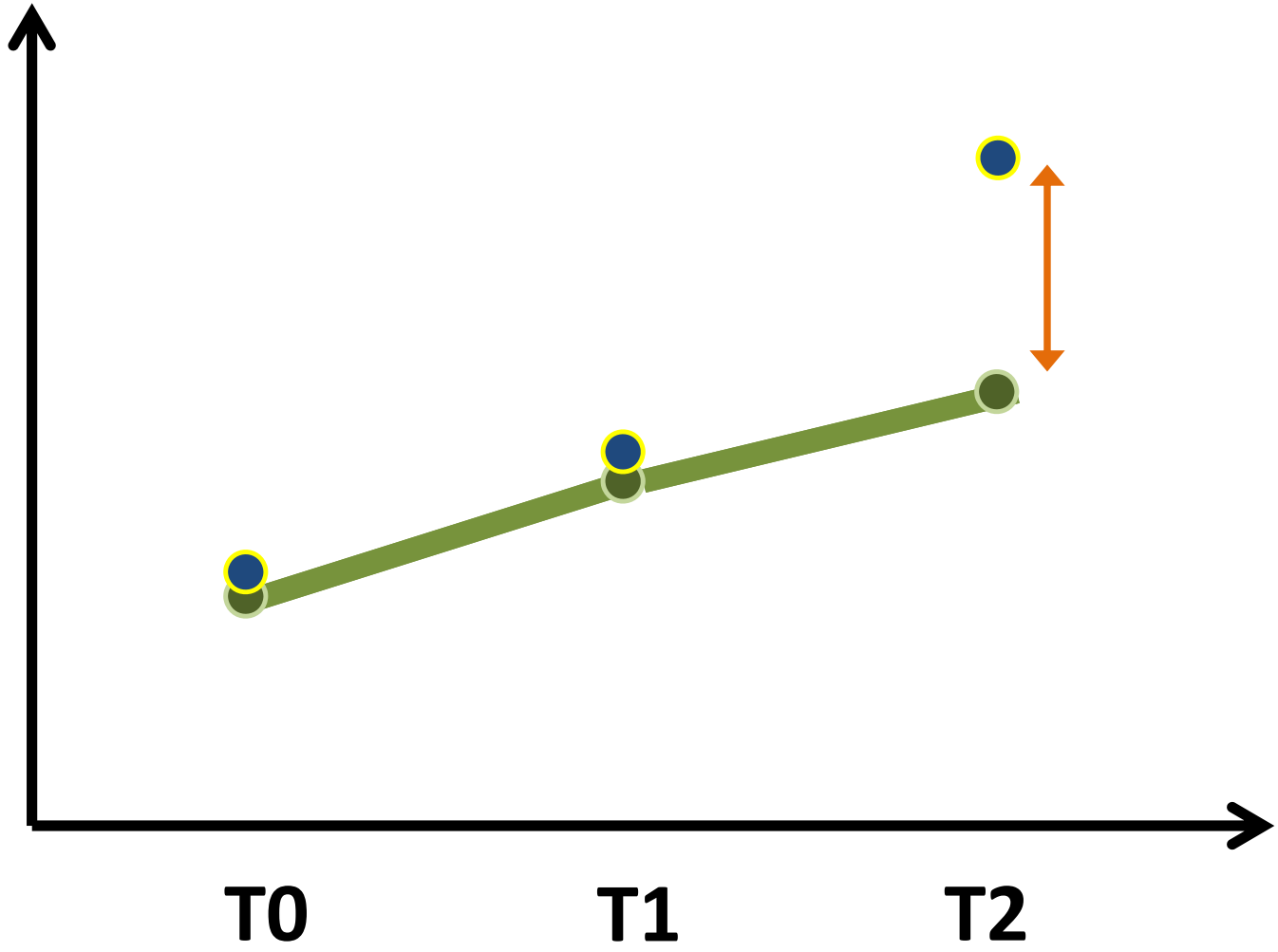


Fig. 1. The range of estimates shown here can make no pretense to accuracy, but they do portray the nature of the Soviet challenge. (Note: All indexes are based upon U.S. real GNP for 1960 = 100 and U.S.S.R. real GNP for 1960 = 50.)

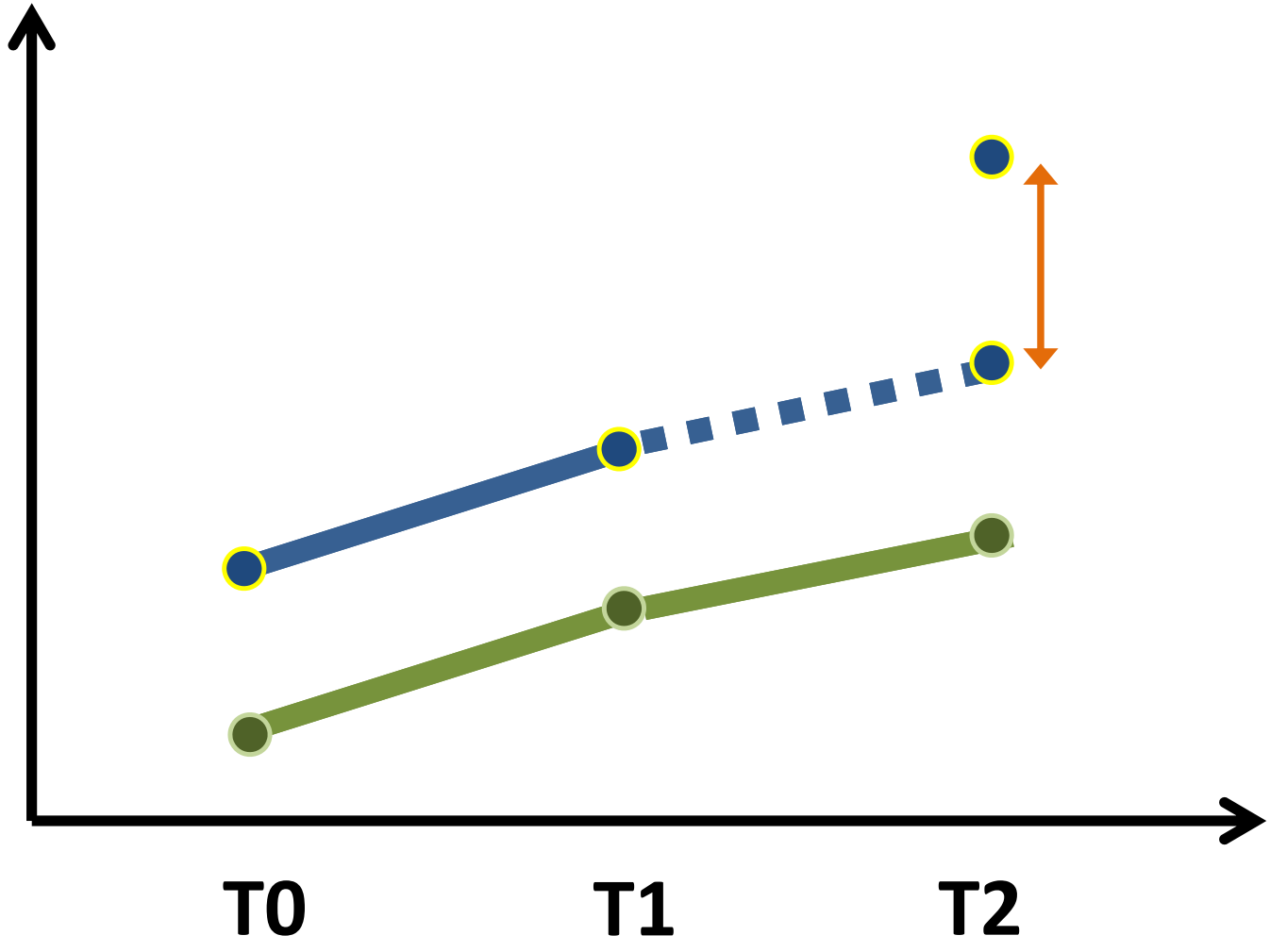


DOT forecasts of road traffic volume

Solution 2: Matching



Solution 3: Find Parallel Worlds



- Statistical solution: SOO, e.g., matching
- Scientific solution: modelling
- Panel data make both easier
  - Matching on lagged outcomes makes SOO more plausible
  - Parallel trends assumption is somewhat "testable"

- Fixed effects and diff-in-diffs

$$Y_{it}(0) = \alpha_i + \delta_t + \epsilon_{it}$$

- Lag dependent variable

$$Y_{it}(0) = \alpha + \theta Y_{i,t-1} + \delta_t + \epsilon_{it}$$

- Synthetic control (non-parametric)

$$Y_{it}(0) = \sum_{j \in \mathcal{C}} w_j^* Y_{jt}$$

- Limitations
  - FE and LDV assume homogeneous treatment effect
  - DID assumes fixed treatment timing
  - Synth allows only one treated unit and inference is less formal

- What if the true model is as complicated as:

$$Y_{it}(0) = X_{it}\beta + Z_i\theta_t + \lambda_t\mu_i + \varepsilon_{it},$$

- Compare with the fixed effects (or DID) model

$$Y_{it}(0) = X_{it}\beta + Z_i\theta_t + \delta_t + \alpha_i + \varepsilon_{it},$$

- $\lambda_t\mu_i$ are fixed effects interacted with time-varying coefficient, in which δ_t and α_i are special cases
- Abadie and Gardeazabal (2003), Abadie, Diamond, and Hainmueller (2010) figured out a way to solve this problem when there is only one treated unit

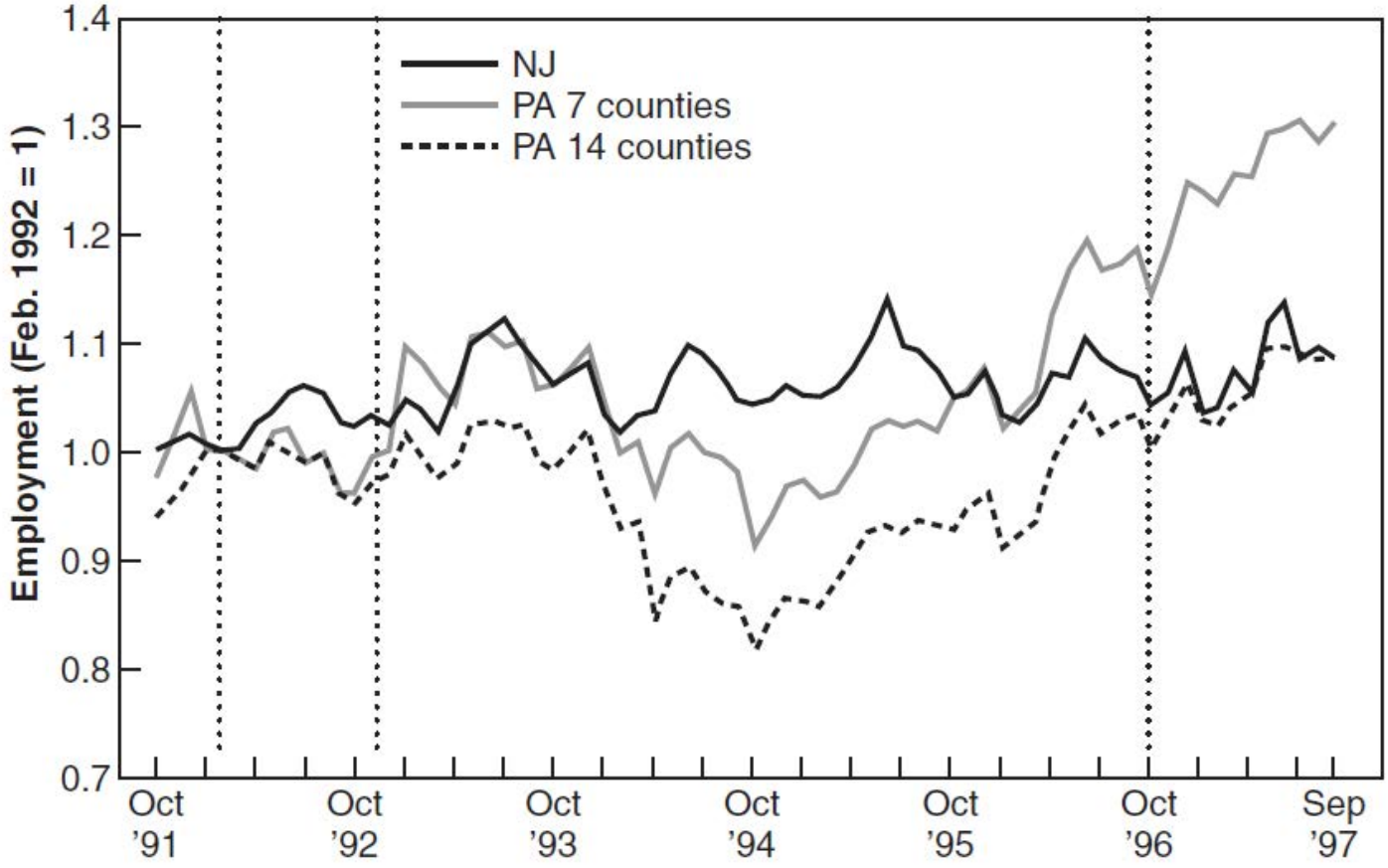


Figure 5.2.2 Employment in New Jersey and Pennsylvania fast food restaurants, October 1991 to September 1997 (from Card and Krueger 2000). Vertical lines indicate dates of the original Card and Krueger (1994) survey and the October 1996 federal minimum wage increase.

Comparative Case Studies:

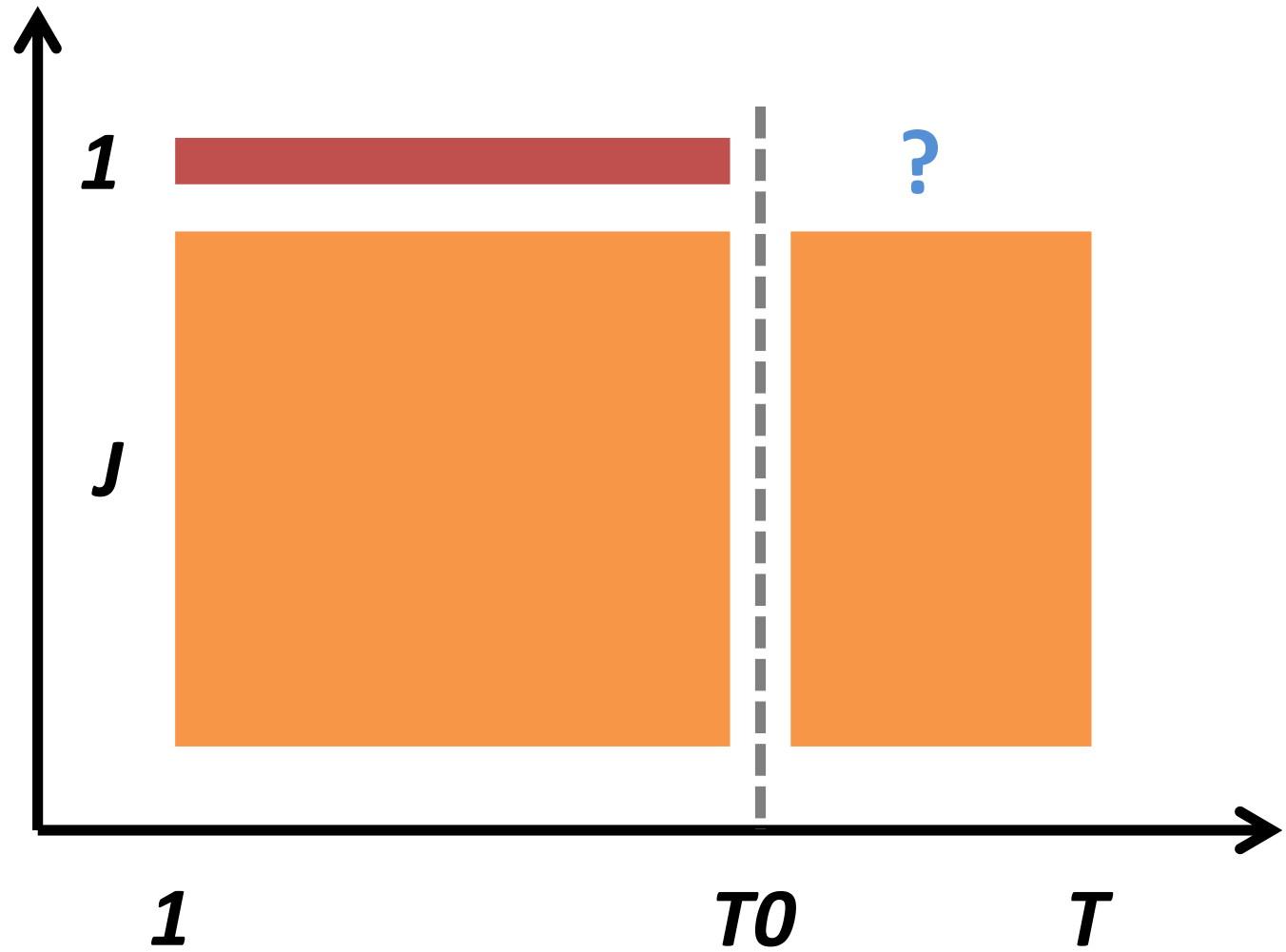
- Compare the evolution of an aggregate outcome for the unit affected by the intervention to the evolution of the same aggregate for some control group (e.g. Card, 1990, Card and Krueger, 1994, Abadie and Gardeazabal, 2003)
- Events or interventions take place at an aggregate level (e.g., cities, states, countries).

Challenges:

- N_{tr} is small by definition
- Selection of control group is often ambiguous
- Standard errors do not reflect uncertainty about the ability of the control group to reproduce the counterfactual of interest

Why synthetic control can be useful?

- Suppose that we observe $J+1$ units in periods $1, 2, \ldots, T$.
- Region "one" is exposed to the intervention during periods $T_0+1, \ldots, T$.
- Let Y_{it}^N be the outcome that would be observed for unit i at time t in the absence of the intervention.
- 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.
- We aim to estimate the effect of the intervention on the treated unit $(\alpha_{1T_0+1}, \ldots, \alpha_{1T})$, where $\alpha_{1t} = Y_{1t}^I - Y_{1t}^N = Y_{1t} - Y_{1t}^N$ for $t > T_0$.


1
?
J
1
T0
T

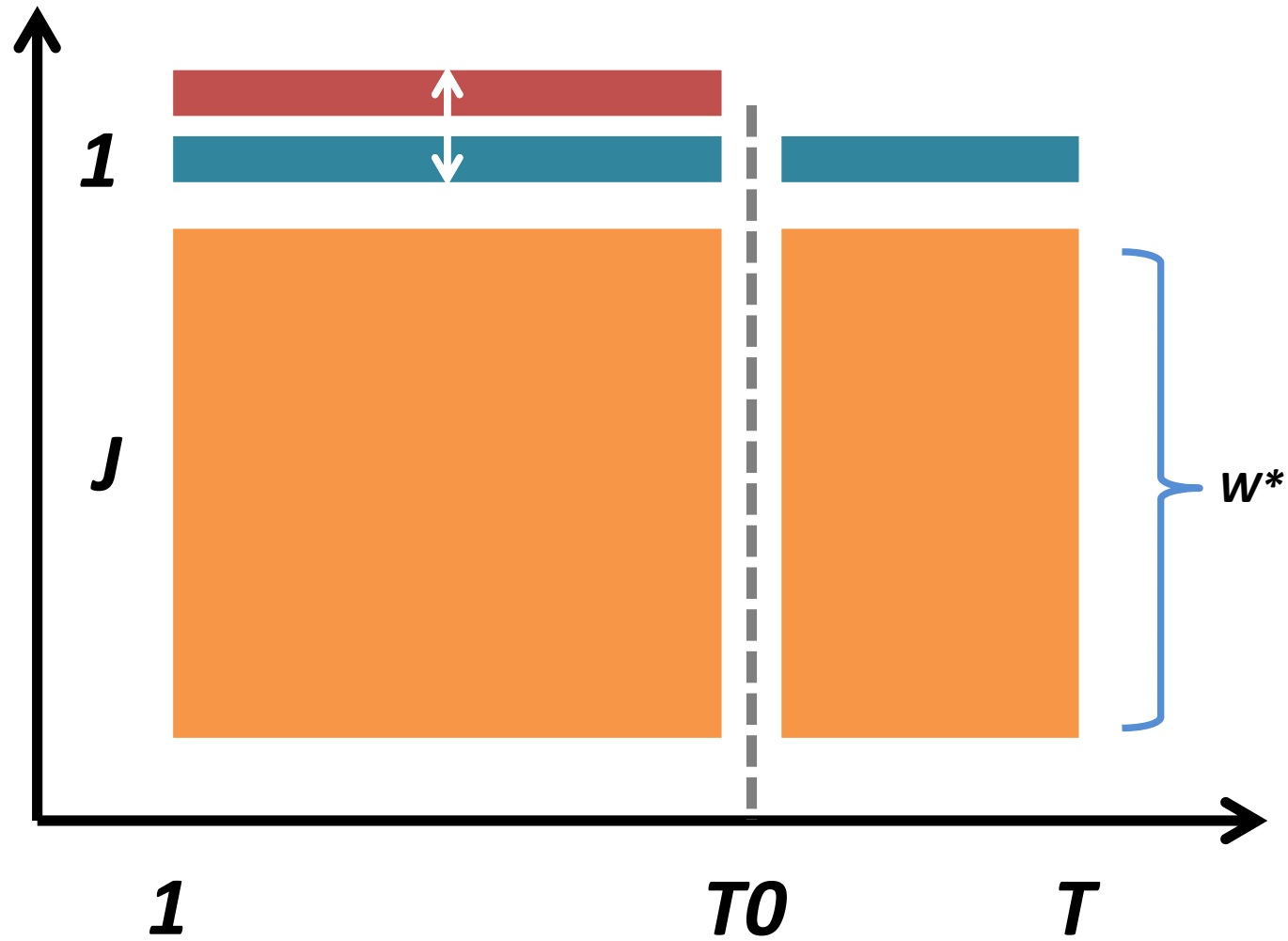
- Suppose that Y_{it}^N is given by a factor model:

$$Y_{it}^N = \delta_t + Z_i\theta_t + \lambda_t\mu_i + \varepsilon_{it},$$

  - δ_t is an unobserved (common) time-dependent factor,
  - Z_i is a $(1 \times r)$ vector of observed covariates,
  - θ_t is a $(r \times 1)$ vector of unknown parameters,
  - λ_t is a $(1 \times F)$ vector of unknown common factors,
  - μ_i is a $(F \times 1)$ vector of unknown factor loadings,
  - ε_{it} are unobserved transitory shocks.

- $\lambda_t\mu_i$: heterogeneous responses to multiple unobserved factors

- Basic idea: reweight the control group such that the synthetic control unit matches Z_i and (some) pre-treatment Y_{it} of the treated unit; as a result, μ_i is automatically matched


1
J
w*
1
T0
T

- Let $W = (w_2, \ldots, w_{J+1})'$ with $w_j \geq 0$ for $j = 2, \ldots, J+1$ and $w_2 + \cdots + w_{J+1} = 1$. Each value of W represents a potential synthetic control
- Let $\bar{Y}_i^{K_1}, \ldots, \bar{Y}_i^{K_M}$ be M linear functions of pre-intervention outcomes ($M \geq F$)
- Suppose that we can choose W^* such that:

$$\sum_{j=2}^{J+1} w_j^* Z_j = Z_1, \sum_{j=2}^{J+1} w_j^* \bar{Y}_j^{K_1} = \bar{Y}_1^{K_1}, \cdots, \sum_{j=2}^{J+1} w_j^* \bar{Y}_j^{K_M} = \bar{Y}_1^{K_M}.$$

- Then (if T_0 is large relative to the scale of ε_{it}), an approximately unbiased estimator of α_{1t} is:

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

for $t \in \{T_0 + 1, \ldots, T\}$

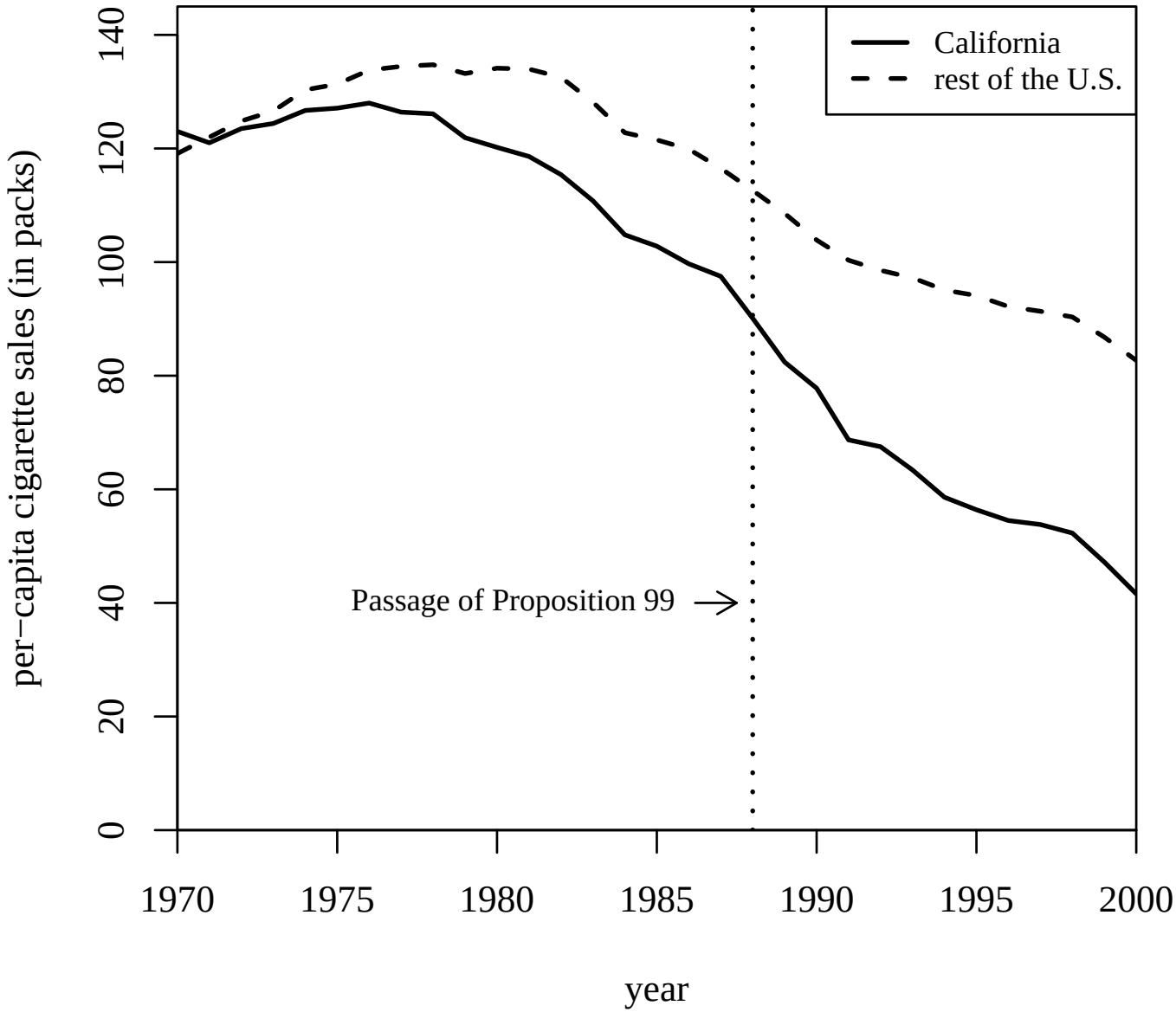
- Let $X_1 = (Z_1, \bar{Y}_1^{K_1}, \ldots, \bar{Y}_1^{K_M})'$ be a $(k \times 1)$ vector of pre-intervention characteristics.
- Similarly, X_0 is a $(k \times J)$ matrix which contains the same variables for the unaffected units.
- The vector W^* is chosen to minimize $\|X_1 - X_0 W\|$, subject to our weight constraints.
- We consider $\|X_1 - X_0 W\|_V = \sqrt{(X_1 - X_0 W)' V (X_1 - X_0 W)}$, where V is some $(k \times k)$ symmetric and positive semidefinite matrix.
- Various ways to choose V (subjective assessment of predictive power of X, regression, minimize MSPE, cross-validation, etc.).

In 1988, California first passed comprehensive tobacco control legislation:

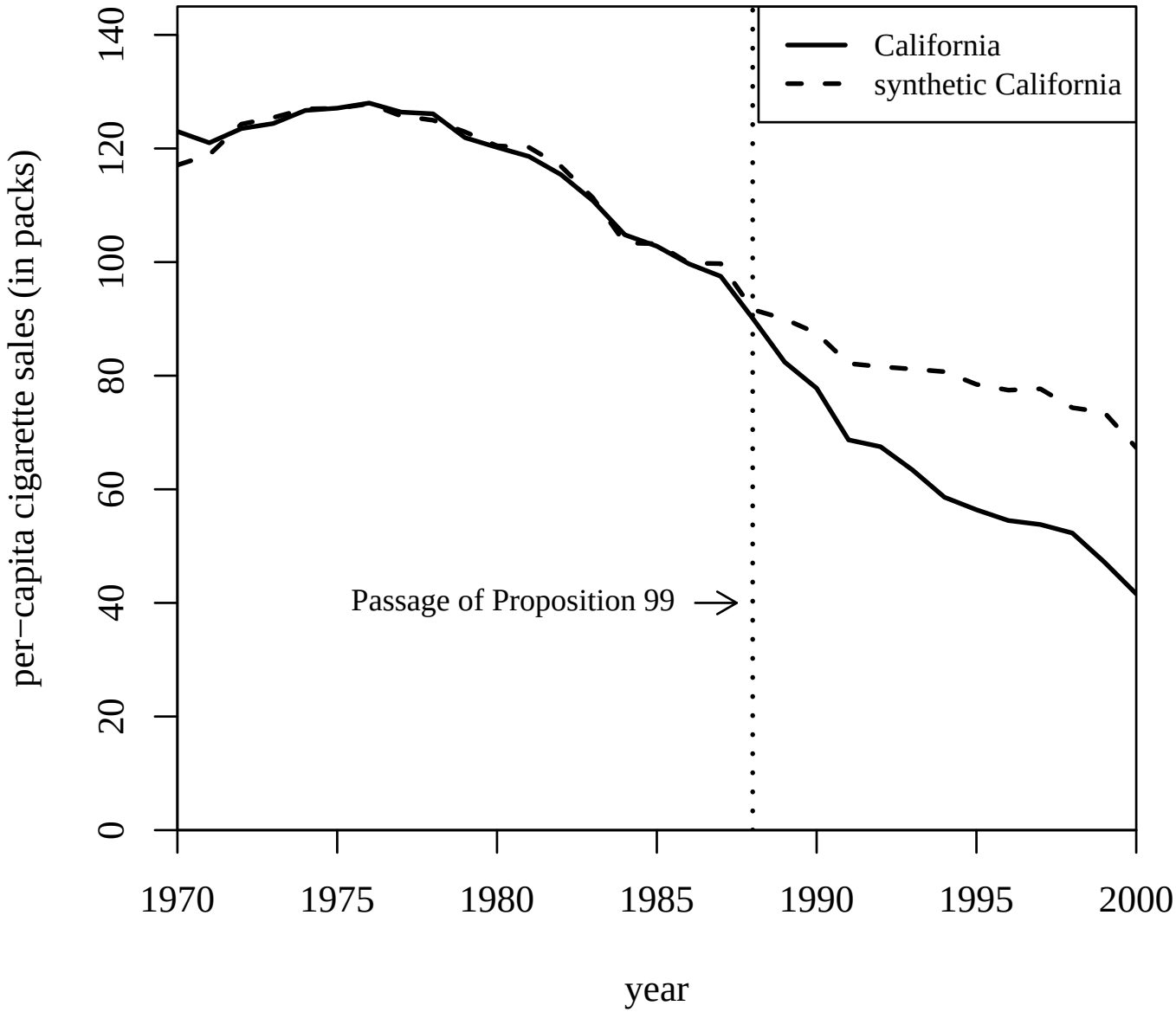
- increased cigarette tax by 25 cents/pack
- earmarked tax revenues to health and anti-smoking budgets
- funded anti-smoking media campaigns
- spurred clean-air ordinances throughout the state
- produced more than $100 million per year in anti-tobacco projects

Other states that subsequently passed control programs are excluded from donor pool of controls (AK, AZ, FL, HA, MA, MD, MI, NJ, NY, OR, WA, DC)

Cigarette Consumption: CA and the Rest of the U.S.



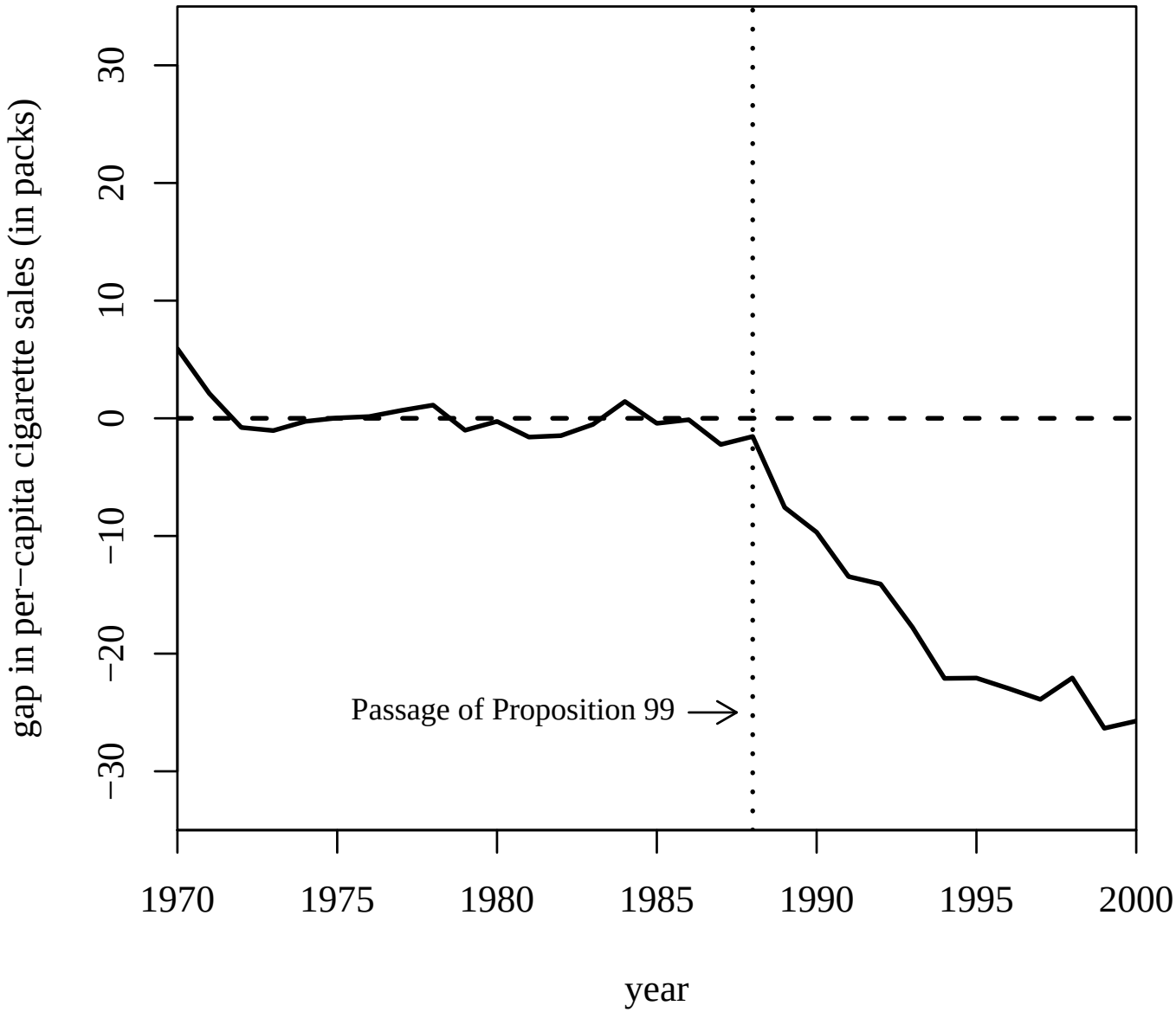
Cigarette Consumption: CA and synthetic CA



Variables	California		Average of
	Real	Synthetic	38 control states
Ln(GDP per capita)	10.08	9.86	9.86
Percent aged 15-24	17.40	17.40	17.29
Retail price	89.42	89.41	87.27
Beer consumption per capita	24.28	24.20	23.75
Cigarette sales per capita 1988	90.10	91.62	114.20
Cigarette sales per capita 1980	120.20	120.43	136.58
Cigarette sales per capita 1975	127.10	126.99	132.81

Note: All variables except lagged cigarette sales are averaged for the 1980-1988 period (beer consumption is averaged 1984-1988).

Smoking Gap Between CA and synthetic CA



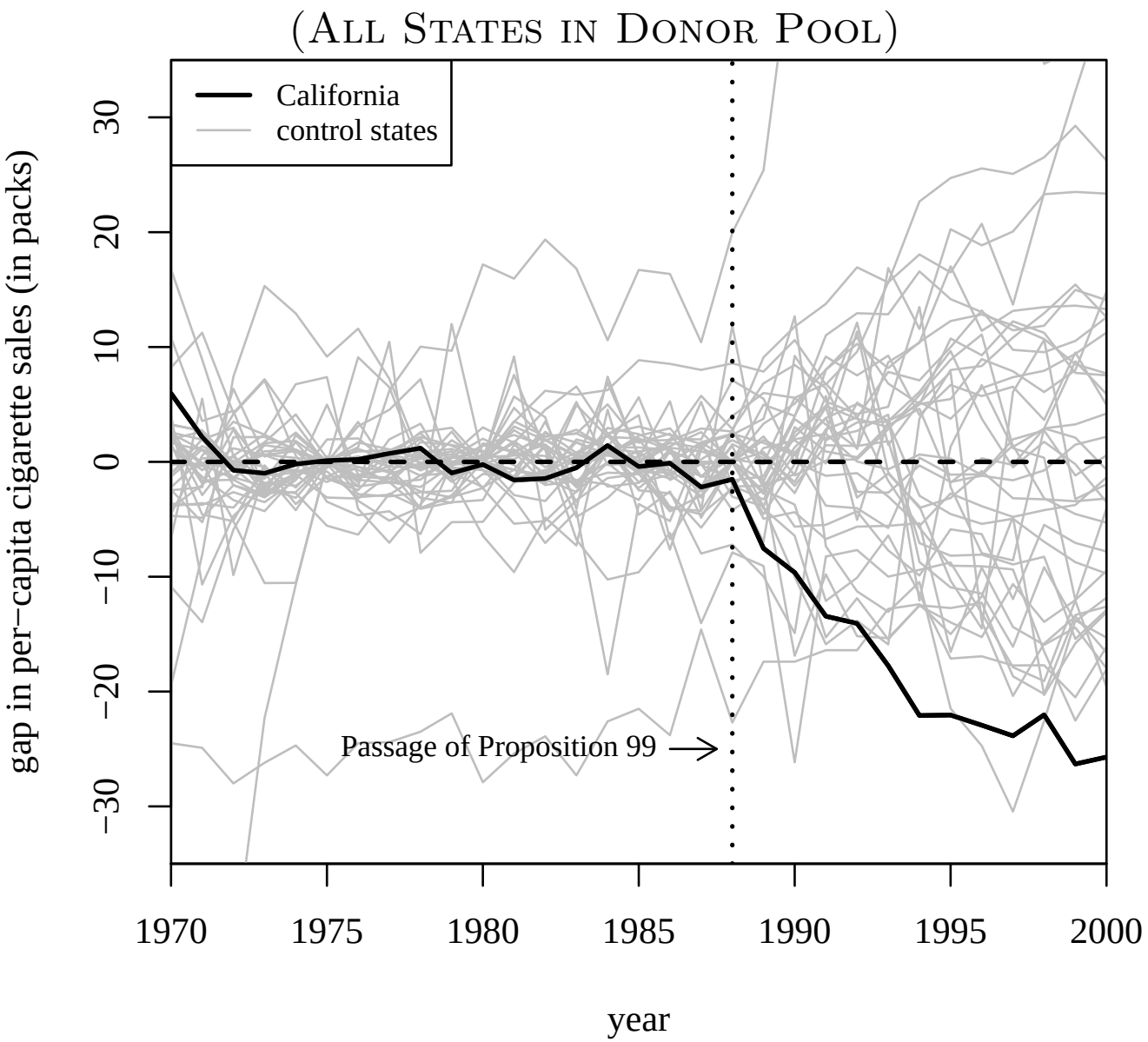
Strategy:

- Whether the effect estimated by the synthetic control for the unit affected by the intervention is large relative to the effect estimated for a unit chosen at random
- Valid regardless of the number of available comparison units, time periods, and whether the data are individual or aggregate

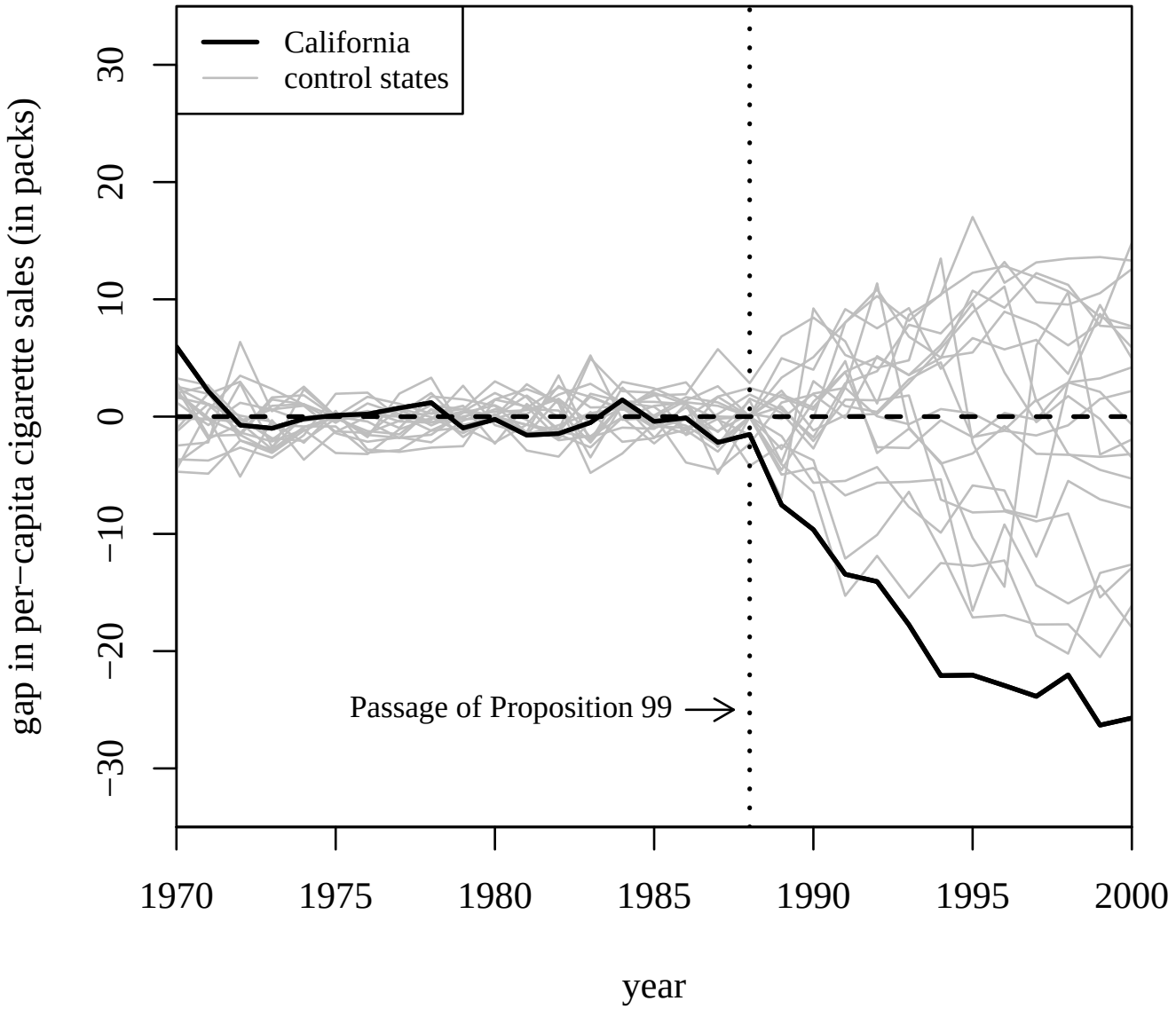
Implementation:

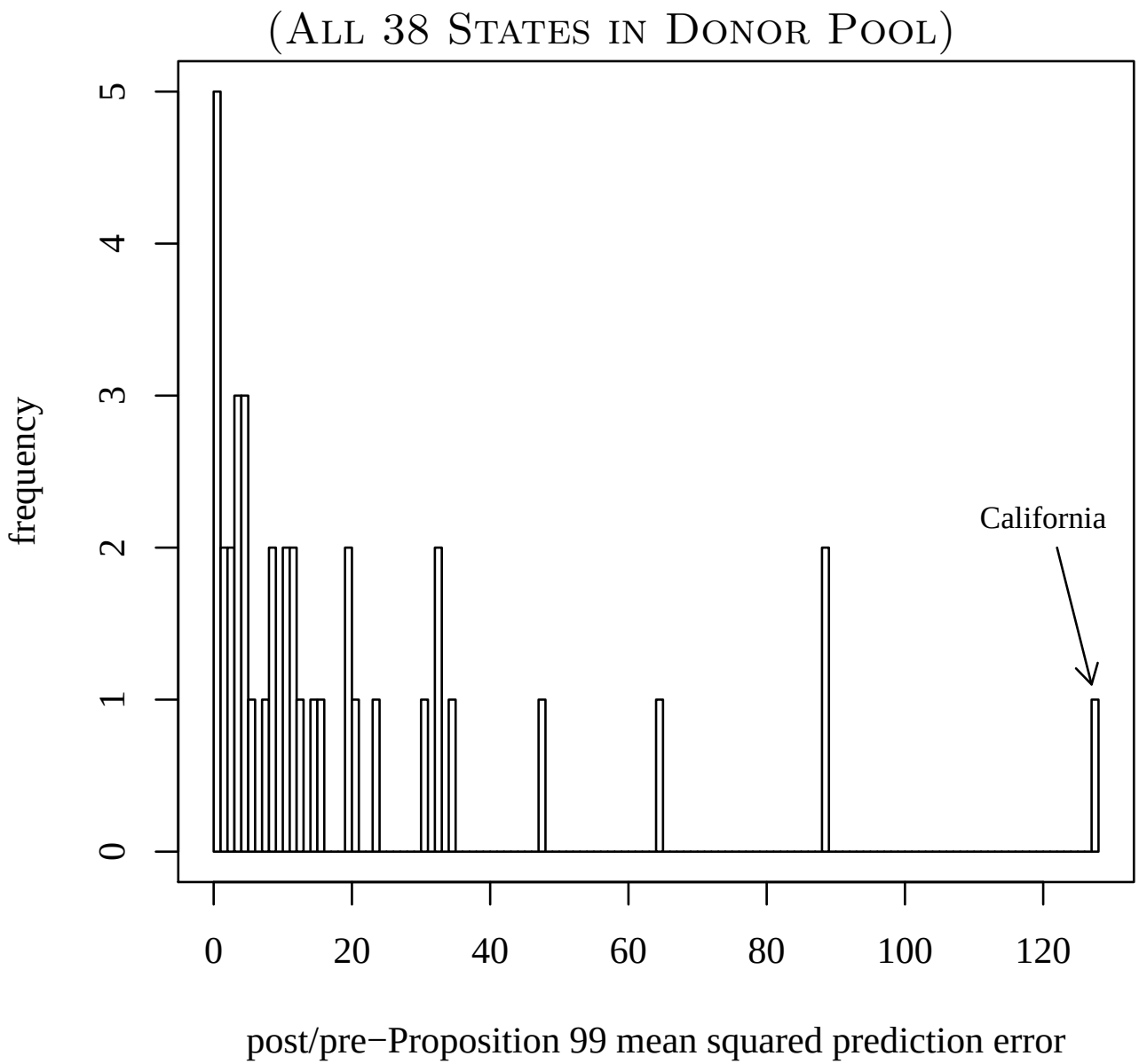
- Iteratively apply the synthetic method to each state in the "donor pool" and obtain a distribution of placebo effects
- Compare the gap for California to the distribution of the placebo gaps

Smoking Gap for CA and 38 control states



(Pre-Prop. 99 MSPE $\leq$ 2 Times Pre-Prop. 99 MSPE for CA)




(All 38 States in Donor Pool)
frequency
0
1
2
3
4
5
California
0
20
40
60
80
100
120
post/pre−Proposition 99 mean squared prediction error

Synth in cross-country studies

- Cross-country regressions are often criticized because they put side-by-side countries of very different characteristics.
- The synthetic control method provides an appealing data-driven procedure to study the effects of events or interventions that take place at the country level.

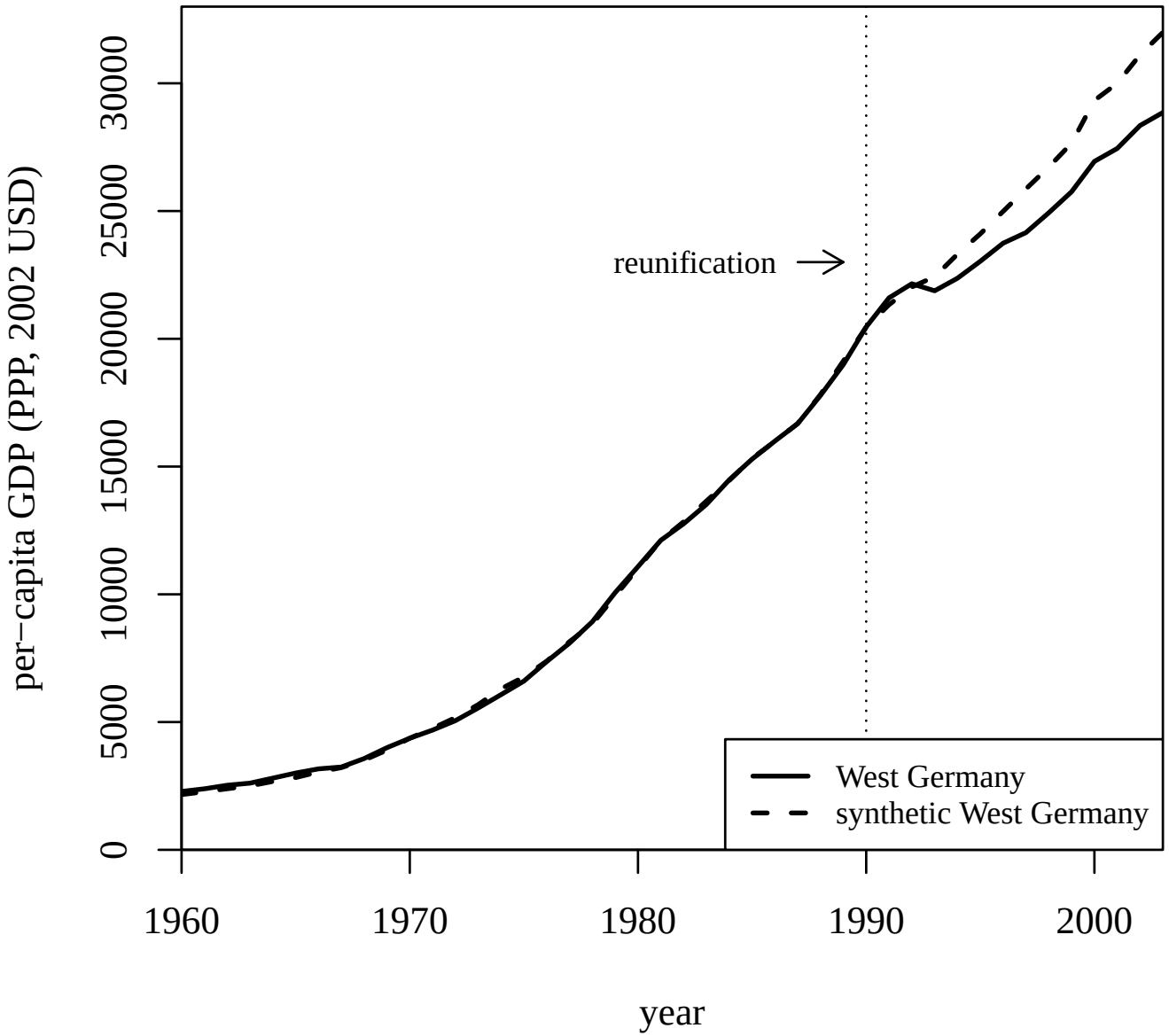
Application:

- The economic impact of the 1990 German unification in West Germany.
- Donor pool is restricted to 21 OECD countries.

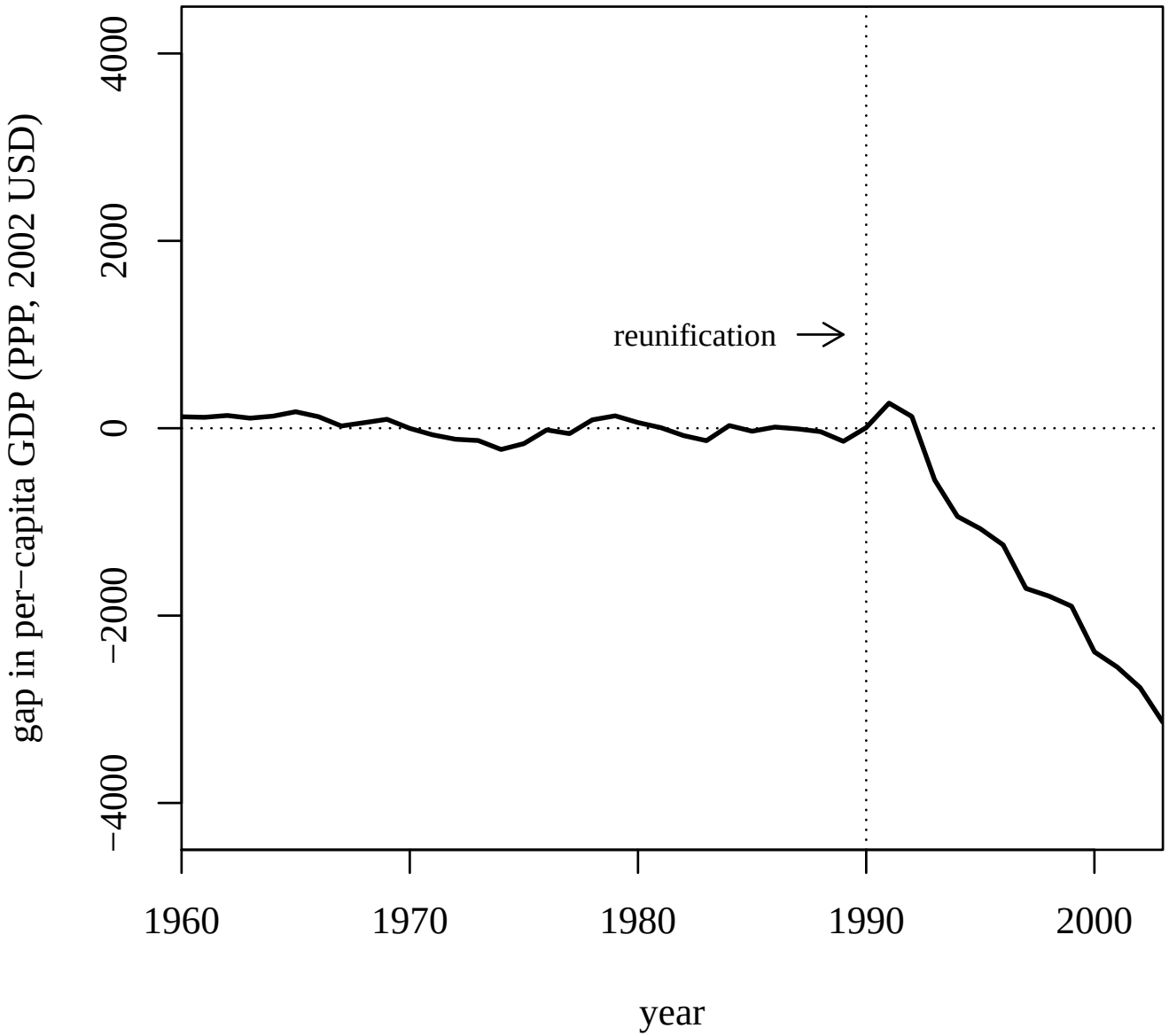
	West Germany	Synthetic West Germany	OECD Sample excl. West Germany
GDP per-capita	8169.8	8163.1	8049.3
Trade openness	45.8	54.4	32.6
Inflation rate	3.4	4.7	7.3
Industry share	34.7	34.7	34.3
Schooling	55.5	55.6	43.8
Investment rate	27.0	27.1	25.9

Note: GDP, inflation rate, and trade openness are averaged for the 1960–1989 period. Industry share is averaged for the 1980–1989 period. Investment rate and schooling are averaged for the 1980–1985 period.

West Germany and synthetic West Germany



GDP Gap: West Germany and synthetic West Germany



Country Weights in the Synthetic West Germany

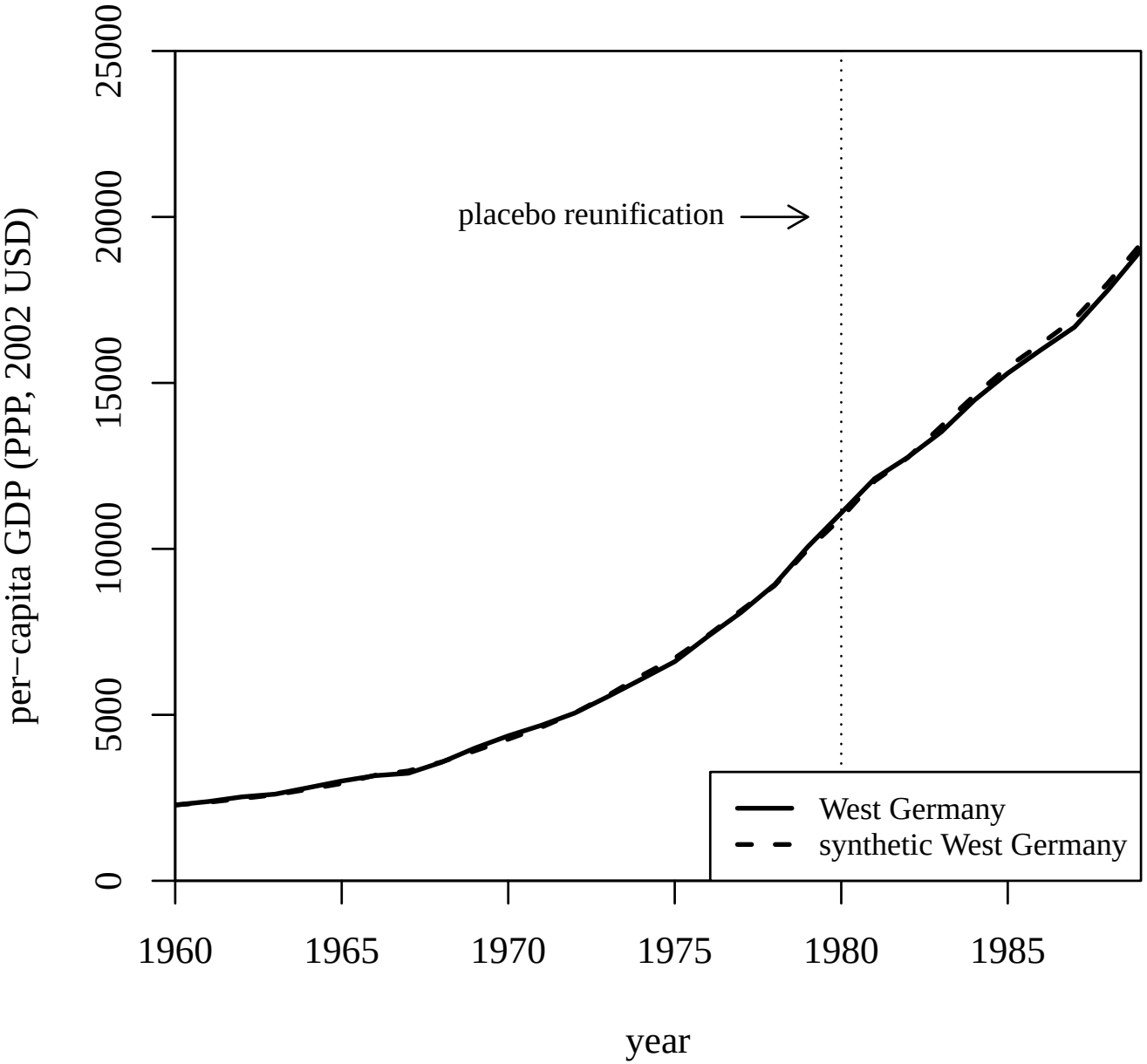
Country	Weight	Country	Weight
Australia	0	Netherlands	0.11
Austria	0.47	New Zealand	0.11
Belgium	0	Norway	0
Canada	0	Portugal	0
Denmark	0	Spain	0
France	0	Sweden	0
Greece	0	Switzerland	0
Ireland	0	United Kingdom	0.17
Italy	0	United States	0
Japan	0		0.14

Country Weights in the Synthetic West Germany

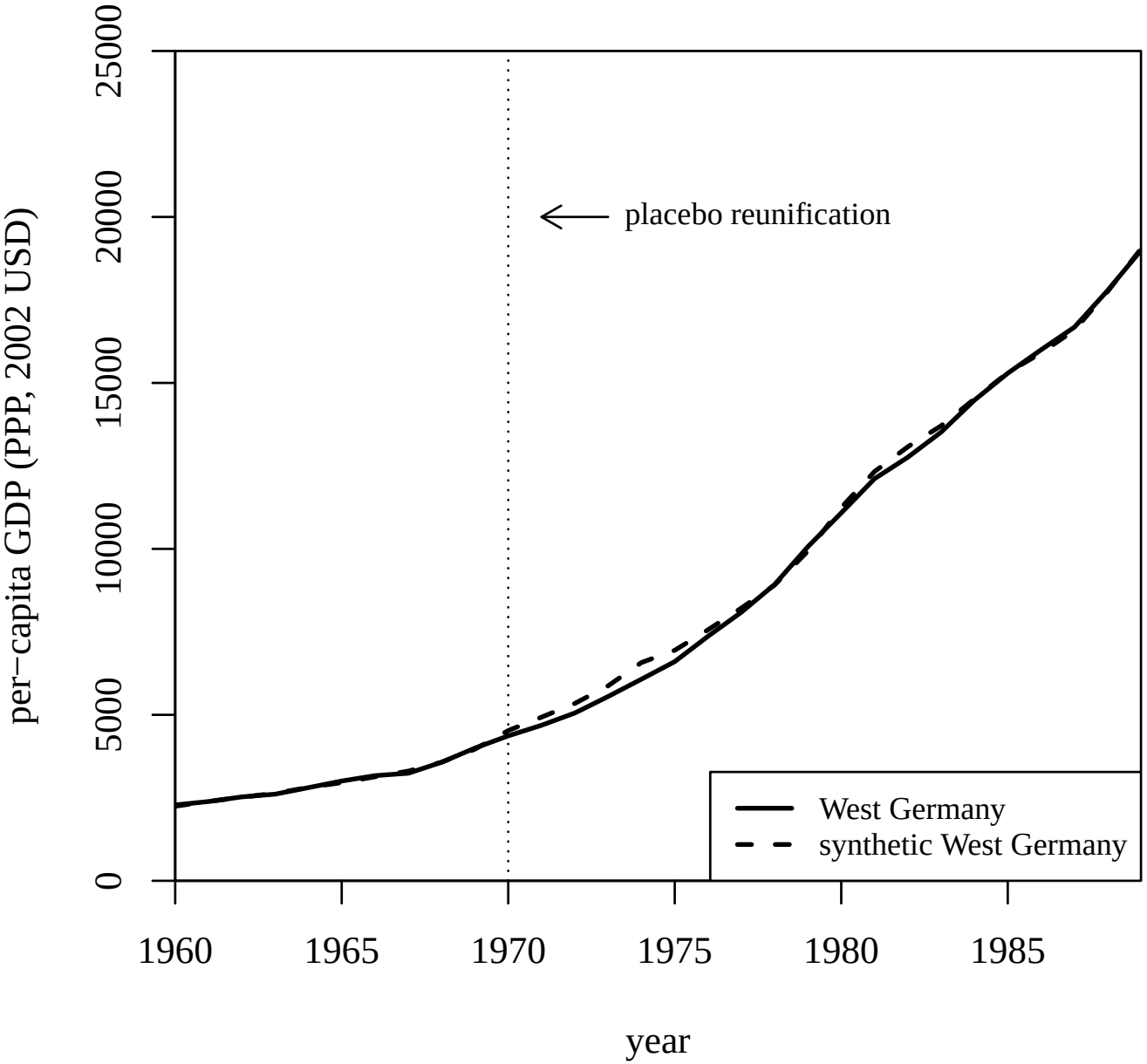
Country	Synthetic Weight	Regression Weight	Country	Synthetic Weight	Regression Weight
Australia	0	0.1	Netherlands	0.11	0.18
Austria	0.47	0.33	New Zealand	0	-0.08
Belgium	0	0.1	Norway	0	-0.07
Canada	0	0.09	Portugal	0	-0.14
Denmark	0	0.04	Spain	0	0
France	0	0.16	Switzerland	0.17	-0.06
Greece	0	0.02	United Kingdom	0	-0.04
Italy	0	-0.17	United States	0.14	0.21
Japan	0.11	0.32			

Note: Synthetic Weight: Unit weight assigned by the synthetic control method. Regression Weight: Unit weight assigned by linear regression.

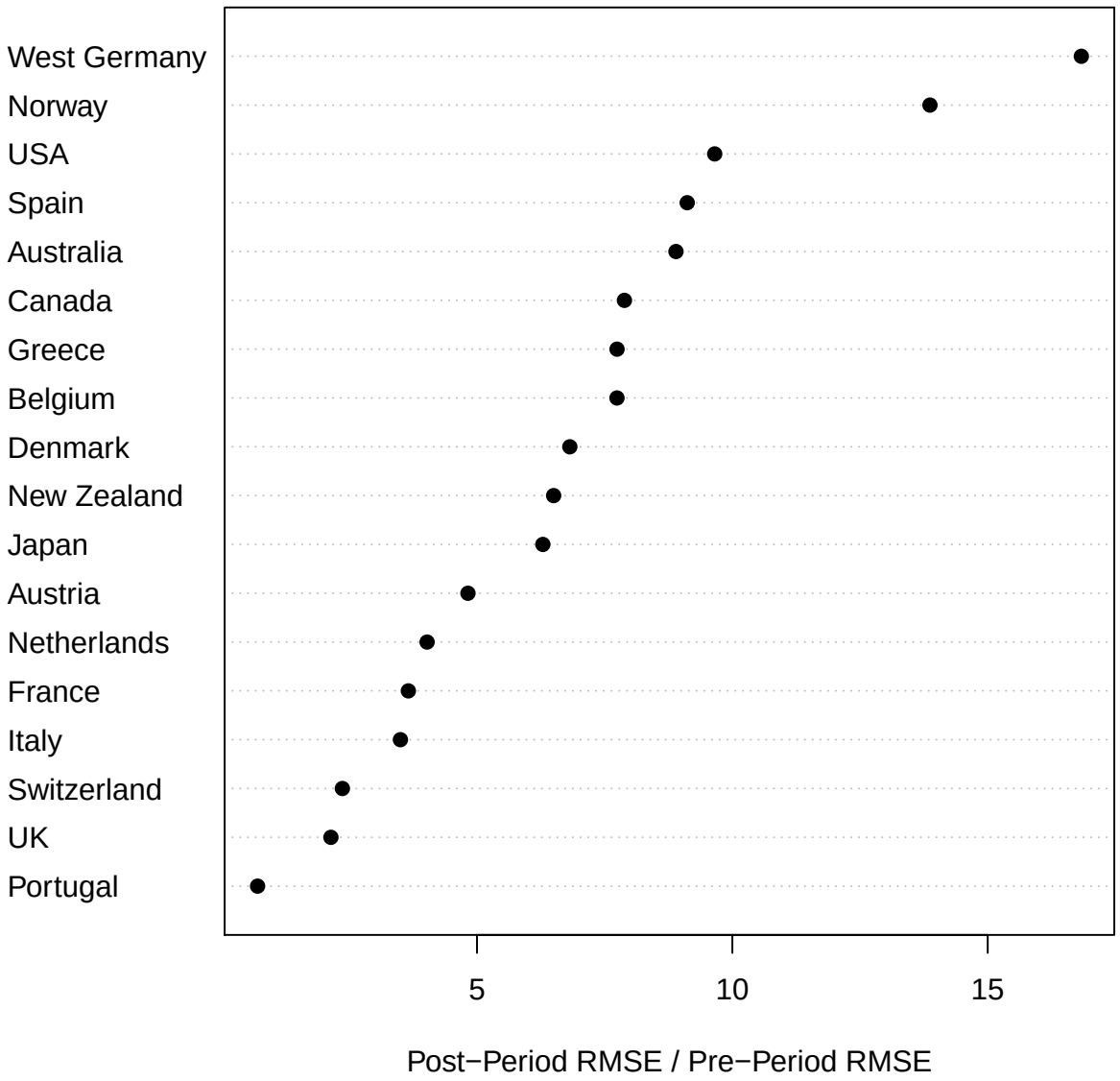
Placebo Reunification 1980



Placebo Reunification 1970



Ratio of post- and pre-reunification MSPE



- Working with panel data makes identification easier
  - Time-invariant confounders are controlled for
  - Selection on unobservables is allowed
  - The identifying assumption is not directly testable, but can be supported by data
- However, diff-in-diffs type of research designs (DID/FE/LDV) are not free from modelling assumptions
- The synthetic control method relaxes the model assumptions, but considers only one treated unit, requires smooth outcomes, and the inference is less formal $\rightarrow$ we are working on it