Making Policy with Data

An Introductory Course on Policy Evaluation

Lecture 5. Difference-in-Differences

Instructor: Prof Yiqing Xu

May 25

Selection on **Unobservables**

- Often there are reasons to believe that treated and untreated units differ in unobservable characteristics that are associated with potential outcomes even after controlling for differences in observed characteristics.

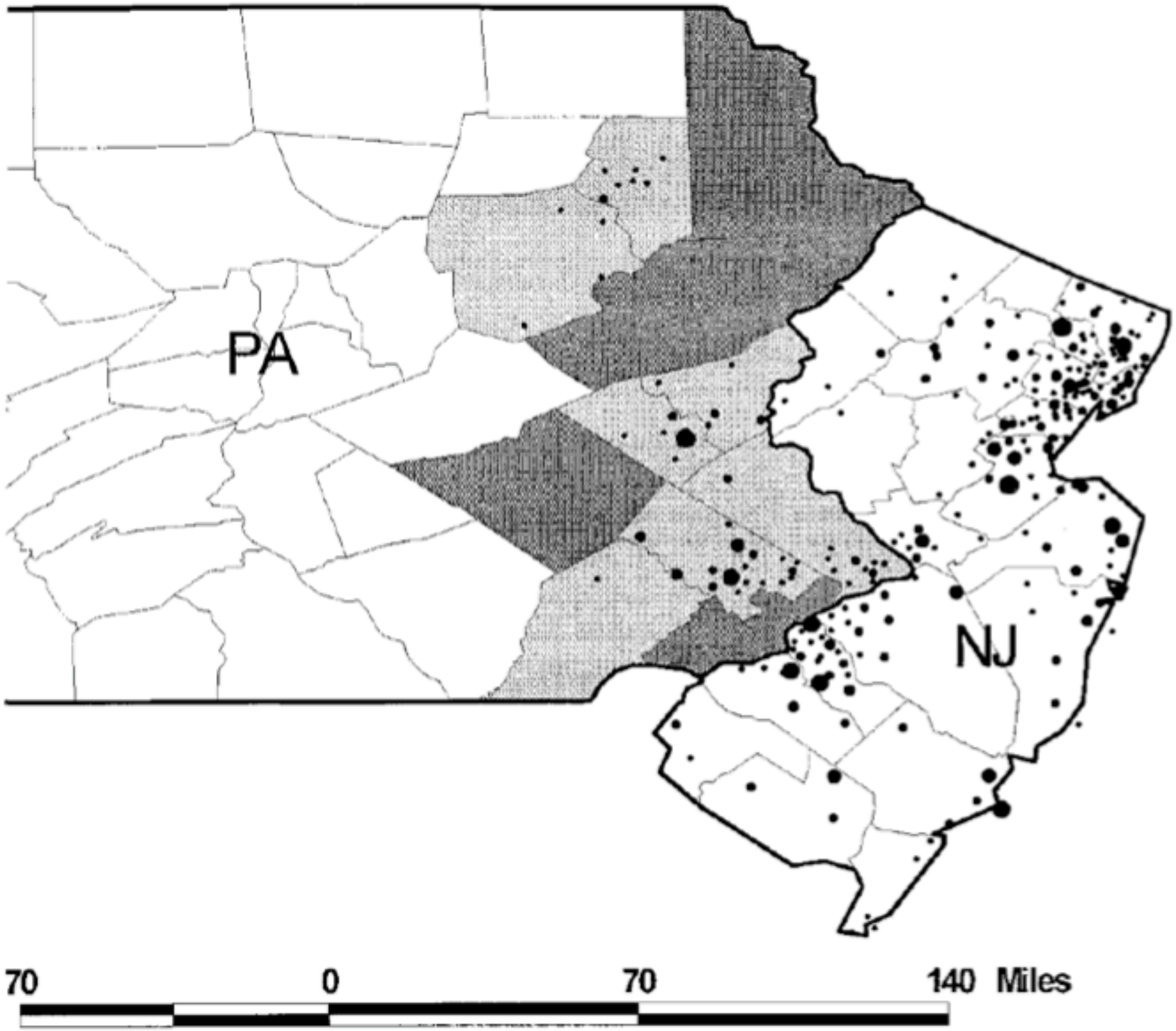
- In such cases, treated and untreated units are not directly comparable. What can we do then?

- If we can trace the research subjects over-time, maybe we can achieve more …

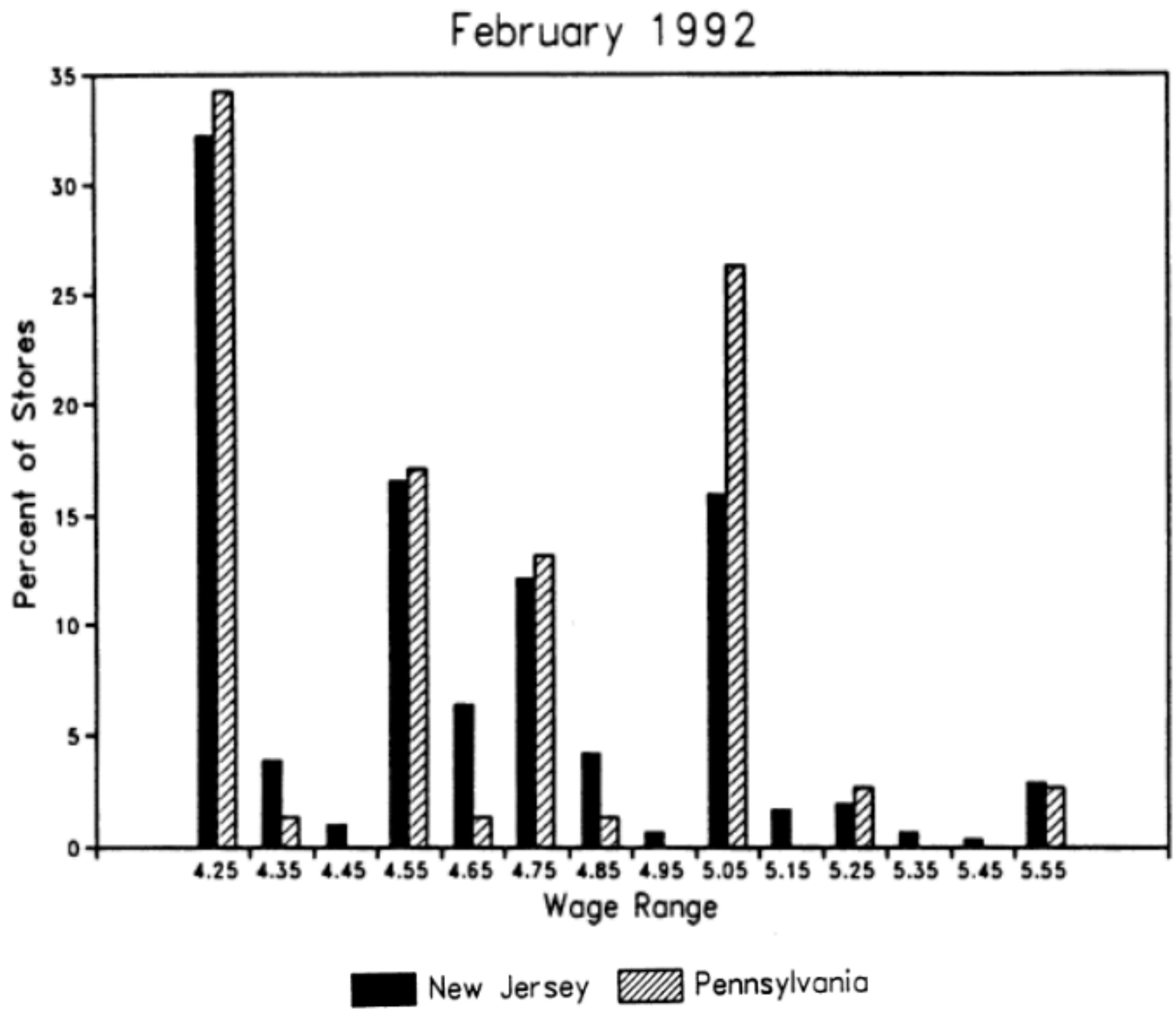
Do Higher Minimum Wages Reduce Employment?

- Difficult to answer. **Why?**
- Card and Krueger (1994) consider impact of New Jersey's 1992 minimum wage increase from $4.25 to $5.05 per hour
- Compare employment in 410 fast-food restaurants in New Jersey and eastern Pennsylvania before and after the rise
- Survey data on wages and employment from two waves:
    - **Wave 1**: March 1992, one month before the minimum wage increase
    - **Wave 2**: December 1992, eight months after increase

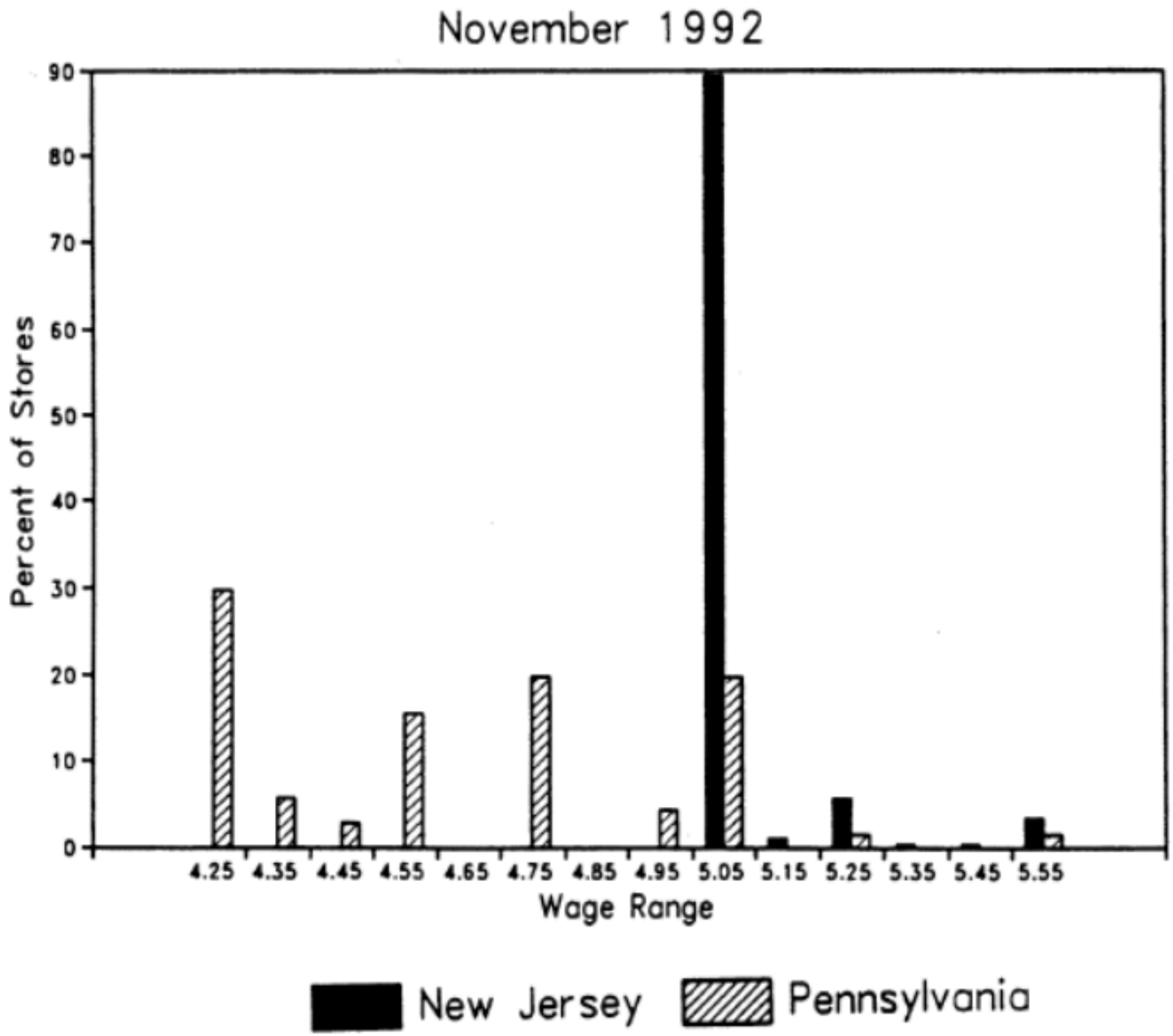
Restaurant Locations



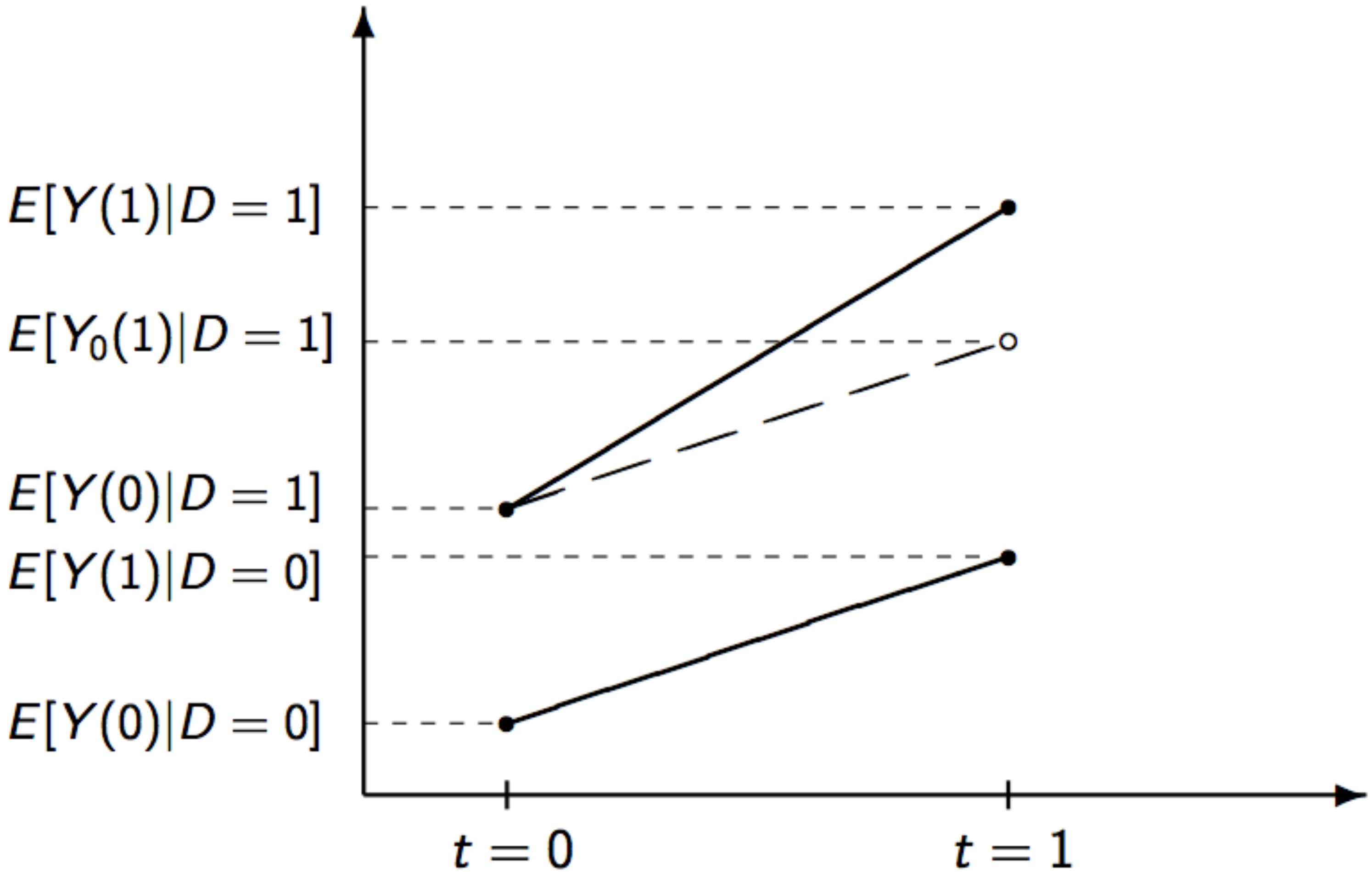
Wage *after* Minimum Wage Increase



Wage *before* Minimum Wage Increase



Difference-in-Differences



Plan

- **Identification**
- Estimation
- Threats to Validity

Setup: Two Groups, Two Periods

Definition

Two groups:

- $D = 1$ Treated units
- $D = 0$ Control units

Two periods:

- $T = 0$ Pre-Treatment period
- $T = 1$ Post-Treatment period

Potential outcomes $Y_d(t)$:

- $Y_{1i}(t)$ potential outcome unit i attains in period t when treated between t and $t - 1$
- $Y_{0i}(t)$ potential outcome unit i attains in period t with control between t and $t - 1$

Setup: Two Groups, Two Periods

Definition

Causal effect for unit i at time t is

- $\tau_{it} = Y_{1i}(t) - Y_{0i}(t)$

Observed outcomes $Y_i(t)$ are realized as

- $Y_i(t) = Y_{0i}(t) \cdot (1 - D_i(t)) + Y_{1i}(t) \cdot D_i(t)$

Fundamental problem of causal inference:

- If D occurs only after $t = 0$ ($D_i = D_i(1)$ and $Y_i(0) = Y_{0i}(0)$) we have: $Y_i(1) = Y_{0i}(1) \cdot (1 - D_i) + Y_{1i}(1) \cdot D_i$

Identification Problem

Estimand (ATT)

Focus on estimating the average effect of the treatment on the treated: $\tau_{ATT} = E[Y_1(1) - Y_0(1)|D = 1]$

	Post-Period (T=1)	**Pre-Period (T=0)**
Treated D=1	$E[Y_1(1)\|D = 1]$	$E[Y_0(0)\|D = 1]$
Control D=0	$E[Y_0(1)\|D = 0]$	$E[Y_0(0)\|D = 0]$

Problem

Missing potential outcome: $E[Y_0(1)|D = 1]$, *ie. what is the average post-period outcome for the treated in the absence of the treatment?*

Identification Problem

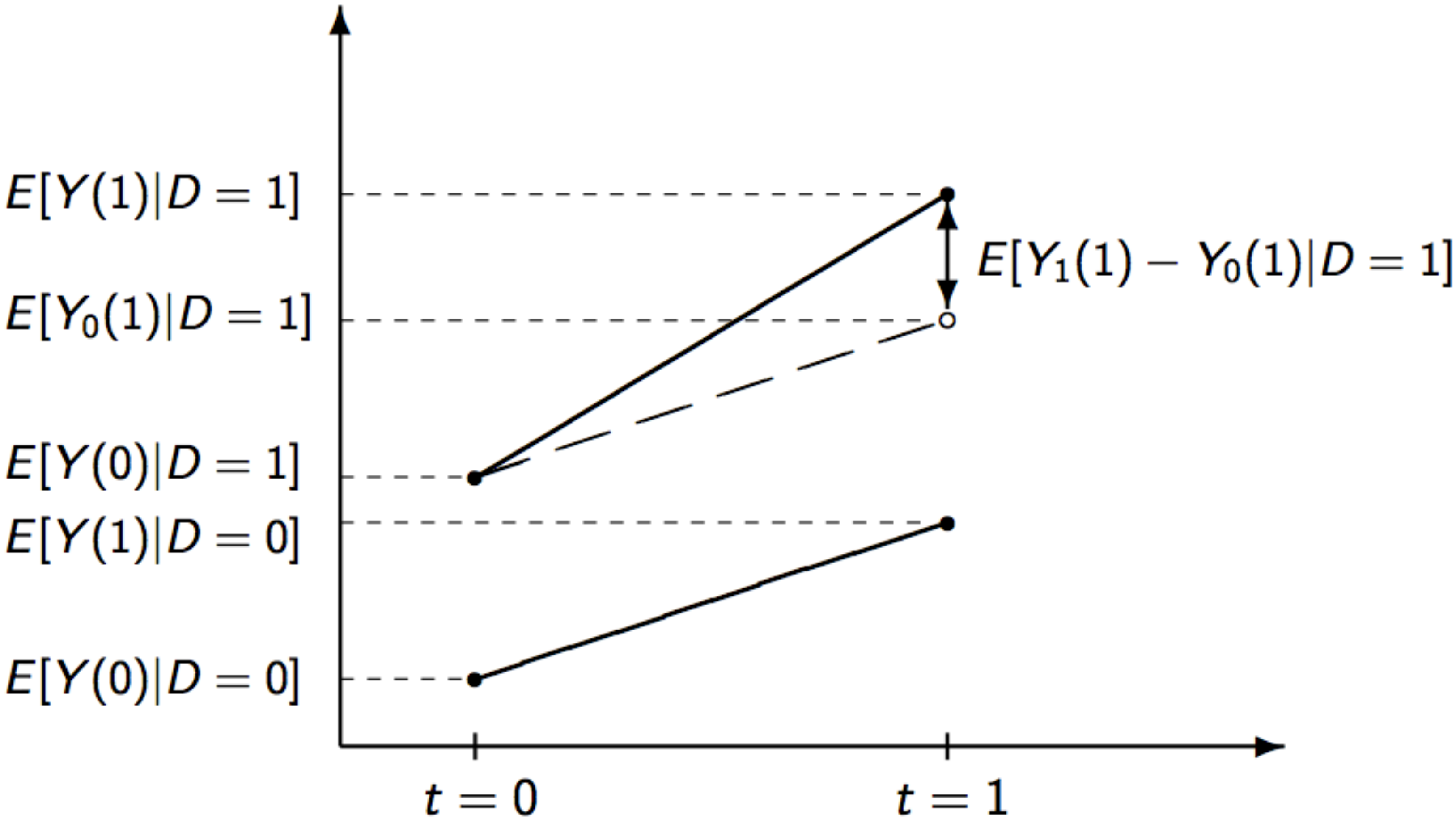
Estimand (ATT)

Focus on estimating the average effect of the treatment on the treated: $\tau_{ATT} = E[Y_1(1) - Y_0(1)|D = 1]$

	Post-Period (T=1)	Pre-Period (T=0)
Treated D=1	$E[Y_1(1)\|D = 1]$	$E[Y_0(0)\|D = 1]$
Control D=0	$E[Y_0(1)\|D = 0]$	$E[Y_0(0)\|D = 0]$

- Assumes: $E[Y_0(1) - Y_0(0)|D = 1] = E[Y_0(1) - Y_0(0)|D = 0]$

Identification Strategy



Identification with Diff-in-Diffs

Identification Assumption (parallel trends)

$E[Y_0(1) - Y_0(0)|D = 1] = E[Y_0(1) - Y_0(0)|D = 0]$

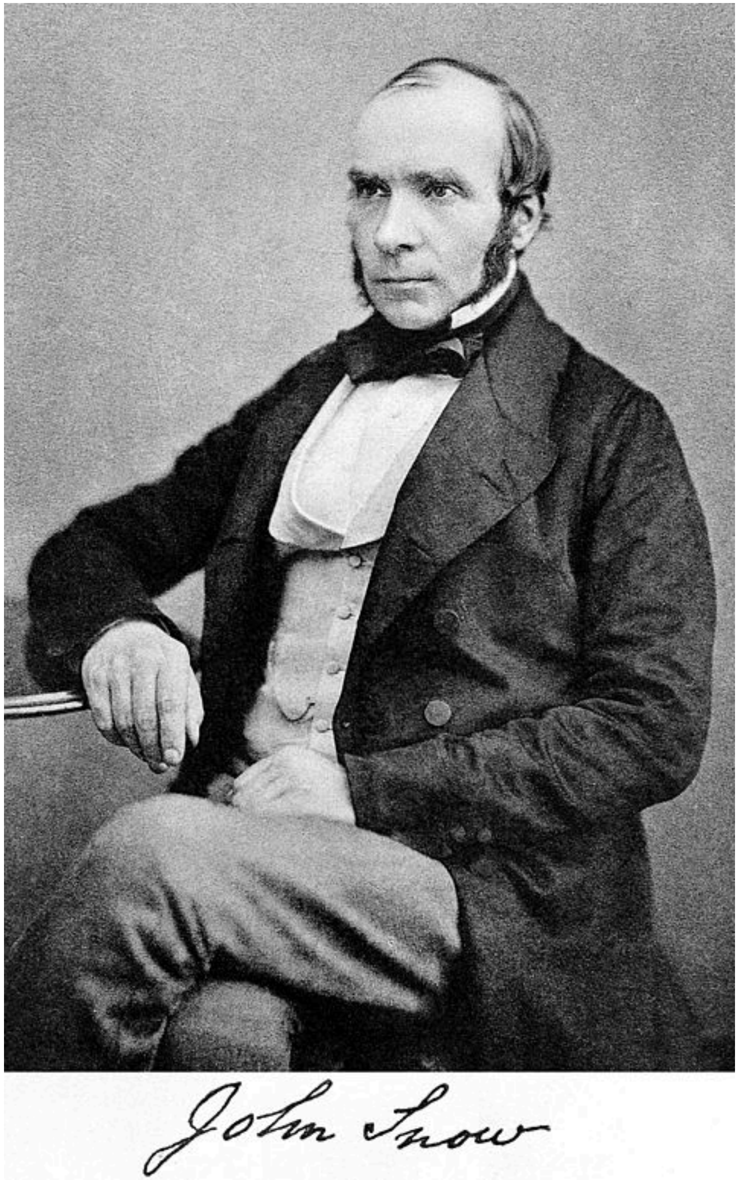
Identification Result

Given parallel trends the ATT is identified as:

$$\begin{aligned} E[Y_1(1) - Y_0(1)|D = 1] &= \Big\{E[Y(1)|D = 1] - E[Y(1)|D = 0]\Big\} \\ &- \Big\{E[Y(0)|D = 1] - E[Y(0)|D = 0]\Big\} \end{aligned}$$

Proof? (optional)

The Birth of Diff-in-Diffs

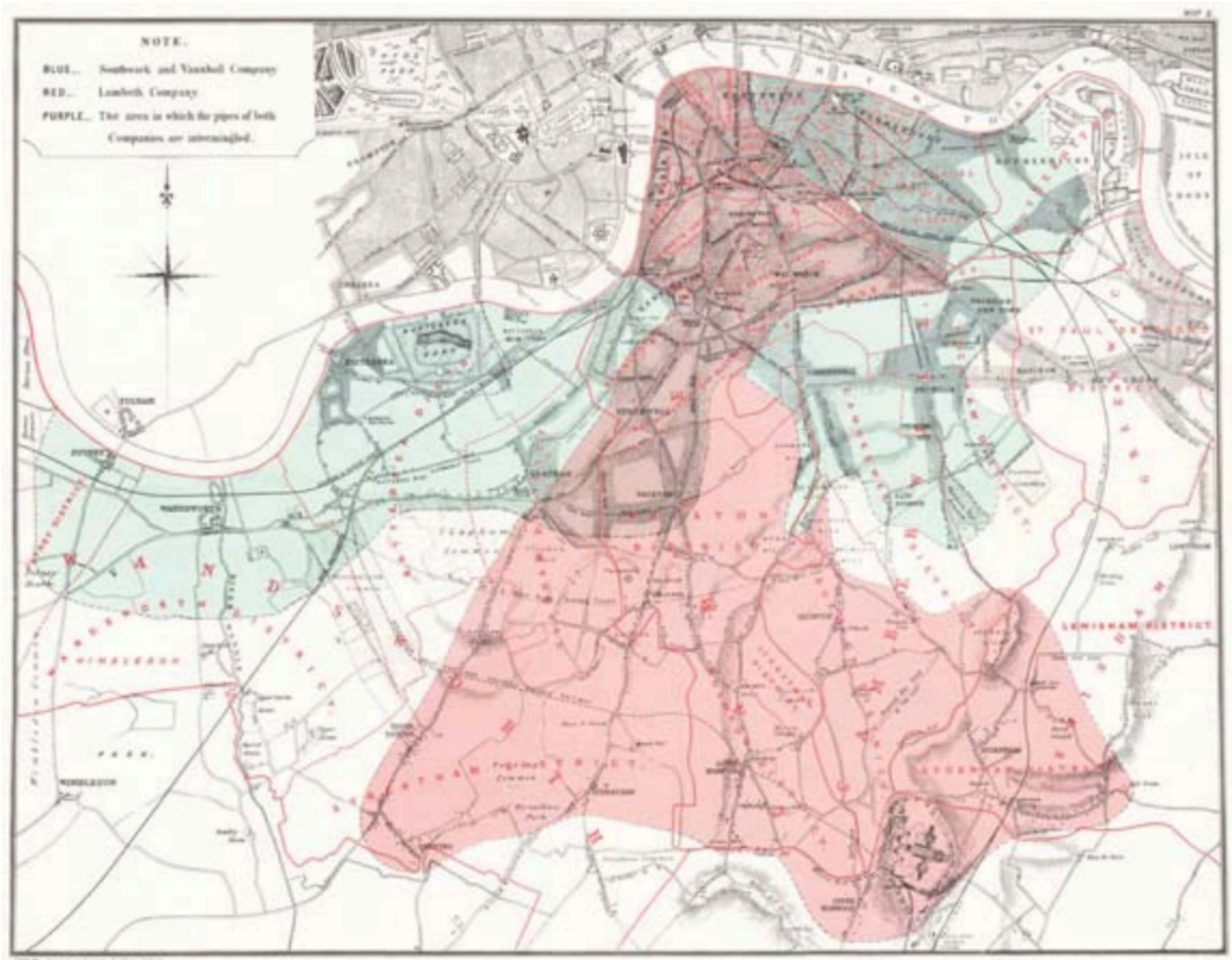


- First developed by British physician John Snow (1813 - 1858)
- Study the cholera epidemic in London in 1849, which claimed over 14,000 lives
- John Snow believed cholera was spread by contaminated water
- But how to prove it?

The Birth of Diff-in-Diffs

- **First Difference:** Water provided by two companies, (1) the Lambeth and (2) the Southwark and Vauxhall. Both got water from the Thames.
- **Second Difference:** Before and after 1852. In 1852, Lambeth moved their intake upriver
- It turned out that Lambeth customers were less likely to get sick afterwards — Difference-in-Differences is born!

The Birth of Diff-in-Diffs



The Birth of Diff-in-Diffs

- **First Difference:** Water provided by two companies, (1) the Lambeth and (2) the Southwark and Vauxhall. Both got water from the Thames.

- **Second Difference:** Before and after 1852. In 1852, Lambeth moved their intake upriver

- It turned out that Lambeth customers were less likely to get sick afterwards — the origin of Difference-in-Differences

- Southwark and Vauxhall: 71 cholera deaths/10,000 homes

- Lambeth after moving water source: 5 cholera deaths/10,000 homes

- As a result, Southwark and Vauxhall moved their intake upriver in 1855 and the epidemic subsided

Plan

- Identification
- **Estimation**
- Threats to Validity

Estimation

Estimand (ATT)

$$E[Y_1(1) - Y_0(1)|D=1] = \left\{E[Y(1)|D=1] - E[Y(1)|D=0]\right\} - \left\{E[Y(0)|D=1] - E[Y(0)|D=0]\right\}$$

Estimator (Sample Means: Panel)

$$\left\{\frac{1}{N_1}\sum_{D_i=1} Y_i(1) - \frac{1}{N_0}\sum_{D_i=0} Y_i(1)\right\} - \left\{\frac{1}{N_1}\sum_{D_i=1} Y_i(0) - \frac{1}{N_0}\sum_{D_i=0} Y_i(0)\right\}$$

$$= \left\{\frac{1}{N_1}\sum_{D_i=1}\{Y_i(1) - Y_i(0)\} - \frac{1}{N_0}\sum_{D_i=0}\{Y_i(1) - Y_i(0)\}\right\},$$

where N_1 and N_0 are the number of treated and control units respectively.

Minimum Wage on Employment

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

Estimation

Estimator (Sample Means: Repeated Cross-Sections)

Let $\{Y_i, D_i, T_i\}_{i=1}^{n}$ *be the pooled sample (the two different cross-sections merged) where* T *is a random variable that indicates the period (0 or 1) in which the individual is observed.*

The difference-in-differences estimator is given by:

$$\left\{ \frac{\sum D_i \cdot T_i \cdot Y_i}{\sum D_i \cdot T_i} - \frac{\sum (1 - D_i) \cdot T_i \cdot Y_i}{\sum (1 - D_i) \cdot T_i} \right\} - \left\{ \frac{\sum D_i \cdot (1 - T_i) \cdot Y_i}{\sum D_i \cdot (1 - T_i)} - \frac{\sum (1 - D_i) \cdot (1 - T_i) \cdot Y_i}{\sum (1 - D_i) \cdot (1 - T_i)} \right\}$$

Estimation

Estimator (Regression: Repeated Cross-Sections)

Alternatively, the same estimator can be obtained using regression techniques. Consider the linear model:

$$Y = \mu + \gamma \cdot D + \delta \cdot T + \tau \cdot (D \cdot T) + \varepsilon,$$

where $E[\varepsilon|D, T] = 0$.

Easy to show that τ estimates the DD effect:

$$\begin{aligned} \tau = & \ \{E[Y|D=1, T=1] - E[Y|D=0, T=1]\} \\ - & \ \{E[Y|D=1, T=0] - E[Y|D=0, T=0]\} \end{aligned}$$

Estimation

Estimator (Regression: Repeated Cross-Sections)

Alternatively, the same estimator can be obtained using regression techniques. Consider the linear model:

$$Y = \mu + \gamma \cdot D + \delta \cdot T + \tau \cdot (D \cdot T) + \varepsilon,$$

where $E[\varepsilon | D, T] = 0$.

	After (T=1)	**Before (T=0)**	**After - Before**
Treated D=1	$\mu + \gamma + \delta + \tau$	$\mu + \gamma$	$\delta + \tau$
Control D=0	$\mu + \delta$	μ	δ
Treated - Control	$\gamma + \tau$	γ	τ

Estimation

Estimator (Regression: Panel Data)

With panel data we can estimate the difference-in-differences effect using a fixed effects regression with unit and period fixed effects:

$$Y_{it} = \mu + \gamma_i + \delta T + \tau D_{it} + X_{it}'\beta + \varepsilon_{it}$$

- *One intercept for each unit* γ_i
- D_{it} *coded as 1 for treated in post-period and 0 otherwise*

Or equivalently we can use regression with the dependent variable in first differences:

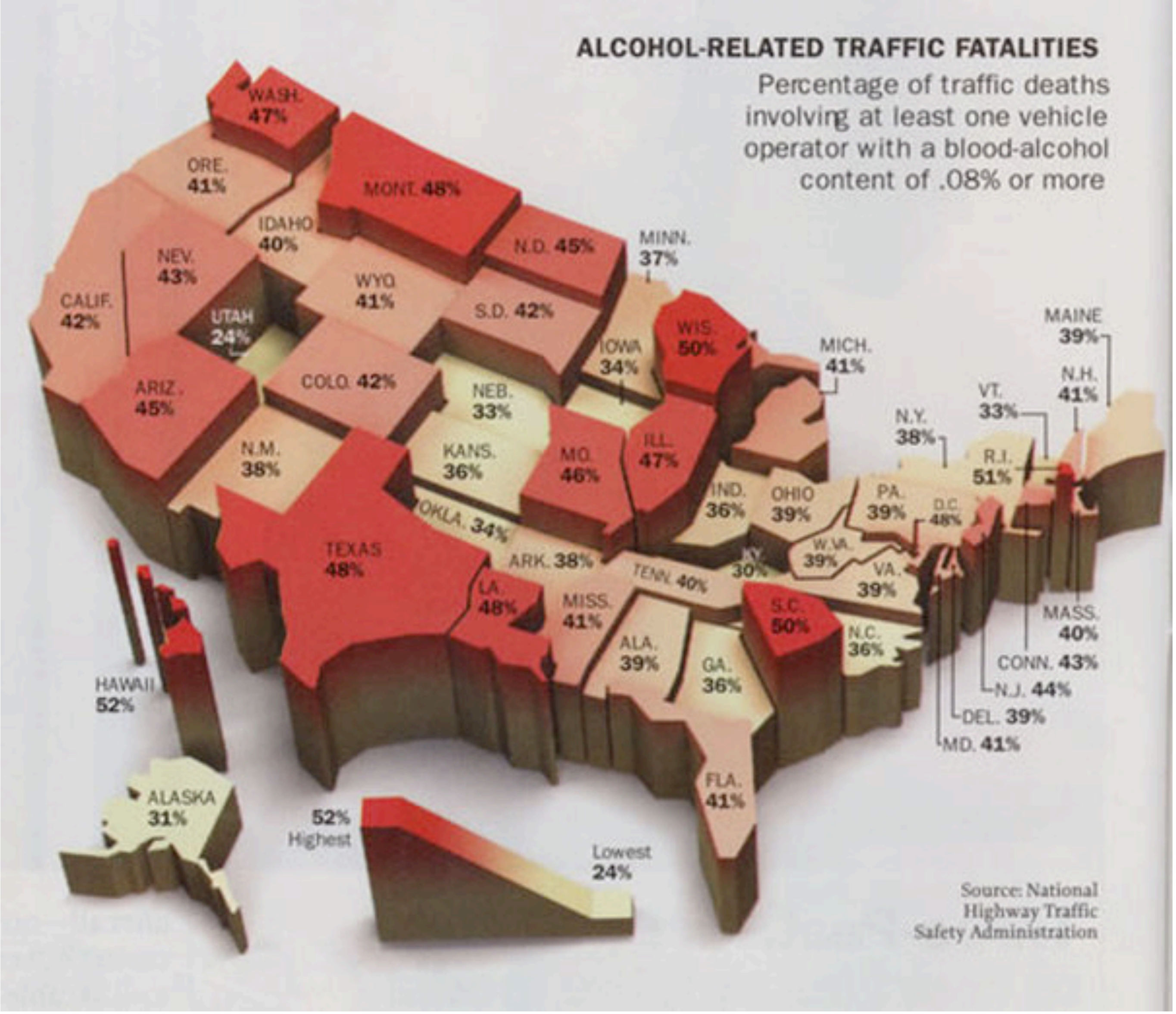
$$\Delta Y_i = \delta + \tau \cdot D_i + u_i,$$

where $\Delta Y_i = Y_i(1) - Y_i(0)$ *and* $u_i = \Delta\varepsilon_i$.

Extension

- More than 2 periods
- More than 2 groups, e.g. treatments with intensity (or a continuous treatment)
- Different treatment timing
- **Example:** The effect of lower minimum legal drinking age on fatalities in traffic accidents (Angrist-Pishcke Chapter 5)

Example: The Effect of Lower Drinking Age



Example: The Effect of Lower Drinking Age

- Since 1933, most states maintain MLDA at 21
- Kansas, New York, North Carolina and a few others allowed drinking at 18
- In 1971, many states lowered the drinking age to 18, but Arkansas, California, and Pennsylvania kept MLDA at 21.
- In 1988, all 50 states and DC opted for an MLDA at 21

Example: The Effect of Lower Drinking Age

$$Y_{st} = \alpha + \delta_{rDD} LEGAL_{st} + \sum_{k=\text{Alaska}}^{\text{Wyoming}} \beta_k STATE_{ks} + \sum_{j=1971}^{1983} \gamma_j YEAR_{jt} + e_{st}. \quad (5.5)$$

- $LEGAL_{st}$: the proportion of 18-20 year-olds allow to drink in state s at time t
- $STATE_{ks}$: a dummy variable (taking values 0 or 1) indicating state s is which of the 50 states
- $YEAR_{jt}$: a dummy variable indicating year *t* is which year from 1971 to 1983.

Example: The Effect of Lower Drinking Age

TABLE 5.2
Regression DD estimates of MLDA effects on death rates

Dependent variable	(1)	(2)	(3)	(4)
All deaths	10.80	8.47	12.41	9.65
	(4.59)	(5.10)	(4.60)	(4.64)
Motor vehicle accidents	7.59	6.64	7.50	6.46
	(2.50)	(2.66)	(2.27)	(2.24)
Suicide	.59	.47	1.49	1.26
	(.59)	(.79)	(.88)	(.89)
All internal causes	1.33	.08	1.89	1.28
	(1.59)	(1.93)	(1.78)	(1.45)
State trends	No	Yes	No	Yes
Weights	No	No	Yes	Yes

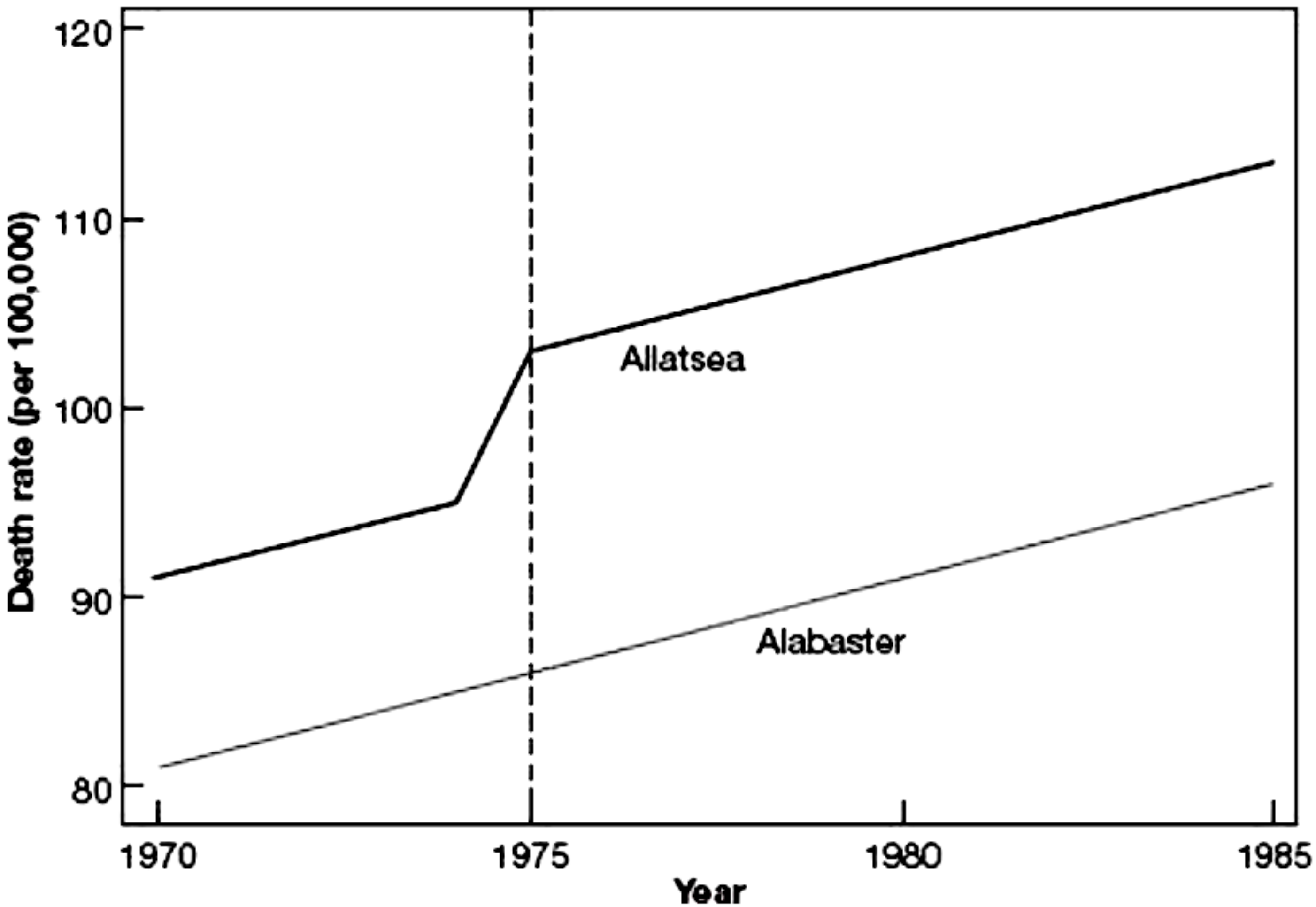
Example: The Effect of Lower Drinking Age

$$Y_{st} = \alpha + \delta_{r\,DD} LEGAL_{st} + \sum_{k=\text{Alaska}}^{\text{Wyoming}} \beta_k STATE_{ks} + \sum_{j=1971}^{1983} \gamma_j YEAR_{jt} + \sum_{k=\text{Alaska}}^{\text{Wyoming}} \theta_k \left(STATE_{ks} \times t\right) + e_{st}. \tag{5.6}$$

- *For each state, fit a state-specific trend*

Example: The Effect of Lower Drinking Age

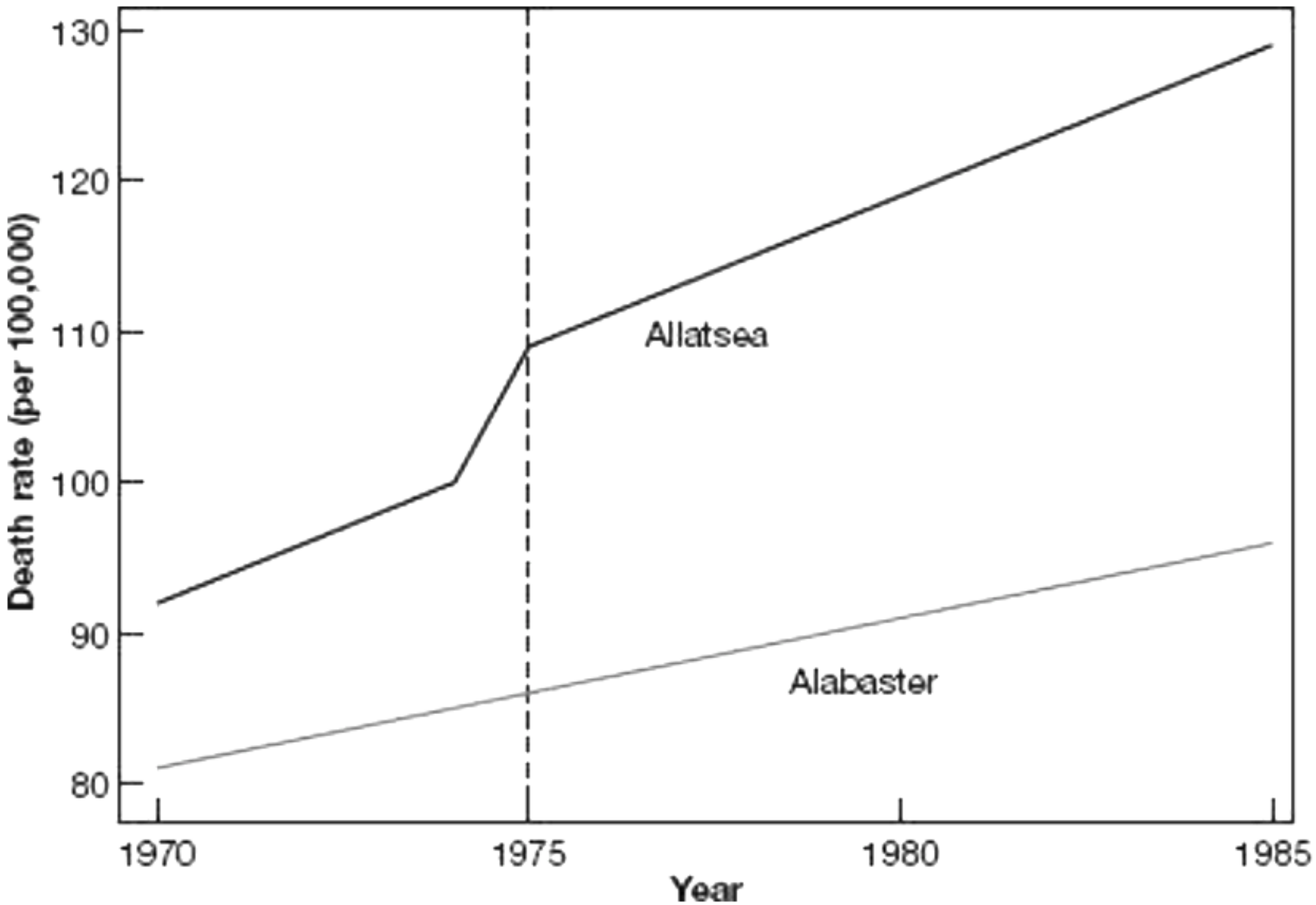
FIGURE 5.4
An MLDA effect in states with parallel trends



Example: The Effect of Lower Drinking Age

FIGURE 5.6

A real MLDA effect, visible even though trends are not parallel



Example: The Effect of Lower Drinking Age

TABLE 5.2
Regression DD estimates of MLDA effects on death rates

Dependent variable	(1)	(2)	(3)	(4)
All deaths	10.80	8.47	12.41	9.65
	(4.59)	(5.10)	(4.60)	(4.64)
Motor vehicle accidents	7.59	6.64	7.50	6.46
	(2.50)	(2.66)	(2.27)	(2.24)
Suicide	.59	.47	1.49	1.26
	(.59)	(.79)	(.88)	(.89)
All internal causes	1.33	.08	1.89	1.28
	(1.59)	(1.93)	(1.78)	(1.45)
State trends	No	Yes	No	Yes
Weights	No	No	Yes	Yes

Correcting Standard Errors

- Case 1. When there are multiple periods, both the treatment status and the outcome variables are likely to be temporally correlated
- Case 2. When there are multiple observations under the same treatment, both the treatment status and the outcome variables are likely to be cross-sectionally correlated
- Treating these observations as independent will inflate the effective sample size and underestimate the uncertainties
- Econometricians develop "clustered" standard errors as a solution

Plan

- Identification
- Estimation
- **Threats to Validity**

Non-parallel Dynamics

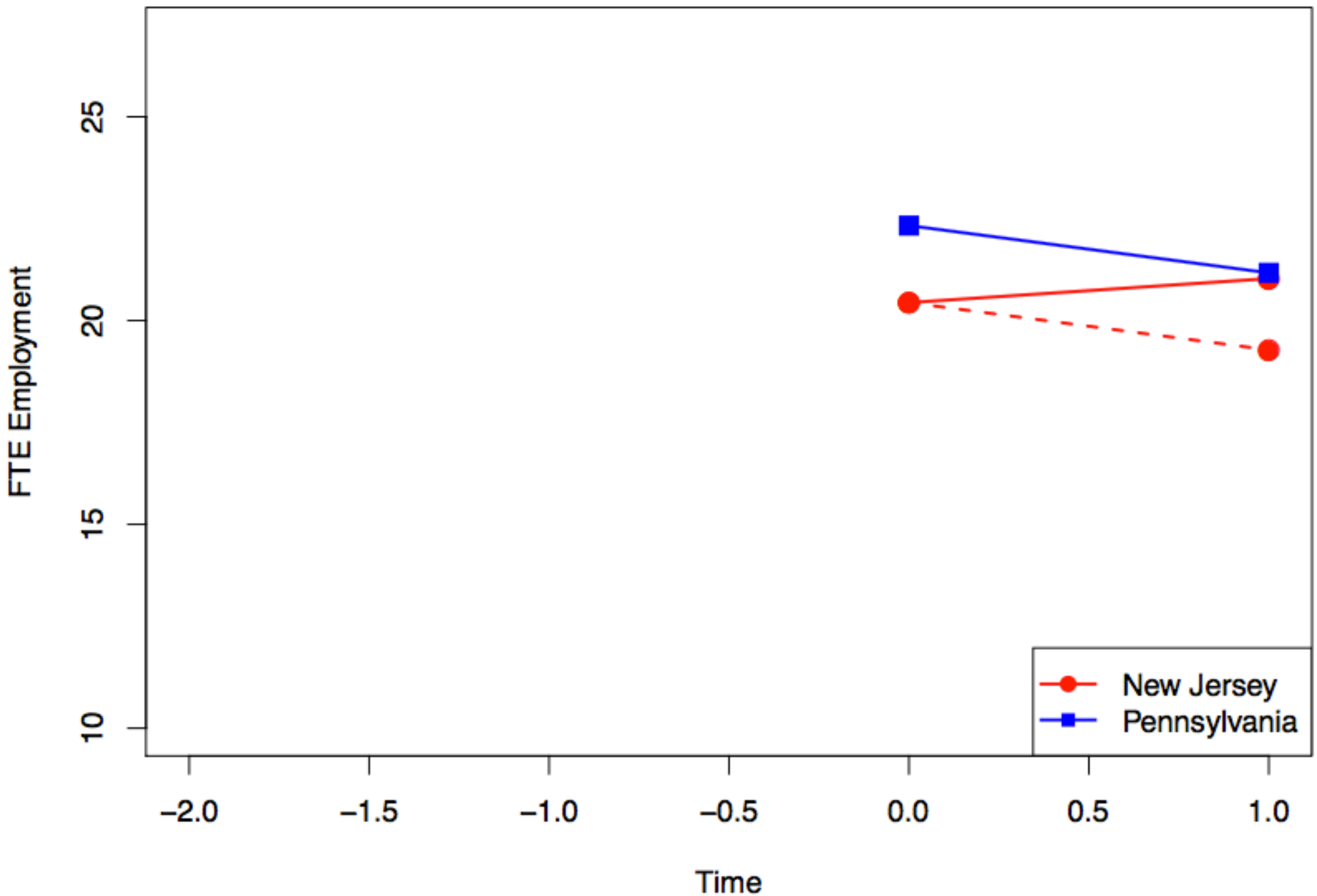
- Often treatments/programs are targeted based on pre-existing differences in outcomes.

  - “Ashenfelter dip”: participants in training programs often experience a dip in earnings just before they enter the program (that may be why they participate).

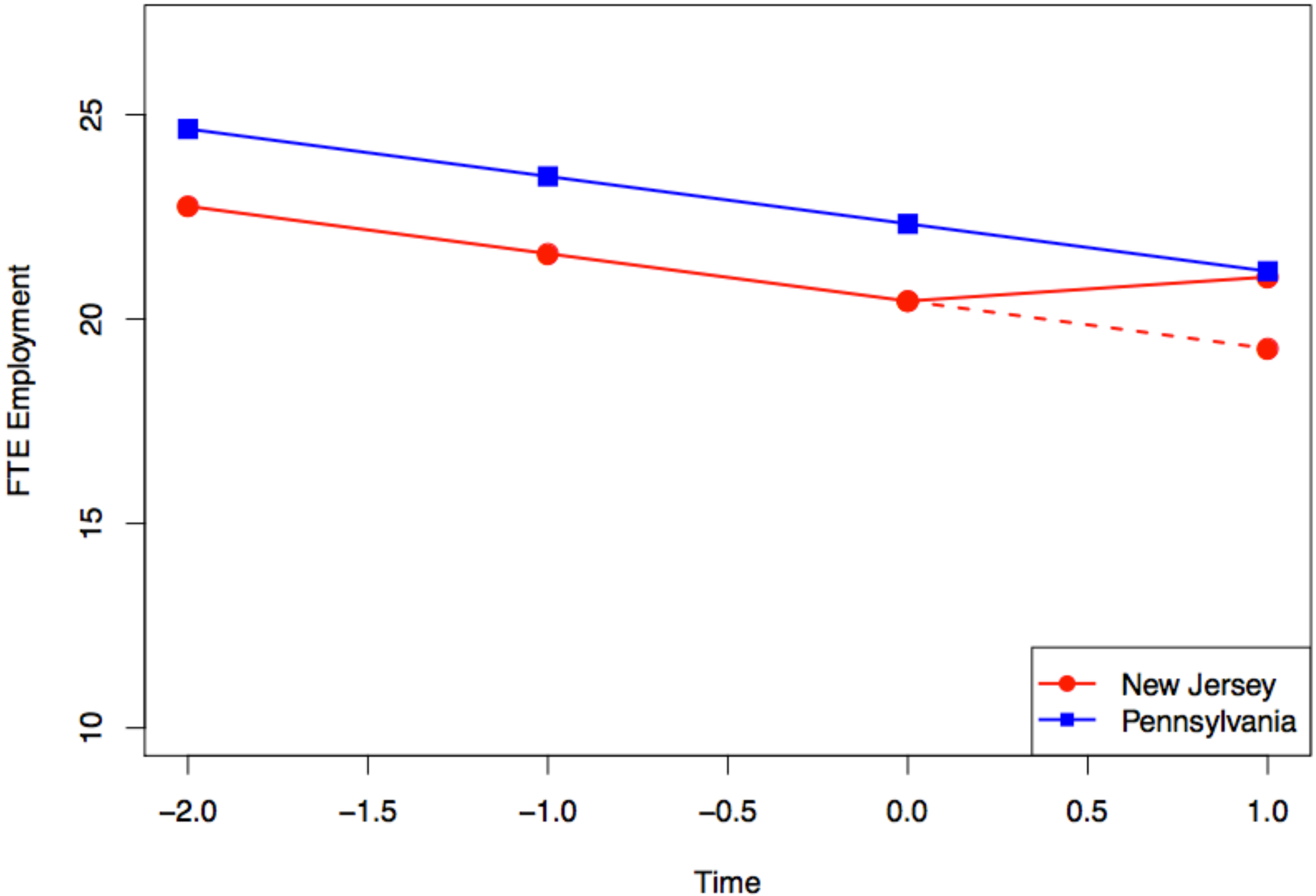
    Since wages have a natural tendency to mean reversion, comparing wages of participants and non-participants using DD leads to an upward biased estimate of the program effect

  - Regional targeting: NGOs may target villages that appear most promising (or worst off)

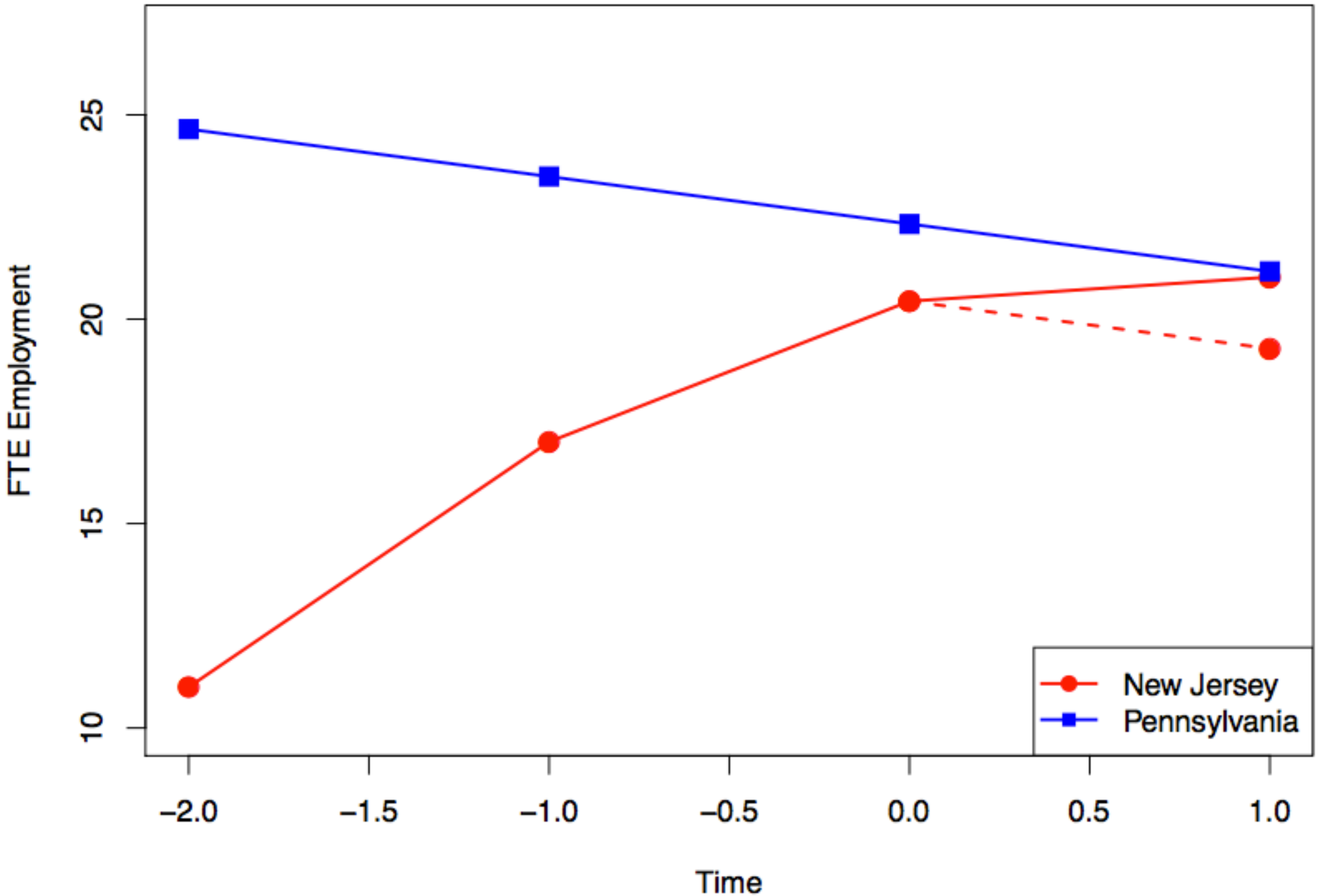
Non-parallel Dynamics



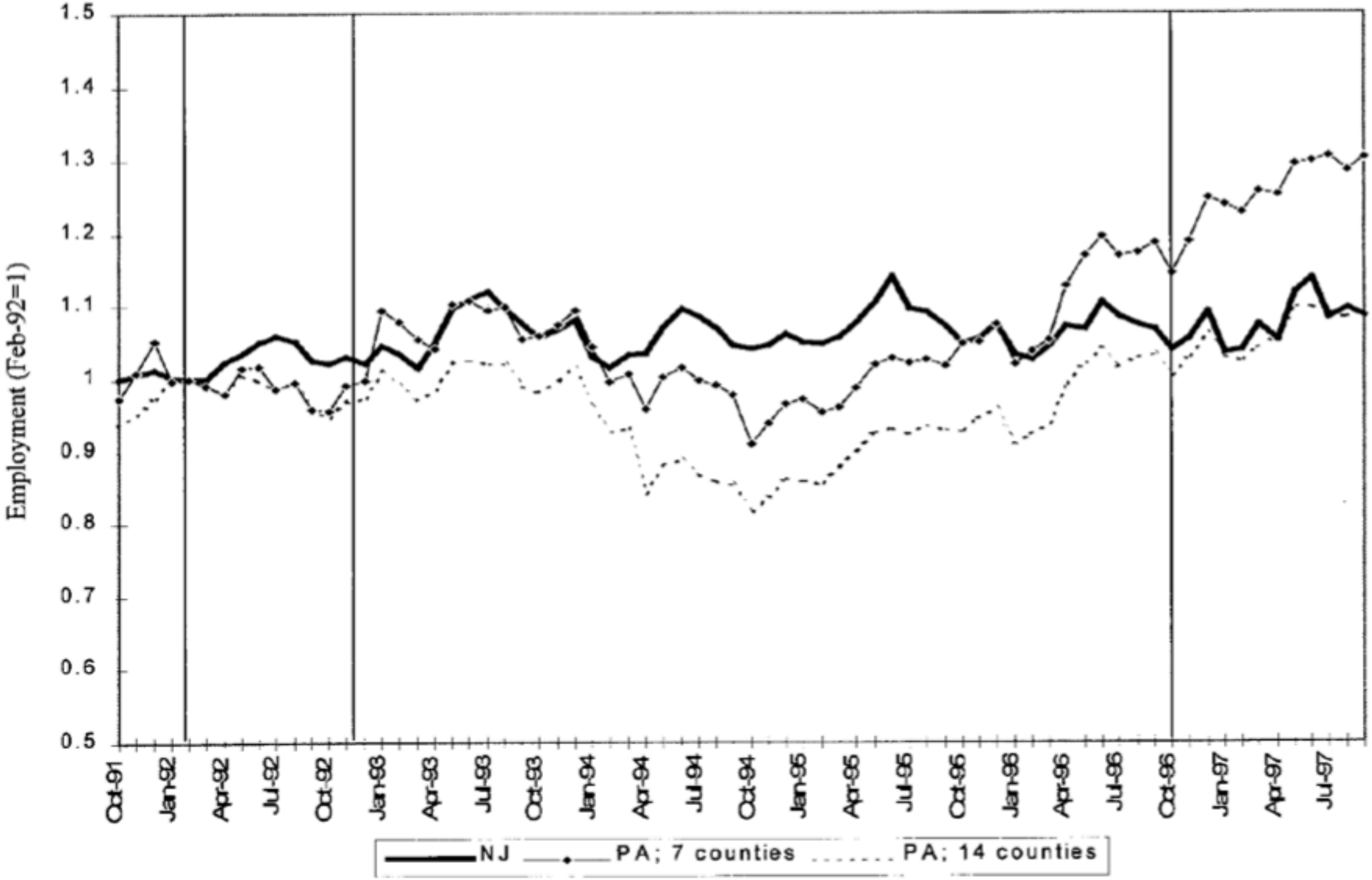
Non-parallel Dynamics



Non-parallel Dynamics



Longer Trends in Employment in NY and PA

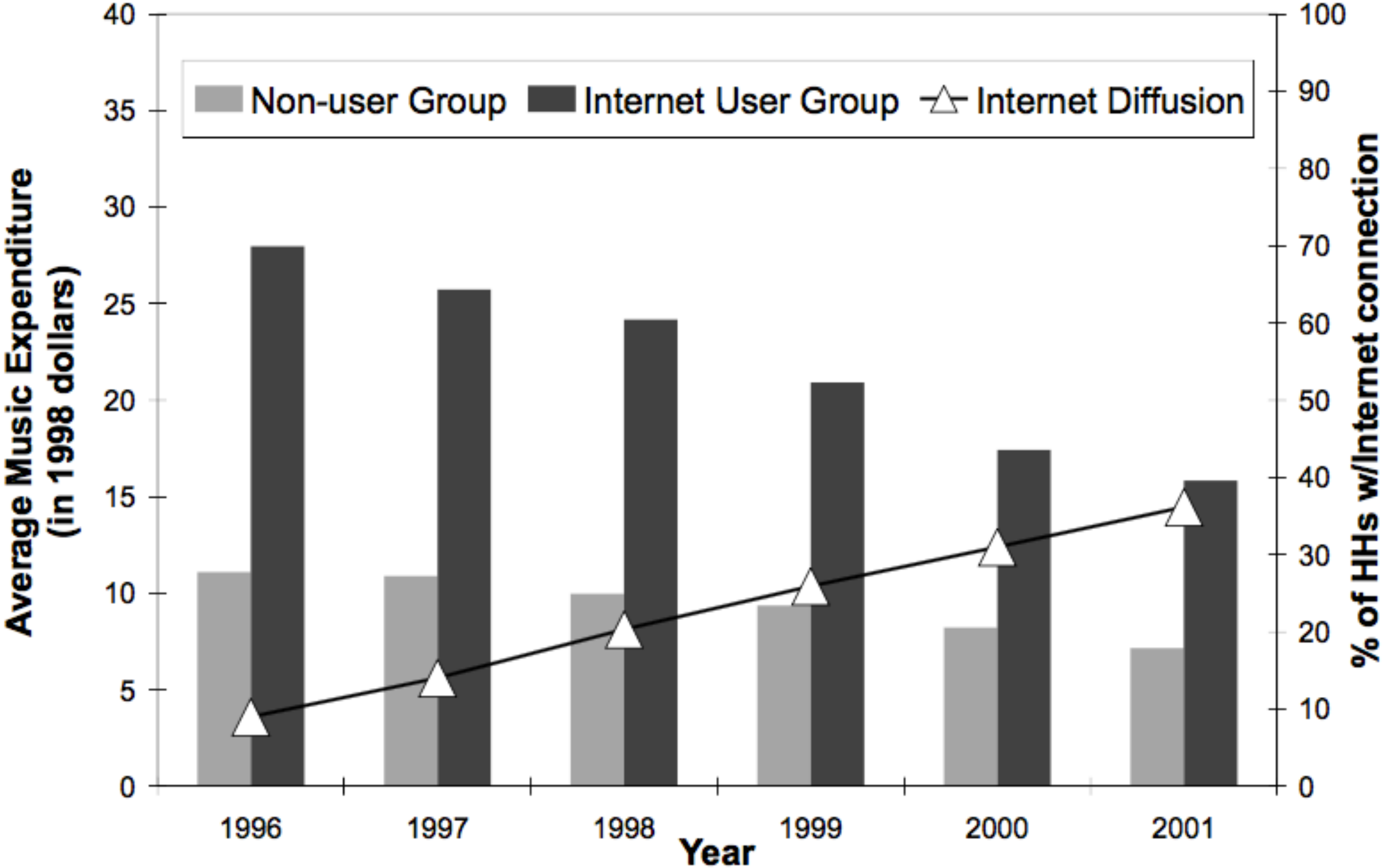


Compositional Differences

- In repeated cross-sections, we do not want that the composition of the sample changes between periods.
- Example:
  - Hong (2011) uses repeated cross-sectional data from Consumer Expenditure Survey (CEX) containing music expenditures and internet use for random samples of U.S. households
  - Study exploits the emergence of Napster (the first sharing software widely used by Internet users) in June 1999 as a natural experiment.
  - Study compares internet users and internet non-users, before and after emergence of Napster

Compositions of Internet Users Change Over Time

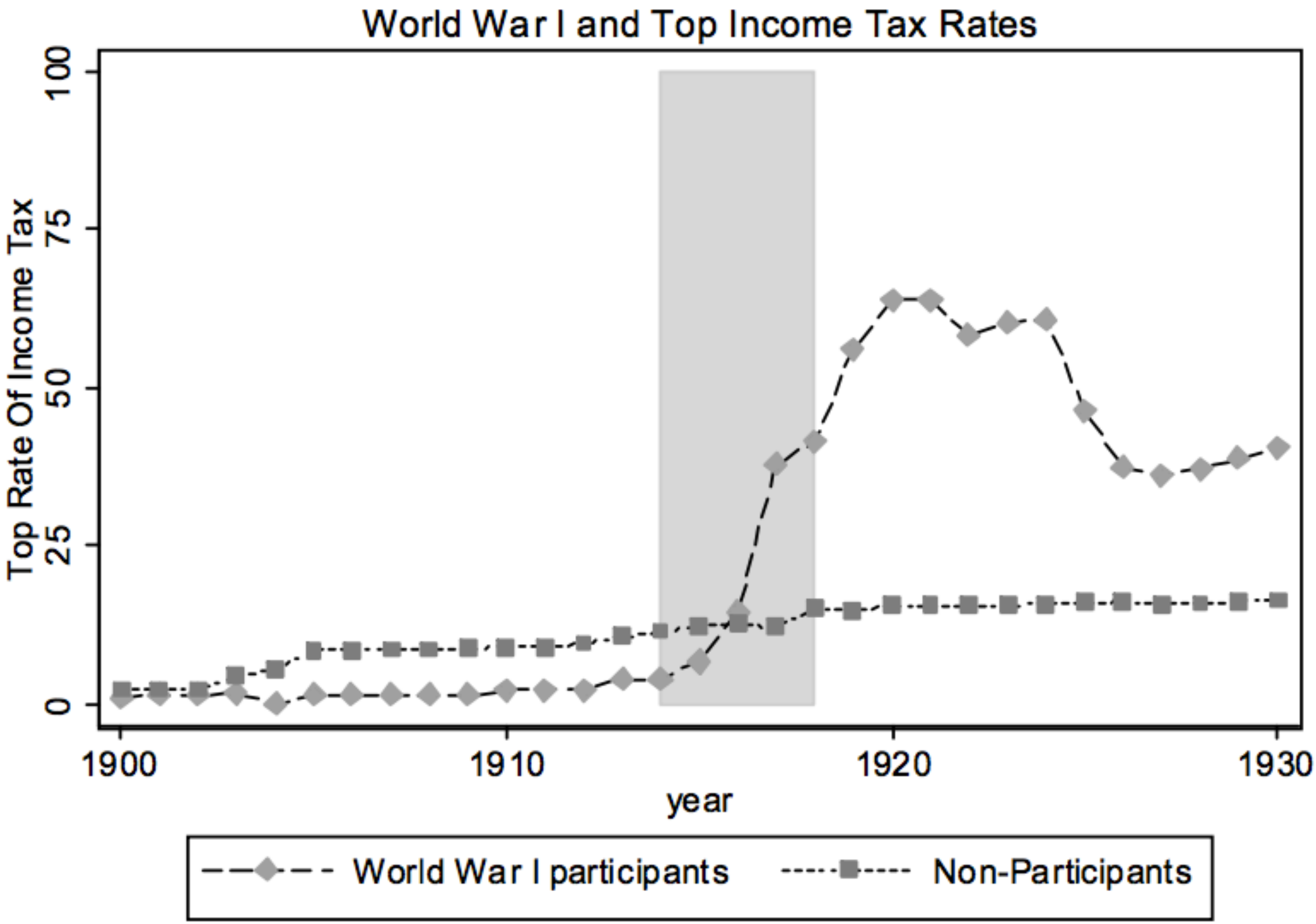
Figure 1: Internet Diffusion and Average Quarterly Music Expenditure in the CEX



Long-term Effects vs. Reliability

- Parallel trends assumption for DD is more likely to hold over a shorter time-window
- In the long-run, many other things may happen that could confound the effect of the treatment
- Should be cautious to extrapolate short-term effects to long-term effects

Effect of War on Tax Rates



Summary

- Diff-in-Diffs: An extremely popular strategy when there is longitudinal data (panel or repeated cross-sections) and the treatment is one-shot
- Parallel trends = a type of ignorability assumption, i.e., unobserved confounding must be additive and time-invariant
- Always be cautious about the assumptions you make. Better to have multiple periods