Lecture 12: Fixed Effect Model and Difference in Differences

Introduction to Econometrics,Fall 2018

Zhaopeng Qu

Nanjing University

12/20/2018

1. Panel Data: What and Why
2. Fixed Effect Model
3. Extension: Regression with Time Fixed Effects
4. Extension: Both Entity and Time Fixed Effects
5. The Fixed Effects Regression Assumptions and Standard Errors
6. Difference in Differences
7. Extensions of DID: Synthetic Controls

Panel Data: What and Why

Introduction

- A panel dataset contains observations on multiple entities, where each entity is observed at two or more points in time.
- If the data set contains observations on the variables X and Y,then the data are denoted

$$(X_{it}, Y_{it}),\ i = 1, ...n\ and\ t = 1, ..., T$$

  - the first subscript,i refers to the entity being observed
  - the second subscript,t refers to the date at which it is observed
- whether some observations are missing
  - balanced panel
  - unbalanced panel

Introduction

TABLE 1.3 Selected Observations on Cigarette Sales, Prices, and Taxes, by State and Year for U.S. States, 1985–1995

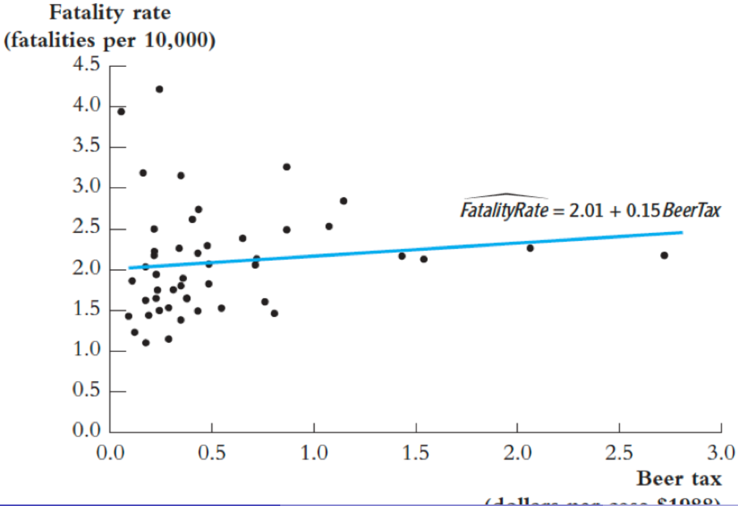
Observation Number	State	Year	Cigarette Sales (packs per capita)	Average Price per Pack (including taxes)	Total Taxes (cigarette excise tax + sales tax)
1	Alabama	1985	116.5	$1.022	$0.333
2	Arkansas	1985	128.5	1.015	0.370
3	Arizona	1985	104.5	1.086	0.362
⋮	⋮	⋮	⋮	⋮	⋮
47	West Virginia	1985	112.8	1.089	0.382
48	Wyoming	1985	129.4	0.935	0.240
49	Alabama	1986	117.2	1.080	0.334
⋮	⋮	⋮	⋮	⋮	⋮
96	Wyoming	1986	127.8	1.007	0.240
97	Alabama	1987	115.8	1.135	0.335
⋮	⋮	⋮	⋮	⋮	⋮

Example: Traffic deaths and alcohol taxes

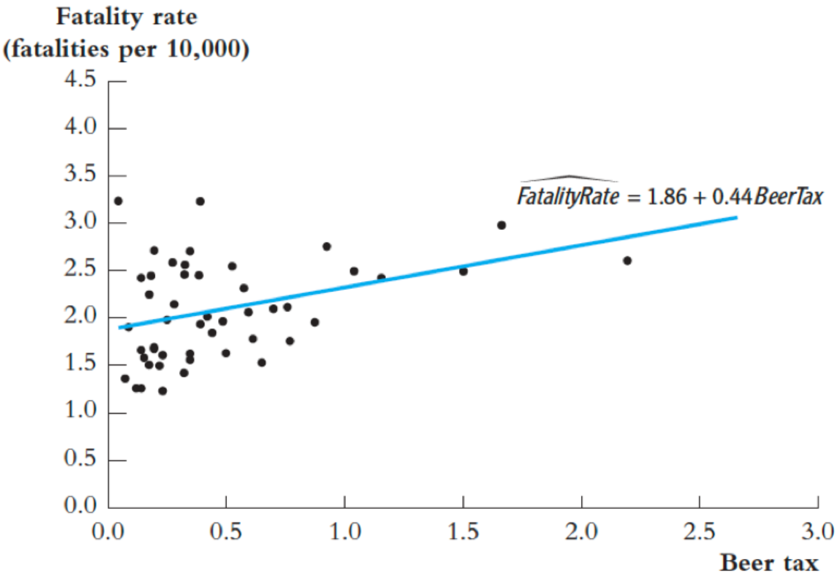
- Observational unit: a year in a U.S. state
- 48 U.S. states, so n = of entities = 48
- 7 years (1982,..., 1988),so T = # of time periods = 7
- Balanced panel, so total # observations $= 7 \times 48 = 336$
- Variables:
  - Dependent Variable: Traffic fatality rate (# traffic deaths in that state in that year, per 10,000 state residents)
  - Independent Variable: Tax on a case of beer
  - Other Controls (legal driving age, drunk driving laws, etc.)

U.S. traffic death data for 1982

- Higher alcohol taxes, more traffic deaths



U.S. traffic death data for 1988



Simple Case: Panel Data with Two Time Periods

- Let Z_i be a factor that determines the fatality rate in the i state but does not change over time.
    - local cultural attitude toward drinking and driving.
- Before and After Model

$$FatalityRate_{it} = \beta_0 + \beta_1 BeerTax_{it} + \beta_2 Z_i + u_{it}$$

where u_{it} is the error term and $i = 1, ...n \; and \; t = 1, ..., T$

- its omission of Z_i might cause omitted variable bias but we don't have data on Z_i.
- The key idea: Any **change** in the fatality rate from 1982 to 1988 cannot be caused by Z_i, because Z_i (by assumption) does not change between 1982 and 1988.

Panel Data with Two Time Periods:Before and After Model

- The math: Consider the regressions for 1982 and 1988...

$$FatalityRate_{i1988} = \beta_0 + \beta_1 BeerTax_{i1988} + \beta_2 Z_i + u_{i1988}$$
$$FatalityRate_{i1982} = \beta_0 + \beta_1 BeerTax_{i1982} + \beta_2 Z_i + u_{i1982}$$

- Then make a difference

$$FatalityRate_{i1988} - FatalityRate_{i1982} =$$
$$\beta_1 (BeerTax_{i1988} - BeerTax_{i1982}) + (u_{i1988} - u_{i1982})$$

- Assumption: if $E(u_it|BeerTax_{it}, Z_{it}) = 0$,then $(u_{i1988} - u_{i1982})$ is uncorrelated with either $BeerTax_{i1988}$ or $BeerTax_{i1982}$

Panel Data with Two Time Periods

- Then this "difference" equation can be estimated by OLS, even though Z_i isn't observed.
- Because the omitted variable Z_i doesn't change, it cannot be a determinant of the change in Y.

Traffic deaths and beer taxes

1982 data:

$$\widehat{FatalityRate} = \underset{(.11)}{1.86} + \underset{(.13)}{0.44}BeerTax \qquad (n = 48)$$

1988 data:

$$\widehat{FatalityRate} = \underset{(.15)}{2.01} + \underset{(.13)}{0.15}BeerTax \qquad (n = 48)$$

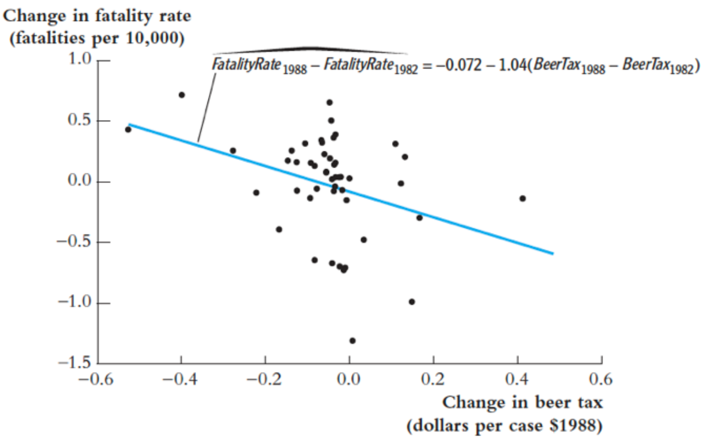
Difference regression ($n = 48$)

$$\widehat{FR_{1988} - FR_{1982}} = \underset{(.065)}{-.072} - \underset{(.36)}{1.04}(BeerTax_{1988} - BeerTax_{1982})$$

Change in traffic deaths and change in beer taxes

FIGURE 10.2 Changes in Fatality Rates and Beer Taxes, 1982–1988

This is a scatterplot of the *change* in the traffic fatality rate and the *change* in real beer taxes between 1982 and 1988 for 48 states. There is a negative relationship between changes in the fatality rate and changes in the beer tax.



Wrap up

- In contrast to the cross-sectional regression results, the estimated effect of a change in the real beer tax is **negative**, as predicted by economic theory.
- By examining changes in the fatality rate over time, the regression in Equation controls for fixed factors such as cultural attitudes toward drinking and driving. But there are many factors that influence traffic safety, and if they change over time and are correlated with the real beer tax, then their omission will produce omitted variable bias.
- This "before and after" analysis works when the data are observed in two different years. Our data set, however, contains observations for seven different years,and it seems foolish to discard those potentially useful additional data. But the "before and after" method does not apply directly when $T > 2$. To analyze all the observations in our panel data set, we use the method of **fixed effects** regression

Fixed Effect Model

Introduction

- Fixed effects regression is a method for controlling for omitted variables in panel data when the omitted variables vary across entities (states) but do not change over time.
- Unlike the "before and after" comparisons,fixed effects regression can be used when there are two or more time observations for each entity.

Fixed Effects Regression Model

- the dependent variable (FatalityRate) and observed regressor (BeerTax) denoted as Y_{it} and X_{it}, respectively:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 Z_i + u_{it} \tag{10.9}$$

- where Z_i is an **unobserved variable** that varies from one state to the next but **does not change over time** (for example, Z_i represents cultural attitudes toward drinking and driving).
- We want to estimate β_1, the effect on Y of X holding constant the unobserved state characteristics Z.

Fixed Effects Regression Model

- Because Z_i varies from one state to the next but is constant over time,then let $\alpha_i = \beta_0 + \beta_1 Z_i$,the Equation becomes

$$Y_{it} = \beta_1 X_{it} + \alpha_i + u_{it} \tag{10.10}$$

- Equation (10.10) is the **fixed effects regression model**, in which α are treated as unknown intercepts to be estimated, one for each state. The interpretation of α_i as a state-specific intercept in Equation (10.10).
- Because the intercept α_i in Equation (10.10) can be thought of as the "effect" of being in entity i (in the current application, entities are states), the terms α_i,an are known as **entity fixed effects**.
- The variation in the entity fixed effects comes from omitted variables that, like Z_i in Equation (10.9), vary across entities but not over time.

Alternative : Fixed Effects by using binary variables

- To develop the fixed effects regression model using binary variables, let $D1_i$ be a binary variable that equals 1 when i = 1 and equals 0 otherwise, let $D2_i$ equal 1 when i = 2 and equal 0 otherwise, and so on.
- Arbitrarily omit the binary variable $D1_i$ for the first group. Accordingly, the fixed effects regression model in Equation (10.10) can be written equivalently as

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \gamma_2 D2_i + \gamma_3 D3_i + ... + \gamma_n Dn_i + u_{it} \quad (10.11)$$

Fixed Effects by using binary variables

- Thus there are two equivalent ways to write the fixed effects regression model, Equations (10.10) and (10.11).
- In Equation (10.10), it is written in terms of n state specific intercepts.

$$Y_{it} = \beta_1 X_{it} + \alpha_i + u_{it} \tag{10.10}$$

- In Equation (10.11), the fixed effects regression model has a common intercept and $n-1$ binary regressors

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \gamma_2 D2_i + \gamma_3 D3_i + ... + \gamma_n Dn_i + u_{it} \tag{10.11}$$

- In both formulations, the slope coefficient on X is the same from one state to the next.

Fixed Effects: Extension to multiple X's.

- The fixed effects regression model is

$$Y_{it} = \beta_1 X_{1,it} + ... + \beta_k X_{k,it} + \alpha_i + u_{it} \tag{10.12}$$

- Equivalently, the fixed effects regression can be expressed in terms of a common intercept

$$\begin{aligned} Y_{it} =& \beta_0 + \beta_1 X_{1,it} + ... + \beta_k X_{k,it} \\ &+ \gamma_2 D2_i + \gamma_3 D3_i + ... + \gamma_n Dn_i + u_{it} \end{aligned}$$

Estimation and Inference

- In principle the binary variable specification of the fixed effects regression model (10.13) can be estimated by OLS.
- But it is tedious to estimate so many fixed effects.If $n = 1000$, then you have to estimate $1000 - 1 = 999$ fixed effects.
- These special routines are equivalent to using OLS on the full binary variable regression, but are faster because they employ some mathematical simplifications that arise in the algebra of fixed effects regression.

Estimation: The "entity-demeaned"

- Computes the OLS fixed effects estimator in two steps
- The first step:
  - take the average across times t of both sides of Equation (10.10);

$$\bar{Y}_i = \beta_1 \bar{X}_i + \alpha_i + \bar{u}_t$$

  - demeaned: let

$$\tilde{Y}_{it} = Y_{it} - \bar{Y}_i$$
$$\tilde{X}_{it} = X_{it} - \bar{X}_i$$
$$\tilde{u}_{it} = u_{it} - \bar{u}_i$$

- The second step: accordingly,estimate

$$\tilde{Y}_{it} = \beta_1 \tilde{X}_{it} + \tilde{u}_{it} \tag{10.14}$$

- In fact, this estimator is identical to the OLS estimator of β_1 without intercept obtained by estimation of the fixed effects model in

OLS estimator without intercept

- OLS estimator without intercept

$$Y_i = \beta_1 X_i + u_i$$

- The least squred term

$$\min_{b_1} \sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} (Y_i - b_1 X_i)^2$$

- F.O.C, thus differentiating with respect to β_1, we get

$$\sum_{i=1}^{n} 2(Y_i - b_1 X_i) X_i = 0$$

- At last,

$$\hat{\beta}_1 = b_1 = \frac{\sum_{i=1}^{n} Y_i X_i}{\sum_{i=1}^{n} X_i^2}$$

Fixed effects estimator

- The second step:

$$\tilde{Y}_{it} = \beta_1 \tilde{X}_{it} + \tilde{u}_{it} \tag{10.14}$$

- then fixed effects estimator based on OLS estimator without intercept

$$\hat{\beta}_{fe} = \frac{\sum_{i=1}^{n} \sum_{t=1}^{T} \tilde{Y}_{it} \tilde{X}_{it}}{\sum_{i=1}^{n} \sum_{t=1}^{T} \tilde{X}_{it}^2}$$

Fixed effect estimator

- Our fixed effects model is

$$Y_{it} = \beta_1 X_{it} + \alpha_i + u_{it} \tag{10.10}$$

- we can think of c_i as fixed effects or "nuisance parameters" to be estimated
- OLS estimation with fixed effects yields

$$(\hat{\beta}, \hat{\alpha}_1, \ldots, \hat{\alpha}_n) = \underset{b, a_1, \ldots, a_n}{\operatorname{argmin}} \sum_{i=1}^{n} \sum_{t=1}^{T} (Y_{it} - bX_{it} - a_i)^2$$

this amounts to including n dummies in regression of Y_{it} on X_{it}

Fixed effect estimator

- The first-order conditions (FOC) for this minimization problem are:

$$\sum_{i=1}^{n}\sum_{t=1}^{T}(Y_{it}-\hat{\beta}X_{it}-\hat{\alpha}_i)X_{it}=0$$

- And

$$\sum_{i=1}^{n}\sum_{t=1}^{T}(Y_{it}-\hat{\beta}X_{it}-\hat{\alpha}_i)=0$$

Fixed effect estimator

- Therefore, for $i = 1, \ldots, N$,

$$\hat{\alpha}_i = \frac{1}{T}\sum_{t=1}^{T}(Y_{it} - \hat{\beta}X_{it}) = \overline{Y}_i - \overline{X}_i\hat{\beta},$$

where

$$\bar{X}_i \equiv \frac{1}{T}\sum_{t=1}^{T}X_{it}; \bar{Y}_i \equiv \frac{1}{T}\sum_{t=1}^{T}Y_{it}$$

Fixed effect estimator

- Plug this result into the first FOC to obtain:

$$
\begin{aligned}
\sum_{i=1}^{n}\sum_{t=1}^{T}(Y_{it}-\hat{\beta}X_{it}-\hat{\alpha}_i)X_{it} &= \sum_{i=1}^{n}\sum_{t=1}^{T}(Y_{it}-X_{it}\hat{\beta}-\overline{Y}_i+\overline{X}_i\hat{\beta})X_{it} \\
&= \left(\sum_{i=1}^{n}\sum_{t=1}^{T}(Y_{it}-\overline{Y}_i)X_{it}\right) \\
&\quad -\hat{\beta}\left(\sum_{i=1}^{n}\sum_{t=1}^{T}(X_{it}-\bar{X}_i)X_{it}\right)=0
\end{aligned}
$$

Fixed effect estimator

- Then we could obtain

$$\hat{\beta} = \left(\sum_{i=1}^{n}\sum_{t=1}^{T}(X_{it}-\bar{X}_i)(X_{it}-\bar{X}_i)\right)^{-1}\left(\sum_{i=1}^{n}\sum_{t=1}^{T}(Y_{it}-\bar{Y})(X_{it}-\bar{X}_i)\right)$$
$$= \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{Y}_{it}}{\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}^2}$$

with time-demeaned variables $\tilde{X}_{it} \equiv X_{it} - \bar{X}, \tilde{Y}_{it} \equiv Y_{it} - \bar{Y}_i$

Statistical property

- Unbiasedness and Consistency

$$\begin{aligned}\hat{\beta}_{fe} &= \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{Y}_{it}}{\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}^2}\\ &= \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}(\beta_1\tilde{X}_{it}+\tilde{u}_{it})}{\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}^2}\\ &= \beta_1 + \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it}}{\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}^2}\end{aligned}$$

- Paralleling the derivation of OLS estimator, we could prove the estimator of fixed effects model is unbiased and consistent.

The Fixed Effects Regression Assumptions

- The model

$$Y_{it} = \beta_1 X_{it} + \alpha_i + u_{it}, i = 1, ...n \; t = 1, ..., T \tag{10.10}$$

1. Assumption 1: u_{it} has conditional mean zero:

$$E(u_{it}|X_{i1}, X_{i2}, ..., X_{iT}, \alpha_i) = 0$$

2. Assumption 2: $(X_{i1}, X_{i2}, ..., X_{iT}, u_{i1}, u_{i2}, ..., u_{iT}), i = 1, 2, ..., n$ are $i.i.d.$
3. Assumption 3: Large outliers are unlikely.
4. Assumption 4: There is no perfect multicollinearity.

- For multiple regressors, X_{it} should be replaced by the full list $X_{1,it}, X_{2,it}, ..., X_{k,it}$

Application to Traffic Deaths

- The OLS estimate of the fixed effects regression based on all 7 years of data (336 observations), is

$$\widehat{FatalityRate} = -\underset{(0.29)}{0.66}BeerTax + StateFixedEffects$$

- the estimated state fixed intercepts are not listed to save space and because they are not of primary interest.
- As predicted by economic theory,higher real beer taxes are associated with fewer traffic deaths, which is the opposite of what we found in the initial cross-sectional regressions of Equations (10.2) and (10.3)
- Because of the additional observations, the standard error is smaller in Equation (10.15) than in Equation (10.8)

Extension: Regression with Time Fixed Effects

Introduction

- Just as fixed effects for each entity can control for variables that are constant over time but differ across entities, so can time fixed effects control for variables that are constant across entities but evolve over time.
- Like safety improvements in new cars as an omitted variable that changes over time but has the same value for all states.

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_3 S_t + u_{it} \tag{10.16}$$

- where S_t is unobserved and where the single t subscript emphasizes that safety changes over time but is constant across states. Because $\beta_3 S_3$ represents variables that determine Y_{it}, if S_t is correlated with X_{it}, then omitting S_t from the regression leads to omitted variable bias.

Time Effects Only

- Although S_t is unobserved, its influence can be eliminated because it varies over time but not across states, just as it is possible to eliminate the effect of Z_i, which varies across states but not over time.
- Similarly, because S_t varies over time but not over states, the presence of S_t leads to a regression model in which each time period has its own intercept,thus

$$Y_{it} = \beta_1 X_{it} + \lambda_t + u_{it} \tag{10.17}$$

- This model has a different intercept, λ_t, for each time period, which are known as time fixed effects.The variation in the time fixed effects comes from omitted variables that, like S_t in Equation (10.16), vary over time but not across entities.

Time Effects Only

- Just as the entity fixed effects regression model can be represented using $n-1$ binary indicators, so, too, can the time fixed effects regression model be represented using $T-1$ binary indicators:

$$Y_{it} = \beta_0 + \beta_1 X_{1,it} + \delta_2 B2_t + ... + \delta_T BT_t + \alpha_i + u_{it} \tag{10.18}$$

  - where $\delta_2, \delta_3, ..., \delta_T$ are unknown coefficients
  - where $B2_t = 1$ if $t=2 and $B2_t = 0$ otherwise and so forth.

Extension: Both Entity and Time Fixed Effects

Both Entity and Time Fixed Effects

- If some omitted variables are constant over time but vary across states (such as cultural norms) while others are constant across states but vary over time (such as national safety standards)
- The combined entity and time fixed effects regression model is

$$Y_{it} = \beta_1 X_{it} + \alpha_i + \lambda_t + u_{it}$$

  - where α_i is the **entity fixed effect** and λ_t is the **time fixed effect**.
- This model can equivalently be represented as follows

$$\begin{aligned} Y_{it} =& \beta_0 + \beta_1 X_{it} + +\gamma_2 D2_i + \gamma_3 D3_i + ... + \gamma_n Dn_i \\ &+ \delta_2 B2_t + \delta_3 B3_t + ... + \delta_T BT_i + u_{it} \end{aligned}$$

Both Entity and Time Fixed Effects: Estimation

- The time fixed effects model and the entity and time fixed effects model are both variants of the multiple regression model.
- First deviating Y and X from their entity and time-period means

Application to traffic deaths

$$\widehat{FatalityRate} = \underset{(0.36)}{-0.64}\,BeerTax + StateFixedEffects + TimeFixedEffects. \quad (10.21)$$

More extension: Allow Unit specific linear time trends:

- Linear time trends that vary by entity,thus allow every entity change in specific trend.

$$Y_{it} = \beta_1 X_{it} + \alpha_i + \lambda_t + \alpha_i \times \lambda_t + u_{it}$$

- the interaction term $\alpha_i \times \lambda_t$ represents the specific trend.

Besley and Burgess(2004)

TABLE 5.2.3
Estimated effects of labor regulation on the performance of firms in Indian states

	(1)	(2)	(3)	(4)
Labor regulation (lagged)	−.186	−.185	−.104	.0002
	(.064)	(.051)	(.039)	(.020)
Log development expenditure per capita		.240	.184	.241
		(.128)	(.119)	(.106)
Log installed electricity capacity per capita		.089	.082	.023
		(.061)	(.054)	(.033)
Log state population		.720	0.310	−1.419
		(.96)	(1.192)	(2.326)
Congress majority			−.0009	.020
			(.01)	(.010)
Hard left majority			−.050	−.007
			(.017)	(.009)
Janata majority			.008	−.020
			(.026)	(.033)
Regional majority			.006	.026
			(.009)	(.023)
State-specific trends	No	No	No	Yes
Adjusted R^2	.93	.93	.94	.95

The Fixed Effects Regression Assumptions and Standard Errors

Autocorrelated in Panel Data

- An important difference between the panel data assumptions in Key Concept 10.3 and the assumptions for cross-sectional data in Key Concept 6.4 is Assumption 2.
  - **Cross-Section**: Assumption 2 holds: i.i.d sample.
  - **Panel data**: independent across entities but no such restricition **within** an entity.
- Like X_{it} can be correlated over time within an entity, thus $Cov(X_t, X_s)$ for some $t \neq s$, then the X_t is said to be **autocorrelated or serially correlated**.

Autocorrelated in Panel Data

- In the traffic fatality example, X_{it}, the beer tax in state i in year t,is autocorrelated:
    - Most of the time, the legislature does not change the beer tax, so if it is high one year relative to its mean value for state i,it will tend to be high the next year,too.

Autocorrelated in Panel Data

- Similarly, u_{it} would be also autocorrelated. It consists of time-varying factors that are determinants of Y_{it} but are not included as regressors, and some of these omitted factors might be autocorrelated. It can formally be expressed as

$$Cov(u_{it}, u_{is}|X_{it}, X_{is}, \alpha_i) \neq 0 \, for \, t \neq s$$

  - eg. a downturn in the local economy and a road improvement project.
  - eg. severe winter driving conditions may not persist for a long time.

Autocorrelated in Panel Data

- If the regression errors are autocorrelated, then the usual heteroskedasticity-robust standard error formula for cross-section regression is not valid.
- The result: an analogy of heteroskedasticity.
- OLS panel data estimators of β are unbiased and consistent but the standard errors will be wrong
  - – usually the OLS standard errors understate the true uncertainty
- This problem can be solved by using **"heteroskedasticity and autocorrelation-consistent(HAC) standard errors"**

Standard Errors for Fixed Effects Regression

- The standard errors used are one type of HAC standard errors, **clustered standard errors**.
- The term **clustered** arises because these standard errors allow the regression errors to have an arbitrary correlation within a cluster, or grouping, but assume that the regression errors are uncorrelated across clusters.
- In the context of panel data,each cluster consists of an entity.Thus **clustered standard errors** allow for heteroskedasticity and for arbitrary autocorrelation *within an entity*, but treat the errors as *uncorrelated across entities*.
- Like **heteroskedasticity-robust standard errors** in regression with cross-sectional data, **clustered standard errors** are valid whether or not there is heteroskedasticity,autocorrelation,or both.

Application: Drunk Driving Laws and Traffic Deaths

- Two ways to cracks down on Drunk Driving

1. toughening driving laws
2. raising taxes

- Both driving laws and economic conditions could be omitted variables

Application: Drunk Driving Laws and Traffic Deaths

TABLE 10.1 Regression Analysis of the Effect of Drunk Driving Laws on Traffic Deaths

Dependent variable: Traffic fatality rate (deaths per 10,000).

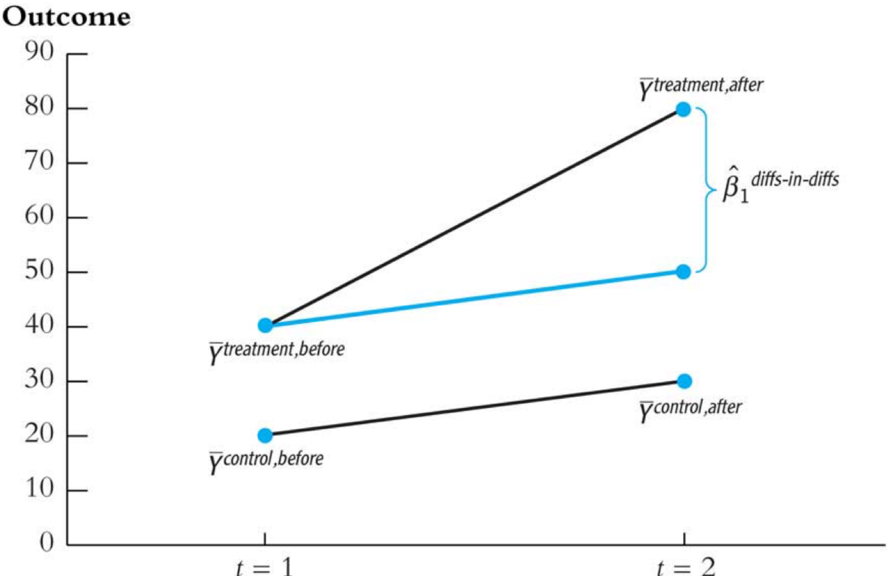
Regressor	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Beer tax	0.36** (0.05)	−0.66* (0.29)	−0.64+ (0.36)	−0.45 (0.30)	−0.69* (0.35)	−0.46 (0.31)	−0.93** (0.34)
Drinking age 18				0.028 (0.070)	−0.010 (0.083)		0.037 (0.102)
Drinking age 19				−0.018 (0.050)	−0.076 (0.068)		−0.065 (0.099)
Drinking age 20				0.032 (0.051)	−0.100+ (0.056)		−0.113 (0.125)
Drinking age						−0.002 (0.021)	
Mandatory jail or community service?				0.038 (0.103)	0.085 (0.112)	0.039 (0.103)	0.089 (0.164)
Average vehicle miles per driver				0.008 (0.007)	0.017 (0.011)	0.009 (0.007)	0.124 (0.049)
Unemployment rate				−0.063** (0.013)		−0.063** (0.013)	−0.091** (0.021)
Real income per capita (logarithm)				1.82** (0.64)		1.79** (0.64)	1.00 (0.68)
Years	1982–88	1982–88	1982–88	1982–88	1982–88	1982–88	1982 & 1988 only
State effects?	no	yes	yes	yes	yes	yes	yes

Difference in Differences

Introduction

- A typical RCT design requires a causal studies to do as follow
  1. Randomly assignment of treatment to divide the population into a "treatment" group and a "control" group.
  2. Collecting the data at the time of post-treatment then comparing them.
- It works because *treatment* and *control* are randomized.
- what if we have the treatment group and the control group, but they are not fully randomized?
- If we have observations across two times at least(one before treatment, the other after treatment), then an easy way to make causal inference is **Difference in Differences(DID)** method.

DID estimator



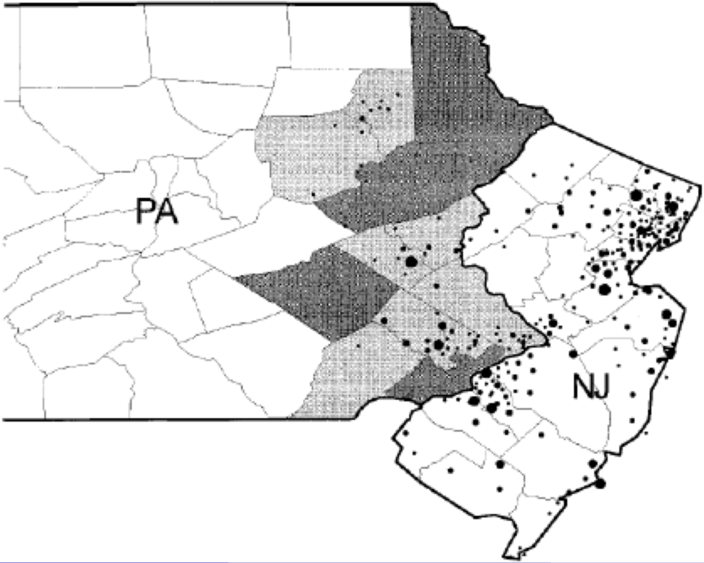
Card and Krueger(1994): minimum wage on employment

- Theoretically,in competitive labor market, increasing binding minimum wage decreases employment.But what about the reality?
- Ideal experiment: randomly assign labor markets to a control group (minimum wage kept constant) and treatment group (minimum wage increased), compare outcomes.
- Policy changes affecting some areas and not others create natural experiments.
    - Unlike ideal experiment, control and treatment groups not randomly assigned.

Card and Krueger(1994): Backgroud

- Policy Change: in April 1992
  - Minimum wage in New Jersey from \$4.25 to \$5.05
  - Minimum wage in Pennsylvania constant at \$4.25
- Research Design:
  - Collecting the data on employment at 400 fast food restaurants in NJ(treatment group) in Feb.1992 (before treatment)and again November 1992(after treatment).
  - Also collecting the data from the same type of restaurants in eastern Pennsylvania(PA) as control group where the minimum wage stayed at \$4.25 throughout this period.

Card & Krueger(1994): Geographic background



Card & Krueger(1994):

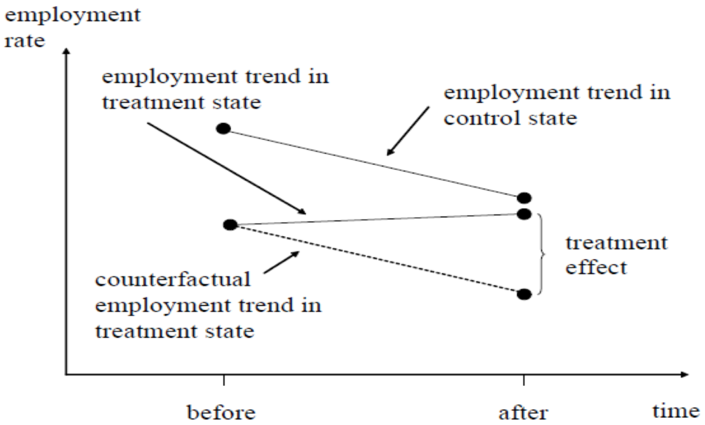


Figure 5.2.1: Causal effects in the differences-in-differences model

Card & Krueger(1994):Result

Table 5.2.1: 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)

Notes: Adapted from Card and Krueger (1994), Table 3. The

DID with Regression

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

DID with Regression

- Formally, an simple DID regression is

$$Y_{it} = \beta_0 + \beta_1 T_i + \beta_2 D_t + \beta_3 (T_i \times D_t) + u_{it}$$

- T_i: a time dummy denotes pre($T_i = 0$) or post treatment($T_i = 1$),thus
- D_t: a treatment dummy denotes in treatment($D_t = 0$) or control group($D_t = 1$)
- $T_i \times D_t$: an interaction term

Regression DD - Card and Krueger

- DID model:

$$Y_{its} = \alpha + \gamma NJ_s + \lambda d_t + \delta (NJ \times d)_{st} + u_{its}$$

  - NJ is a dummy equal to 1 if the observation is from NJ,otherwise equal to 0(from Penny)
  - d is a dummy equal to 1 if the observation is from November (the post period),otherwise equal to 0(Feb. the pre period)

- This equation takes the following values

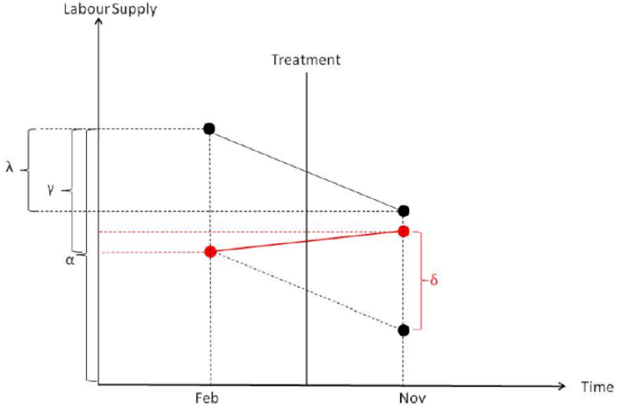
  - PA Pre: α
  - PA Post: $\alpha + \lambda$
  - NJ Pre: $\alpha + \gamma$
  - NJ Post: $\alpha + \gamma + \lambda + \delta$

- then DID estimator

$$\begin{aligned}\hat{\beta}_{DID} &= (\bar{Y}_{treat,after} - \bar{Y}_{treat,before}) - (\bar{Y}_{control,after} - \bar{Y}_{control,before}) \\ &= (NJ\ Post - NJ\ Pre) - (PA\ Post - PA\ Pre)\end{aligned}$$

Regression DD - Card and Krueger

$$Y_{ist} = \alpha + \gamma NJ_s + \lambda d_t + \delta(NJ \times d)_{st} + \varepsilon_{ist}$$

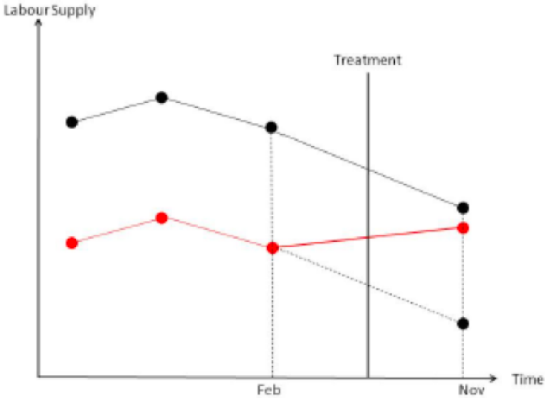


Key Assumption For DID

- A key identifying assumption for DID is: **Common trends** or **Parallel trends**
  - Treatment would be the same "trend" in both groups in the absence of treatment.
- This doesn't mean that they have to have the same mean of the outcome.
- It is difficult to verify because technically one of the parallel trends is an unobserved counterfactual
- There are some unobservable factors affected on outcomes of both group. But as long as the effects have the same trends on both groups, then DID will eliminate the factors.
- But one often will check using pre-treatment data to show that the trends are the same

Assessing natural experiment

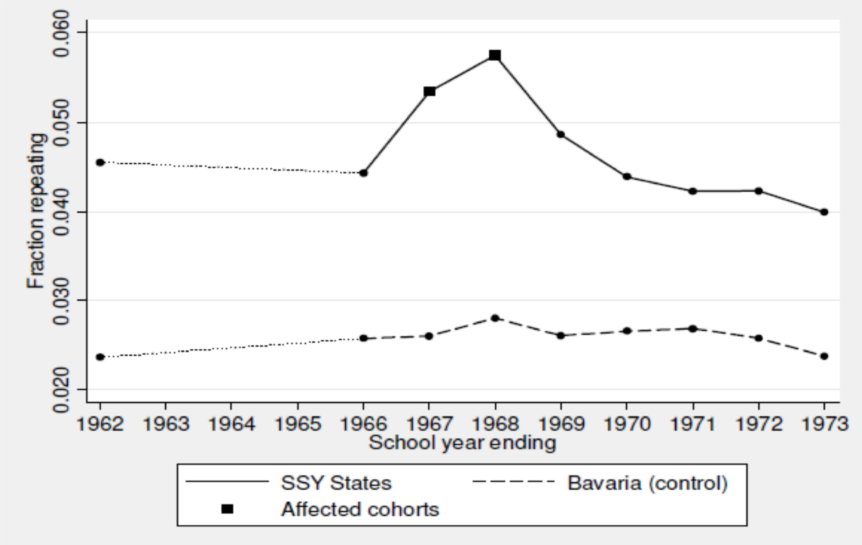
- Common Trend: Common trend assumption is difficult to verify but one often uses pre-treatment data to show that the trends are the same.



An Encouraging Example: Pischeke(2007)

- Topic: the length of school year on student performance
- Backgroud:
  - Until the 1960s, children in all German states except Bavaria started school in the Spring. In 1966-1967 school year, the Spring moved to Fall.
  - It make two shorter school years for affected cohort, 24 weeks long instead of 37.
- Reseach Design:
  - Dependent Variable: Retreating rate
  - Independent Variable: spending time on school
  - Treatment group: Students in the German **States except Bavaria**.
  - Control group: Students in **Bavaria**.

An Encouraging Example: Pischeke(2007)



Pischke(2007)

- This graph provides strong visual evidence of treatment and control states with a common underlying trend.
- A treatment effect that induces a sharp but transitory deviation from this trend.
- It seems to be clear that a short school years have increased repetition rates for affected cohorts.

Regression DD Including Leads and Lags

- If you have a multiple years panel data, then Including leads into the DD model is an easy way to analyze pre-treatment trends.
- Lags can be included to analyze whether the treatment effect changes over time after assignment
- The estimated regression would be

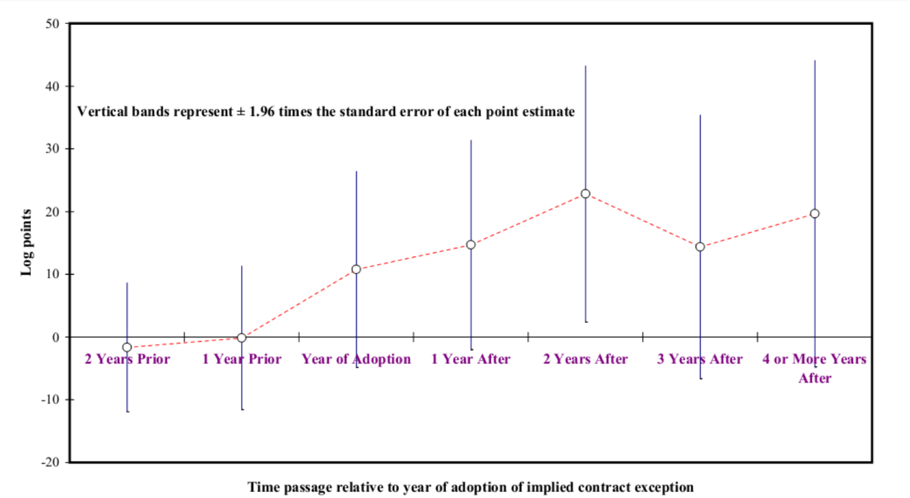
$$Y_{its} = \gamma_s + \lambda_t + \sum_{\tau=-q}^{-1} \theta_\tau D_{st} + \sum_{\tau=0}^{p} \delta_\tau D_{st} + X_{ist} + u_{its}$$

- Treatment occurs in year 0
- Includes q leads or anticipatory effects
- Includes p leads or post treatment effects

Study including leads and lags – Autor (2003)

- Autor (2003) includes both leads and lags in a DD model analyzing the e ect of increased employment protection on the firm's use of temporary help workers.
- In the US employers can usually hire and fire workers at will.
- U.S labor law allows "employment at will" but in some state courts have allowed a number of exceptions to the doctrine, leading to lawsuits for "unjust dismissal"
- The employment of temporary workers in a state to dummy variables indicating state court rulings that allow exceptions to the employment-at-will doctrine.
- The standard thing to do is normalize the adoption year to 0
- Autor then analyzes the effect of these exemptions on the use of temporary help workers.

Study including leads and lags – Autor (2003)



Standard errors in DD strategies

- Many paper using DD strategies use data from many years: not just 1 pre and 1 post period.
- The variables of interest in many of these setups only vary at a group level (say a state level) and outcome variables are often serially correlated
- In the Card and Krueger study, it is very likely that employment in each state is not only correlated within the state but also serially correlated.
- As Bertrand, Duflo and Mullainathan (2004) point out, conventional standard errors often severely *understate* the standard deviation of the estimators – standard errors are biased downward.

Standard errors in DD strategies

- Block bootstrapping standard errors (if you analyze states the block should be the states and you would sample whole states with replacement for bootstrapping)
- Clustering standard errors at the group level.

Threats to validity

- Non-parallel trends
- Compositional differences
- Long-term effects v.s reliability
- Functional form dependence

Non-parallel trends

- Often policymakers will select the treatment and controls based on pre-existing di erences in outcomes – practically guaranteeing the parallel trends assumption will be violated.
- "Ashenfelter dip"
    - Participants in job trainings program often experience a "dip" in earnings just prior to entering the program.
    - 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.

Checks for DD Design

- Very common for readers and others to request a variety of "robustness checks" from a DID design
- Think of these as along the same lines as the leads and lags
  - Falsification test using data for prior periods (already discussed)
  - Falsification test using data for alternative control group
  - Falsification test using alternative "placebo" outcome that should not be affected by the treatment

Alternative control group – DDD

- More convincing analysis sometime comes from higher-order contrasts: **DDD** or **Triple D** design.
  - Build the third dimension of contrast to eliminate the potential bias.
- DDD in Regression

$$Y_{ijt} = \alpha + \beta_1 X_{ijt} + \beta_2 \lambda_t + \beta_3 \delta_j + \beta_4 D_i + \beta_5 (\delta \times \lambda)_{jt}$$
$$+\beta_6 (\lambda \times D)_{ti} + \beta_7 (\delta \times D)_{ij} + \beta_8 (\lambda \times \delta \times D)_{ijt} + u_{its}$$

Alternative control group – DDD

- e.g: Health Plan for elderly
  - Treatment group: Elderly aged above 65 in treatment place.
  - Control group 1: Elderly aged between 55-65 in treatment place.
  - Assumption 1: the elderly on different age would have the same trends of health status if there were not the plan.
  - Control group 2: Elderly aged above 65 in control place.
  - Assumption 2: the elderly in different places would have the same trends of health status if there were not the plan.
- It can loose the simple *common trend* assumption in simple DID.

Extensions of DID: Synthetic Controls

Extensions of DID: Synthetic Controls Method

- The synthetic control method is based on the observation that, when the units of analysis are a few aggregate entities.
- The basic idea behind synthetic controls is that a combination of units often provides a better comparison for the unit exposed to the intervention than any single unit alone.
- It is a data-driven procedure to use a small number of non-treated units to build the suitable counterfactuals.

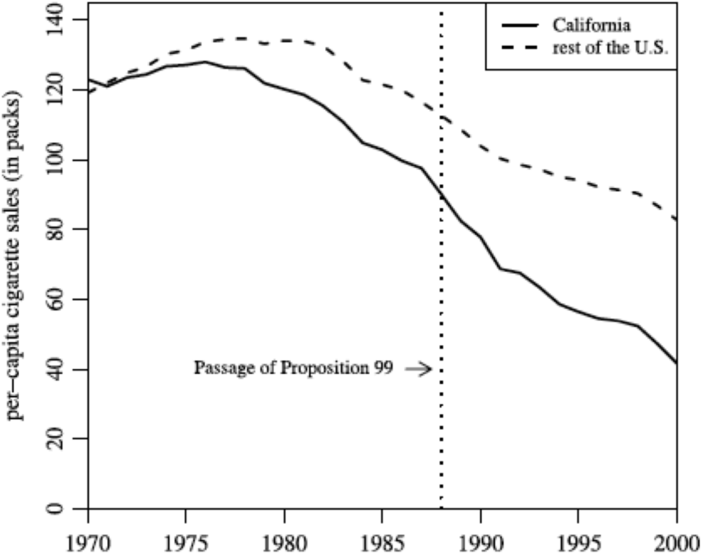
Extensions of DID: Synthetic Controls Method

- Use (long) longitudinal data to build the weighted average of non-treated units that best reproduces characteristics of the treated unit over time in pre-treatment period.
- The weighted average of non-treated units is the synthetic cohort.
- Causal effect of treatment can be quantified by a simple difference after treatment: treated vs synthetic cohort.

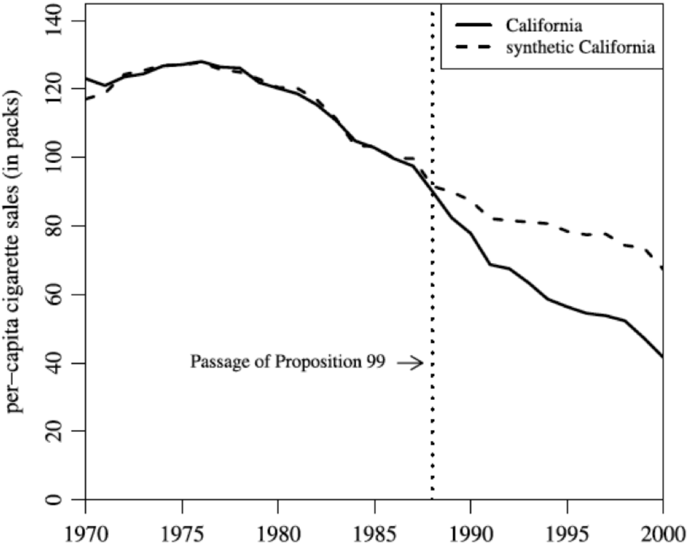
Abadie et. al (2010): tobacco tax on cigarette consumption

- In 1988, California passed comprehensive tobacco control legislation: Increased cigarette taxes by $0.25 per pack ordinances.
- estimate the effect of the policy on cigarette consumption in California.

Abadie et. al (2010): tobacco tax on ciga-consumption



Abadie et. al (2010): tobacco tax on ciga-consumption

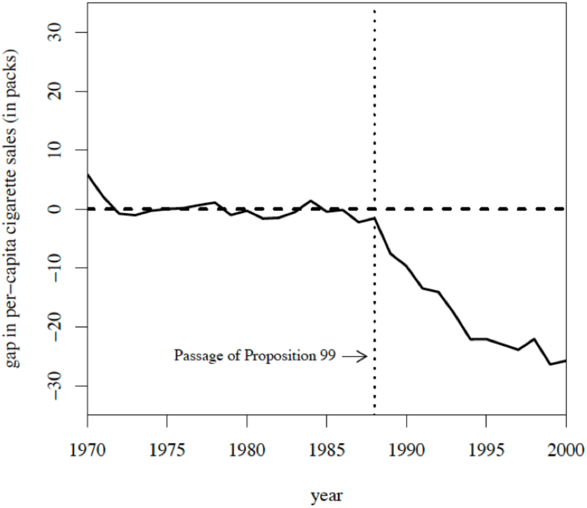


Predictor Means: Actual vs Synthetic California

Variables	California Real	California Synthetic	Average of 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).

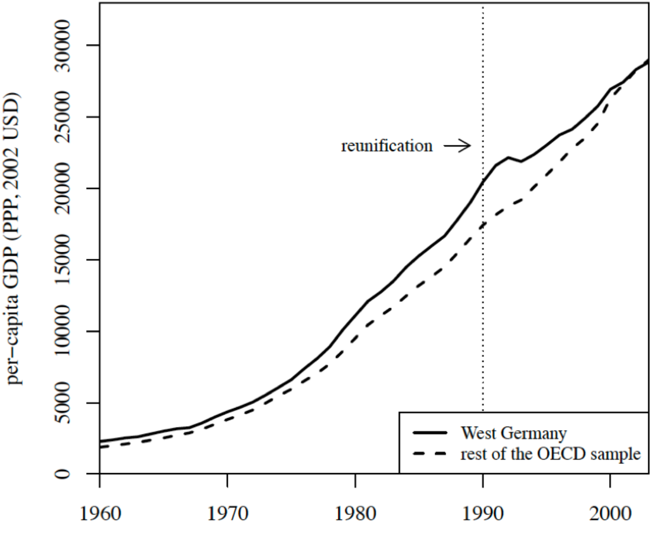
The Application: Actual vs Synthetic California



The Application: The 1990 German Reunification

- Cross-country regressions are often criticized because they put side-by-side countries of very di erent characteristics.
  - “What do Thailand, the Dominican Republic, Zimbabwe,Greece and Bolivia have in common that merits their being put in the same regression analysis? Answer: For most purposes, nothing at all.” (Harberger 1987)
- Application: the economic impact of the 1990 German reunification in West Germany
- Donor pool is restricted to 16 OECD countries

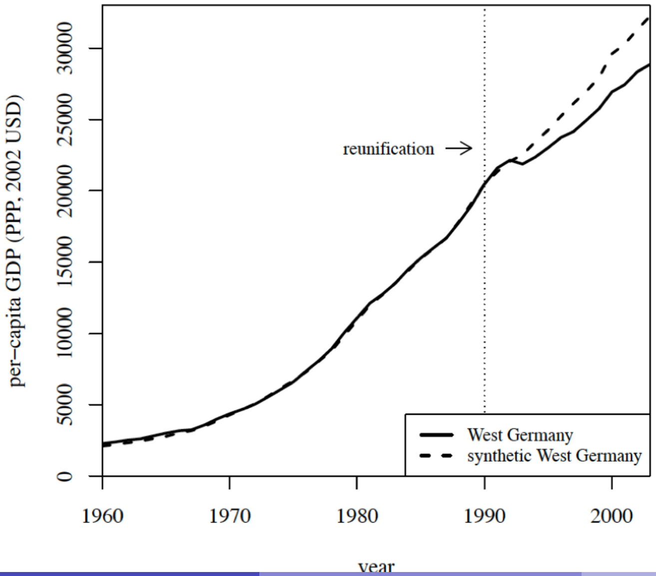
The Application: The 1990 German Reunification



The Application: The 1990 German Reunification

	West Germany	Synthetic West Germany	OECD Sample
GDP per-capita	15808.9	15800.9	8021.1
Trade openness	56.8	56.9	31.9
Inflation rate	2.6	3.5	7.4
Industry share	34.5	34.4	34.2
Schooling	55.5	55.2	44.1
Investment rate	27.0	27.0	25.9

The Application: The 1990 German Reunification



Final Thoughts

- A good research design is one you are excited to tell people about – that's basically what characterizes all research designs, whether instrumental variable or regression discontinuity designs or others(**Seven Magic Weapons**).
- Causality is easy and hard. Don't get confused which is the hard part and which is the easy part.
- Always understand what assumptions you must make, be clear which parameters you are and are not identifying, and don't be afraid of your answers.
- Remember: Good question is always the first priority. Along with good research design is in the second place.