

Week 12: Repeated Observations and Panel Data

Brandon Stewart¹

Princeton

December 5–7, 2022

¹Many slides drawn from Matt Blackwell, Adam Glynn, Jens Hainmueller, Erin Hartman, and Ian Lundberg.

Where We've Been and Where We're Going...

- Last Week
 - ▶ causal inference with unmeasured confounding
- This Week
 - ▶ difference-in-differences (DID/'diff-in-diff')
 - ▶ synthetic controls
 - ▶ fixed effects
 - ▶ wrap-up
- The Following Week
 - ▶ reading period and review session
- Long Run
 - ▶ probability $\rightarrow$ inference $\rightarrow$ regression $\rightarrow$ causality

- 1 Difference-in-Differences
- 2 Synthetic Controls
- 3 Fixed Effects
- 4 Non-parametric Identification and Fixed Effects
- 5 Wrap-Up
 - Questions
 - Concluding Thoughts for the Course

- 1 Difference-in-Differences
- 2 Synthetic Controls
- 3 Fixed Effects
- 4 Non-parametric Identification and Fixed Effects
- 5 Wrap-Up
 - Questions
 - Concluding Thoughts for the Course

Motivation: Studying the Minimum Wage

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

By DAVID CARD AND ALAN B. KRUEGER*

On April 1, 1992, New Jersey's minimum wage rose from \$4.25 to \$5.05 per hour. To evaluate the impact of the law we surveyed 410 fast-food restaurants in New Jersey and eastern Pennsylvania before and after the rise. Comparisons of employment growth at stores in New Jersey and Pennsylvania (where the minimum wage was constant) provide simple estimates of the effect of the higher minimum wage. We also compare employment changes at stores in New Jersey that were initially paying high wages (above \$5) to the changes at lower-wage stores. We find no indication that the rise in the minimum wage reduced employment. (JEL J30, J23)

<https://www.jstor.org/stable/2118030>

Motivation: Studying the Minimum Wage

- Economics conventional wisdom: higher minimum wages decrease low-wage jobs.

Motivation: Studying the Minimum Wage

- Economics conventional wisdom: higher minimum wages decrease low-wage jobs.
- Card and Krueger (1994) study a 1992 NJ minimum wage increase (\$4.25 to \$5.05).

Motivation: Studying the Minimum Wage

- Economics conventional wisdom: higher minimum wages decrease low-wage jobs.
- Card and Krueger (1994) study a 1992 NJ minimum wage increase (\$4.25 to \$5.05).
- Idea: compare employment rates in 410 fast-food restaurants in NJ and eastern PA (where there wasn't a wage increase) both before and after the change.

Motivation: Studying the Minimum Wage

- Economics conventional wisdom: higher minimum wages decrease low-wage jobs.
- Card and Krueger (1994) study a 1992 NJ minimum wage increase (\$4.25 to \$5.05).
- Idea: compare employment rates in 410 fast-food restaurants in NJ and eastern PA (where there wasn't a wage increase) both before and after the change.
- Based on survey data:
 - ▶ Wave 1: March 1992, one month before the minimum wage increased
 - ▶ Wave 2: December 1992, eight months after increase

Motivation: Studying the Minimum Wage

- Economics conventional wisdom: higher minimum wages decrease low-wage jobs.
- Card and Krueger (1994) study a 1992 NJ minimum wage increase (\$4.25 to \$5.05).
- Idea: compare employment rates in 410 fast-food restaurants in NJ and eastern PA (where there wasn't a wage increase) both before and after the change.
- Based on survey data:
 - ▶ Wave 1: March 1992, one month before the minimum wage increased
 - ▶ Wave 2: December 1992, eight months after increase
- “What would a skeptic consider convincing evidence?” David Card
- “There was a time when we thought econometric techniques would solve a lot of the data problems. Now I think the feeling is that there are a lot of problems for which it is easier to get better data.” Alan Krueger

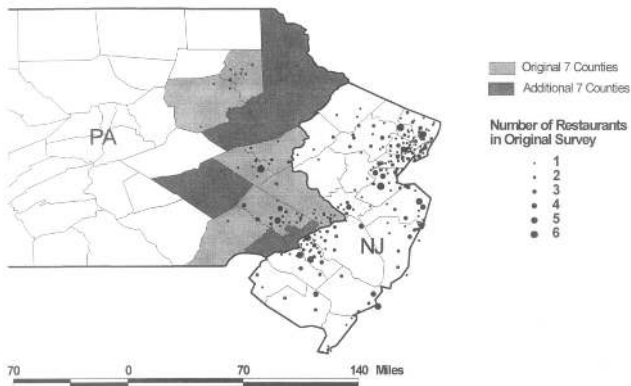


FIGURE 1. AREAS OF NEW JERSEY AND PENNSYLVANIA COVERED BY ORIGINAL SURVEY AND BLS DATA

Source: Card and Krueger 2000

Two Economists Catch Clinton's Eye By Bucking the Common Wisdom

© 2004 Blackwell Publishing Ltd

A CONVENTION OF THE 100,000-member National Wildlife Federation (NWF) in Washington, D.C., last week passed a resolution to support the use of a single federal agency to regulate the environment. The resolution calls for the creation of a new cabinet-level department to coordinate the activities of all federal agencies that regulate the environment. The resolution also calls for the creation of a new cabinet-level department to coordinate the activities of all federal agencies that regulate the environment. The resolution also calls for the creation of a new cabinet-level department to coordinate the activities of all federal agencies that regulate the environment.

The two, Daniel Court, 37 years old, and Alan Kravitz, 35, are economists at Princeton University and prominent in a new kind of market research: technocracy. Together, technocrats work outside universities like Cornell, MIT, Harvard's public-health economists — have found that leaving the non-ventilated windows on heavily polluted sites like the downtown mega-

If the two volumes, which are in no sense "being sold here," said Professor Clark, a Canadian who creates many of the best-selling contemporary books of the hour, community is to grow up to it. People cannot be taken too far, and the book is not the only one that is meant to be read.

If the two were merely academic lectures, teaching the elegant discourse that history of Communist attempts without creating credible alternatives, nobody in the field would read them. But they are not. They are not just for the people of the world, as Professor Clark says, but in the country to serve of nations in the most serious journals, and after the publication of their progress, the author of the book, Professor Clark, is a strong critic, mostly in the report, a leading war between the two.

The New Jersey Test

Then, another of the country's wage hags attracted the New England States' economic attention. In 1932, the lowest annual average wage in the country of 19 cents was paid to workers, but Illinois attracted no such, its lowest pay companies who cannot afford to pay them. Profoundly, Card and Kasper decided to find that when New Jersey, in the middle of a recession, raised the minimum wage to the then 10 cents from \$1.05 - 40 cents higher than neighboring Pennsylvania.

They happened not far from Pennsylvania and found that those in New Jersey actually raised 23 workers after the minimum wage

[illegible]

21 JAN 1987



Publication programme: David Ford, 1998, and Alan Watts, 1999.

They use control groups in their research, and test minimum-wage theories by surveying the managers of fast-food restaurants.

They also caused a stir throughout the news-debating press, insisting that the law encourages negligent employers to offer their workers less than the minimum wage, and that it could hurt them and the law.

Professor Krueger studied pay and income in the marketing industry and concluded that the market — not the establishment — set the pace for determining pay for the beneficiary. When benefit costs are pushed up by either market forces or government insurance, wages tend to go up more slowly. In states with high market competition coverage, pay tended to be correspondingly lower.

And, once economists hit pay and Not a penny

[illegible]

have themselves a capital and enjoyed their share in your economic "miracles" more of the most vibrant, changing in the national and more again.

Examiners have this "classical" economic "Professor Card" said: "The big selling point for our biggest Volkswagen dealer is your master degree certificate. But it's like a Card, rapidly - our strength is the one with the most, it's a huge number of people believe that the economic crisis work and that you can't have to strengthen our facts.

'Let's Not Assume'

"The joke used to be that if the stars didn't die in space, that must be something wrong with your make. Now there's a younger generation that says 'Let's not assume the worst!'"

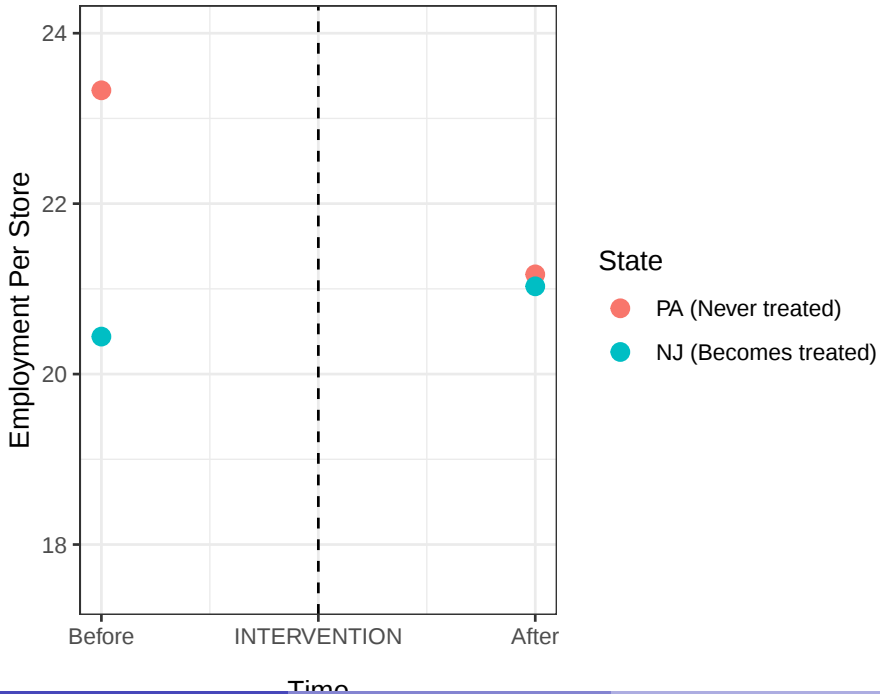
“What attracted us to control project is that we wanted to develop our own fish stock. It was difficult,” Professor Clark said. “What would a simple, rapidly expanding industry?” Although they always stock large gamefishes, salmon, they think they are better suited for releasing their own fish — so they did by introducing a variety of rainbow trout parr.

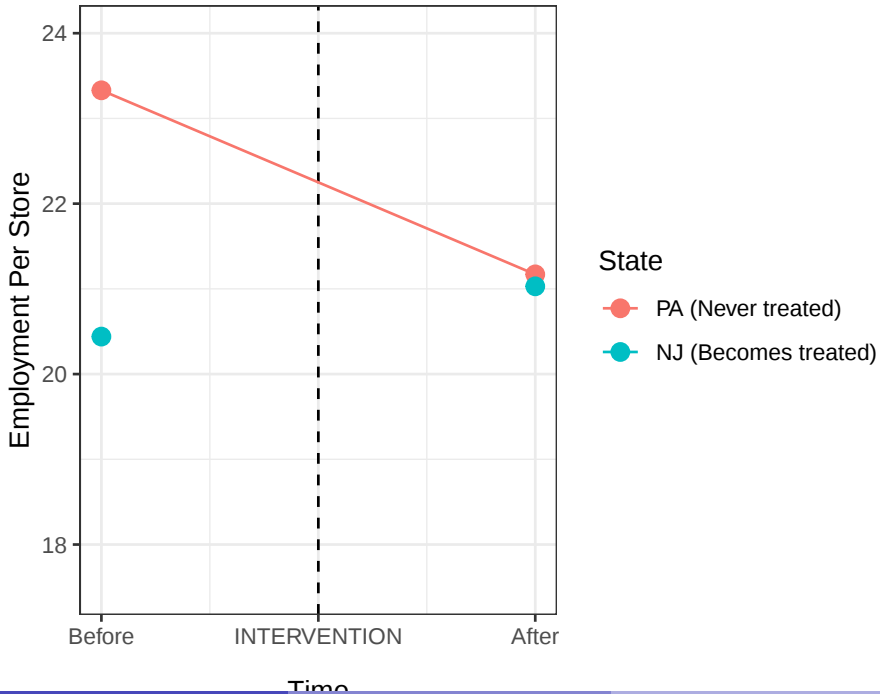
That situation — to be the greatest threat, that kind of crisis that governments have taken pains in this Clinton administration — is their scientific detachment. They mean, for example, that they are not prepared to let the President, say Mr. Bush, find it's a good idea to put the Payton's minimum wage on automatic. Or the making changes in the keeping of all these numbers is a bad idea. Keeping order over the entire economy is a bad idea. The numbers are so important that they must be made the way the scientist say. I really don't know what that has to do with politics, unless these persons insist on more positive politics. Professor Card told, "Suppose we say he believes that the numbers ought, should be higher. A lot of people would say

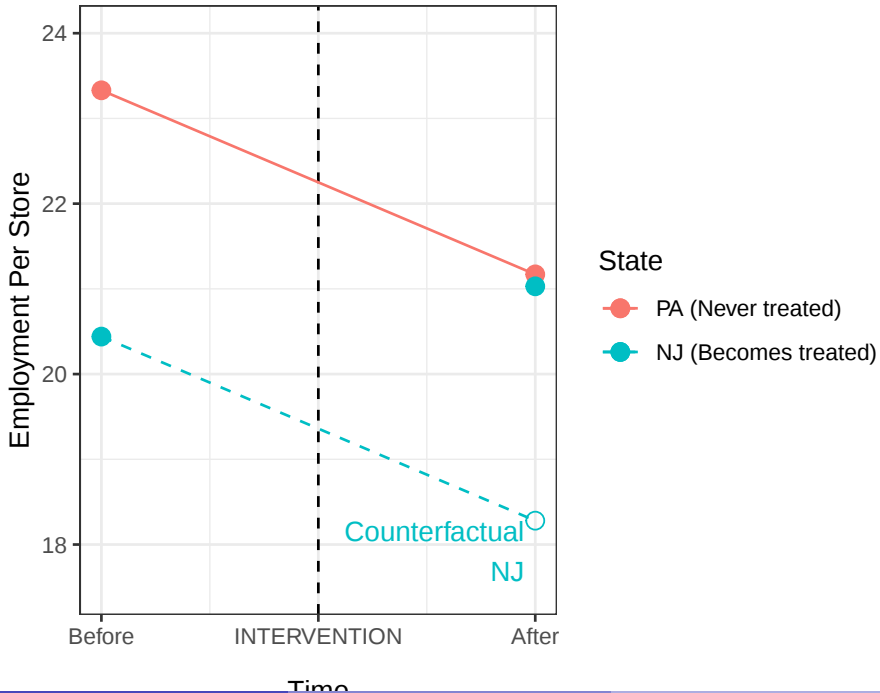
"Julius Friedman, who never was C.E.A. chairman or Fed chairman or anything like that," Friedman's younger son, "is the most influential economist alive today."

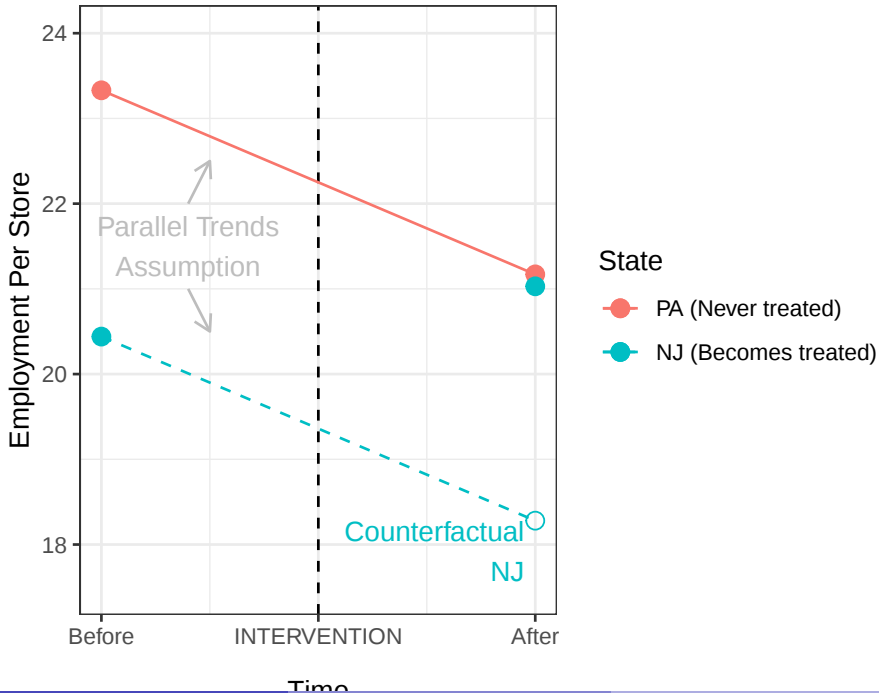
They also caused a stir throughout the news-debating press, insisting that the law encourages negligent employers to offer their workers less than the minimum wage, and that it could hurt them and the law.

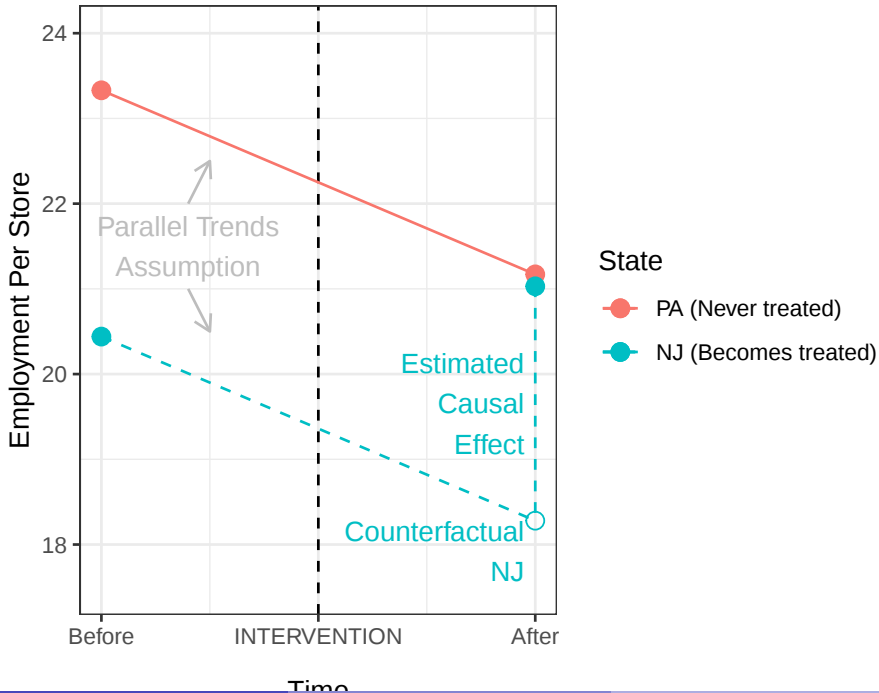
Professor Krueger studied pay and income in the marketing industry and concluded that the market — not the establishment — set the wage substantially higher for the beneficiary. When firms' costs are pushed up by either market forces or government measures, wages tend to go up more slowly. In states with high market competition coverage, pay tended to be correspondingly lower.











Example: Card and Krueger (2000)

- Do increases to the minimum wage depress employment at fast-food restaurants?

$$\text{employment}_{it} = \beta_0 + \beta_1 \text{minimum wage}_{it} + a_i + u_{it}$$

Example: Card and Krueger (2000)

- Do increases to the minimum wage depress employment at fast-food restaurants?

$$\text{employment}_{it} = \beta_0 + \beta_1 \text{minimum wage}_{it} + a_i + u_{it}$$

- Each i here is a different fast food restaurant in either New Jersey or Pennsylvania

Example: Card and Krueger (2000)

- Do increases to the minimum wage depress employment at fast-food restaurants?

$$\text{employment}_{it} = \beta_0 + \beta_1 \text{minimum wage}_{it} + a_i + u_{it}$$

- Each i here is a different fast food restaurant in either New Jersey or Pennsylvania
- Between $t = 1$ and $t = 2$ NJ raised its minimum wage

Example: Card and Krueger (2000)

- Do increases to the minimum wage depress employment at fast-food restaurants?

$$\text{employment}_{it} = \beta_0 + \beta_1 \text{minimum wage}_{it} + a_i + u_{it}$$

- Each i here is a different fast food restaurant in either New Jersey or Pennsylvania
- Between $t = 1$ and $t = 2$ NJ raised its minimum wage
- Employment in fast food might be driven by other state-level policies correlated with minimum wage

Example: Card and Krueger (2000)

- Do increases to the minimum wage depress employment at fast-food restaurants?

$$\text{employment}_{it} = \beta_0 + \beta_1 \text{minimum wage}_{it} + a_i + u_{it}$$

- Each i here is a different fast food restaurant in either New Jersey or Pennsylvania
- Between $t = 1$ and $t = 2$ NJ raised its minimum wage
- Employment in fast food might be driven by other state-level policies correlated with minimum wage
- Diff-in-diff approach: regress changes in employment on store being in NJ

$$\Delta \text{employment}_i = \beta_0 + \beta_1 \text{NJ}_i + \Delta u_i$$

Example: Card and Krueger (2000)

- Do increases to the minimum wage depress employment at fast-food restaurants?

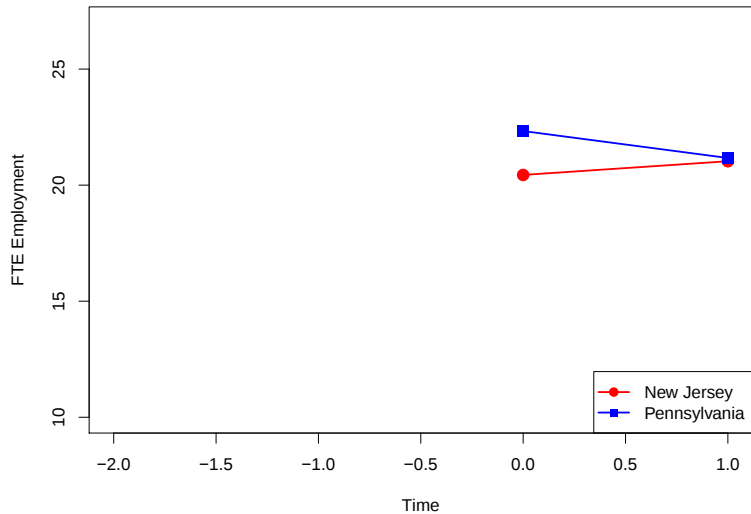
$$\text{employment}_{it} = \beta_0 + \beta_1 \text{minimum wage}_{it} + a_i + u_{it}$$

- Each i here is a different fast food restaurant in either New Jersey or Pennsylvania
- Between $t = 1$ and $t = 2$ NJ raised its minimum wage
- Employment in fast food might be driven by other state-level policies correlated with minimum wage
- Diff-in-diff approach: regress changes in employment on store being in NJ

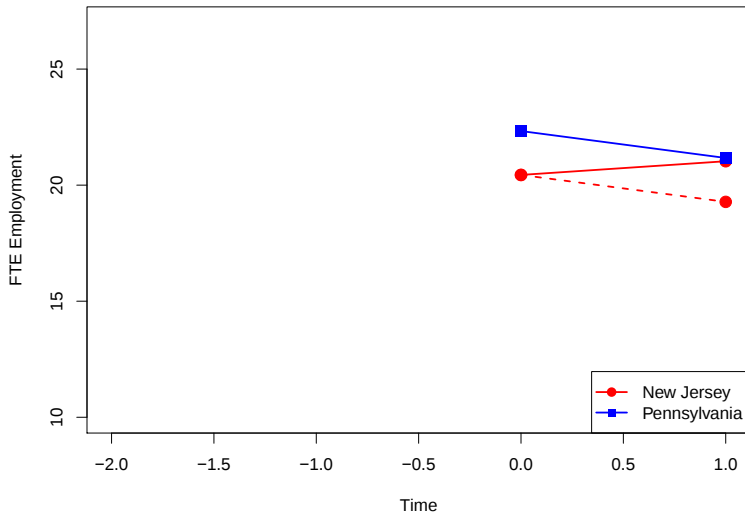
$$\Delta \text{employment}_i = \beta_0 + \beta_1 \text{NJ}_i + \Delta u_i$$

- NJ_i indicates which stores received the treatment of a higher minimum wage at time period $t = 2$

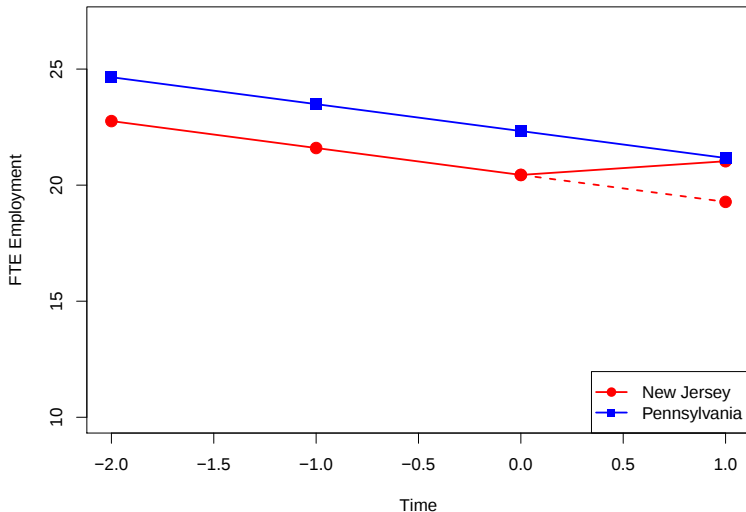
Parallel Trends?



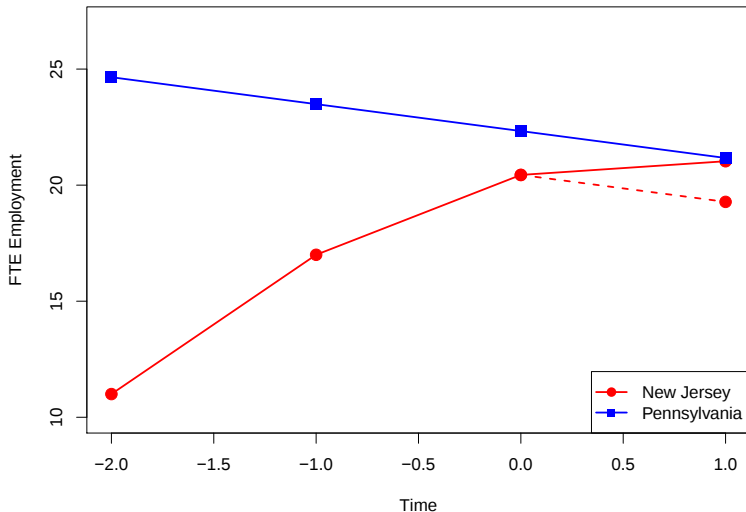
Parallel Trends?



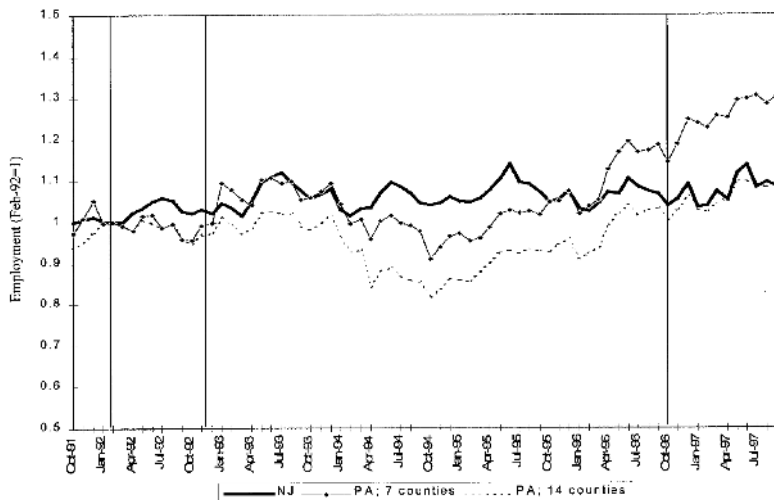
Parallel Trends?



Parallel Trends?



Longer Trends in Employment (Card and Krueger 2000)



First two vertical lines indicate the dates of the Card-Krueger survey. October 1996 line is the federal minimum wage hike which was binding in PA but not NJ

Threats to Identification

Threats to Identification

1) Failure of Exogeneity

Treatment needs to be independent of the idiosyncratic shocks

Threats to Identification

1) Failure of Exogeneity

Treatment needs to be independent of the idiosyncratic shocks

2) Non-parallel dynamics

variation in the outcome over time is the same for the treated and control groups (i.e. no omitted time-varying confounders). e.g. Ashenfelter's dip: people who enroll in job training programs see their earnings decline prior to that training (presumably why they are entering)

Threats to Identification

- 1) Failure of Exogeneity
Treatment needs to be independent of the idiosyncratic shocks
- 2) Non-parallel dynamics
variation in the outcome over time is the same for the treated and control groups (i.e. no omitted time-varying confounders). e.g. Ashenfelter's dip: people who enroll in job training programs see their earnings decline prior to that training (presumably why they are entering)
- 3) Changes in Composition of Treatment/Control Groups
we don't want composition of sample to change between periods. what if workers move from eastern PA to NJ in search of higher paying jobs?

Threats to Identification

- 1) Failure of Exogeneity
Treatment needs to be independent of the idiosyncratic shocks
- 2) Non-parallel dynamics
variation in the outcome over time is the same for the treated and control groups (i.e. no omitted time-varying confounders). e.g. Ashenfelter's dip: people who enroll in job training programs see their earnings decline prior to that training (presumably why they are entering)
- 3) Changes in Composition of Treatment/Control Groups
we don't want composition of sample to change between periods. what if workers move from eastern PA to NJ in search of higher paying jobs?
- 4) Long-term vs. Short-term Effects
parallel trends are less credible over a long time horizon, but many policies need time to take effect.

Threats to Identification

Threats to Identification

- 5) Functional Form Dependence
difference in levels and difference in logs can be quite different
(example via Justin Grimmer)

Threats to Identification

5) Functional Form Dependence

difference in levels and difference in logs can be quite different
(example via Justin Grimmer)

- ▶ imagine a training program for the young

Threats to Identification

5) Functional Form Dependence

difference in levels and difference in logs can be quite different
(example via Justin Grimmer)

- ▶ imagine a training program for the young
- ▶ employment for the young increases from 20% to 30%
- ▶ employment for the old increases from 5% to 10%

Threats to Identification

5) Functional Form Dependence

difference in levels and difference in logs can be quite different
(example via Justin Grimmer)

- ▶ imagine a training program for the young
- ▶ employment for the young increases from 20% to 30%
- ▶ employment for the old increases from 5% to 10%
- ▶ positive DiD effect: $(30 - 20) - (10 - 5) = 5\%$

Threats to Identification

5) Functional Form Dependence

difference in levels and difference in logs can be quite different
(example via Justin Grimmer)

- ▶ imagine a training program for the young
- ▶ employment for the young increases from 20% to 30%
- ▶ employment for the old increases from 5% to 10%
- ▶ positive DiD effect: $(30 - 20) - (10 - 5) = 5\%$
- ▶ but if you consider changes in the log:
$$[\log(30) - \log(20)] - [\log(10) - \log(5)] = \log(1.5) - \log(2) < 0$$

Threats to Identification

5) Functional Form Dependence

difference in levels and difference in logs can be quite different
(example via Justin Grimmer)

- ▶ imagine a training program for the young
- ▶ employment for the young increases from 20% to 30%
- ▶ employment for the old increases from 5% to 10%
- ▶ positive DiD effect: $(30 - 20) - (10 - 5) = 5\%$
- ▶ but if you consider changes in the log:
 $[\log(30) - \log(20)] - [\log(10) - \log(5)] = \log(1.5) - \log(2) < 0$
- ▶ how do we tell which (if either) yields parallel trends?

Threats to Identification

5) Functional Form Dependence

difference in levels and difference in logs can be quite different
(example via Justin Grimmer)

- ▶ imagine a training program for the young
- ▶ employment for the young increases from 20% to 30%
- ▶ employment for the old increases from 5% to 10%
- ▶ positive DiD effect: $(30 - 20) - (10 - 5) = 5\%$
- ▶ but if you consider changes in the log:
 $[\log(30) - \log(20)] - [\log(10) - \log(5)] = \log(1.5) - \log(2) < 0$
- ▶ how do we tell which (if either) yields parallel trends?

- 1 Difference-in-Differences
- 2 Synthetic Controls
- 3 Fixed Effects
- 4 Non-parametric Identification and Fixed Effects
- 5 Wrap-Up
 - Questions
 - Concluding Thoughts for the Course

- 1 Difference-in-Differences
- 2 Synthetic Controls**
- 3 Fixed Effects
- 4 Non-parametric Identification and Fixed Effects
- 5 Wrap-Up
 - Questions
 - Concluding Thoughts for the Course

Synthetic control

Abadie, A., Diamond, A., & 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(490), 493-505.

Synthetic control

Abadie, A., Diamond, A., & 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(490), 493-505.

In 1988, California implemented a tobacco control program

Synthetic control

Abadie, A., Diamond, A., & 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(490), 493-505.

In 1988, California implemented a tobacco control program

- New tax: 25 cents per pack

Synthetic control

Abadie, A., Diamond, A., & 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(490), 493-505.

In 1988, California implemented a tobacco control program

- New tax: 25 cents per pack
- Money earmarked for smoking-reduction programs

Synthetic control

Abadie, A., Diamond, A., & 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(490), 493-505.

In 1988, California implemented a tobacco control program

- New tax: 25 cents per pack
- Money earmarked for smoking-reduction programs

How much did it reduce CA cigarette sales in 1990? 1995? 2000?

Synthetic control

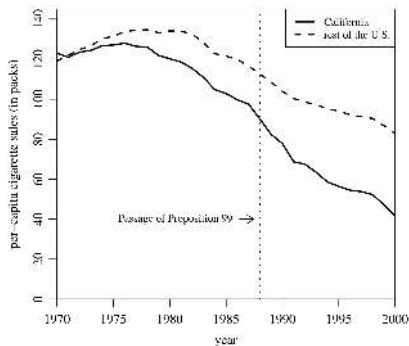


Figure 1. Trends in per-capita cigarette sales: California vs. the rest of the United States.

Can't use RDD

— Effect at 1988 not of interest

Can't use ITS

— Hard to extrapolate Y^0 trend

Can't use DID

— No other state like CA

Idea: Create a
synthetic CA

to estimate

$Y_{CA,t}^0$ for $t \geq 1988$

A Synthetic California

Table 1. Cigarette sales predictor means

Variables	California		Average of 38 control states
	Real	Synthetic	
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). GDP per capita is measured in 1997 dollars, retail prices are measured in cents, beer consumption is measured in gallons, and cigarette sales are measured in packs.

A Synthetic California

Table 2. State weights in the 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

Synthetic control

Theoretical Estimand: $\tau(t) = Y_{CA,t}^1 - Y_{CA,t}^0 \quad t \geq 1988$

Synthetic control

Theoretical Estimand: $\tau(t) = Y_{CA,t}^1 - Y_{CA,t}^0 \quad t \geq 1988$

Empirical Estimand: $\theta(t) = Y_{CA,t}^1 - Y_{\text{SyntheticCA},t}^0 \quad t \geq 1988$

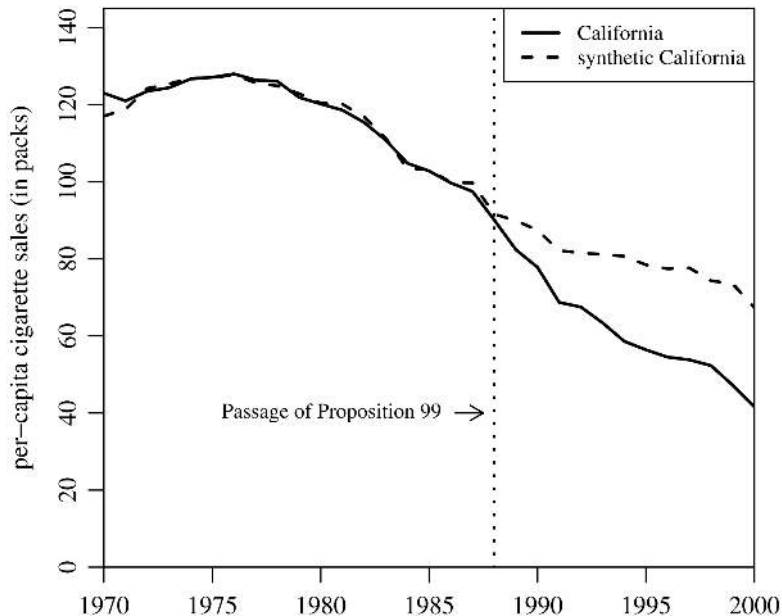
Synthetic control

Theoretical Estimand: $\tau(t) = Y_{CA,t}^1 - Y_{CA,t}^0 \quad t \geq 1988$

Empirical Estimand: $\theta(t) = Y_{CA,t}^1 - Y_{\text{SyntheticCA},t}^0 \quad t \geq 1988$

Identifying Assumption: $\underbrace{Y_{CA,t}^0}_{\text{Counterfactual}} = \underbrace{Y_{\text{SyntheticCA},t}^0}_{\text{Factual}} \quad t \geq 1988$

Synthetic control



Justifications of Synthetic Control

Justifications of Synthetic Control

- Two **model-based** justifications for synthetic control.

Justifications of Synthetic Control

- Two **model-based** justifications for synthetic control.
 - ▶ Model 1: structured, unobserved confounding with **interactive fixed effects**

Justifications of Synthetic Control

- Two **model-based** justifications for synthetic control.
 - ▶ Model 1: structured, unobserved confounding with **interactive fixed effects**
 - ▶ Model 2: **autoregressive model** without fixed effects

Justifications of Synthetic Control

- Two **model-based** justifications for synthetic control.
 - ▶ Model 1: structured, unobserved confounding with **interactive fixed effects**
 - ▶ Model 2: **autoregressive model** without fixed effects
 - ▶ Works for fixed effects or lagged dependent variables, but not both.

Justifications of Synthetic Control

- Two **model-based** justifications for synthetic control.
 - ▶ Model 1: structured, unobserved confounding with **interactive fixed effects**
 - ▶ Model 2: **autoregressive model** without fixed effects
 - ▶ Works for fixed effects or lagged dependent variables, but not both.
- Strategies that generalize to more treated units:
 - ▶ interactive fixed effects (e.g. Xu 2017)
 - ▶ matrix completion (e.g. Athey et al 2021)

Justifications of Synthetic Control

- Two **model-based** justifications for synthetic control.
 - ▶ Model 1: structured, unobserved confounding with **interactive fixed effects**
 - ▶ Model 2: **autoregressive model** without fixed effects
 - ▶ Works for fixed effects or lagged dependent variables, but not both.
- Strategies that generalize to more treated units:
 - ▶ interactive fixed effects (e.g. Xu 2017)
 - ▶ matrix completion (e.g. Athey et al 2021)

To read more about core SCM: Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. "Comparative politics and the synthetic control method." *American Journal of Political Science* 59.2 (2015): 495-510.

- 1 Difference-in-Differences
- 2 Synthetic Controls
- 3 Fixed Effects
- 4 Non-parametric Identification and Fixed Effects
- 5 Wrap-Up
 - Questions
 - Concluding Thoughts for the Course

- 1 Difference-in-Differences
- 2 Synthetic Controls
- 3 Fixed Effects**
- 4 Non-parametric Identification and Fixed Effects
- 5 Wrap-Up
 - Questions
 - Concluding Thoughts for the Course

Basic Model

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + a_i + u_{it}$$

- Consider the above basic model where a_i is an intercept associated with unit i and u_{it} satisfy strict exogeneity.

Basic Model

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + a_i + u_{it}$$

- Consider the above basic model where a_i is an intercept associated with unit i and u_{it} satisfy strict exogeneity.
- The **fixed effects model** is a way to remove time-invariant unmeasured confounding from a_i

Basic Model

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + a_i + u_{it}$$

- Consider the above basic model where a_i is an intercept associated with unit i and u_{it} satisfy strict exogeneity.
- The **fixed effects model** is a way to remove time-invariant unmeasured confounding from a_i
- We will start by assuming the model and discussing properties and in the next section, we will consider non-parametric identification.

Fixed Effects Models

- Core idea is to focus on **within-unit comparisons**: changes in y_{it} and x_{it} relative to their within-group means

Fixed Effects Models

- Core idea is to focus on **within-unit comparisons**: changes in y_{it} and x_{it} relative to their within-group means
- First note that taking the average of the y 's over time for a given unit leaves us with a very similar model:

$$\bar{y}_i = \frac{1}{T} \sum_{t=1}^T [\mathbf{x}'_{it}\boldsymbol{\beta} + a_i + u_{it}]$$

Fixed Effects Models

- Core idea is to focus on **within-unit comparisons**: changes in y_{it} and x_{it} relative to their within-group means
- First note that taking the average of the y 's over time for a given unit leaves us with a very similar model:

$$\begin{aligned}\bar{y}_i &= \frac{1}{T} \sum_{t=1}^T [\mathbf{x}'_{it}\boldsymbol{\beta} + a_i + u_{it}] \\ &= \left(\frac{1}{T} \sum_{t=1}^T \mathbf{x}'_{it} \right) \boldsymbol{\beta} + \frac{1}{T} \sum_{t=1}^T a_i + \frac{1}{T} \sum_{t=1}^T u_{it}\end{aligned}$$

Fixed Effects Models

- Core idea is to focus on **within-unit comparisons**: changes in y_{it} and x_{it} relative to their within-group means
- First note that taking the average of the y 's over time for a given unit leaves us with a very similar model:

$$\begin{aligned}\bar{y}_i &= \frac{1}{T} \sum_{t=1}^T [\mathbf{x}'_{it}\boldsymbol{\beta} + a_i + u_{it}] \\ &= \left(\frac{1}{T} \sum_{t=1}^T \mathbf{x}'_{it} \right) \boldsymbol{\beta} + \frac{1}{T} \sum_{t=1}^T a_i + \frac{1}{T} \sum_{t=1}^T u_{it} \\ &= \bar{\mathbf{x}}'_i \boldsymbol{\beta} + a_i + \bar{u}_i\end{aligned}$$

Fixed Effects Models

- Core idea is to focus on **within-unit comparisons**: changes in y_{it} and x_{it} relative to their within-group means
- First note that taking the average of the y 's over time for a given unit leaves us with a very similar model:

$$\begin{aligned}\bar{y}_i &= \frac{1}{T} \sum_{t=1}^T [\mathbf{x}'_{it}\boldsymbol{\beta} + a_i + u_{it}] \\ &= \left(\frac{1}{T} \sum_{t=1}^T \mathbf{x}'_{it} \right) \boldsymbol{\beta} + \frac{1}{T} \sum_{t=1}^T a_i + \frac{1}{T} \sum_{t=1}^T u_{it} \\ &= \bar{\mathbf{x}}'_i \boldsymbol{\beta} + a_i + \bar{u}_i\end{aligned}$$

- Key fact: because it is **time-constant** the mean of a_i is just a_i

Fixed Effects Models

- Core idea is to focus on **within-unit comparisons**: changes in y_{it} and x_{it} relative to their within-group means
- First note that taking the average of the y 's over time for a given unit leaves us with a very similar model:

$$\begin{aligned}\bar{y}_i &= \frac{1}{T} \sum_{t=1}^T [\mathbf{x}'_{it}\boldsymbol{\beta} + a_i + u_{it}] \\ &= \left(\frac{1}{T} \sum_{t=1}^T \mathbf{x}'_{it} \right) \boldsymbol{\beta} + \frac{1}{T} \sum_{t=1}^T a_i + \frac{1}{T} \sum_{t=1}^T u_{it} \\ &= \bar{\mathbf{x}}'_i \boldsymbol{\beta} + a_i + \bar{u}_i\end{aligned}$$

- Key fact: because it is **time-constant** the mean of a_i is just a_i
- This regression is sometimes called the “between regression”

Within Transformation

Within Transformation

- The “fixed effects,” “within,” or “time-demeaning” transformation is when we subtract off the over-time means from the original data:

$$(y_{it} - \bar{y}_i) = (\mathbf{x}'_{it} - \bar{\mathbf{x}}'_i)\boldsymbol{\beta} + (u_{it} - \bar{u}_i)$$

Within Transformation

- The “fixed effects,” “within,” or “time-demeaning” transformation is when we subtract off the over-time means from the original data:

$$(y_{it} - \bar{y}_i) = (\mathbf{x}'_{it} - \bar{\mathbf{x}}'_i)\boldsymbol{\beta} + (u_{it} - \bar{u}_i)$$

- If we write $\ddot{y}_{it} = y_{it} - \bar{y}_i$, then we can write this more compactly as:

$$\ddot{y}_{it} = \ddot{\mathbf{x}}'_{it}\boldsymbol{\beta} + \ddot{u}_{it}$$

Within Transformation

- The “fixed effects,” “within,” or “time-demeaning” transformation is when we subtract off the over-time means from the original data:

$$(y_{it} - \bar{y}_i) = (\mathbf{x}'_{it} - \bar{\mathbf{x}}'_i)\boldsymbol{\beta} + (u_{it} - \bar{u}_i)$$

- If we write $\ddot{y}_{it} = y_{it} - \bar{y}_i$, then we can write this more compactly as:

$$\ddot{y}_{it} = \ddot{\mathbf{x}}'_{it}\boldsymbol{\beta} + \ddot{u}_{it}$$

- Degrees of freedom: $nT - n - k - 1$, which accounts for within transformation (i.e. either use a package like `p1m` or adjust the degrees of freedom manually).

Within Transformation

- The “fixed effects,” “within,” or “time-demeaning” transformation is when we subtract off the over-time means from the original data:

$$(y_{it} - \bar{y}_i) = (\mathbf{x}'_{it} - \bar{\mathbf{x}}'_i)\boldsymbol{\beta} + (u_{it} - \bar{u}_i)$$

- If we write $\ddot{y}_{it} = y_{it} - \bar{y}_i$, then we can write this more compactly as:

$$\ddot{y}_{it} = \ddot{\mathbf{x}}'_{it}\boldsymbol{\beta} + \ddot{u}_{it}$$

- Degrees of freedom: $nT - n - k - 1$, which accounts for within transformation (i.e. either use a package like `p1m` or adjust the degrees of freedom manually).
- We are now modeling observations as deviation from their group mean.

Within Transformation

- The “fixed effects,” “within,” or “time-demeaning” transformation is when we subtract off the over-time means from the original data:

$$(y_{it} - \bar{y}_i) = (\mathbf{x}'_{it} - \bar{\mathbf{x}}'_i)\boldsymbol{\beta} + (u_{it} - \bar{u}_i)$$

- If we write $\ddot{y}_{it} = y_{it} - \bar{y}_i$, then we can write this more compactly as:

$$\ddot{y}_{it} = \ddot{\mathbf{x}}'_{it}\boldsymbol{\beta} + \ddot{u}_{it}$$

- Degrees of freedom: $nT - n - k - 1$, which accounts for within transformation (i.e. either use a package like `plm` or adjust the degrees of freedom manually).
- We are now modeling observations as deviation from their group mean.
- NB: you **must** demean the X variables not just the Y variables.

Fixed Effects Example with data from Ross

```
fe.mod <- plm(log(kidmort_unicef) ~ democracy + log(GDPcur), data = ross, index = c("id", "year"),
  model = "within")
summary(fe.mod)

## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = log(kidmort_unicef) ~ democracy + log(GDPcur),
##      data = ross, model = "within", index = c("id", "year"))
##
## Unbalanced Panel: n=166, T=1-7, N=649
##
## Residuals :
##      Min.    1st Qu.    Median    3rd Qu.    Max.
## -0.70500 -0.11700   0.00628   0.12200   0.75700
##
## Coefficients :
##              Estimate Std. Error t-value Pr(>|t|)
## democracy    -0.143233   0.033500  -4.2756 2.299e-05 ***
## log(GDPcur)  -0.375203   0.011328 -33.1226 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    81.711
## Residual Sum of Squares: 23.012
## R-Squared      : 0.71838
##      Adj. R-Squared : 0.53242
## F-statistic: 613.481 on 2 and 481 DF, p-value: < 2.22e-16
```

Strict Exogeneity

- FE models are valid if $E[\mathbf{u}|\mathbf{X}] = 0$: all errors are uncorrelated with covariates in every period:

$$E[\ddot{u}_{it}|\ddot{\mathbf{X}}] = E[u_{it}|\ddot{\mathbf{X}}] - E[\bar{u}_i|\ddot{\mathbf{X}}] = 0 - 0 = 0$$

Strict Exogeneity

- FE models are valid if $E[\mathbf{u}|\mathbf{X}] = 0$: all errors are uncorrelated with covariates in every period:

$$E[\ddot{u}_{it}|\ddot{\mathbf{X}}] = E[u_{it}|\ddot{\mathbf{X}}] - E[\bar{u}_i|\ddot{\mathbf{X}}] = 0 - 0 = 0$$

- This is because the composite errors, $\ddot{u}_{it}$ are function of the errors in every time period through the average, $\bar{u}_i$

Strict Exogeneity

- FE models are valid if $E[\mathbf{u}|\mathbf{X}] = 0$: all errors are uncorrelated with covariates in every period:

$$E[\ddot{u}_{it}|\ddot{\mathbf{X}}] = E[u_{it}|\ddot{\mathbf{X}}] - E[\bar{u}_i|\ddot{\mathbf{X}}] = 0 - 0 = 0$$

- This is because the composite errors, $\ddot{u}_{it}$ are function of the errors in every time period through the average, $\bar{u}_i$
- This rules out, for instance, lagged dependent variables, since $y_{i,t-1}$ has to be correlated with $u_{i,t-1}$. Thus it can't be a covariate for y_{it} .

Fixed Effects and Time-Invariant Covariates

- What if there is a covariate that doesn't vary over time?

Fixed Effects and Time-Invariant Covariates

- What if there is a covariate that doesn't vary over time?
- Then $x_{it} = \bar{x}_i$ and $\ddot{x}_{it} = 0$ for all periods t .

Fixed Effects and Time-Invariant Covariates

- What if there is a covariate that doesn't vary over time?
- Then $x_{it} = \bar{x}_i$ and $\ddot{x}_{it} = 0$ for all periods t .
- If the time-demeaned covariate is always 0, then it will be perfectly collinear with the intercept and will violate full rank. R/Stata will **drop** it from the regression.

Fixed Effects and Time-Invariant Covariates

- What if there is a covariate that doesn't vary over time?
- Then $x_{it} = \bar{x}_i$ and $\ddot{x}_{it} = 0$ for all periods t .
- If the time-demeaned covariate is always 0, then it will be perfectly collinear with the intercept and will violate full rank. R/Stata will **drop** it from the regression.
- Basic message: any time-constant variable gets “absorbed” by the fixed effect. It has nothing to contribute because the comparison is **within the units**.

Fixed Effects and Time-Invariant Covariates

- What if there is a covariate that doesn't vary over time?
- Then $x_{it} = \bar{x}_i$ and $\ddot{x}_{it} = 0$ for all periods t .
- If the time-demeaned covariate is always 0, then it will be perfectly collinear with the intercept and will violate full rank. R/Stata will **drop** it from the regression.
- Basic message: any time-constant variable gets “absorbed” by the fixed effect. It has nothing to contribute because the comparison is **within the units**.
- Can include interactions between time-constant and time-varying variables, but lower order term of the time-constant variables get absorbed by fixed effects too

Time-constant variables

- Pooled model with a time-constant variable, proportion Islamic:

```
library(lmtest)
p.mod <- plm(log(kidmort_unicef) ~ democracy + log(GDPcur) + islam,
             data = ross, index = c("id", "year"), model = "pooling")
coeftest(p.mod)

##
## t test of coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.30607817  0.35951939  28.6663 < 2.2e-16 ***
## democracy   -0.80233845  0.07766814 -10.3303 < 2.2e-16 ***
## log(GDPcur) -0.25497406  0.01607061 -15.8659 < 2.2e-16 ***
## islam        0.00343325  0.00091045   3.7709 0.0001794 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Time-constant variables

- FE model, where the islam variable drops out, along with the intercept:

```
fe.mod2 <- plm(log(kidmort_unicef) ~ democracy + log(GDPcur) + islam,  
               data = ross, index = c("id", "year"), model = "within")  
coeftest(fe.mod2)
```

```
##  
## t test of coefficients:  
##  
##           Estimate Std. Error  t value  Pr(>|t|)  
## democracy   -0.129693   0.035865  -3.6162 0.0003332 ***  
## log(GDPcur) -0.379997   0.011849 -32.0707 < 2.2e-16 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Alternate Computation: Least Squares Dummy Variable

- As an alternative to the within transformation, we can also include a series of $n - 1$ dummy variables for each unit:

$$y_{it} = \mathbf{x}'_{it}\beta + d_i^{(1)}\alpha_1 + d_i^{(2)}\alpha_2 + \cdots + d_i^{(n)}\alpha_n + u_{it}$$

Alternate Computation: Least Squares Dummy Variable

- As an alternative to the within transformation, we can also include a series of $n - 1$ dummy variables for each unit:

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + d_i^{(1)}\alpha_1 + d_i^{(2)}\alpha_2 + \cdots + d_i^{(n)}\alpha_n + u_{it}$$

- Here, $d_i^{(1)}$ is a binary variable which is 1 if $i = 1$ and 0 otherwise—just a unit dummy.

Alternate Computation: Least Squares Dummy Variable

- As an alternative to the within transformation, we can also include a series of $n - 1$ dummy variables for each unit:

$$y_{it} = \mathbf{x}'_{it}\beta + d_i^{(1)}\alpha_1 + d_i^{(2)}\alpha_2 + \cdots + d_i^{(n)}\alpha_n + u_{it}$$

- Here, $d_i^{(1)}$ is a binary variable which is 1 if $i = 1$ and 0 otherwise—just a unit dummy.
- Gives the **exact** same estimates/standard errors as with time-demeaning

Alternate Computation: Least Squares Dummy Variable

- As an alternative to the within transformation, we can also include a series of $n - 1$ dummy variables for each unit:

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + d_i^{(1)}\alpha_1 + d_i^{(2)}\alpha_2 + \cdots + d_i^{(n)}\alpha_n + u_{it}$$

- Here, $d_i^{(1)}$ is a binary variable which is 1 if $i = 1$ and 0 otherwise—just a unit dummy.
- Gives the **exact** same estimates/standard errors as with time-demeaning
 - Advantage: easy to implement in base R (so is the de-meaning but you have to recompute standard errors by changing the degrees of freedom manually).

Alternate Computation: Least Squares Dummy Variable

- As an alternative to the within transformation, we can also include a series of $n - 1$ dummy variables for each unit:

$$y_{it} = \mathbf{x}'_{it}\beta + d_i^{(1)}\alpha_1 + d_i^{(2)}\alpha_2 + \cdots + d_i^{(n)}\alpha_n + u_{it}$$

- Here, $d_i^{(1)}$ is a binary variable which is 1 if $i = 1$ and 0 otherwise—just a unit dummy.
- Gives the **exact** same estimates/standard errors as with time-demeaning
 - Advantage: easy to implement in base R (so is the de-meaning but you have to recompute standard errors by changing the degrees of freedom manually).
 - Disadvantage: computationally difficult with large data sets, since we have to run a regression with $n + k$ variables.

Alternate Computation: Least Squares Dummy Variable

- As an alternative to the within transformation, we can also include a series of $n - 1$ dummy variables for each unit:

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + d_i^{(1)}\alpha_1 + d_i^{(2)}\alpha_2 + \cdots + d_i^{(n)}\alpha_n + u_{it}$$

- Here, $d_i^{(1)}$ is a binary variable which is 1 if $i = 1$ and 0 otherwise—just a unit dummy.
- Gives the **exact** same estimates/standard errors as with time-demeaning
 - Advantage: easy to implement in base R (so is the de-meaning but you have to recompute standard errors by changing the degrees of freedom manually).
 - Disadvantage: computationally difficult with large data sets, since we have to run a regression with $n + k$ variables.
- Why are these equivalent? (remember partialing out strategy and Frisch-Waugh-Lovell theorem)

Example with Ross data

```
library(lmtest)
lsdv.mod <- lm(log(kidmort_unicef) ~ democracy + log(GDPcur) +
               as.factor(id), data = ross)
coeftest(lsdv.mod)[1:6,]
coeftest(fe.mod)[1:2,]
```


##		Estimate	Std. Error	t value	Pr(> t)
##	(Intercept)	13.7644887	0.26597312	51.751427	1.008329e-198
##	democracy	-0.1432331	0.03349977	-4.275644	2.299393e-05
##	log(GDPcur)	-0.3752030	0.01132772	-33.122568	3.494887e-126
##	as.factor(id)AGO	0.2997206	0.16767730	1.787485	7.448861e-02
##	as.factor(id)ALB	-1.9309618	0.19013955	-10.155498	4.392512e-22
##	as.factor(id)ARE	-1.8762909	0.17020738	-11.023558	2.386557e-25

##		Estimate	Std. Error	t value	Pr(> t)
##	democracy	-0.1432331	0.03349977	-4.275644	2.299393e-05
##	log(GDPcur)	-0.3752030	0.01132772	-33.122568	3.494887e-126

Applying Fixed Effects

- We can use fixed effects for other data structures to restrict comparisons to within unit variation

Applying Fixed Effects

- We can use fixed effects for other data structures to restrict comparisons to within unit variation
 - ▶ Matched pairs
 - ★ Twin fixed effects to control for unobserved effects of family background

Applying Fixed Effects

- We can use fixed effects for other data structures to restrict comparisons to within unit variation
 - ▶ Matched pairs
 - ★ Twin fixed effects to control for unobserved effects of family background
 - ▶ Cluster fixed effects in hierarchical data
 - ★ School fixed effects to control for unobserved effects of school

Moving Beyond Linear Separable Confounding

Moving Beyond Linear Separable Confounding

- One reason we like DAGs is that the identification results don't have to start with a statement like, assume the following linear model:

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + a_i + u_{it}$$

Moving Beyond Linear Separable Confounding

- One reason we like DAGs is that the identification results don't have to start with a statement like, assume the following linear model:

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + a_i + u_{it}$$

- What assumptions have we made so far?

Moving Beyond Linear Separable Confounding

- One reason we like DAGs is that the identification results don't have to start with a statement like, assume the following linear model:

$$y_{it} = \mathbf{x}'_{it}\beta + a_i + u_{it}$$

- What assumptions have we made so far?
 - ▶ constant effects
 - ▶ linearity
 - ▶ strict exogeneity

Moving Beyond Linear Separable Confounding

- One reason we like DAGs is that the identification results don't have to start with a statement like, assume the following linear model:

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + a_i + u_{it}$$

- What assumptions have we made so far?
 - ▶ constant effects
 - ▶ linearity
 - ▶ strict exogeneity
- We've seen the trouble with constant effects before, it goes back to Lecture 10 and results on regression with heterogeneous treatment effects more generally.

Contemporaneous, Cumulative and Dynamic Effects

- Another assumption we have been making is that our interest is in a single contemporaneous effect: $\mathbf{x}'_{it}\beta$

Contemporaneous, Cumulative and Dynamic Effects

- Another assumption we have been making is that our interest is in a single contemporaneous effect: $\mathbf{x}'_{it}\beta$
- What if we want to consider the history of a treatment or the effect of a treatment regime (i.e. a treatment that varies over time)?

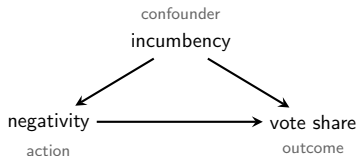
Contemporaneous, Cumulative and Dynamic Effects

- Another assumption we have been making is that our interest is in a single contemporaneous effect: $\mathbf{x}'_{it}\beta$
- What if we want to consider the history of a treatment or the effect of a treatment regime (i.e. a treatment that varies over time)?
- Opens up new estimands, but have to think about how time-varying confounders affect treatment assignment.

Contemporaneous, Cumulative and Dynamic Effects

- Another assumption we have been making is that our interest is in a single contemporaneous effect: $\mathbf{x}'_{it}\beta$
- What if we want to consider the history of a treatment or the effect of a treatment regime (i.e. a treatment that varies over time)?
- Opens up new estimands, but have to think about how time-varying confounders affect treatment assignment.

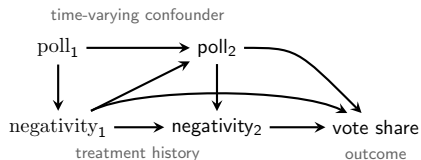
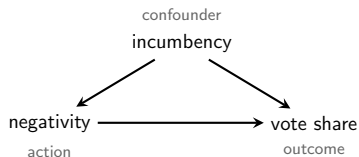
Examples of static and dynamic causal inference problems:



Contemporaneous, Cumulative and Dynamic Effects

- Another assumption we have been making is that our interest is in a single contemporaneous effect: $\mathbf{x}'_{it}\beta$
- What if we want to consider the history of a treatment or the effect of a treatment regime (i.e. a treatment that varies over time)?
- Opens up new estimands, but have to think about how time-varying confounders affect treatment assignment.

Examples of static and dynamic causal inference problems:



Core Conundrum

There is a (possibly irresolvable) tension: modeling **causal dynamics** between treatment and outcomes OR addressing **unobserved time-invariant confounders**.

Core Conundrum

There is a (possibly irresolvable) tension: modeling **causal dynamics** between treatment and outcomes OR addressing **unobserved time-invariant confounders**. Three great recent papers:

A Framework for Dynamic Causal Inference in Political Science

Nathaniel Blackhall *University of Tennessee*

Abstract: This paper develops a framework for dynamic causal inference in political science. It begins by reviewing the literature on causal inference in political science, and then discusses the challenges of dynamic causal inference. The framework is then presented, and its application to a case study is discussed. The framework is designed to be flexible and adaptable to a wide range of research questions and data structures.

When we think about the challenges of causal inference in political science, we often think about the challenges of unobserved confounders. But there are other challenges, too. One of the most important is the challenge of dynamic causal inference. In this paper, I develop a framework for dynamic causal inference in political science. The framework is designed to be flexible and adaptable to a wide range of research questions and data structures.

The framework is designed to be flexible and adaptable to a wide range of research questions and data structures. It begins by reviewing the literature on causal inference in political science, and then discusses the challenges of dynamic causal inference.

One of the most important challenges of dynamic causal inference is the challenge of unobserved confounders. In this paper, I develop a framework for dynamic causal inference in political science. The framework is designed to be flexible and adaptable to a wide range of research questions and data structures.

The framework is designed to be flexible and adaptable to a wide range of research questions and data structures. It begins by reviewing the literature on causal inference in political science, and then discusses the challenges of dynamic causal inference.

There is a (possibly irresolvable) tension: modeling **causal dynamics** between treatment and outcomes OR addressing **unobserved time-invariant confounders**. Three great recent papers:

Nathan Blackwell is author of *Unsettled*.

1. The authors are grateful to the members of the Department of Mathematics, University of Illinois at Chicago, for their hospitality and to the members of the Department of Mathematics, University of California at Berkeley, for their hospitality and to the members of the Department of Mathematics, University of Michigan, for their hospitality.

[illegible]

There are a number of factors which may be important in determining the extent of the effect of the above variables. For example, the effect of the above variables may be different for different types of firms, or for different types of countries. For example, the effect of the above variables may be different for different types of firms, or for different types of countries.

the same. But it is important to note that the same will be true if you are not a member of the same group. The only way to avoid this is to make sure that you are a member of the same group. This is why it is important to make sure that you are a member of the same group.

This article aims to discuss the possibility of a new role for science in our society in response to knowledge and epistemic crises. Following the work of 1980s and 1990s science and technology studies (STS) scholars, I defend a constructivist view of science and technology, arguing that it is not determined by epistemic or technical factors, but that it is shaped by social and cultural factors. I argue that the effects of the new science and technology on society are not simply determined by the technology, but by the way it is used.

Journal of Clinical Sleep Medicine 2006;11:4, April 2006
doi:10.1093/sleep/11.4.400

MATTHEW BLACKWELL, *Harvard University*
ADAM N. GLYNN, *Duquesne University*

Respondent measurements of the same construct, repeated at groups (or time) that are relevant in many fields of political science. These nonexperimental, nonexperimental (time-series) cross-sectional (TSCS) data, also known collectively as *repeated cross-sections*, are widely used in political science to study public opinion, and other effects of lagged variables. Unfortunately, popular methods for TSCS data analysis (e.g., panel data random-effects for lagged effects) suffer from some strong assumptions. In this paper, we use potential outcomes to define causal questions of interest in these settings and clarify how standard models likely to be estimated without this model can produce biased estimates of their quantities and in post-treatment considerations. We then describe two estimation strategies that avoid these post-treatment biases—*inverse probability weighting* using only causal model assumptions, and *event study* estimators that use the complete set of assumptions to match causal models. We illustrate these methods in a study of how welfare spending affects taxation.

Many attempts to estimate the impact of an intervention on a population of interest, such as the effect of a vaccine on the incidence of a disease, are based on data from observational studies. These studies are often subject to confounding, which is the effect of a third variable on the relationship between the intervention and the outcome. For example, in a study of the effect of a vaccine on the incidence of a disease, the age of the subjects may be a confounding variable. Older subjects may be more likely to receive the vaccine and also more likely to develop the disease. This can lead to a biased estimate of the effect of the vaccine. To address this problem, researchers have developed various methods for analyzing observational data, including propensity score matching, inverse probability weighting, and doubly robust estimation. These methods aim to reduce the bias caused by confounding and provide more accurate estimates of the treatment effect. However, these methods often rely on strong assumptions, such as the absence of unmeasured confounding, which may not be realistic in many situations. Therefore, it is important to carefully evaluate the assumptions underlying these methods and to consider alternative approaches when appropriate.

This paper makes three contributions to the study of TMCs data. Our first contribution is re-defining some

smaller than classical effects and discuss the among tests needed to identify them nonparametrically. We also relate these quantities of interest to common tests of the null hypothesis of no treatment effect. Responses, and show how to derive them from the parameters of a nonlinear TOST model. The nonparametric analysis is then applied to a study of the effect of a 10-day short-course treatment on latent tuberculous infection, unfortunately, however, many caution that the use of a TOST model is inappropriate in this case, including a lack of causal feedback between the treatment and test reaping responses. This feedback, for example, might mean a patient's level of disease at the time of testing is a function of the treatment which in turn might affect future levels of response. We argue that this type of feedback is common in TOST experiments and, in the absence of a causal feedback mechanism in this paper, we discuss the feasibility with this choice compared to standard fixed-effects models. In the latter case, one also needs to consider the possibility of a causal feedback.

This second contribution is to provide an introduction to two methods from Bayesian statistics that take into account the effect of treatment heterogeneity without bias and under weaker assumptions than common TSCs models. We focus on two methods: (i) structural semi-infinite models or SSIMs (Rubin 1997) and (ii) marginal structural models (MSMs) with inverse probability of treatment weighting (IPTW) (Robins, Hernan, and

There is a (possibly irresolvable) tension: modeling **causal dynamics** between treatment and outcomes OR addressing **unobserved time-invariant confounders**. Three great recent papers:

36 / 68

Core Conundrum

There is a (possibly irresolvable) tension: modeling **causal dynamics** between treatment and outcomes OR addressing **unobserved time-invariant confounders**. Three great recent papers:

A Framework for Dynamic Causal Inference
in Political Science

Nathan Blackwell is author of *Unsettled*.

1. The authors of the paper have not provided a clear definition of the term "cognitive bias" and have not provided a clear definition of the term "cognitive bias". The authors have not provided a clear definition of the term "cognitive bias" and have not provided a clear definition of the term "cognitive bias".

the first time that the United States has been able to produce a significant surplus in the export of goods and services. The surplus is the result of a combination of factors, including a strong domestic economy, a weak dollar, and a strong export sector.

© 1999 by The American Psychological Association
0893-3200/99/\$12.00
DOI: 10.1037/0893-3200.13.1.103

the same. There is no apparent reason that the two authors of *Applied Statistics* and *Statistics for Management Decision Making* should have chosen to use different notations for the same concepts. The authors of *Statistics for Management Decision Making* have chosen to use the notation σ^2 for the variance of a random variable X , while the authors of *Applied Statistics* have chosen to use the notation σ^2_X . This is a very unfortunate choice of notation, as it is not clear what σ^2_X represents. It is not clear if it represents the variance of X , or if it represents the variance of the sample mean $\bar{X}$. The authors of *Statistics for Management Decision Making* have chosen to use the notation σ^2 for the variance of a random variable X , while the authors of *Applied Statistics* have chosen to use the notation σ^2_X . This is a very unfortunate choice of notation, as it is not clear what σ^2_X represents. It is not clear if it represents the variance of X , or if it represents the variance of the sample mean $\bar{X}$.

The article assesses the climate policy impact of a switch to the *business as usual* scenario, through its literature and epidemiology studies. For years, 1990 and 2002, the *business as usual* scenario was the only one considered for social assessment of greenhouse gases, as a result of the lack of available data on climate change.

[illegible]

Journal of Clinical Research in Nursing 1998, 1(1), 4, April 1998
doi:10.1017/S1049268898000077

How to Make Causal Inferences with Time-Series Cross-Sectional Data under Selection on Observables

MATTHEW BLACKWELL *Harvard*
ADAM N. GLYNN *Durham University*

Rational measurement of the same construct, repeated in groups over time, often yields many false or spurious results. These measurements, sometimes called time series or panel data, are often expected to estimate a broad set of constructs, including compensation effects and other effects of lagged variables. Unfortunately, popular methods for TSCS data can only provide valid inferences for lagged effects under some strong assumptions. In this paper, we use potential outcomes to define causal questions of interest in these settings and clarify how standard models may be sensitive to their underlying true model. We produce direct estimates of these questions and use post-treatment considerations. We then describe two estimation strategies that avoid these post-treatment biases—inverse probability weighting using an event causal model, and event causal models using the two-stage least squares method applied to event causal models. We illustrate these methods in a study of how salespeople perceive effects of pay.

INTRODUCTION

Much attention to political science has to do with the study of representative institutions of the state—governments, people, or groups at several points in time. This type of data, sometimes called time-series cross-sectional (TSCS) data, allows researchers to look on a large range of phenomena when estimating causal effects. TSCS data also give researchers the power to add a delay unit of questions (data with a single time lag) or a shock unit (for example, see Black and Katz [2011]). However, this type of data does not capture the universal contemporaneous (quasi-)causal effects of a single event—and instead ask how the delay of a process affects the political world. Unfortunately, the most common approaches to modeling TSCS data require strict assumptions to estimate the effect of treatment histories without bias and make it difficult to understand the nature of the causal relationships.

Stephen M. Edelstein is an Assistant Professor, Department of Economics, University of California, San Diego.

[illegible]

This second contribution is to provide an alternative to two methods from the literature that can estimate the effect of treatment heterogeneity without bias and under weaker assumptions than common TVC models. We focus on two methods: (1) structural nested mean models or SNMMs (Robins 1997) and (2) marginal structural models (MSMs) with inverse probability of treatment weighting (IPTW) (Robins, Hernan, and

AJPS AMERICAN JOURNAL
OF PSYCHIATRY

When Should We Use Unit Fixed Effects Regression Models for Causal Inference with Longitudinal Data?

Keywords: *work, stress, coping, organizational commitment, organizational citizenship behavior*

[illegible]

Figure 1. Effect of the type of the substrate on the adsorption of Cu(II) ions. The adsorption was carried out at 25 °C, pH 5.5, and 100 mg/L of Cu(II) ions. The adsorbent was 0.5 g of the substrate. The adsorption time was 24 h. The error bars represent the standard deviation.

Under the influence of previous studies, we have found that the use of the Internet in the workplace is not only a means of communication, but also a means of learning. The use of the Internet in the workplace is not only a means of communication, but also a means of learning. The use of the Internet in the workplace is not only a means of communication, but also a means of learning.

Dr. M. J. Griffin, *University of Southampton*, is the author of *Engineering Vibration* (Addison-Wesley, 1985), *Engineering Vibration: Theory and Practice* (Addison-Wesley, 1990), *Engineering Vibration: Analysis and Design* (Addison-Wesley, 1993), and *Engineering Vibration: Analysis and Design* (Addison-Wesley, 1995).

Downloaded At: 11:53 11 September 2009

The author is indebted to the staff of the Department of Statistics, University of Toronto, for their assistance in the preparation of this manuscript. The author is also indebted to the staff of the Department of Statistics, University of Toronto, for their assistance in the preparation of this manuscript.

© 2006 Blackwell Publishing Ltd *Journal of Internal Medicine* 260: 395–403

We are going to focus on addressing unobserved time-invariant confounders using the last paper.

Next several slides are based on slides graciously provided by In Song Kim and Kosuke Imai.

Directed Acyclic Graph (DAG)

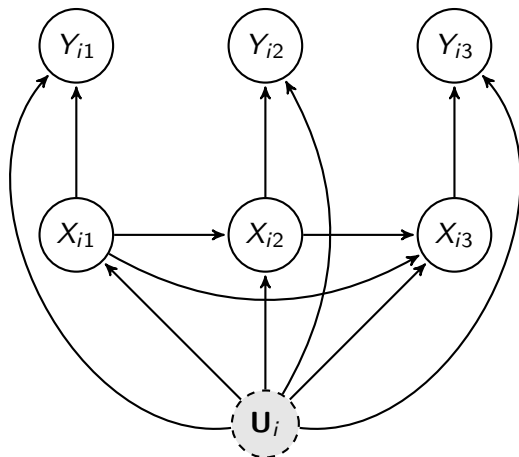
Non-parametric identification assumptions for fixed effects:

$$Y_{it} = g(X_{it}, \mathbf{U}_i, \epsilon_{it}) \quad \text{and} \quad \epsilon_{it} \perp\!\!\!\perp \{\mathbf{X}_i, \mathbf{U}_i\}$$

Directed Acyclic Graph (DAG)

Non-parametric identification assumptions for fixed effects:

$$Y_{it} = g(X_{it}, \mathbf{U}_i, \epsilon_{it}) \quad \text{and} \quad \epsilon_{it} \perp\!\!\!\perp \{\mathbf{X}_i, \mathbf{U}_i\}$$

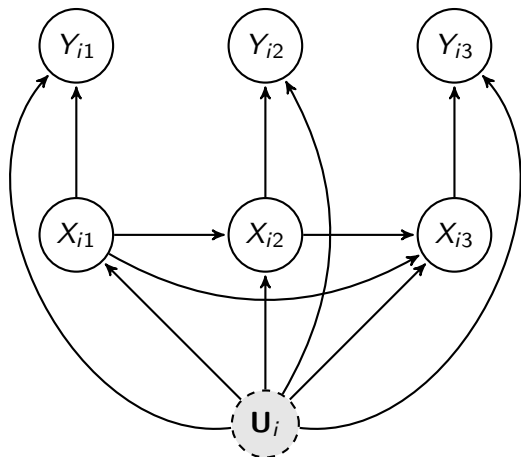


Assumptions:

Directed Acyclic Graph (DAG)

Non-parametric identification assumptions for fixed effects:

$$Y_{it} = g(X_{it}, \mathbf{U}_i, \epsilon_{it}) \quad \text{and} \quad \epsilon_{it} \perp\!\!\!\perp \{\mathbf{X}_i, \mathbf{U}_i\}$$



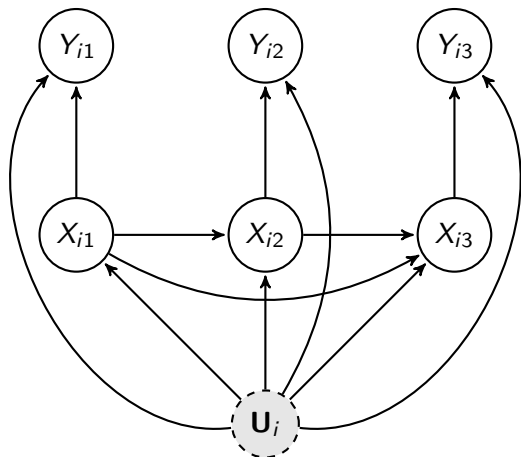
Assumptions:

- 1 No unobserved time-varying confounders

Directed Acyclic Graph (DAG)

Non-parametric identification assumptions for fixed effects:

$$Y_{it} = g(X_{it}, \mathbf{U}_i, \epsilon_{it}) \quad \text{and} \quad \epsilon_{it} \perp\!\!\!\perp \{\mathbf{X}_i, \mathbf{U}_i\}$$



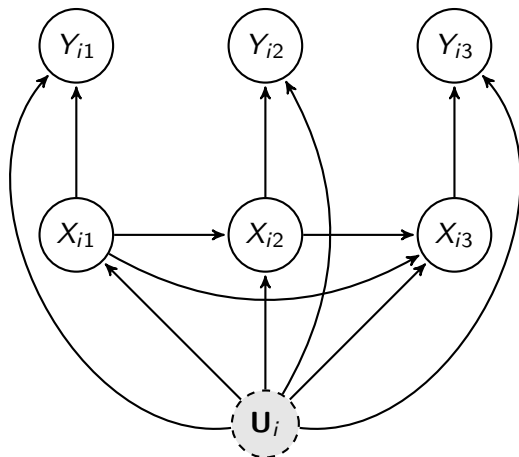
Assumptions:

- 1 No unobserved time-varying confounders
- 2 Past outcomes do not directly affect current outcome

Directed Acyclic Graph (DAG)

Non-parametric identification assumptions for fixed effects:

$$Y_{it} = g(X_{it}, \mathbf{U}_i, \epsilon_{it}) \quad \text{and} \quad \epsilon_{it} \perp\!\!\!\perp \{\mathbf{X}_i, \mathbf{U}_i\}$$



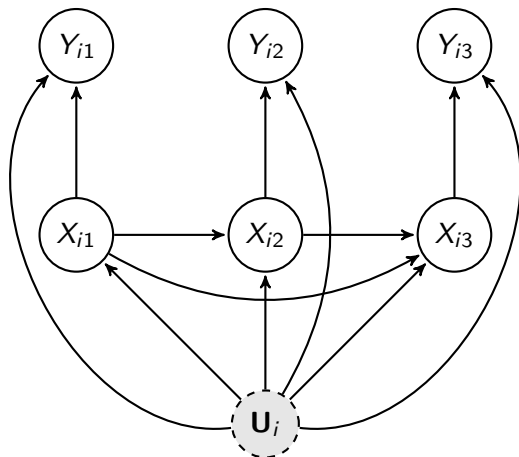
Assumptions:

- 1 No unobserved time-varying confounders
- 2 Past outcomes do not directly affect current outcome
- 3 Past outcomes do not directly affect current treatment

Directed Acyclic Graph (DAG)

Non-parametric identification assumptions for fixed effects:

$$Y_{it} = g(X_{it}, \mathbf{U}_i, \epsilon_{it}) \quad \text{and} \quad \epsilon_{it} \perp\!\!\!\perp \{\mathbf{X}_i, \mathbf{U}_i\}$$



Assumptions:

- 1 No unobserved time-varying confounders
- 2 Past outcomes do not directly affect current outcome
- 3 Past outcomes do not directly affect current treatment
- 4 Past treatments do not directly affect current outcome

*the result implies that the **counterfactual** outcome for a treated observation in a given time period is estimated using the **observed outcomes of different time periods of the same unit**. Since such a comparison is **valid only when no causal dynamics exist**, this finding underscores the important limitation of linear regression models with unit fixed effects.*

- Imai and Kim (2019)

What Ideal Experiment Corresponds to the Fixed Effects Model?

- Experiment that satisfies the model assumptions:

What Ideal Experiment Corresponds to the Fixed Effects Model?

- Experiment that satisfies the model assumptions:
 - 1 randomize X_{i1} given $\mathbf{U}_i$

What Ideal Experiment Corresponds to the Fixed Effects Model?

- Experiment that satisfies the model assumptions:
 - 1 randomize X_{i1} given $\mathbf{U}_i$
 - 2 randomize X_{i2} given X_{i1} and $\mathbf{U}_i$

What Ideal Experiment Corresponds to the Fixed Effects Model?

- Experiment that satisfies the model assumptions:
 - 1 randomize X_{i1} given $\mathbf{U}_i$
 - 2 randomize X_{i2} given X_{i1} and $\mathbf{U}_i$
 - 3 randomize X_{i3} given X_{i2} , X_{i1} , and $\mathbf{U}_i$
 - 4 and so on

What Ideal Experiment Corresponds to the Fixed Effects Model?

- Experiment that satisfies the model assumptions:
 - ① randomize X_{i1} given $\mathbf{U}_i$
 - ② randomize X_{i2} given X_{i1} and $\mathbf{U}_i$
 - ③ randomize X_{i3} given X_{i2} , X_{i1} , and $\mathbf{U}_i$
 - ④ and so on
- Experiment that does not satisfy the model assumptions:

What Ideal Experiment Corresponds to the Fixed Effects Model?

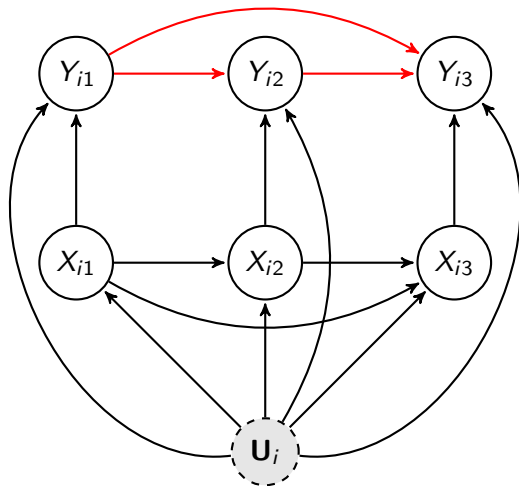
- Experiment that satisfies the model assumptions:
 - ① randomize X_{i1} given $\mathbf{U}_i$
 - ② randomize X_{i2} given X_{i1} and $\mathbf{U}_i$
 - ③ randomize X_{i3} given X_{i2} , X_{i1} , and $\mathbf{U}_i$
 - ④ and so on
- Experiment that does not satisfy the model assumptions:
 - ① randomize X_{i1}
 - ② randomize X_{i2} given X_{i1} and Y_{i1}
 - ③ randomize X_{i3} given X_{i2} , X_{i1} , Y_{i1} , and Y_{i2}
 - ④ and so on

What Ideal Experiment Corresponds to the Fixed Effects Model?

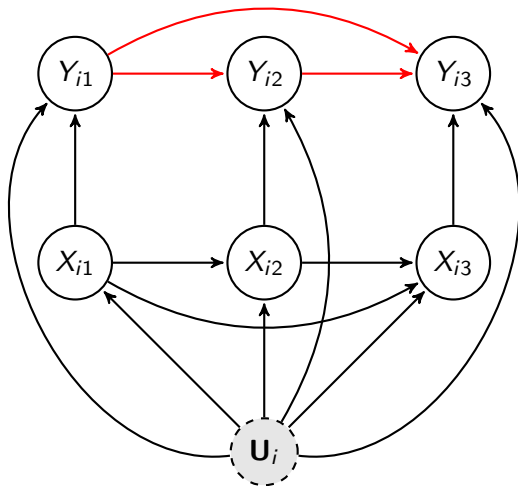
- Experiment that satisfies the model assumptions:
 - ① randomize X_{i1} given $\mathbf{U}_i$
 - ② randomize X_{i2} given X_{i1} and $\mathbf{U}_i$
 - ③ randomize X_{i3} given X_{i2} , X_{i1} , and $\mathbf{U}_i$
 - ④ and so on
- Experiment that does not satisfy the model assumptions:
 - ① randomize X_{i1}
 - ② randomize X_{i2} given X_{i1} and Y_{i1}
 - ③ randomize X_{i3} given X_{i2} , X_{i1} , Y_{i1} , and Y_{i2}
 - ④ and so on
- Now let's consider each assumption in turn.

Past Outcomes Don't Directly Affect Current Outcome (A2)

- Strict exogeneity still holds.

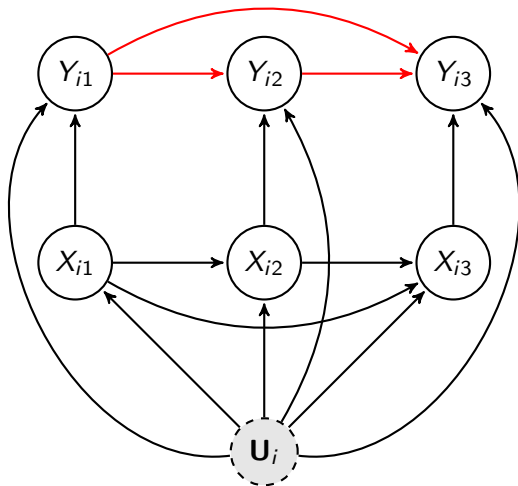


Past Outcomes Don't Directly Affect Current Outcome (A2)



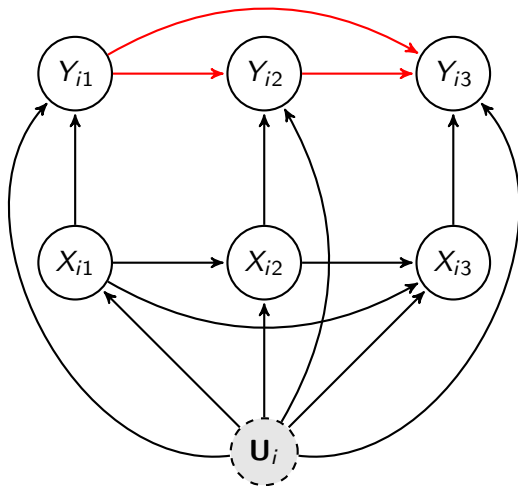
- Strict exogeneity still holds.
- Past outcomes do not confound $X_{it} \rightarrow Y_{it}$ given U_i .

Past Outcomes Don't Directly Affect Current Outcome (A2)



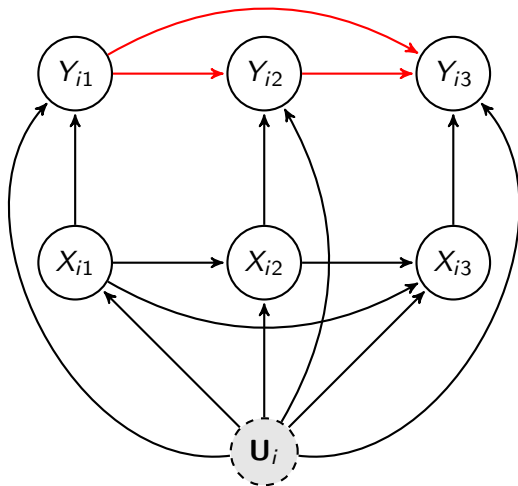
- Strict exogeneity still holds.
- Past outcomes do not confound $X_{it} \rightarrow Y_{it}$ given U_i .
- No need to adjust for past outcomes.

Past Outcomes Don't Directly Affect Current Outcome (A2)



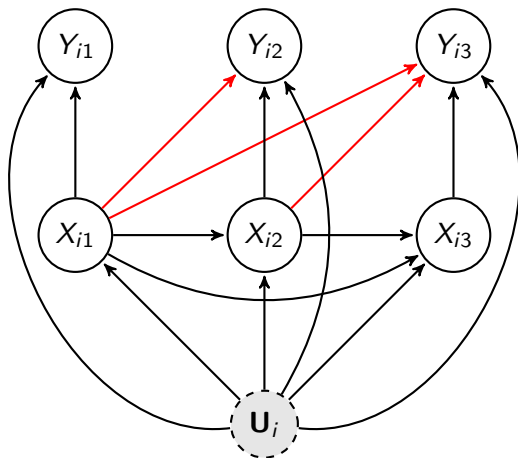
- Strict exogeneity still holds.
- Past outcomes do not confound $X_{it} \rightarrow Y_{it}$ given U_i .
- No need to adjust for past outcomes.
- Should use cluster robust standard errors for inference.

Past Outcomes Don't Directly Affect Current Outcome (A2)



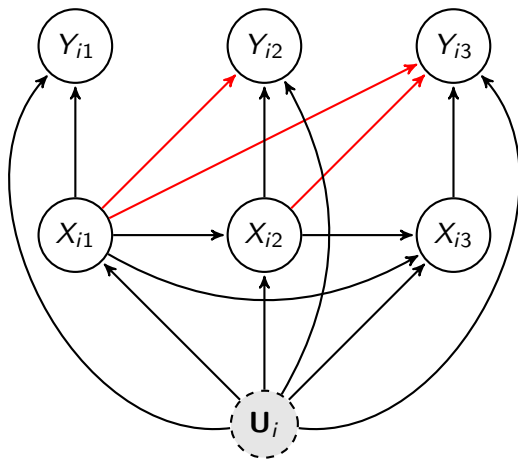
- Strict exogeneity still holds.
- Past outcomes do not confound $X_{it} \rightarrow Y_{it}$ given U_i .
- No need to adjust for past outcomes.
- Should use cluster robust standard errors for inference.
- Conclusion: **The assumption can be relaxed**

Past Treatments Don't Directly Affect Current Outcome (A4)



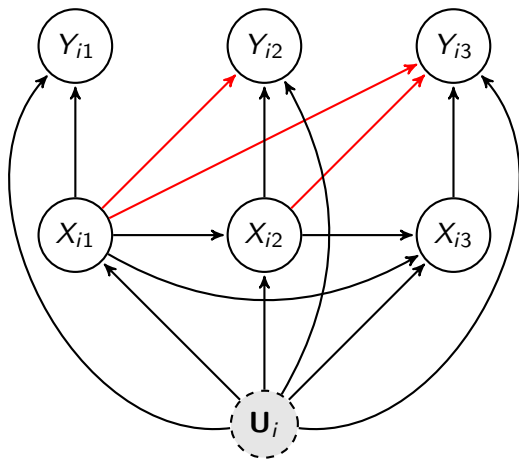
- Need to adjust for past treatments

Past Treatments Don't Directly Affect Current Outcome (A4)



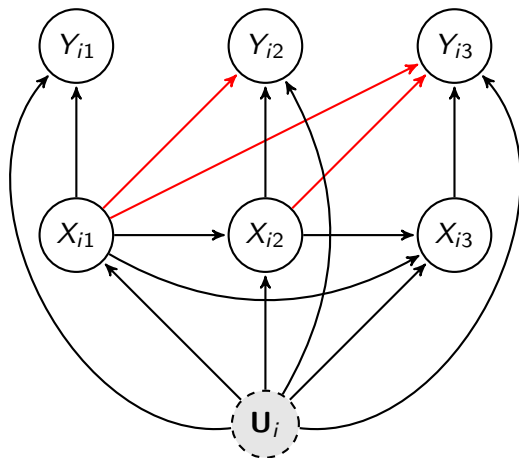
- Need to adjust for past treatments
- Strict exogeneity holds given past treatments and U_i

Past Treatments Don't Directly Affect Current Outcome (A4)



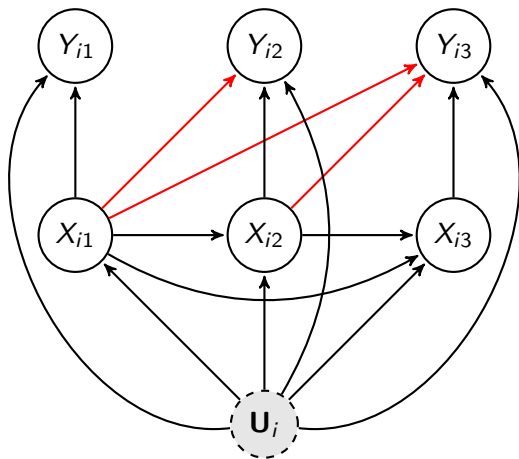
- Need to adjust for past treatments
- Strict exogeneity holds given past treatments and U_i
- Impossible to adjust for an entire treatment history and U_i at the same time

Past Treatments Don't Directly Affect Current Outcome (A4)



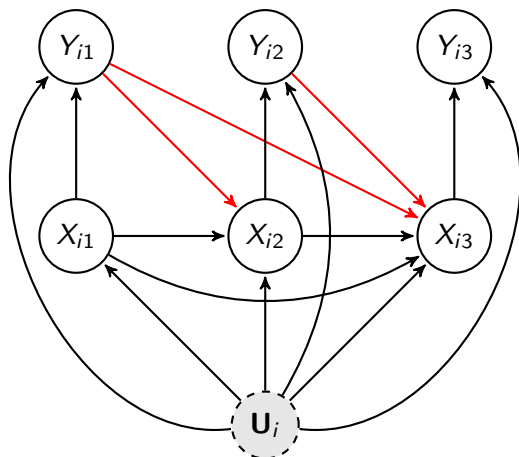
- Need to adjust for past treatments
- Strict exogeneity holds given past treatments and U_i
- Impossible to adjust for an entire treatment history and U_i at the same time
- Adjust for a small number of past treatments $\rightsquigarrow$ often arbitrary

Past Treatments Don't Directly Affect Current Outcome (A4)



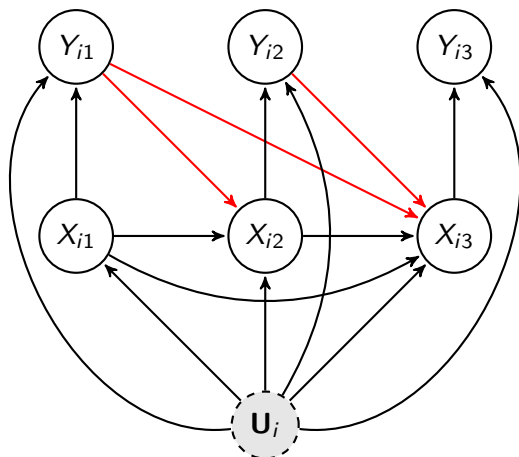
- Need to adjust for past treatments
- Strict exogeneity holds given past treatments and U_i
- Impossible to adjust for an entire treatment history and U_i at the same time
- Adjust for a small number of past treatments $\rightsquigarrow$ often arbitrary
- Conclusion: The assumption can be partially relaxed

Past Outcomes Don't Directly Affect Current Treatment (A3)

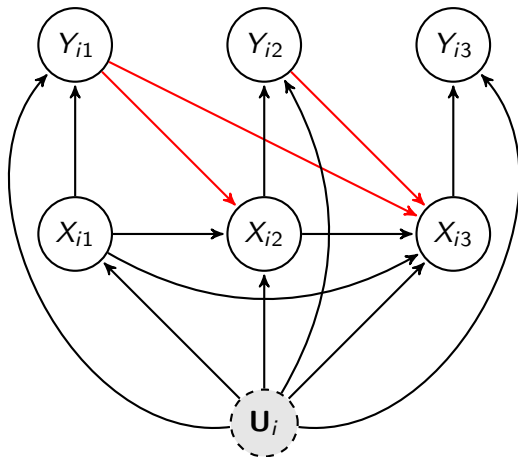


Past Outcomes Don't Directly Affect Current Treatment (A3)

- Correlation between error term and future treatments

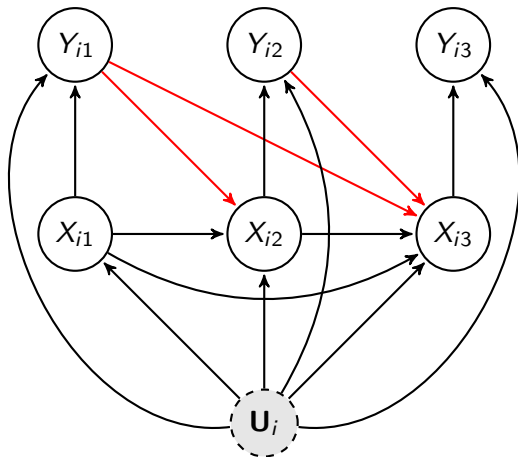


Past Outcomes Don't Directly Affect Current Treatment (A3)



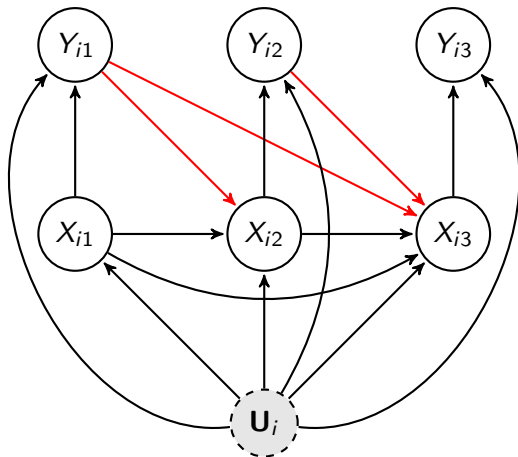
- Correlation between error term and future treatments
- Violation of strict exogeneity

Past Outcomes Don't Directly Affect Current Treatment (A3)



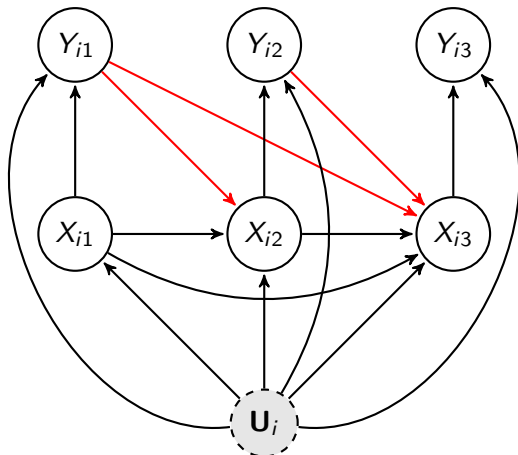
- Correlation between error term and future treatments
- Violation of strict exogeneity
- No adjustment is sufficient

Past Outcomes Don't Directly Affect Current Treatment (A3)



- Correlation between error term and future treatments
- Violation of strict exogeneity
- No adjustment is sufficient
- Implication: No dynamic causal relationships between treatment and outcome variables

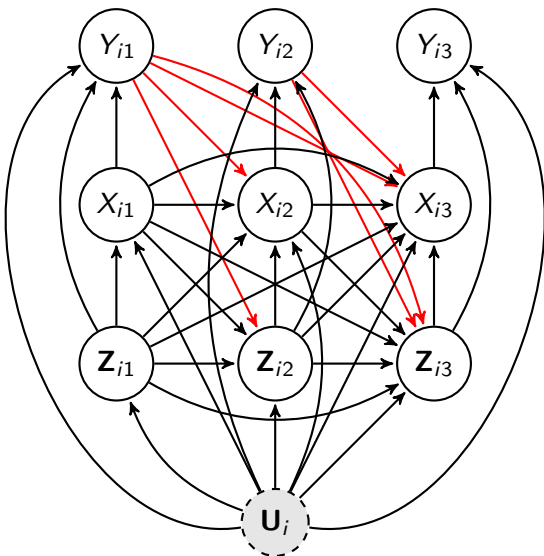
Past Outcomes Don't Directly Affect Current Treatment (A3)



- Correlation between error term and future treatments
- Violation of strict exogeneity
- No adjustment is sufficient
- Implication: No dynamic causal relationships between treatment and outcome variables
- Conclusion: **The assumption cannot be relaxed**

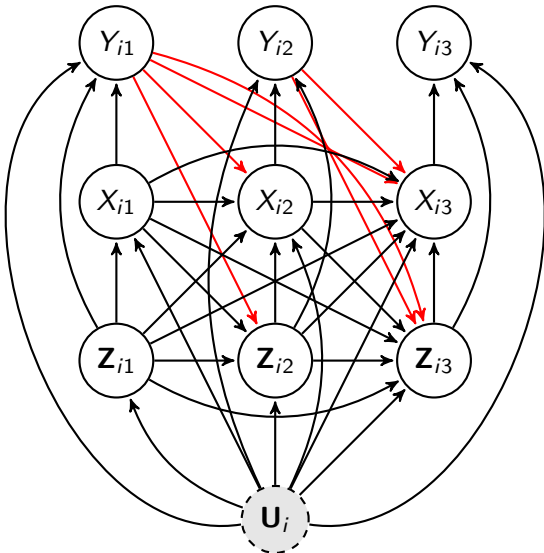
Can't We Just Adjust for Time-Varying Confounders?

Can't We Just Adjust for Time-Varying Confounders?



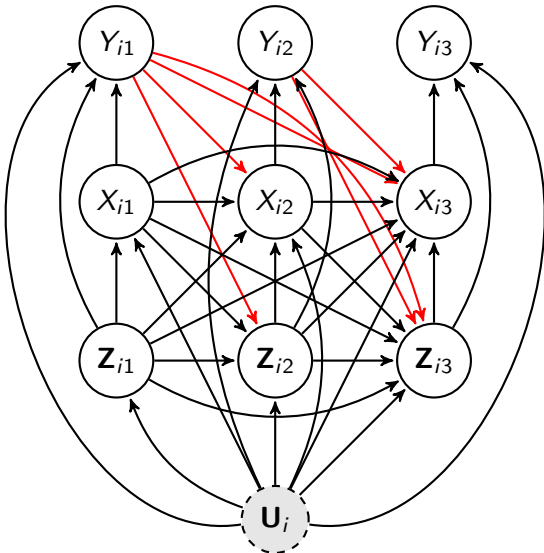
- $Y_{it} = \alpha_i + \beta X_{it} + \gamma^\top \mathbf{Z}_{it} + \epsilon_{it}$

Can't We Just Adjust for Time-Varying Confounders?



- $Y_{it} = \alpha_i + \beta X_{it} + \gamma^\top \mathbf{Z}_{it} + \epsilon_{it}$
- past outcomes cannot directly affect current treatment

Can't We Just Adjust for Time-Varying Confounders?



- $Y_{it} = \alpha_i + \beta X_{it} + \gamma^\top \mathbf{Z}_{it} + \epsilon_{it}$
- past outcomes cannot directly affect current treatment
- past outcomes cannot *indirectly* affect current treatment through $\mathbf{Z}_{it}$

But What If I Have Causal Dynamics?

But What If I Have Causal Dynamics?

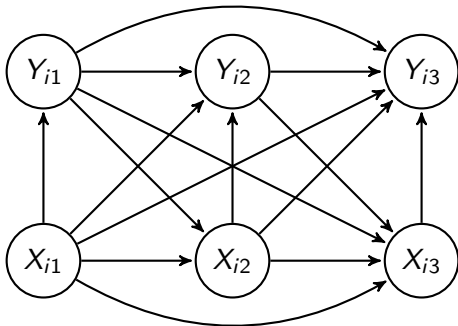
Alternative: **Marginal Structural Models** (Robins, Hernán and Brumback, 2000)

But What If I Have Causal Dynamics?

Alternative: **Marginal Structural Models** (Robins, Hernán and Brumback, 2000) — see Blackwell 2013 and Blackwell and Glynn 2018 for accessible introductions.

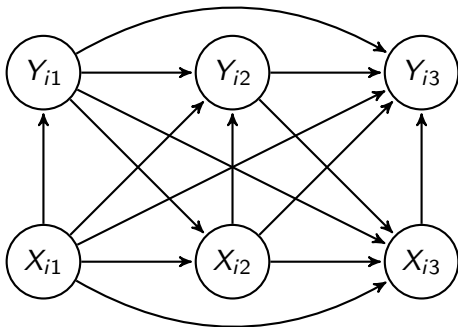
But What If I Have Causal Dynamics?

Alternative: **Marginal Structural Models** (Robins, Hernán and Brumback, 2000) — see Blackwell 2013 and Blackwell and Glynn 2018 for accessible introductions.



But What If I Have Causal Dynamics?

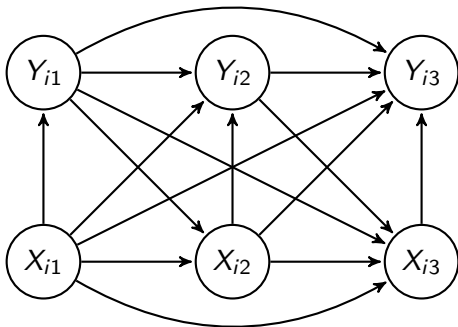
Alternative: **Marginal Structural Models** (Robins, Hernán and Brumback, 2000) — see Blackwell 2013 and Blackwell and Glynn 2018 for accessible introductions.



- Absence of unobserved time-invariant confounders $\mathbf{U}_i$

But What If I Have Causal Dynamics?

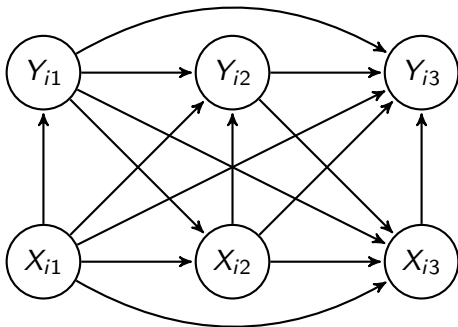
Alternative: **Marginal Structural Models** (Robins, Hernán and Brumback, 2000) — see Blackwell 2013 and Blackwell and Glynn 2018 for accessible introductions.



- Absence of unobserved time-invariant confounders $\mathbf{U}_i$
- past treatments can directly affect current outcome

But What If I Have Causal Dynamics?

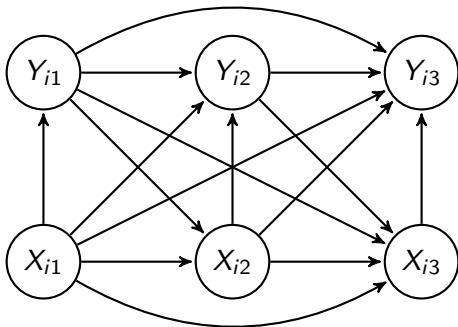
Alternative: **Marginal Structural Models** (Robins, Hernán and Brumback, 2000) — see Blackwell 2013 and Blackwell and Glynn 2018 for accessible introductions.



- Absence of unobserved time-invariant confounders $\mathbf{U}_i$
- past treatments can directly affect current outcome
- past outcomes can directly affect current treatment

But What If I Have Causal Dynamics?

Alternative: **Marginal Structural Models** (Robins, Hernán and Brumback, 2000) — see Blackwell 2013 and Blackwell and Glynn 2018 for accessible introductions.

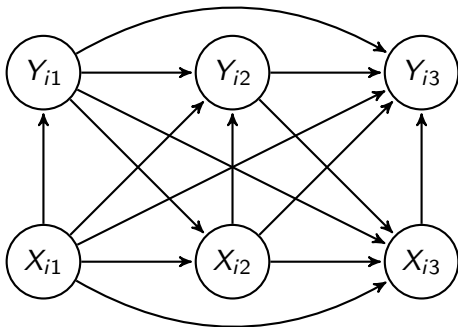


- Absence of unobserved time-invariant confounders $\mathbf{U}_i$
- past treatments can directly affect current outcome
- past outcomes can directly affect current treatment

- Comparison across units within the same time rather than across different time periods within the same unit

But What If I Have Causal Dynamics?

Alternative: **Marginal Structural Models** (Robins, Hernán and Brumback, 2000) — see Blackwell 2013 and Blackwell and Glynn 2018 for accessible introductions.

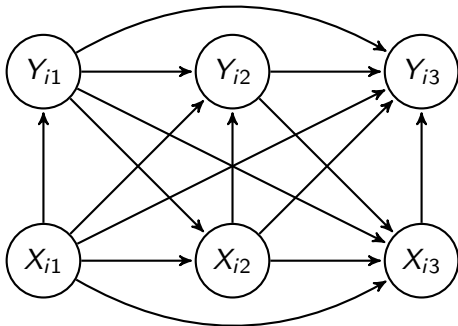


- Absence of unobserved time-invariant confounders $\mathbf{U}_i$
- past treatments can directly affect current outcome
- past outcomes can directly affect current treatment

- Comparison across units within the same time rather than across different time periods within the same unit
- Can identify the average effect of an entire treatment sequence

But What If I Have Causal Dynamics?

Alternative: **Marginal Structural Models** (Robins, Hernán and Brumback, 2000) — see Blackwell 2013 and Blackwell and Glynn 2018 for accessible introductions.



- Absence of unobserved time-invariant confounders $\mathbf{U}_i$
- past treatments can directly affect current outcome
- past outcomes can directly affect current treatment

- Comparison across units within the same time rather than across different time periods within the same unit
- Can identify the average effect of an entire treatment sequence
- Trade-off $\rightsquigarrow$ no free lunch

Conclusions and Nonparametric Estimation

Conclusions and Nonparametric Estimation

- Imai and Kim (2019) offer a matching framework for fixed effects models which exploits an equivalence to weighted unit fixed effects estimators (see `wfe` package in R as well).

Conclusions and Nonparametric Estimation

- Imai and Kim (2019) offer a matching framework for fixed effects models which exploits an equivalence to weighted unit fixed effects estimators (see `wfe` package in R as well).
- The paper clarifies assumptions for fixed effects and first difference estimators.

Conclusions and Nonparametric Estimation

- Imai and Kim (2019) offer a matching framework for fixed effects models which exploits an equivalence to weighted unit fixed effects estimators (see `wfe` package in R as well).
- The paper clarifies assumptions for fixed effects and first difference estimators.
- Follow-up papers Imai and Kim (2021) and Imai, Kim and Wang (2022+) extends to two-way fixed effects estimator.

Conclusions and Nonparametric Estimation

- Imai and Kim (2019) offer a matching framework for fixed effects models which exploits an equivalence to weighted unit fixed effects estimators (see `wfe` package in R as well).
- The paper clarifies assumptions for fixed effects and first difference estimators.
- Follow-up papers Imai and Kim (2021) and Imai, Kim and Wang (2022+) extends to two-way fixed effects estimator.
- Tradeoff:
 - 1) unobserved time-invariant confounders $\rightsquigarrow$ fixed effects

Conclusions and Nonparametric Estimation

- Imai and Kim (2019) offer a matching framework for fixed effects models which exploits an equivalence to weighted unit fixed effects estimators (see `wfe` package in R as well).
- The paper clarifies assumptions for fixed effects and first difference estimators.
- Follow-up papers Imai and Kim (2021) and Imai, Kim and Wang (2022+) extends to two-way fixed effects estimator.
- Tradeoff:
 - 1) unobserved time-invariant confounders $\rightsquigarrow$ fixed effects
 - 2) causal dynamics between treatment and outcome $\rightsquigarrow$ selection-on-observables

Summary Table (Imai and Kim 2019)

TABLE 1 Identification Assumptions of Various Estimators

	Linearity	Time-Invariant Unobservables	Past Outcomes Affect Current Treatment	Past Treatments Affect Current Outcome
$Y_{it} = \alpha_i + \beta X_{it} + \epsilon_{it}$	Yes	Allowed	Not allowed	Not allowed
$Y_{it} = \alpha_i + \beta X_{it} + \rho Y_{i,t-1} + \epsilon_{it}$	Yes	Allowed	Allowed	Not allowed
$Y_{it} = \alpha_i + \beta_1 X_{it} + \beta_2 X_{i,t-1} + \epsilon_{it}$	Yes	Allowed	Not allowed	Allowed
$Y_{it} = \alpha_i + \beta_1 X_{it} + \beta_2 X_{i,t-1} + \rho Y_{i,t-1} + \epsilon_{it}$	Yes	Allowed	Partially allowed	Partially allowed
Marginal structural models	No	Not allowed	Allowed	Allowed

What to read next?

- Morgan and Winship Chapter 11 Repeated Observations and the Estimation of Causal Effects
- Imai and Kim (2019) "When Should We Use Unit Fixed Effects Regression Models for Causal Inference with Longitudinal Data?" *American Journal of Political Science*,
<http://dx.doi.org/10.1111/ajps.12417>
- Liu, Wang and Xu "A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data" *American Journal of Political Science*.

- 1 Difference-in-Differences
- 2 Synthetic Controls
- 3 Fixed Effects
- 4 Non-parametric Identification and Fixed Effects
- 5 Wrap-Up
 - Questions
 - Concluding Thoughts for the Course

- 1 Difference-in-Differences
- 2 Synthetic Controls
- 3 Fixed Effects
- 4 Non-parametric Identification and Fixed Effects
- 5 **Wrap-Up**
 - Questions
 - Concluding Thoughts for the Course

Q: What conditions do we need to infer causality?

Q: What conditions do we need to infer causality?

A: A clear estimand, an identification strategy and an estimation strategy.

Identification Strategies in This Class

Identification Strategies in This Class

- Experiments (ignorability via randomization)

Identification Strategies in This Class

- Experiments (ignorability via randomization)
- Selection on Observables (conditional ignorability)

Identification Strategies in This Class

- Experiments (ignorability via randomization)
- Selection on Observables (conditional ignorability)
- Natural Experiments (ignorability via quasi-randomization)

Identification Strategies in This Class

- Experiments (ignorability via randomization)
- Selection on Observables (conditional ignorability)
- Natural Experiments (ignorability via quasi-randomization)
- Interrupted Time Series (forecasting)

Identification Strategies in This Class

- Experiments (ignorability via randomization)
- Selection on Observables (conditional ignorability)
- Natural Experiments (ignorability via quasi-randomization)
- Interrupted Time Series (forecasting)
- Instrumental Variables (instrument + exclusion restriction)

Identification Strategies in This Class

- Experiments (ignorability via randomization)
- Selection on Observables (conditional ignorability)
- Natural Experiments (ignorability via quasi-randomization)
- Interrupted Time Series (forecasting)
- Instrumental Variables (instrument + exclusion restriction)
- Regression Discontinuity (continuity assumption)

Identification Strategies in This Class

- Experiments (ignorability via randomization)
- Selection on Observables (conditional ignorability)
- Natural Experiments (ignorability via quasi-randomization)
- Interrupted Time Series (forecasting)
- Instrumental Variables (instrument + exclusion restriction)
- Regression Discontinuity (continuity assumption)
- Difference-in-Differences / Synthetic Controls (parallel trends)

Identification Strategies in This Class

- Experiments (ignorability via randomization)
- Selection on Observables (conditional ignorability)
- Natural Experiments (ignorability via quasi-randomization)
- Interrupted Time Series (forecasting)
- Instrumental Variables (instrument + exclusion restriction)
- Regression Discontinuity (continuity assumption)
- Difference-in-Differences / Synthetic Controls (parallel trends)
- Fixed Effects (time-invariant unobserved heterogeneity, strict ignorability)

Identification Strategies in This Class

- Experiments (ignorability via randomization)
- Selection on Observables (conditional ignorability)
- Natural Experiments (ignorability via quasi-randomization)
- Interrupted Time Series (forecasting)
- Instrumental Variables (instrument + exclusion restriction)
- Regression Discontinuity (continuity assumption)
- Difference-in-Differences / Synthetic Controls (parallel trends)
- Fixed Effects (time-invariant unobserved heterogeneity, strict ignorability)

Almost everything assumes: consistency/SUTVA (no interference between units, variation in the treatment is irrelevant) and positivity (there is some chance of all getting treatment)

Some Estimation Strategies

Some Estimation Strategies

- Stratification

Some Estimation Strategies

- Stratification
- Regression (and relatives)

Some Estimation Strategies

- Stratification
- Regression (and relatives)
- Matching (not covered)
- Weighting (not covered)

Q: What perspective do you want us to take away from the class?

Q: What perspective do you want us to take away from the class?

- State your **target estimand** before getting into estimation.

Q: What perspective do you want us to take away from the class?

- State your **target estimand** before getting into estimation.
- Make explicit how your **evidence** relates to your **argument** both in terms of what the unit-specific quantity it is and which units it applies to.

Q: What perspective do you want us to take away from the class?

- State your **target estimand** before getting into estimation.
- Make explicit how your **evidence** relates to your **argument** both in terms of what the unit-specific quantity it is and which units it applies to.
- Know what you can **defend with data** and what you have to **argue for by assumption** and don't confuse the two!

Q: What perspective do you want us to take away from the class?

- State your **target estimand** before getting into estimation.
- Make explicit how your **evidence** relates to your **argument** both in terms of what the unit-specific quantity it is and which units it applies to.
- Know what you can **defend with data** and what you have to **argue for by assumption** and don't confuse the two!
- Be at least a bit skeptical of **model-based** inference: think about the non-parametric object you are approximating.

Q: What perspective do you want us to take away from the class?

- State your **target estimand** before getting into estimation.
- Make explicit how your **evidence** relates to your **argument** both in terms of what the unit-specific quantity it is and which units it applies to.
- Know what you can **defend with data** and what you have to **argue for by assumption** and don't confuse the two!
- Be at least a bit skeptical of **model-based** inference: think about the non-parametric object you are approximating.
- You can look under the hood for why/under what conditions things work, rules of thumb or textbook guidance do not universally: there is generally **no single right answer** to how you should analyze some data.

Q: What perspective do you want us to take away from the class?

- State your **target estimand** before getting into estimation.
- Make explicit how your **evidence** relates to your **argument** both in terms of what the unit-specific quantity it is and which units it applies to.
- Know what you can **defend with data** and what you have to **argue for by assumption** and don't confuse the two!
- Be at least a bit skeptical of **model-based** inference: think about the non-parametric object you are approximating.
- You can look under the hood for why/under what conditions things work, rules of thumb or textbook guidance do not universally: there is generally **no single right answer** to how you should analyze some data.

See also 'What is Your Estimand?' with Lundberg and Johnson *ASR*, 2021.

Q: What can make standard errors larger or smaller?

Q: What can make standard errors larger or smaller?

A: Let's consider the anatomy of a standard error.

Anatomy of the Standard Error

Imagine we have a regression of Y on a variable of interest X and a vector of other variables $\mathbf{Z}$. Let's consider the homoskedastic case for simplicity:

$$\widehat{\text{Var}}(\hat{\beta}_X) = \frac{\frac{1}{(n-k-1)} \sum_{i=1}^n \hat{u}_i^2}{(1 - R_{X \sim \mathbf{Z}}^2) \sum_{i=1}^n (X_i - \bar{X})^2}$$

Anatomy of the Standard Error

Imagine we have a regression of Y on a variable of interest X and a vector of other variables $\mathbf{Z}$. Let's consider the homoskedastic case for simplicity:

$$\widehat{\text{Var}}(\widehat{\beta}_X) = \frac{\frac{1}{(n-k-1)} \sum_{i=1}^n \hat{u}_i^2}{(1 - R_{X \sim \mathbf{Z}}^2) \sum_{i=1}^n (X_i - \bar{X})^2}$$

- the numerator is our estimator for σ_u^2 the unknown error variance. It is formed by the degrees of freedom correction times the sum of the squared residuals.

Anatomy of the Standard Error

Imagine we have a regression of Y on a variable of interest X and a vector of other variables $\mathbf{Z}$. Let's consider the homoskedastic case for simplicity:

$$\widehat{\text{Var}}(\widehat{\beta}_X) = \frac{\frac{1}{(n-k-1)} \sum_{i=1}^n \hat{u}_i^2}{(1 - R_{X \sim \mathbf{Z}}^2) \sum_{i=1}^n (X_i - \bar{X})^2}$$

- the numerator is our estimator for σ_u^2 the unknown error variance. It is formed by the degrees of freedom correction times the sum of the squared residuals.
- the denominator includes one minus the R^2 from the regression of X_i on $\mathbf{Z}_i$

Anatomy of the Standard Error

Imagine we have a regression of Y on a variable of interest X and a vector of other variables $\mathbf{Z}$. Let's consider the homoskedastic case for simplicity:

$$\widehat{\text{Var}}(\widehat{\beta}_X) = \frac{\frac{1}{(n-k-1)} \sum_{i=1}^n \hat{u}_i^2}{(1 - R_{X \sim \mathbf{Z}}^2) \sum_{i=1}^n (X_i - \bar{X})^2}$$

- the numerator is our estimator for σ_u^2 the unknown error variance. It is formed by the degrees of freedom correction times the sum of the squared residuals.
- the denominator includes one minus the R^2 from the regression of X_i on $\mathbf{Z}_i$
- we complete the denominator by multiplying a measure of the variation in X_i , the sum of squared deviations from the mean.

Anatomy of the Standard Error

Imagine we have a regression of Y on a variable of interest X and a vector of other variables $\mathbf{Z}$. Let's consider the homoskedastic case for simplicity:

$$\widehat{\text{Var}}(\widehat{\beta}_X) = \frac{\frac{1}{(n-k-1)} \sum_{i=1}^n \hat{u}_i^2}{(1 - R_{X \sim \mathbf{Z}}^2) \sum_{i=1}^n (X_i - \bar{X})^2}$$

- the numerator is our estimator for σ_u^2 the unknown error variance. It is formed by the degrees of freedom correction times the sum of the squared residuals.
- the denominator includes one minus the R^2 from the regression of X_i on $\mathbf{Z}_i$
- we complete the denominator by multiplying a measure of the variation in X_i , the sum of squared deviations from the mean.

$$\widehat{\text{SE}}(\widehat{\beta}_X) = \sqrt{\widehat{\text{Var}}(\widehat{\beta}_X)}$$

Q: When conducting an experiment, should we avoid OLS and always go for difference in means?

Q: When conducting an experiment, should we avoid OLS and always go for difference in means?

A: Regression adjustment of experiments can be helpful for improving precision. We don't need it for confounding, but it can improve our standard errors to adjust for pre-treatment covariates that are highly predictive of the output. If done correctly and in moderate-to-large samples, this can dramatically improve your standard errors. Even better though is blocking which is adjustment by design.

Q: When conducting an experiment, should we avoid OLS and always go for difference in means?

A: Regression adjustment of experiments can be helpful for improving precision. We don't need it for confounding, but it can improve our standard errors to adjust for pre-treatment covariates that are highly predictive of the output. If done correctly and in moderate-to-large samples, this can dramatically improve your standard errors. Even better though is blocking which is adjustment by design.

Further Reading:

- Lin, W., 2013. 'Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique.' *The Annals of Applied Statistics*
- Athey, S. and Imbens, G.W., 2017. 'The Econometrics of Randomized Experiments.' In *Handbook of Economic Field Experiments* (Vol. 1, pp. 73-140).
- Egap Methods Guide: 10 things you need to know about covariate adjustment.
<https://egap.org/methods-guides/10-things-know-about-covariate-adjustment>

Q: What are your favorite resources for learning tricky concepts?

Q: What are your favorite resources for learning tricky concepts?

I've used the following procedure many times:

Q: What are your favorite resources for learning tricky concepts?

I've used the following procedure many times:

- 1 Identify approx. the best textbook (often can do this via syllabi hunting)

Q: What are your favorite resources for learning tricky concepts?

I've used the following procedure many times:

- 1 Identify approx. the best textbook (often can do this via syllabi hunting)
- 2 Read the relevant textbook material

Q: What are your favorite resources for learning tricky concepts?

I've used the following procedure many times:

- 1 Identify approx. the best textbook (often can do this via syllabi hunting)
- 2 Read the relevant textbook material
- 3 Derive the equations/math

Q: What are your favorite resources for learning tricky concepts?

I've used the following procedure many times:

- 1 Identify approx. the best textbook (often can do this via syllabi hunting)
- 2 Read the relevant textbook material
- 3 Derive the equations/math
- 4 Try to explain it to someone else

Q: In the real world of research, when you have your data,
how do you know which method to use?

Q: In the real world of research, when you have your data,
how do you know which method to use?

A: Let's chat.

Q: What would you have covered with more time?

Q: What would you have covered with more time?

1 Missing Data

Q: What would you have covered with more time?

- 1 Missing Data
- 2 Sensitivity Analysis

Q: What would you have covered with more time?

- 1 Missing Data
- 2 Sensitivity Analysis
- 3 Mediation (and Front Door Adjustment)

Q: What would you have covered with more time?

- 1 Missing Data
- 2 Sensitivity Analysis
- 3 Mediation (and Front Door Adjustment)
- 4 Bounds and Partial Identification

Q: What would you have covered with more time?

- 1 Missing Data
- 2 Sensitivity Analysis
- 3 Mediation (and Front Door Adjustment)
- 4 Bounds and Partial Identification
- 5 Weighting/Matching Approaches

Q: What would you have covered with more time?

- 1 Missing Data
- 2 Sensitivity Analysis
- 3 Mediation (and Front Door Adjustment)
- 4 Bounds and Partial Identification
- 5 Weighting/Matching Approaches
- 6 Quantile Regression

Q: What would you have covered with more time?

- 1 Missing Data
- 2 Sensitivity Analysis
- 3 Mediation (and Front Door Adjustment)
- 4 Bounds and Partial Identification
- 5 Weighting/Matching Approaches
- 6 Quantile Regression
- 7 Machine Learning and Generalizations of Regression

Q: What would you have covered with more time?

- 1 Missing Data
- 2 Sensitivity Analysis
- 3 Mediation (and Front Door Adjustment)
- 4 Bounds and Partial Identification
- 5 Weighting/Matching Approaches
- 6 Quantile Regression
- 7 Machine Learning and Generalizations of Regression
- 8 On the Ground Decision Making

- 1 Difference-in-Differences
- 2 Synthetic Controls
- 3 Fixed Effects
- 4 Non-parametric Identification and Fixed Effects
- 5 Wrap-Up
 - Questions
 - Concluding Thoughts for the Course

- 1 Difference-in-Differences
- 2 Synthetic Controls
- 3 Fixed Effects
- 4 Non-parametric Identification and Fixed Effects
- 5 **Wrap-Up**
 - Questions
 - Concluding Thoughts for the Course

Where are you?

Where are you?

You've been given a powerful set of tools



Your New Weapons

Your New Weapons

- Basic probability theory

- ▶ Probability axioms, random variables, marginal and conditional probability, building a probability model
- ▶ Expected value, variances, independence
- ▶ CDF and PDF (discrete and continuous)

Your New Weapons

- Basic probability theory

- ▶ Probability axioms, random variables, marginal and conditional probability, building a probability model
- ▶ Expected value, variances, independence
- ▶ CDF and PDF (discrete and continuous)

- Properties of Estimators

- ▶ Bias, Efficiency, Consistency
- ▶ Central limit theorem

Your New Weapons

- Basic probability theory

- ▶ Probability axioms, random variables, marginal and conditional probability, building a probability model
- ▶ Expected value, variances, independence
- ▶ CDF and PDF (discrete and continuous)

- Properties of Estimators

- ▶ Bias, Efficiency, Consistency
- ▶ Central limit theorem

- Univariate Inference

- ▶ Interval estimation (normal and non-normal Population)
- ▶ Confidence intervals, hypothesis tests, p-values
- ▶ Practical versus statistical significance

Your New Weapons

Your New Weapons

- Simple Regression

- ▶ regression to approximate the conditional expectation function
- ▶ idea of conditioning
- ▶ kernel and loess regressions
- ▶ OLS estimator for bivariate regression
- ▶ Variance decomposition, goodness of fit, interpretation of estimates, transformations

Your New Weapons

• Simple Regression

- ▶ regression to approximate the conditional expectation function
- ▶ idea of conditioning
- ▶ kernel and loess regressions
- ▶ OLS estimator for bivariate regression
- ▶ Variance decomposition, goodness of fit, interpretation of estimates, transformations

• Multiple Regression

- ▶ OLS estimator for multiple regression in matrix notation
- ▶ Regression assumptions (classical and agnostic)
- ▶ Properties: Bias, Efficiency, Consistency
- ▶ Standard errors, testing, p-values, and confidence intervals
- ▶ Polynomials, Interactions, Dummy Variables
- ▶ F-tests

Your New Weapons

Your New Weapons

- Diagnosing and Fixing Regression Problems
 - ▶ Non-linear transformations
 - ▶ Classical diagnostics
 - ▶ QQ-plots
 - ▶ Clustering

Your New Weapons

- Diagnosing and Fixing Regression Problems
 - ▶ Non-linear transformations
 - ▶ Classical diagnostics
 - ▶ QQ-plots
 - ▶ Clustering
- Causal Inference
 - ▶ Frameworks: potential outcomes and DAGs
 - ▶ Measured Confounding
 - ▶ Unmeasured Confounding
 - ▶ Methods for repeated data

Your New Weapons

- Diagnosing and Fixing Regression Problems
 - ▶ Non-linear transformations
 - ▶ Classical diagnostics
 - ▶ QQ-plots
 - ▶ Clustering
- Causal Inference
 - ▶ Frameworks: potential outcomes and DAGs
 - ▶ Measured Confounding
 - ▶ Unmeasured Confounding
 - ▶ Methods for repeated data
- And you learned how to use R: you're not afraid of trying something new!

Using these Tools

Using these Tools

So, Admiral Ackbar, now that you've learned how to run these regressions we can just use them blindly, right?



IT'S A TRAP!



Beyond Linear Regressions

You need more training



Beyond Linear Regressions

Beyond Linear Regressions

There is so much more to learn! Take classes, read books!

Beyond Linear Regressions

There is so much more to learn! Take classes, read books!

- Take SOC 504 or the POL sequence!

Beyond Linear Regressions

There is so much more to learn! Take classes, read books!

- Take SOC 504 or the POL sequence!
- Pursue the SML certificate!

Beyond Linear Regressions

There is so much more to learn! Take classes, read books!

- Take SOC 504 or the POL sequence!
- Pursue the SML certificate!
- Go to workshops and seminars etc.!

Beyond Linear Regressions

There is so much more to learn! Take classes, read books!

- Take SOC 504 or the POL sequence!
- Pursue the SML certificate!
- Go to workshops and seminars etc.!
- Read the suggested books in the syllabus!

Beyond Linear Regressions

There is so much more to learn! Take classes, read books!

- Take SOC 504 or the POL sequence!
- Pursue the SML certificate!
- Go to workshops and seminars etc.!
- Read the suggested books in the syllabus!

Thanks!

Thanks so much for an amazing semester.

