

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

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.
- 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

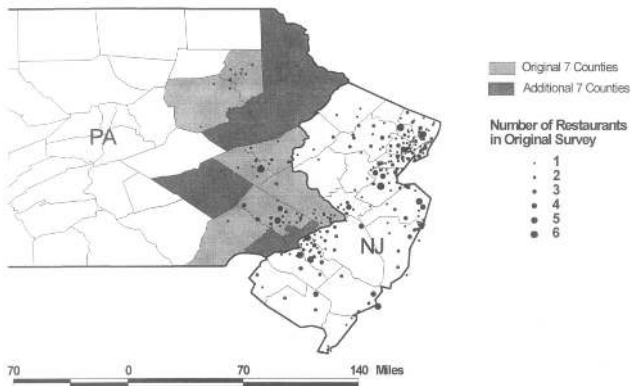


FIGURE 1. AREAS OF NEW JERSEY AND PENNSYLVANIA COVERED BY ORIGINAL SURVEY AND BLS DATA

Source: Card and Krueger 2000

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

[illegible][illegible]

The two Harvard PhD students, Alan Kravitz and his mentor at Princeton University, and professor is a new kind of survey research, sometimes together, sometimes with others, within the Larry Katz, Harvard's right-hand man—have found that having the conventional wisdom on trends top ten like the minimum wage, minimum benefits and immigration can yield a rich harvest of unexpected results. Quite often, they argue, the research ends

If the case were merely academic, Lindholm, teaching the elegant biochemical chemistry of conventional xenotransplants without creating artificial alternatives, might be one factor in your decision. But Lindholm is not just a teacher. Professor Lindholm wants to see his patients. He has been invited to give lectures to groups of students in top overseas journals, and edits the publication's most prestigious medical-ethical journal, *Transplantation*, in which other, similarly top-flight, is a leading way to publish

The New Jersey Test

Trade unions of the continent have been attacked by the Soviet authorities. Chinese workers in the Soviet Far East are being paid less than their Russian colleagues. In the ports of some non-Soviet countries, non-Communist sailors are not being paid by companies who cannot afford to pay them. Professors Card and Kasper, 20 students in and out from New Jersey, 20 the middle of a semester, raised the minimum wage in the state to more than \$10.00 - 10 cents higher than non-unioning Pennsylvania.

They organized all four food restaurants and found that those in New Jersey actually raised 13 workers after the minimum wage.

Some doctors believe that many women in the study may have been taking the two women were specifically told that any fear that pregnancy typically generates with a couple of symptoms is not a real disease, despite I can assure.

"I told her the treatment is safe," says Friedman. "I told her she was going to be fine, so she was looking for pain."

Friedman's findings add the study shows that the fetus doesn't get a worse prognosis as well. "There are lots of concerns about the



Publication supported by David Ford, J.D., and Alan Weiss.

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 challenging the conventional wisdom that the free enterprise system empowers the individual to rise from poverty to wealth. "The conventional economic model ignores these concerns. It assumes people make decisions about where to work, how hard to work or how to invest capital based on purely economic goals."

Professor Krueger studied pay and income in the teaching industry and concluded that the market — not the institutions — set the pace for determining pay for the benefits. When benefits costs are pushed up by either market forces or government institutions, wages tend to go up more slowly. In states with high market competition, accordingly, pay tended to be contemporaneously lower.

Stronger support depresses the wage and raises employment, but the American working poor, the Catholic Card study found, is the least mobile in Florida. In an effort to establish credit records from the bottom of over 100,000 records.

These wage data from FLORIDA, which say that it's always not true, in comparison, and the other place the minimum wage, CARD began to think "It's really, my observations. Florida, not Card alone." "It's always possible that there's some idiosyncratic thing that just occurred."

The 4000 (1978) (1978) (1978)

have increased its capital and allowed their clients to grow more secure. Despite more of the most recent strategies, the national set moves here.

Examiners have the financial situation. Professor Card said: "The big selling point for an efficient Examiners is that you must improve the situation. But it's like a Card, rapidly — our strength is the use of technology. A huge number of people believe that the regulations really work and that you can't have a strategy that can't be."

'Let's Not Assume'

"The joke went to me that if the dance didn't fix the money, that means he's something wrong with your mind. Then there's a younger generation that says 'Let's not assume the answer.'"

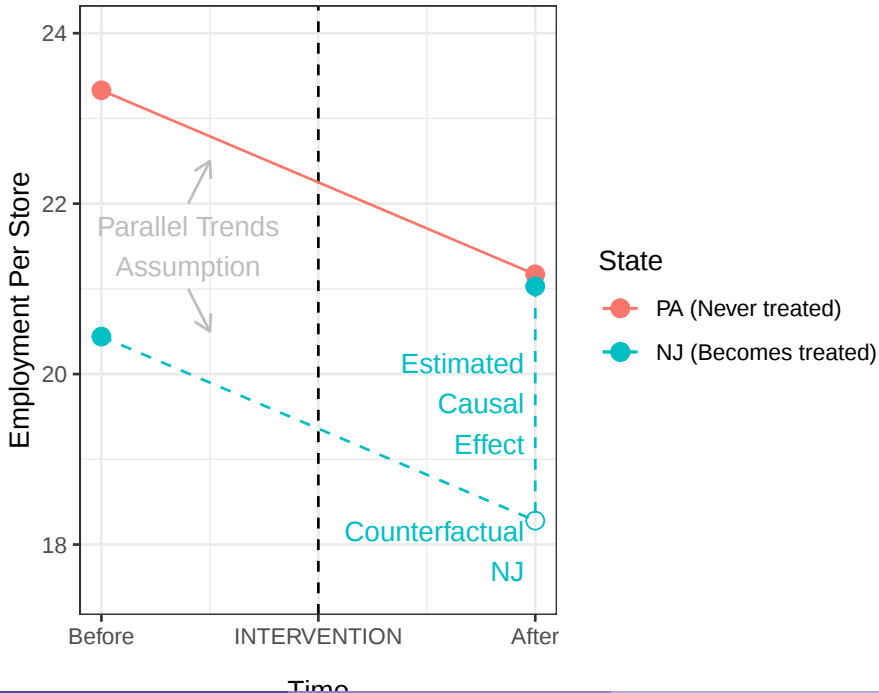
"That statement is in contrast to what we wanted to do with this research," Professor Clark said. "What would a skeptic's response concerning evidence?" Although they always seek large genetic effects, however, they think they are better equipped for following them down. "It's not that they are not interested in working in this field," he said. "They are just a little bit different."

Through continued backtracking, Clark would say is the main problem. "Professor Farmer said, 'Now I think the feeling is that there are a lot of problems. The whole is in fact not as good as it is.'"

That's obvious — to a the greatest degree that some of their governors who have taken pains in the Clinton administration — their second-day detachments. They mean, for example, that they are not intended to tell the world that they are not in the Clinton administration, but that they are in the Clinton administration. Or that making employees who become for all their workers is a bad idea.

Taking sides would conflict with their main mission — making education the labor market ready market the way the world is. I really don't want to see the government's role in education reduced to the point where it is no longer a priority. President George W. Bush's support for education is a good example of what the government should be doing. A lot of people would be interested in that, but they would be wrong.

William Fendelman, who covers the C.E.A. beat for the *Post* (National or anything like that), yesterday blundered into "the most important comment since today."



Example: Card and Krueger (2000)

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

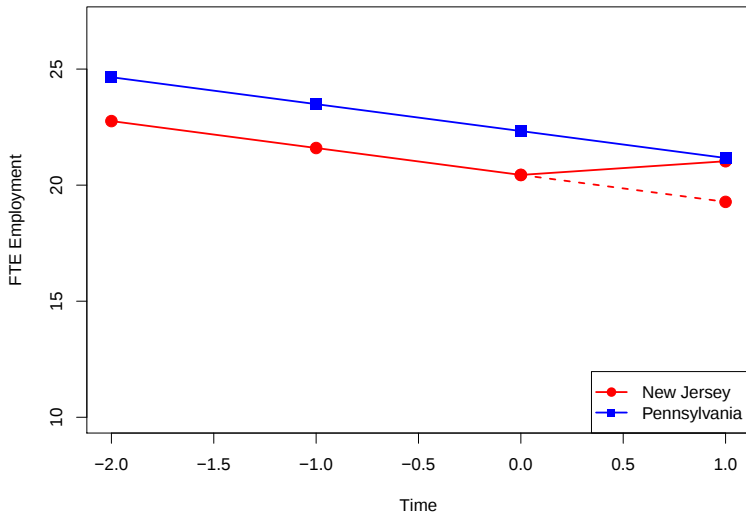
$$\text{employment}_{it} = \beta_0 + \beta_1 \text{minimum wage}_{it} + a_i + u_{it}$$

- 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

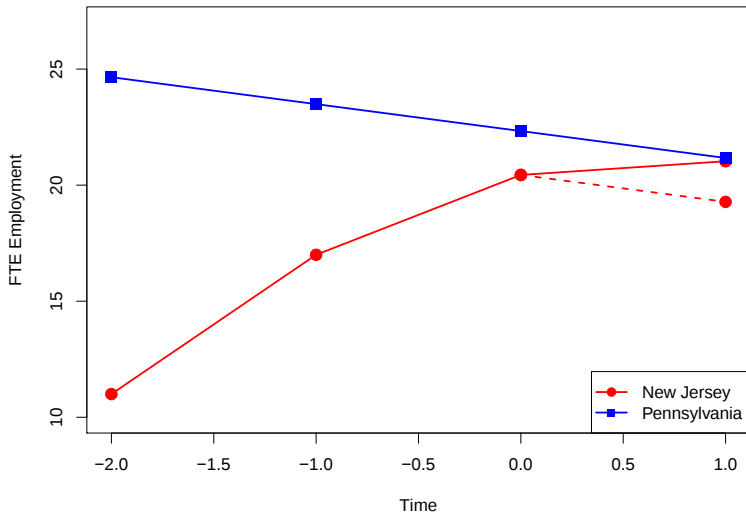
$$\Delta \text{employment}_i = \beta_0 + \beta_1 \text{NJ}_i + \Delta u_i$$

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

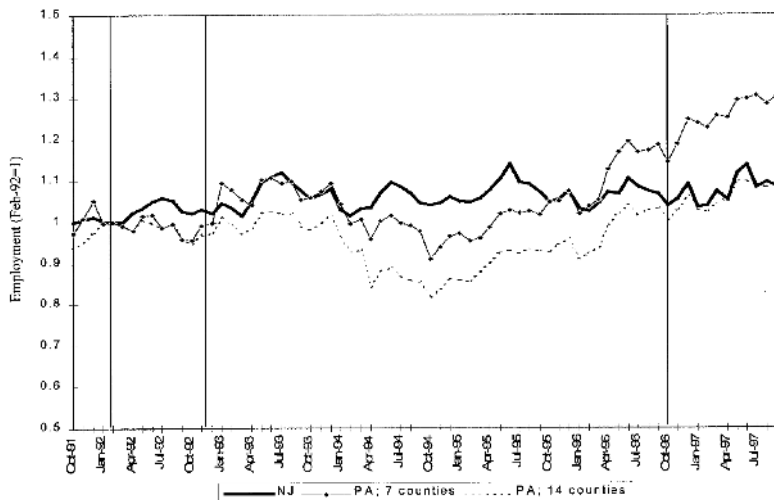
Parallel Trends?



Parallel Trends?



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

- 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

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

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

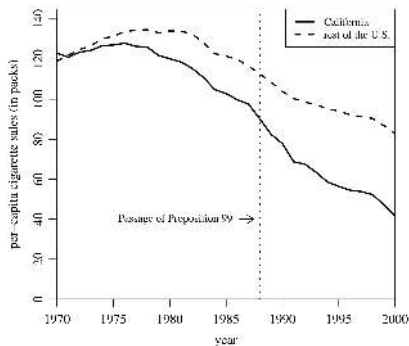


Figure 1. Trends in per-capita cigarette sales: California vs. the rest of the United States.

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

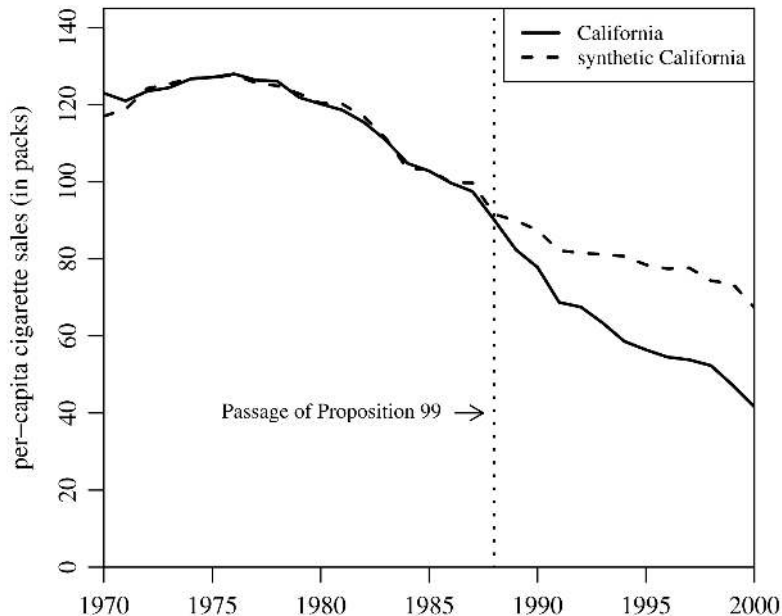
Synthetic control

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

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

- 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

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
- 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

- 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:

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

- 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?
- 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:

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

- 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
 - ▶ 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

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

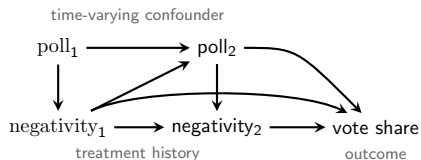
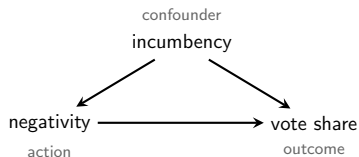
$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + a_i + u_{it}$$

- 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: $\mathbf{x}'_{it}\beta$
- 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:



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 editorial board of the journal for their valuable comments and suggestions.

When there are no other alternatives available, a small, well-ventilated, and well-lit room seems to be the only option for sleeping at work. The author has found that having a change of clothes and a change of shoes in the room can help reduce the risk of infection. The author also found that having a change of clothes and a change of shoes in the room can help reduce the risk of infection.

the same. There is no apparent reason for this, but the authors do point out that the 100% correlation between the two studies is probably due to the small number of cases. The authors also note that the 100% correlation between the two studies is probably due to the small number of cases.

This study joins the literature by pointing out the need for a more in-depth analysis, focused on language and epistemology studies, to improve the use of ICTs in the classroom. In addition, it provides a theoretical framework for the development of innovative educational practices, based on the use of ICTs in the classroom.

...the results of the study suggest that the effects of the intervention on the use of the intervention are not statistically significant. The results of the study suggest that the effects of the intervention on the use of the intervention are not statistically significant. The results of the study suggest that the effects of the intervention on the use of the intervention are not statistically significant.

How to Make Causal Inferences with Time-Series Cross-Sectional Data under Selection on Observables
MATTHEW BLACKWELL *Harvard University*
ADAM N. GLYNN *Duquesne University*

Recursive measurement of the same construct, given at groups over time, can be used to identify levels of political exposure. These measurements, sometimes called time-varying covariates (TVCSs), can allow researchers to estimate a broad set of causal questions, including compensation effects and other effects of lagged treatments. Unfortunately, popular methods for TVCS data can only produce estimates for lagged treatments, and they are not designed to estimate the effects of time-varying treatments. We show how to use TVCSs to estimate the effects of time-varying treatments, and we use them to define causal questions of interest in three settings and, finally, how standard models like the two-treatment difference-in-differences model can produce biased estimates of these quantities due to measurement error. We show that these models can be extended to handle time-varying treatments, and we probabilistically weight out structural nested mean models—and show via simulation that they can compare standard approaches to causal settings. We illustrate these ideas in a study of how workers

INTRODUCTION

Misapplication to political science involves the study of repeated measurements on the same individuals (countries, people, or groups) at several points in time. This type of data, sometimes called time-series-cross-sectional (TSCS) data, allows researchers to look at a larger pool of information when estimating causal effects. TSCS data also give researchers the power to add a richer set of variables than data with a single time point (a cross-section) (for example, see Hoxby and Kolesár [2011]). Moreover, TSCS data allow researchers to adjust the observed contemporaneous (quasi-)treatment as the effects of a single event?—and instead ask how the history 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

Markus Mäkelä is an Assistant Professor, Department of Management and Institute for Quantitative Social Science, Harvard University.

multifactorial causal effects and discuss the sampling error needed to identify them nonparametrically. We also relate these quantities of interest to common causal models, such as the structural equation model, and, in particular, show how to derive them from the parameters of a common DSGE model. The autoregressive AR(1) model of Aizerman (1972) is used as an example. The results show that the identification of the short-run values of the parameters can be obtained nonparametrically, but that the long-run values are only identified under additional assumptions. However, many common DSGE models are not identified, and the identification of the short-run values is often not unique. The results also provide a link of causality breakdown between the treatment and the target macroeconomic. This breakdown, for example, might result in a future level of unemployment that is higher than the current level of unemployment, which is in contrast with all future levels of spending. We argue that this type of feedback is common in DSGE models. We finish with a discussion of the implications for the treatment of the feedback and the feedback with the choice compared to standard fixed of DSGE models, noting that the latter may also not be identified.

This second contribution is to provide an alternative to the two methods (but not substitutes) that can inform the effect of treatment transfer without bias and under weaker assumptions than common TGCS models. We focus on two methods: (1) structural nested event models or SNEMs (Robins 1997) and (2) marginal structural models (MSMs) with inverse probability of treatment weighting (IPTW) (Robins, Hernan, and

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

Kenneth Grant has written widely
in English literature.

[illegible]

Figure 1. Effect of the 30-day pre-treatment on the growth of *Salmonella* in the presence of 100 µg/ml of chloramphenicol. The mean values of $\log_{10}$ CFU/ml of *Salmonella* were determined after 24 h of incubation at 37°C in the presence of 100 µg/ml of chloramphenicol. The error bars represent the standard deviation.

U.S. and Canadian governments have agreed to a new pact to combat the growing threat of terrorism. The agreement, signed in Ottawa on Tuesday, calls for increased cooperation between the two countries' intelligence and law enforcement agencies. It also establishes a joint task force to investigate and prevent terrorist activities. The pact is a significant step in strengthening transatlantic security ties in the wake of the September 11 attacks.

Journal of the American Statistical Association 4(46): 1-10, 1951. Reprinted in *ITS Library of Theoretical Statistics*, Vol. 1, Number 10, 1951. Copyright 1951 by the American Statistical Association. Reprinted by permission of the American Statistical Association.

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Received: 10 June 2014; Accepted: 10 July 2014; Published: 11 July 2014

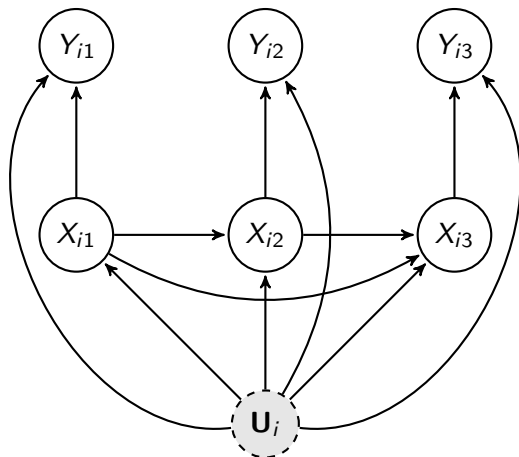
We are going to focus on addressing unobserved time-invariant confounders using the last paper.

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:

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



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

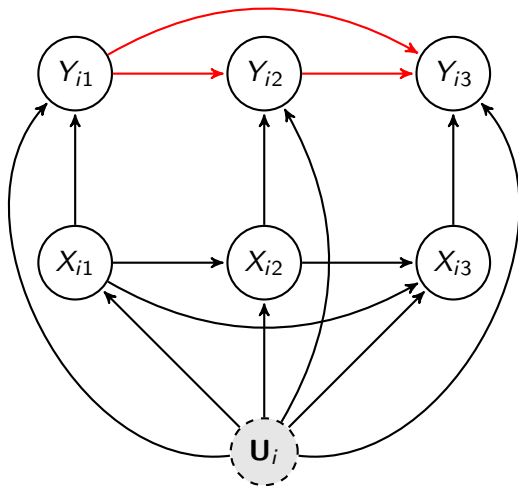
*the result implies that the **counterfactual** outcome for a treated observation in a given time period is estimated using the **observed outcomes of different time periods of the same unit**. Since such a comparison is **valid only when no causal dynamics exist**, this finding underscores the important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

- Imai and Kim (2019)

What Ideal Experiment Corresponds to the Fixed Effects Model?

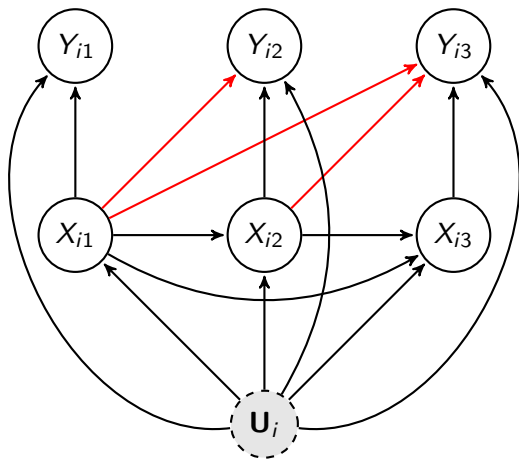
- Experiment that satisfies the model assumptions:
 - ① randomize X_{i1} given $\mathbf{U}_i$
 - ② randomize X_{i2} given X_{i1} and $\mathbf{U}_i$
 - ③ randomize X_{i3} given X_{i2} , X_{i1} , and $\mathbf{U}_i$
 - ④ and so on
- Experiment that does not satisfy the model assumptions:
 - ① randomize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



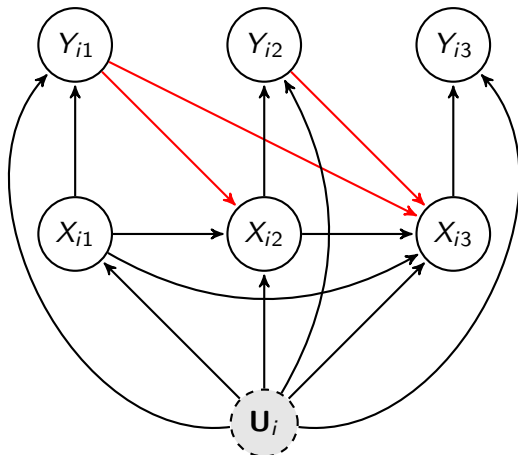
- Strict exogeneity still holds.
- Past outcomes do not confound $X_{it} \rightarrow Y_{it}$ given U_i .
- No need to adjust for past outcomes.
- Should use cluster robust standard errors for inference.
- Conclusion: **The assumption can be relaxed**

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



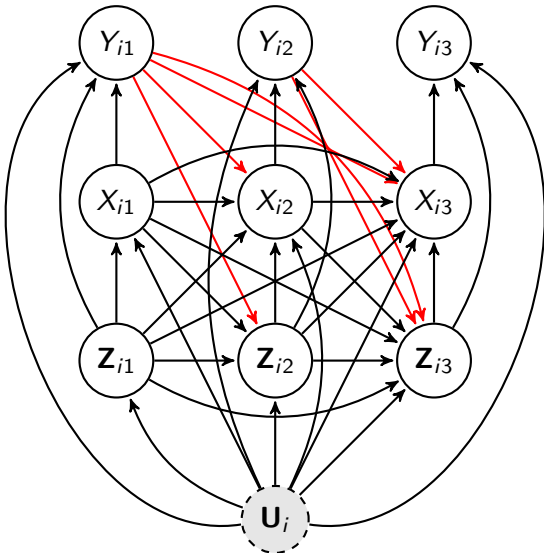
- Need to adjust for past treatments
- Strict exogeneity holds given past treatments and U_i
- Impossible to adjust for an entire treatment history and U_i at the same time
- Adjust for a small number of past treatments $\rightsquigarrow$ often arbitrary
- Conclusion: The assumption can be partially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

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



- 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**

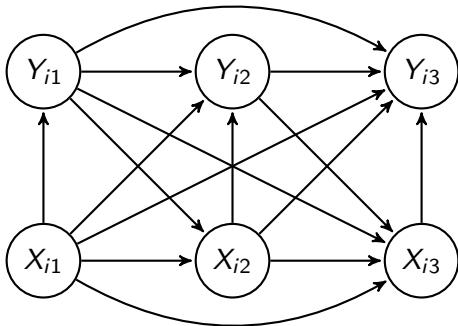
Can't We Just Adjust for Time-Varying Confounders?



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

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

- 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

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

- 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

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

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?

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:

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

- 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?

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?

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?

A: Let's chat.

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

Where are you?

You've been given a powerful set of tools



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

• 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

- 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

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

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.

