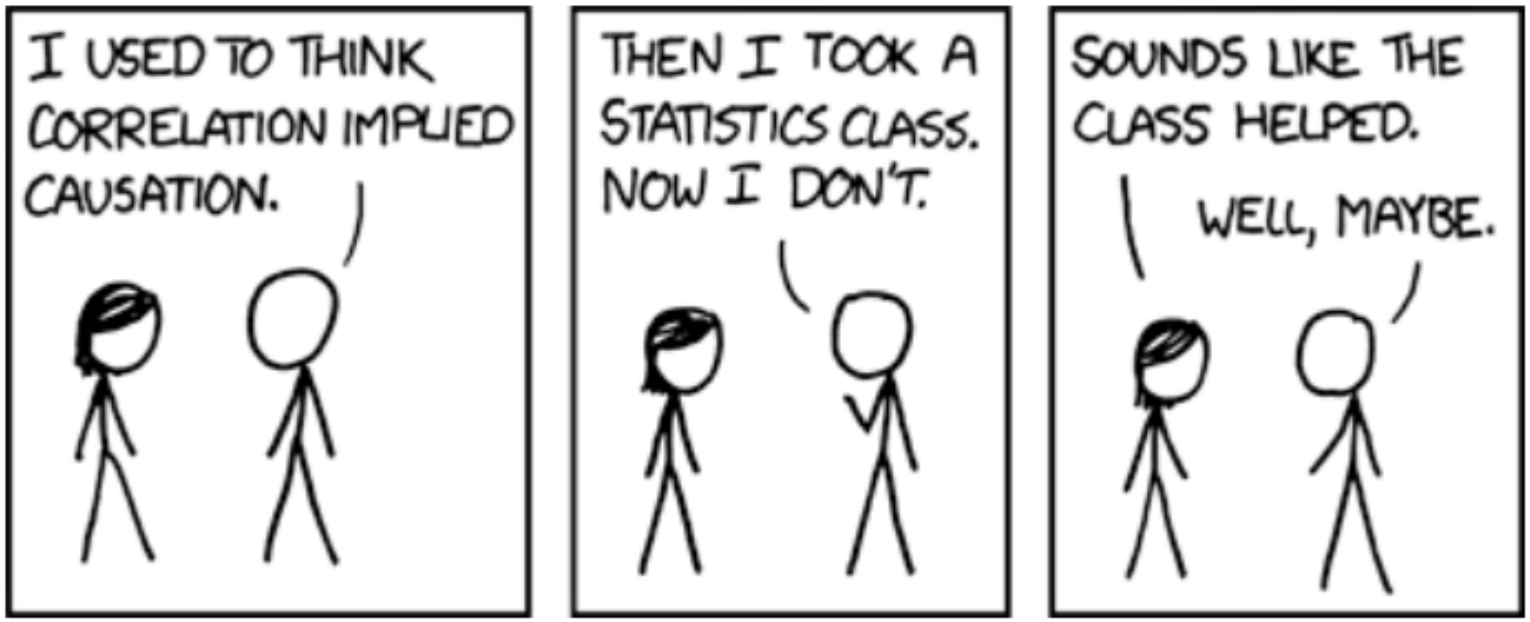
Econometrics I

Lecture 6: Experiments and Quasi-Experiments

Paul T. Scott
NYU Stern

Fall 2021



Source: xkcd.com

Treatment Effects

- As usual, let Y_i refer to an outcome variable of interest for individual i, but now let

$$\begin{aligned} Y_{i1} &: \text{ Outcome if } i \text{ receives treatment} \\ Y_{i0} &: \text{ Outcome if } i \text{ does not receive treatment} \end{aligned}$$

  noting that for a given individual, only one of these outcomes is actually observed.

- Define the **average treatment effect** as follows:

$$E\left[Y_{i1} - Y_{i0}\right].$$

  The ATE is the average difference that the treatment makes in the outcome, averaging over all individuals in the population.

Treatment Effects

Y_{i1} : Outcome if i receives treatment
Y_{i0} : Outcome if i does not receive treatment

- Define the **average treatment effect on the treated** (ATT or ATET) as follows:

$$E\left[Y_{i1} - Y_{i0} | T_i = 1\right] = E\left[Y_{i1} | T_i = 1\right] - E\left[Y_{i0} | T_i = 1\right]$$

where T_i is an indicator for treatment status. Note that the second term is not observed. In general, the ATT may differ from the ATE if the sub-population that receives the treatment is special somehow (e.g., the people who end up receiving treatment are those who find it more effective).

Selection Bias

- We can observe the following difference in the population:

$$\begin{aligned} E[Y_{i1}|T_i=1]-E[Y_{i0}|T_i=0] &= E[Y_{i1}|T_i=1]-E[Y_{i0}|T_i=1] \\ &+ E[Y_{i0}|T_i=1]-E[Y_{i0}|T_i=0] \end{aligned}$$

- The first term on the RHS, $E[Y_{i1}|T_i=1]-E[Y_{i0}|T_i=1]$, is the average treatment effect on the treated. This is potentially an object of interest — it tells us how much the treatment improved outcomes for those that received treatment.
- The second term, $E[Y_{i0}|T_i=1]-E[Y_{i0}|T_i=0]$, is **selection bias**. It tells us how the treatment ($T_i=1$) and control ($T_i=0$) groups differ even in the absence of treatment.

The Magic of Randomization

$$
\begin{aligned}
E\left[Y_{i1}|T_i=1\right]-E\left[Y_{i0}|T_i=0\right] &= \\
E\left[Y_{i1}|T_i=1\right]-E\left[Y_{i0}|T_i=1\right] &+ \underbrace{E\left[Y_{i0}|T_i=1\right]-E\left[Y_{i0}|T_i=0\right]}_{\text{selection bias}}
\end{aligned}
$$

- If treatment is assigned randomly, then Y_{i0} and T_i should be independent. Consequently,

$$E\left[Y_{i0}|T_i=1\right]=E\left[Y_{i0}|T_i=0\right],$$

  recalling that if Y_{i0} and T_i are independent, $E\left[Y_{i0}|T_i\right]=E\left[Y_{i0}\right]$.
- Thus, randomization of treatment eliminates selection bias.

The Magic of Randomization II

- We've just shown that randomization gives us the average effect of treatment on the treated (ATT) without selection bias.

$$E\left[Y_{i1}|T_i=1\right]-E\left[Y_{i0}|T_i=0\right]=E\left[Y_{i1}|T_i=1\right]-E\left[Y_{i0}|T_i=1\right]$$

- Furthermore, randomization of treatment also implies that the ATT equals the ATE. If T_i is independent of Y_{i0} and Y_{i1}, then

$$E\left[Y_{i1}|T_i=1\right]-E\left[Y_{i0}|T_i=1\right]=E\left[Y_{i1}\right]-E\left[Y_{i0}\right]=E\left[Y_{i1}-Y_{i0}\right].$$

What Does Randomization Do?

1. Randomization of treatment eliminates selection bias.
2. Randomization of treatment ensures that the ATE=ATT.

Randomization and Endogeneity I

- Selection bias has to do with the fact that *baseline* outcomes for the treated and untreated groups may differ. Example: schoolchildren who get the treatment of having small class sizes (private schools) are also children who have access to private tutors and well-educated parents.
- This is a version of an *endogeneity* problem, and it's a fundamental problem for causal inference. Randomizing treatment solves the problem, for it means that baseline outcomes should no longer be correlated with the treatment.

Randomization and Endogeneity II

- Simplifying the situation by assuming $\beta = Y_{i1} - Y_{i0}$ for all i, we can put this back in the regression equation framework,

$$Y_i = \beta_0 + \beta_1 T_i + \varepsilon_i$$

  where $\beta_0 = E[Y_{i0}]$ and $\varepsilon_i = Y_{i0} - \beta_0$.

- If heterogeneity in baseline outcomes Y_{i0} is correlated with treatment status T_i, then the error term ε_i is correlated with the regressor ε_i, violating the strict exogeneity assumption, and leading to biased estimates of β.

- In theory, randomization of regressors is a way of ensuring that the strict exogeneity assumption holds: we randomly control T_i so that it won't be related to whatever is in ε_i.

Randomization and Endogeneity III

- In practice, randomization doesn't *guarantee* there is no endogeneity problem:
  - Random control of treatment may be imperfect: attrition and manipulation.
  - Treatment status may directly influence ε in some cases: placebo effects and behavioral responses to treatment.
  - Blind and double-blind studies aim to mitigate these concerns.
- This requires more notation to discuss within the potential outcome framework – we need to distinguish between whether subject *believes* they are being treated or not, as well as whether they are *actually* being treated or not – but it is easier to talk about in the regression framework.

Randomization and Endogeneity IV

- What does it mean for the disturbance to be influenced by a regressor status? Can't we just consider this an effect of the regressor?
- It depends: often there will be an external validity issue, in the sense that the effect won't extrapolate outside of the study.
- Examples:
  - With placebo effect, we could understand this as treatment influencing the disturbance term, but maybe we're happy to enjoy the placebo effect. On the other hand, what if we're considering whether to add a nutrient to a pill or not, rather than deciding whether or not to give the subjects a pill or not? Perhaps we won't get the placebo effect twice.
  - If treatment influences error term because of manipulation on the part of the researcher implementing the study, then the relationship between treatment and outcomes we see typically won't extrapolate outside of the study.

Randomization and Heterogeneity

- The fact that ATE$\neq$ ATT is a separate issue having to do with *heterogeneity* of treatment effects. This issue is also "solved" by randomization in a way, but is this an issue we want to solve?
- Example: the people who take anti-depressants benefit more than the people who don't. The ATE for a given drug in the whole population might be low, but that doesn't mean the drug is ineffective. If the ATE within the group of people diagnosed with depression is high, and the people who end up taking the drug fall within that group, the ATT in practice might correspond closely to the sub-population ATE.
- Bottom line: differences between ATE and ATT don't reflect problems of causal inference, but they reflect the importance of understanding the population of interest. There's a reason clinical trials for new cancer drugs focus on people that have cancer.

Experiment with Regression Controls

- Consider the model

$$Y_i = \beta_0 + \beta_1 T_i + \boldsymbol{\beta}_2' \mathbf{X}_i + \varepsilon_i$$

  where T_i is assigned randomly.

- Does controlling for $\mathbf{X}$ matter in the experimental context? Note that even if $\mathbf{X}$ is omitted from the regression model, there is **no problem of omitted variables bias** because T and $\mathbf{X}$ are uncorrelated.

- However, controlling for covariates may improve precision of the estimates, especially in small sample sizes where $\mathbf{X}$ may not be balanced across the treatment and control groups.

Randomization with Small Samples

- In a finite sample, there's always a chance that we end up with subjects that look very different across the treatment and control groups.
  - For observable characteristics, it is customary to check that the two groups have similar means and medians. This is often Table 1 in experimental papers.
- What would you do if you randomly assigned subjects to the two groups, and, before running the experiment, you notice that characteristics are not balanced? Would it be bad to re-randomize?
- Related ideas:
  - Student (1938), the t-test guy, argued against randomization in agricultural trials. Simple random sampling vs. systematic sampling.
  - Stratified sampling, clustered sampling, sampling theory.
  - Chassang et al (2012) explore the idea of selective trials.

Experiments in the Social Science

- Randomized Controlled Trials have long been the gold standard for research in the natural sciences.
- In social sciences, many important questions don't lend themselves well to randomization.
  - Macroeconomic policy, other large-scale policy issues, especially when there are spillovers across markets/countries. E.g., what is global impact of EU's decision to implement carbon pricing?
  - Mergers and antitrust, other situations where policy questions are very context-specific.
- However, experiments are becoming increasingly popular in some fields
  - Lab experiments: behavioral economics
  - Field experiments: development, labor, education

Example: Banerjee et al (2007)

- Background: getting kids into schools in India seemingly had unimpressive impacts on educational attainment. School quality (educational inputs) is also important.
- Experimental treatment: remedial education. Third and fourth grade students identified as at risk for falling behind are assigned an extra teacher for two hours/day.
  - Group A (50% of schools): third grade classrooms treated in 2001-2, fourth grade classrooms treated in 2002-3
  - Group B (50% of schools): fourth grade classrooms treated in 2001-2, third grade classrooms treated in 2002-3.

Differences

- Consider using OLS to estimate

$$Y_i = \beta_0 + \beta_1 T_i + \varepsilon_i$$

when $T_i \in \{0, 1\}$ is assigned randomly.

- From the OLS formula and a little algebra, we can show that

$$\hat{b}_1 = \bar{Y}_1 - \bar{Y}_0$$

where $\bar{Y}_1$ is the sample mean of Y_i conditional on $T_i = 1$, and similarly $\bar{Y}_1$ is the sample mean conditional on $T_i = 0$.

- Note: this loops us back to the treatment effects setup, where we were comparing conditional means.

Differences-in-Differences I

- **Differences-in-differences** is a **quasi-experimental** design that attempts to deal with selection by focusing how treatment and control groups change between two periods
- Consider the model

$$Y_{it} = \beta_0 + \beta_1 T_{it} + \boldsymbol{\beta}_2' \mathbf{X}_i + \beta_{post} Post_t + \varepsilon_{it},$$

and furthermore suppose that

  - Some variables in $\mathbf{X}$ may be unobserved
  - T_{it} is not random and may be correlated with $\mathbf{X}_i$
  - $Post_t$ is a dummy variable for the post-treatment period $t = 2$
  - Each individual i is observed (at least) twice
  - For some individuals, T_{it} changes over time (**natural experiment**)
  - $\mathbf{X}_i$ does not change over time

Differences-in-Differences II

$$Y_{it} = \beta_0 + \beta_1 T_{it} + \boldsymbol{\beta}_2' \mathbf{X}_i + \beta_{post} Post_t + \varepsilon_{it},$$

- Let ΔY_i represent the change in Y_{it} between two time periods:

$$\Delta Y_i = Y_{i2} - Y_{i1},$$

  and define ΔT_i and $\Delta \varepsilon_i$ similarly.

- A differenced regression equation:

$$\Delta Y_i = \beta_{post} + \beta_1 \Delta T_i + \Delta \varepsilon_i$$

Differences-in-Differences III

$$\Delta Y_i = \beta_{post} + \beta_1 \Delta T_i + \Delta \varepsilon_i$$

- Applying OLS to the differenced regression equation will provide unbiased estimates as long as

$$E\left[\Delta \varepsilon_i | \Delta T_i\right] = 0$$

strict exogeneity assumptions of this form are known as the **parallel trends assumption**.

Differences-in-Differences IV

$$\Delta Y_i = \beta_{post} + \beta_1 \Delta T_i + \Delta \varepsilon_i$$

- From before, recall that the OLS estimator for a simple experiment amounted to the difference in conditional means of the outcome variable.
- The same is true here, but now we start with an outcome variable that is differenced over time. Suppose that for some observations, $\Delta T_i = 1$ (treatment group), and for others $T_{i1} = T_{i2} = 0$ (untreated group).

$$\hat{b}_{1,DID} = \overline{\Delta Y}_T - \overline{\Delta Y}_U$$

where $\overline{\Delta Y}_T$ is the conditional mean of ΔY_i for the treatment group, and $\overline{\Delta Y}_U$ is the conditional mean of ΔY_i for the control group.

The Appeal of DID

- The DID strategy is robust to a form of selection bias (when, the selection is related to persistent characteristics). Simple differences across treatment and control groups would not be.
- DID estimates are also robust to aggregate shocks or time effects. Simple differences over time for the treatment group would not be.
- DID can be used to study many "natural experiments," where something changes for one group but not for an otherwise similar group.

Example: Card and Krueger (1994)

- Background: New Jersey increased minimum wage from \$4.25 to \$5.05 per hour, effective April 1, 1992.
- At the same time, the minimum wage across the border in Pennsylvania did not change.
- Population of interest: fast food stores in NJ and PA.
- A standard competitive model predicts that an increase in the minimum wage should cause employment to drop.

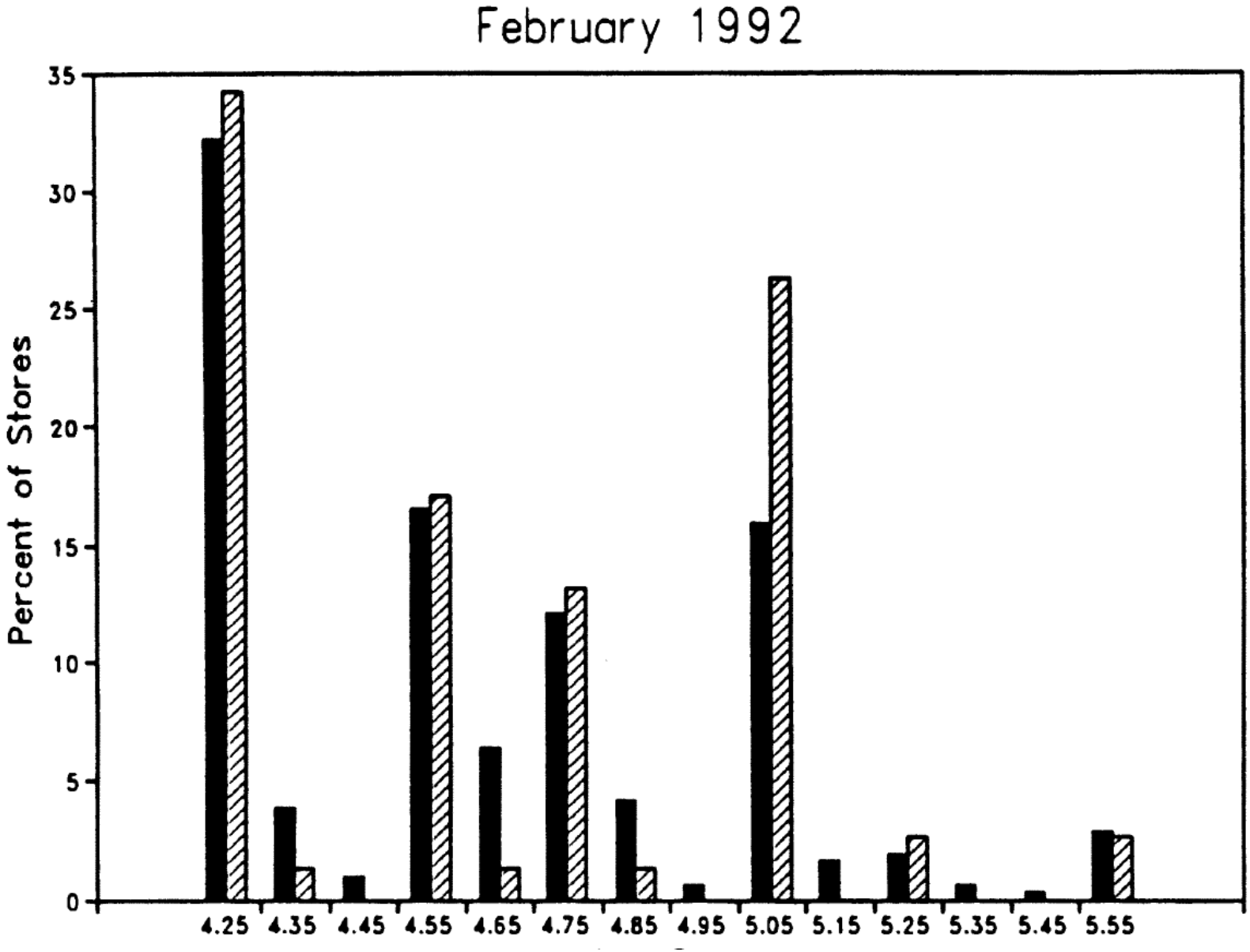
TABLE 1—SAMPLE DESIGN AND RESPONSE RATES

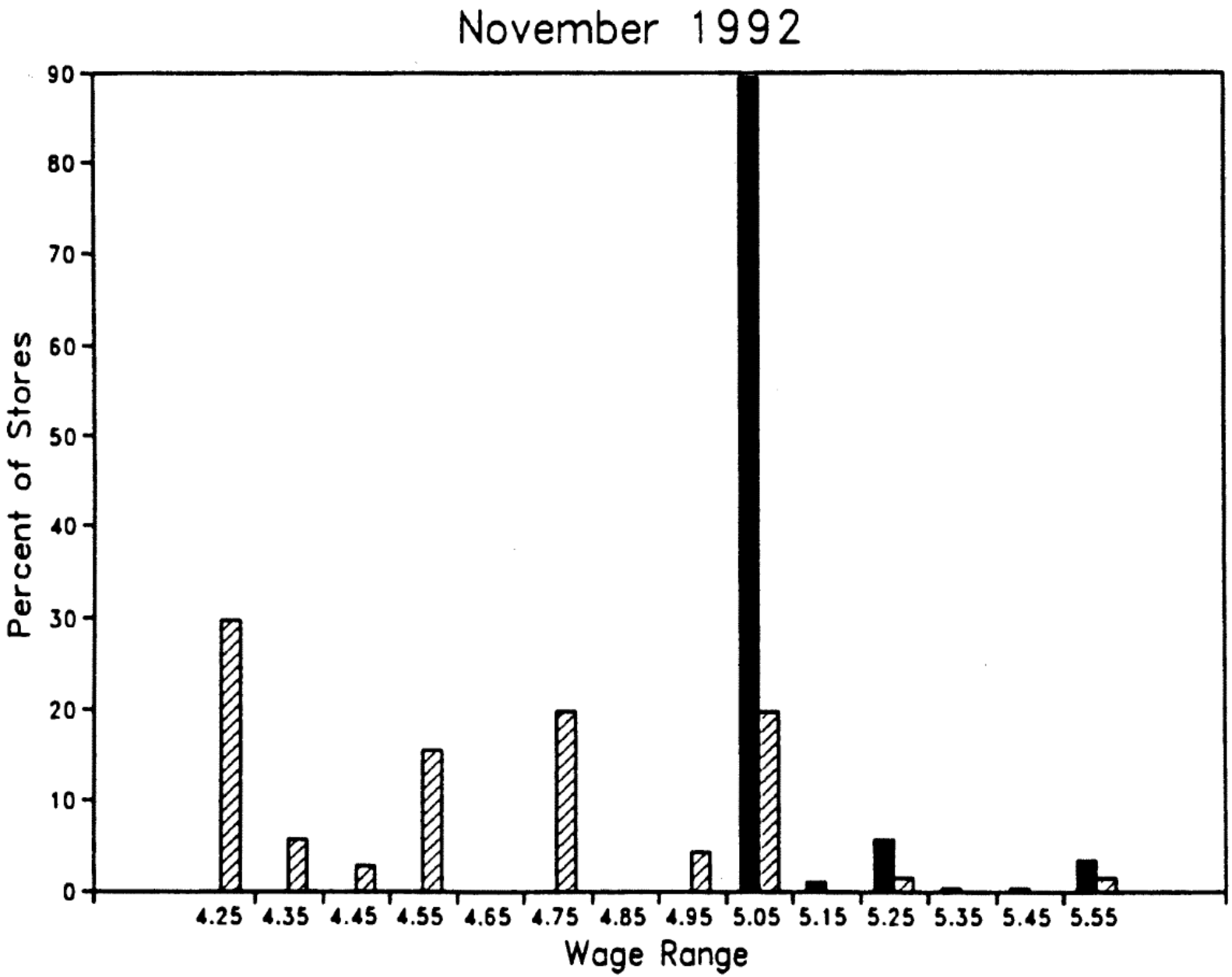
	All	Stores in: NJ	PA
Wave 1, February 15–March 4, 1992:			
Number of stores in sample frame:[a]	473	364	109
Number of refusals:	63	33	30
Number interviewed:	410	331	79
Response rate (percentage):	86.7	90.9	72.5
Wave 2, November 5–December 31, 1992:			
Number of stores in sample frame:	410	331	79
Number closed:	6	5	1
Number under rennovation:	2	2	0
Number temporarily closed:[b]	2	2	0
Number of refusals:	1	1	0
Number interviewed:[c]	399	321	78

[a]Stores with working phone numbers only; 29 stores in original sample frame had disconnected phone numbers.

[b]Includes one store closed because of highway construction and one store closed because of a fire.

[c]Includes 371 phone interviews and 28 personal interviews of stores that refused an initial request for a phone interview.


February 1992
Percent of Stores
35
30
25
20
15
10
5
0
4.25
4.35
4.45
4.55
4.65
4.75
4.85
4.95
5.05
5.15
5.25
5.35
5.45
5.55


November 1992
Percent of Stores
90
80
70
60
50
40
30
20
10
0
4.25
4.35
4.45
4.55
4.65
4.75
4.85
4.95
5.05
5.15
5.25
5.35
5.45
5.55
Wage Range

Variable	Stores by state		
	PA (i)	NJ (ii)	Difference, NJ − PA (iii)
1. FTE employment before, all available observations	23.33 (1.35)	20.44 (0.51)	−2.89 (1.44)
2. FTE employment after, all available observations	21.17 (0.94)	21.03 (0.52)	−0.14 (1.07)
3. Change in mean FTE employment	−2.16 (1.25)	0.59 (0.54)	2.76 (1.36)
4. Change in mean FTE employment, balanced sample of stores[c]	−2.28 (1.25)	0.47 (0.48)	2.75 (1.34)
5. Change in mean FTE employment, setting FTE at temporarily closed stores to 0[d]	−2.28 (1.25)	0.23 (0.49)	2.51 (1.35)

NJ Minimum Wage Follow-Up

- The study has been controversial, and it has had a big impact on policy discussions.
- Card and Krueger update their conclusion in a 2000 follow up:

  > *The increase in New Jersey's minimum wage probably had no effect on total employment in New Jersey's fast-food industry, and possibly had a small positive effect.*

- There have been *many* criticisms of the paper, including studies showing that the result is not robust to the sample of stores and data source used, and concerns about the external validity of the focus on only fast food stores.

DID with Time-Varying Covariates

- Suppose individuals have time-varying observable covariates:

$$Y_{it} = \beta_0 + \beta_1 D_{it} + \boldsymbol{\beta}_2' \mathbf{X}_{it} + \varepsilon_{it},$$

- We can still estimate β_1 with a differenced linear regression,

$$\Delta Y_i = \beta_1 \Delta T_i + \boldsymbol{\beta}_2' \Delta \mathbf{X}_{it} + \Delta \varepsilon_i$$

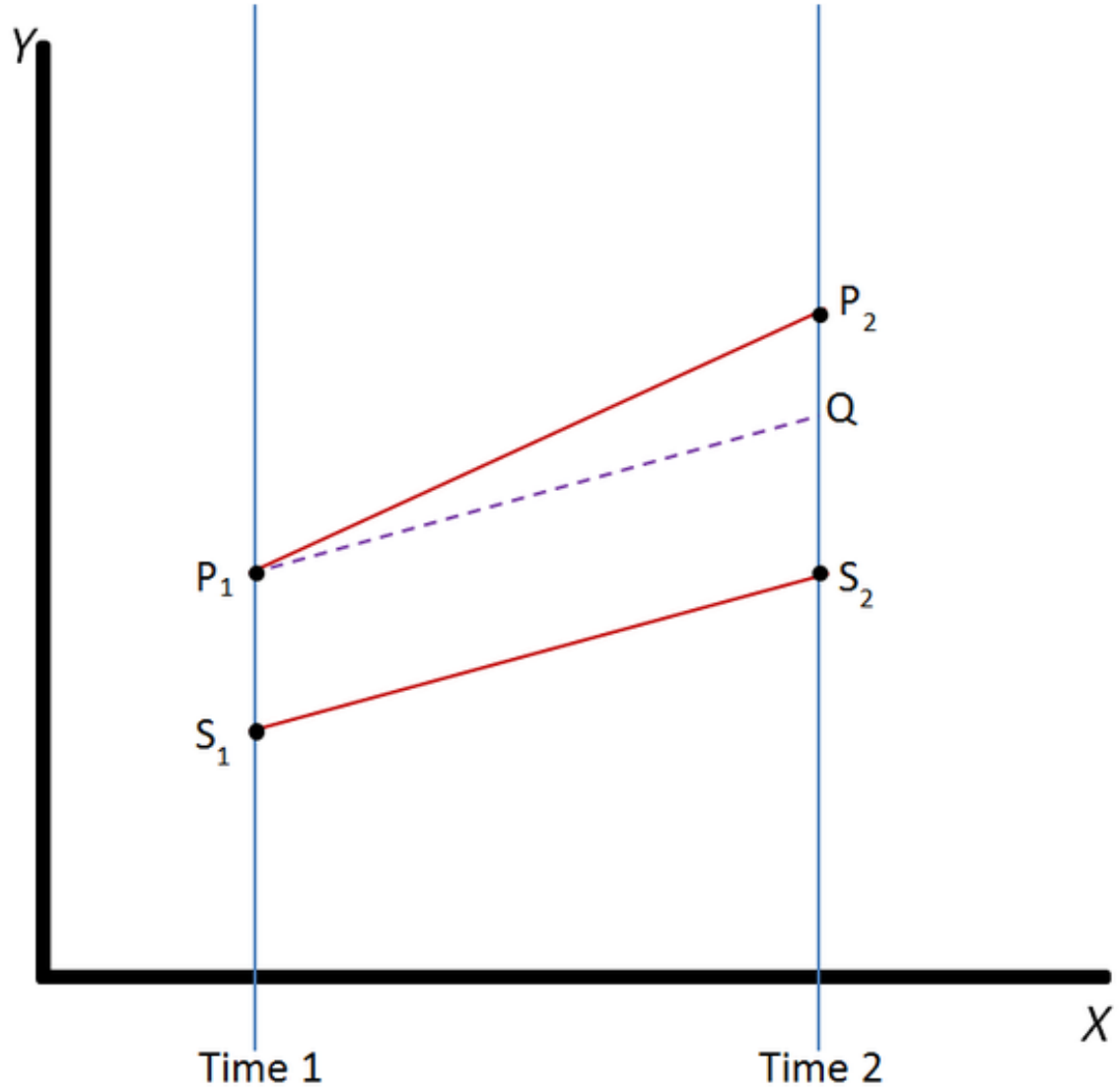
noting that the exogeneity assumption becomes

$$E\left[\Delta \varepsilon_i \middle| \begin{pmatrix} \Delta T_i \\ \Delta \mathbf{X}_i \end{pmatrix}\right] = 0$$

- The estimate of β_1 here does not correspond to a simple difference in differences, so I'm not sure this should still be called "DID", but some people still describe this sort of estimation strategy as DID. Given Frisch-Waugh, it's still DID after controlling for $\mathbf{X}$.

Endogeneity Problems for DID

- DID is robust to endogeneity of treatment to *time-invariant* $\mathbf{X}$, even if those $\mathbf{X}$ are unobserved.
- By using a differenced linear regression, "DID" estimates like the above can also be robust to endogeneity of treatment to *time-varying* $\mathbf{X}$, but such $\mathbf{X}$ must be included in the differenced regression and therefore observed.
- Time-varying unobservables create endogeneity problems for DID estimators, i.e. violations of the parallel trends assumption.


Y
P2
Q
P1
S2
S1
X
Time 1
Time 2

Parallel trends vs pre-trends

- When the data is available, researchers often check **pre-trends** as a way of motivating the parallel trends assumption.
  - That is, look at how the control and treatment groups evolve in several periods preceding the time of treatment to see if they were changing in similar ways over time.
- This is a worthwhile check, and we should certainly worry about the plausibility of the parallel trends assumption if the pre-trends are not parallel.
- However, the parallel trends assumption is about what *would have happened* immediately at the time of treatment in the absence of treatment, not about what happened before treatment. Thus, establishing parallel pre-trends **does not** establish the parallel trends assumption.
  - Just as we can't test exogeneity within the OLS framework, we can't test the parallel trends assumption within the DiD framework.

Parallel trends vs pre-trends: example

Consider the NJ minimum wage study. Here's a concern that would not be assuaged by seeing parallel pre-trends:

- If the minimum wage changed because a new administration was elected in NJ and they implemented a portfolio of new policies, then the minimum wage change might have been just one of several ways in which NJ and PA were diverging at the time.
- Some of the other policy changes might have led to greater labor force participation and/or incentivized businesses to hire more.
- Because of these other changes, we would expect NJ's restaurant employment to move differently from PA's in 1992 even if the minimum wage change had not happened.

DID Experiments

- Recall that before we argued that experimental effects can be estimated with regressions using additional covariates (controls) even though they don't need to be.
- Similarly, they can be estimated using differences-in-differences rather than simple differences. They don't need to be, as with randomized treatment, the treatment and control groups should have similar baselines.

Regression Discontinuity I

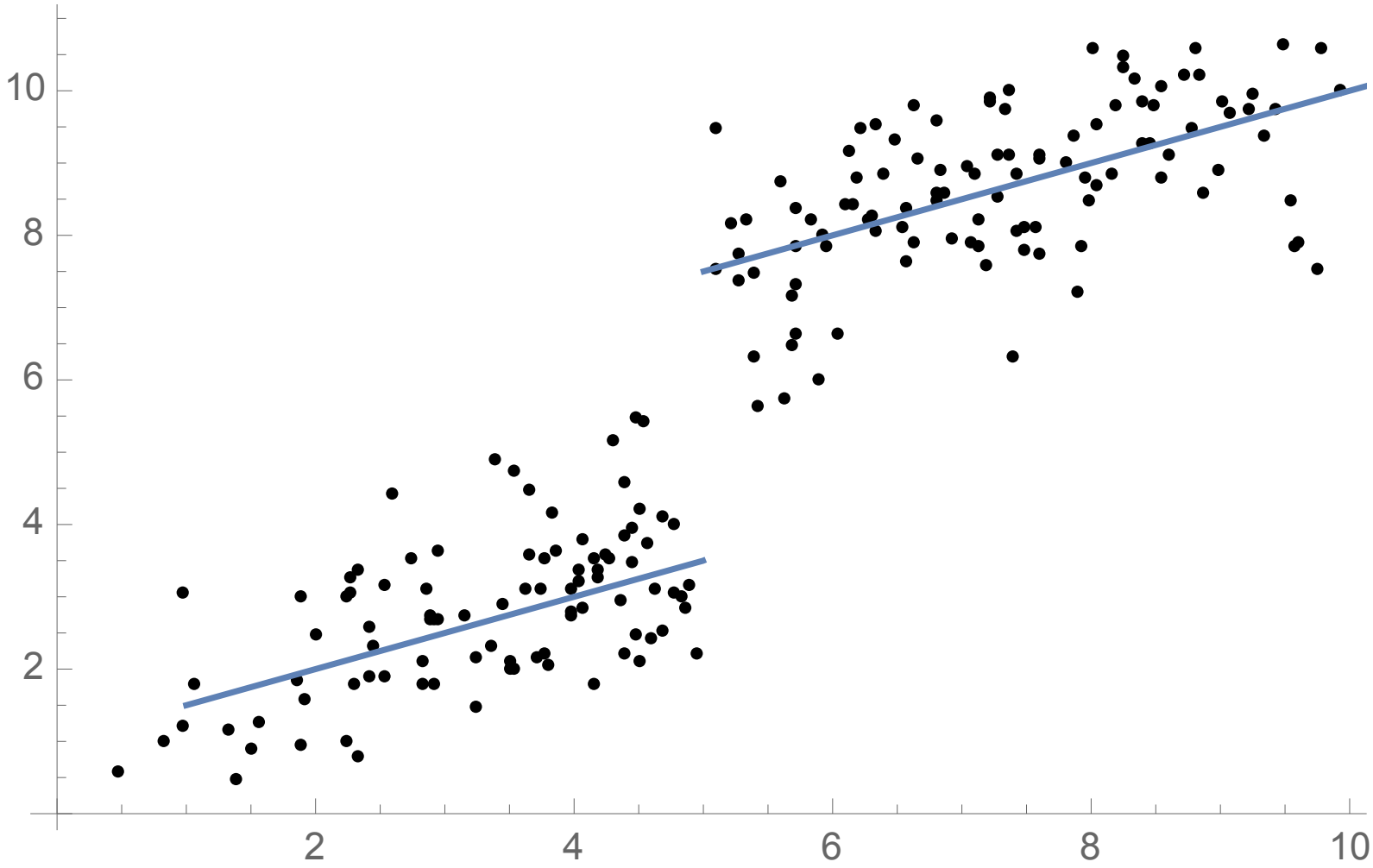
- Regression discontinuity models involve a treatment that is based on a threshold rule for a observed variable:

$$T_i = \mathbf{1}\left[X_i > X^*\right].$$

- For instance, if students who score above a given threshold on an entrance exam, they might get admission to a special school.
- We'd like to know the impact of going to the special school. However, comparing achievement by students who go to that school to those that don't plausibly involves a lot of selection bias.

Regression Discontinuity II

- Idea: if we were to compare students just above and just below the cutoff on the entrance exam, the selection bias would be small (vanishingly small if we could compare students with $A^* - \epsilon$ and $A^* + \epsilon$ for tiny ϵ.


10
8
6
4
2
2
4
6
8
10

Regression Discontinuity III

- If the regression discontinuity is *sharp* (meaning the treatment is assigned deterministically based on the cutoff) and the impact of the X variable is linear, RD can be implemented within the linear regression framework:

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 X_i + \varepsilon_i$$

  where $T_i = \mathbf{1}[X_i > X^*]$.

- Identifying assumption:

$$E[\varepsilon | X, T] = E[\varepsilon | X].$$

  The major part of RD's appeal is that it *does not* require $E[\varepsilon|X] = 0$ to get a consistent estimate of β_1. (But we would need this for β_2.)

Regression Discontinuity IV

- Often RD estimation allows for non-linear direct effects of X:

$$Y_i = \beta_0 + \beta_1 T_i + f(X_i) + \varepsilon_i$$

  where $f(\cdot)$ is some continuous function that can be approximated with a flexible functional form.

- Alternatively, RD estimates can be obtained non-parametrically by just computing the mean of Y for data just above and just below the cutoff.

- A **fuzzy RD** involves probabilistic treatment, is more complicated.

Matching: The Basic Idea

- Suppose we compare the treated and untreated groups conditional on particular value of $X_i = x$:

$$\delta_x = E\left[Y_i | T_i = 1, X_i = x\right] - E\left[Y_i | T_i = 0, X_i = x\right]$$

- Assuming **conditional independence**,

$$(Y_{i1}, Y_{i0}) \perp T_i | X_i,$$

then δ_x will be an unbiased estimate of the treatment effect for $X_i = x$.

- An estimate of the ATT can be constructed as follows:

$$\sum_x \delta_x Pr\left(T_i = 1 | X_i = x\right)$$

Matching and Regression I

- For simplicity, let's assume $\forall i : Y_{i1} - Y_{i0} = \beta_1$.

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 X_i + \varepsilon_i$$

- Recall how we map between the hypothetical values (Y_{i0}, Y_{i1}) and the regression framework:

$$\varepsilon_i = Y_{i0} - E[Y_{i0}]$$

- The conditional indpendence assumption implies

$$\varepsilon_i \perp T_i | X$$

- Note that this doesn't give us strict exogeneity, but it has some of the same flavor. If a mean-zero error term ϵ_i is independent of a regressor τ_i, then we have $E(\epsilon_i | \tau_i) = 0$.

Matching and Regression II

- Given

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 X_i + \varepsilon_i$$
$$\varepsilon_i \perp T_i | X,$$

  it's effectively the case that we have exogeneity between T and ε (but not necessarily X and ε).
- Intuitively, we might expect OLS to provide an unbiased estimate of β_1 (but not necessarily β_2).
- We can show that's indeed the case as long as expectations conditional on X are linear in X. For example, see Angrist and Pischke, pp. 74-75.
- In other words, the assumptions we use to justify the matching estimator are similar to the assumptions that justify linear regression estimator, but regression does rely on additional functional form assumptions.

Matching and Regression III

- Bottom line: matching and regression are not terribly different identification strategies. They both give you an estimate of the impact of T while controlling for X.
- That said, matching and regression **do not** deliver the same results. The are both weighted averages of the x-specific treatment effects δ_X, but each places different weights on different values of x.

What Makes a Match?

- Computing

$$\delta_x = E\left[Y_i | T_i = 1, X_i = x\right] - E\left[Y_i | T_i = 0, X_i = x\right]$$

  can be difficult to do in practice if we don't have treated and untreated observations for a given value of x.
- In practice, we can group observations that are "close enough." See: Kernel bandwidth selection.
- To make this easier, we can also do **propensity score matching**, meaning instead of matching based on values of X_i, we match based on values of $Pr\left(T_i = 1 | X_i\right)$. This makes it much easier to form matches if X is high-dimensional.

Propensity Score Theorem

Propensity Score Theorem

Assuming Conditional Independence

$$(Y_{i1}, Y_{i0}) \perp T_i | X_i,$$

it follows that

$$(Y_{i1}, Y_{i0}) \perp T_i | p(X_i),$$

where $p(X_i) = Pr(T_i = 1 | X_i)$

- In words: if there is no selection bias after controlling for X, then there is no selection bias after controlling only for $p(X)$.
- There are serious concerns about propensity score matching in practice. See King and Nielsen (2019) and Desai and Franklin (2019)

Quasi-Experiments

- Always remember that quasi-experimental approaches to causal inference don't establish causality for free; they only establish causality given assumptions that are often questionable.
- A healthy dose of skepticism is probably called for any time somebody shows you a study based on difference-in-differences, matching, or OLS without experimental variation.
- Some validation studies have established strong grounds for skepticism. We will have a look at LaLonde (1986). See also Bertrand, Duflo, and Mullainathan (2004), "How Much Should We Trust Differences-In-Differences Estimates"?

Evaluating the Econometric Evaluations of Training Programs with Experimental Data

By ROBERT J. LALONDE*

This paper compares the effect on trainee earnings of an employment program that was run as a field experiment where participants were randomly assigned to treatment and control groups with the estimates that would have been produced by an econometrician. This comparison shows that many of the econometric procedures do not replicate the experimentally determined results, and it suggests that researchers should be aware of the potential for specification errors in other nonexperimental evaluations.

TABLE 2—ANNUAL EARNINGS OF NSW TREATMENTS, CONTROLS, AND EIGHT CANDIDATE COMPARISON GROUPS FROM THE *PSID* AND THE *CPS-SSA*

Year	Treat-ments	Controls	Comparison Group[a,b]							
			PSID-1	*PSID*-2	*PSID*-3	*PSID*-4	*CPS-SSA*-1	*CPS-SSA*-2	*CPS-SSA*-3	*CPS-SSA*-4
1975	$895	$877	7,303	2,327	937	6,654	7,788	3,748	4,575	2,049
	(81)	(90)	(317)	(286)	(189)	(428)	(63)	(250)	(135)	(333)
1976	$1,794	$646	7,442	2,697	665	6,770	8,547	4,774	3,800	2,036
	(99)	(63)	(327)	(317)	(157)	(463)	(65)	(302)	(128)	(337)
1977	$6,143	$1,518	7,983	3,219	891	7,213	8,562	4,851	5,277	2,844
	(140)	(112)	(335)	(376)	(229)	(484)	(68)	(317)	(153)	(450)
1978	$4,526	$2,885	8,146	3,636	1,631	7,564	8,518	5,343	5,665	3,700
	(270)	(244)	(339)	(421)	(381)	(480)	(72)	(365)	(166)	(593)
1979	$4,670	$3,819	8,016	3,569	1,602	7,482	8,023	5,343	5,782	3,733
	(226)	(208)	(334)	(381)	(334)	(462)	(73)	(371)	(170)	(543)
Number of Observations	600	585	595	173	118	255	11,132	241	1,594	87

- Note that even comparison groups who are composed similarly to the eligible population for the treatment (PSID-3 and CPS-SSA-4) seem to have quite a different composition from the treatment and control group.

TABLE 4—EARNINGS COMPARISONS AND ESTIMATED TRAINING EFFECTS FOR THE NSW AFDC PARTICIPANTS USING COMPARISON GROUPS FROM THE *PSID* AND THE *CPS-SSA*[a,b]

Name of Comparison Group[d]	Comparison Group Earnings Growth 1975–79 (1)	NSW Treatment Earnings Less Comparison Group Earnings: Pre-Training Year, 1975 Unadjusted (2)	Pre-Training Year, 1975 Adjusted[c] (3)	Post-Training Year, 1979 Unadjusted (4)	Post-Training Year, 1979 Adjusted[c] (5)	Difference in Differences: Difference in Earnings Growth 1975–79 Treatments Less Comparisons Without Age (6)	With Age (7)	Unrestricted Difference in Differences: Quasi Difference in Earnings Growth 1975–79 Unadjusted (8)	Adjusted[c] (9)	Controlling for All Observed Variables and Pre-Training Earnings: Without AFDC (10)	With AFDC (11)
	2,942	−17	−22	851	861	833	883	843	864	854	–
Controls	(220)	(122)	(122)	(307)	(306)	(323)	(323)	(308)	(306)	(312)	
PSID-1	713	−6,443	−4,882	−3,357	−2,143	3,097	2,657	1746	1,354	1664	2,097
	(210)	(326)	(336)	(403)	(425)	(317)	(333)	(357)	(380)	(409)	(491)
PSID-2	1,242	−1,467	−1,515	1,090	870	2,568	2,392	1,764	1,535	1,826	–
	(314)	(216)	(224)	(468)	(484)	(473)	(481)	(472)	(487)	(537)	
PSID-3	665	−77	−100	3,057	2,915	3,145	3,020	3,070	2,930	2,919	–
	(351)	(202)	(208)	(532)	(543)	(557)	(563)	(531)	(543)	(592)	
PSID-4	928	−5,694	−4,976	−2,822	−2,268	2,883	2,655	1,184	950	1,406	2,146
	(311)	(306)	(323)	(460)	(491)	(417)	(434)	(483)	(503)	(542)	(652)
CPS-SSA-1	233	−6,928	−5,813	−3,363	−2,650	3,578	3,501	1,214	1,127	536	1,041
	(64)	(272)	(309)	(320)	(365)	(280)	(282)	(272)	(309)	(349)	(503)
CPS-SSA-2	1,595	−2,888	−2,332	−683	−240	2,215	2,068	447	620	665	–
	(360)	(204)	(256)	(428)	(536)	(438)	(446)	(468)	(554)	(651)	
CPS-SSA-3	1,207	−3,715	−3,150	−1,122	−812	2,603	2,615	814	784	−99	1,246
	(166)	(226)	(325)	(311)	(452)	(307)	(328)	(305)	(429)	(481)	(720)
CPS-SSA-4	1,684	−1,189	−780	926	756	2,126	1,833	1,222	952	827	–
	(524)	(249)	(283)	(630)	(716)	(654)	(663)	(637)	(717)	(814)	

- For the experimental estimates, does it matter whether we use simple differences or DID?

Treatment effects with imperfect compliance

- Often, researchers can't perfectly control treatment. Subjects not intended to receive treatment may receive it anyway; subjects intended to receive treatment may not receive it.

Treatment effects with imperfect compliance

- Often, researchers can't perfectly control treatment. Subjects not intended to receive treatment may receive it anyway; subjects intended to receive treatment may not receive it.
  - In experiments such as medical trials, efforts are typically taken to limit non-compliance. We could avoid selection issues by estimating the effect of *intent to treat*, but this will frequently underestimate the causal effect of treatment with non-compliance.
  - Next semester, Professor Conlon will discuss treatment effects in more detail and introduce the LATE concept (Local Average Treatment Effect).

Treatment effects with imperfect compliance

- Often, researchers can't perfectly control treatment. Subjects not intended to receive treatment may receive it anyway; subjects intended to receive treatment may not receive it.
  - In experiments such as medical trials, efforts are typically taken to limit non-compliance. We could avoid selection issues by estimating the effect of *intent to treat*, but this will frequently underestimate the causal effect of treatment with non-compliance.
  - Next semester, Professor Conlon will discuss treatment effects in more detail and introduce the LATE concept (Local Average Treatment Effect).
- Imperfect compliance is closely related to *instrumental variables estimators*. In observational studies, there's sometimes a variable that we think is completely randomized. Example: draft lottery outcomes (randomized instrument) and military service (treatment status).