Statistics II – Lab
(Unofficial) Final Exam Review

TA: Gabriel Tarriba

18 April 2019

Hertie School of Governance

Which topic(s) is/are least clear?

- Basic concepts (POF, DAGs, ATE/NATE/LATE)
- OLS
- Matching
- Instrumental Variables
- Differences in Differences
- Fixed Effects
- Regression Discontinuity
- Mediators and moderators

Basic concepts (I)

- Causal inference
- Potential outcomes framework (POF)
- Individual treatment effect (ITE)
- Counterfactuals
- Directed Acyclical Graphs (DAGs)
- Mediators, colliders and confounders
- Back-door paths & Back-door criterion
- Average Treatment Effect (ATE)
- Naïve Average Treatment Effect (NATE)
- Average Treatment on the Treated (ATT)
- Decomposing the bias of NATE: baseline bias and differential treatment effect bias

Basic concepts (II)

- Stable Unit Treatment Value Assumption (SUTVA)
- Independence/Ignorability of treatment assignment
- Conditional Independence/Conditional of treatment assignment
- Selection bias
- Randomized treatment assignment
- Observational vs experimental studies
- Natural experiments
- Exogeneity vs Endogeneity

Regression

- Mechanics of OLS
- Slope of the line of best fit: cov(x,y)/var(x)
- Interpretation of the coefficients
- Omitted Variable Bias (OVB)
- Estimation of the bias of a coefficient when there is OVB
- Conditions under which regression can be used to estimate causal effects
- Criteria for selecting relevant covariates
- Post-treatment bias

Matching

- Conditional vs unconditional randomization
- Stratification
- Exact matching
- Common support
- Propensity score matching
- Balance tests
- Selection on observables
- Interpreting matching output in R

Instrumental variables

- Intention to treat (ITT) effect
- Instrumental variable (Z) versus treatment receipt indicator (D)
- IV assumptions:
  - Relevance
  - Exogeneity/ignorability of the instrument
  - Exclusion restriction
  - Monotonicity
- Principal strata (or compliance types): Compliers, Always-takers, never-takers, defiers
- Meaning of Local Average Treatment Effect (LATE)
- Estimation of LATE when Z, D and Y are binary variables (Cov(Y, Z)/Cov(D, Z))
- Estimation of LATE with Two-Stage Least-Squares
- Causal graphs perspective on IV
- Which of the assumptions are testable?

Differences-in-Differences

- Assumption of parallel trends – why is it important?
- Visual approach to DD
- Algebraic approach to DD
- Implementing DD in regression: what is the estimator of the treatment effect?
- Potential sources of parallel trends violations

Fixed Effects

- Panel data vs pooled cross section
- Error term decomposition in panel setup: unit fixed effects, time fixed effects, idiosyncratic error
- Time demeaning
- Within vs between variation
- Least Squares Dummy Variables (LSDV) estimation
- Two-way FE model
- How to include time-invariant variables in a FE model
- Limitations of FE models

Regression Discontinuity

- Meaning of forcing (or running) and treatment variables
- Treatment threshold (cut-off point)
- Local randomization
- Sharp RD (SRD) vs fuzzy RD (FRD)
- Assumption of continuity of average potential outcomes
- LATE at the threshold
- Bandwidth choice and bias-variance trade-off
- Coefficient interpretation in different types of model:
  - linear with common slope
  - linear with different slopes
  - non-linear
- Falsification checks: sensitivity, balance checks, placebo thresholds, sorting
- Fuzzy RD estimation: What it is and how to interpret threshold
- Interrupted time-series (ITS): how does it compare to sharp RD?
- Internal and external validity of RD

Mediators and moderators

- Distinction between mediators and moderators
- Conceptual decomposition of total causal effect into:
  - Direct effects (varying Treatment but holding mediator constant)
  - Indirect effects (holding Treatment constant, varying mediator)

Research design and method selection

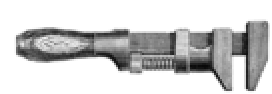
- **The toolbox is yours, but which tool to use?**
- Which methods allow you to estimate causal effects when there is selection on the observables?
- Which methods do not require selection on the observables?
- Which methods can be used for observational data?

Regression

Matching

IV

DID



FE

RD

Some important tables and graphs…

Types of bias: baseline and differential treatment effect (W02)

<u>Example: EU Funding and economic growth rate</u>

- Imagine we could actually observe the counterfactual:

Funding status (D)	Share of regions (%)	$E(Y^1)$	$E(Y^0)$
1	20	5	**4**
0	80	**2**	2

- **Given these (partly unobserved) values, what is the baseline bias?**
- **What is the differential treatment effect bias?**
- **How do we interpret these quantities?**

OLS estimation (W03)

- With multiple regression, we choose the line that minimizes:

$$\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \ldots - \hat{\beta}_k x_{ik})^2$$

- (Note the last 3 terms = $\hat{y}_i$)
- In bivariate regression, the slope of the line of best fit is conveniently

$$\hat{\beta}_1 = \frac{cov(x, y)}{var(x)} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

- The intercept can then be derived as

$$\hat{\beta}_0 = \bar{y} - \beta_1 \bar{x}$$

OVB - (W03)

- Suppose I estimate the following model:

$$ln(\text{CEO salary}) = \beta_0 + \beta_1 \text{Sales} + u,$$

- However, I believe that another control, CEO tenure, should have been included.
- **Is β_1 likely positively or negatively biased? Why/why not?**

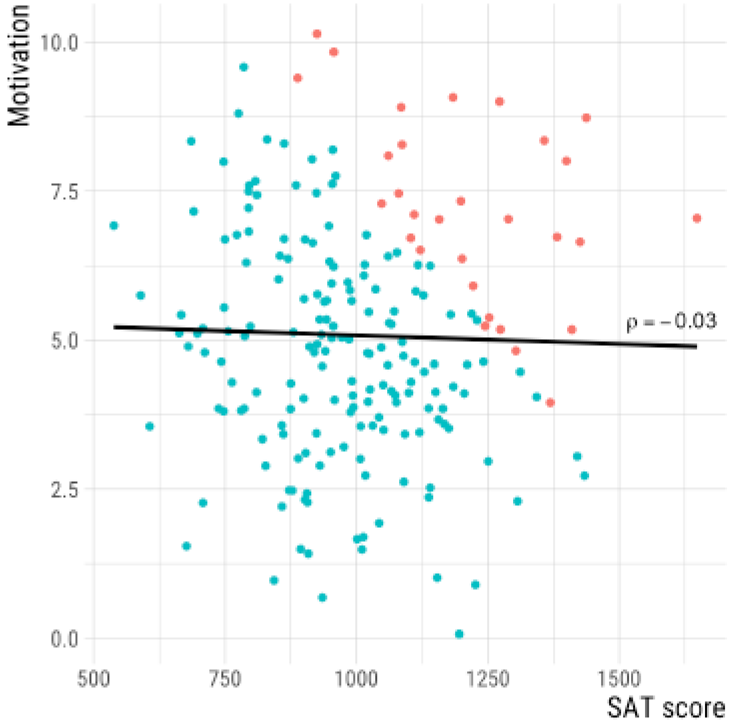
TABLE 3.2

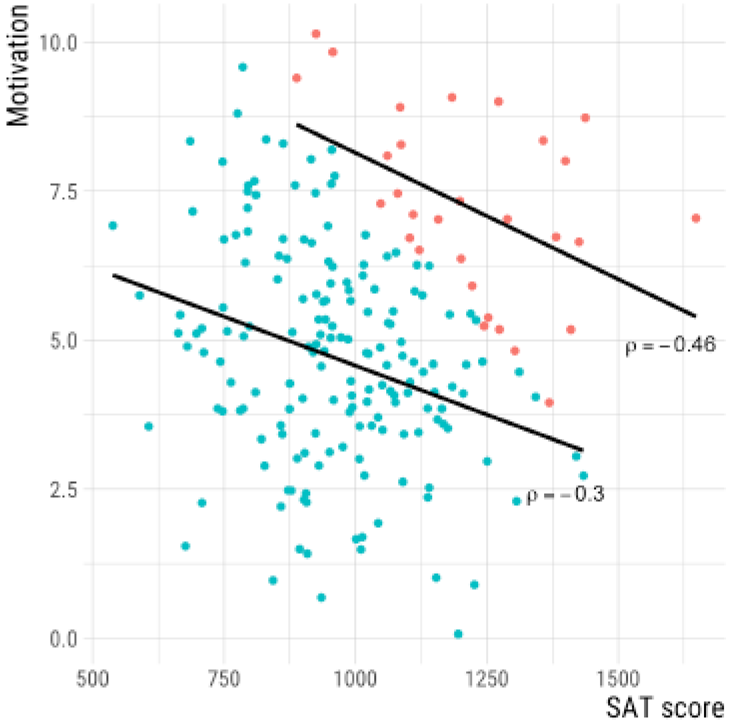
Summary of Bias in $\tilde{\beta}_1$ when x_2 Is Omitted in Estimating Equation (3.40)

	$\text{Corr}(x_1,x_2) > 0$	$\text{Corr}(x_1,x_2) < 0$
$\beta_2 > 0$	Positive bias	Negative bias
$\beta_2 < 0$	Negative bias	Positive bias

(See Wooldridge 5th ed. Ch. 3, pp. 86/87, "Omitted Variable Bias: The Simple Case")

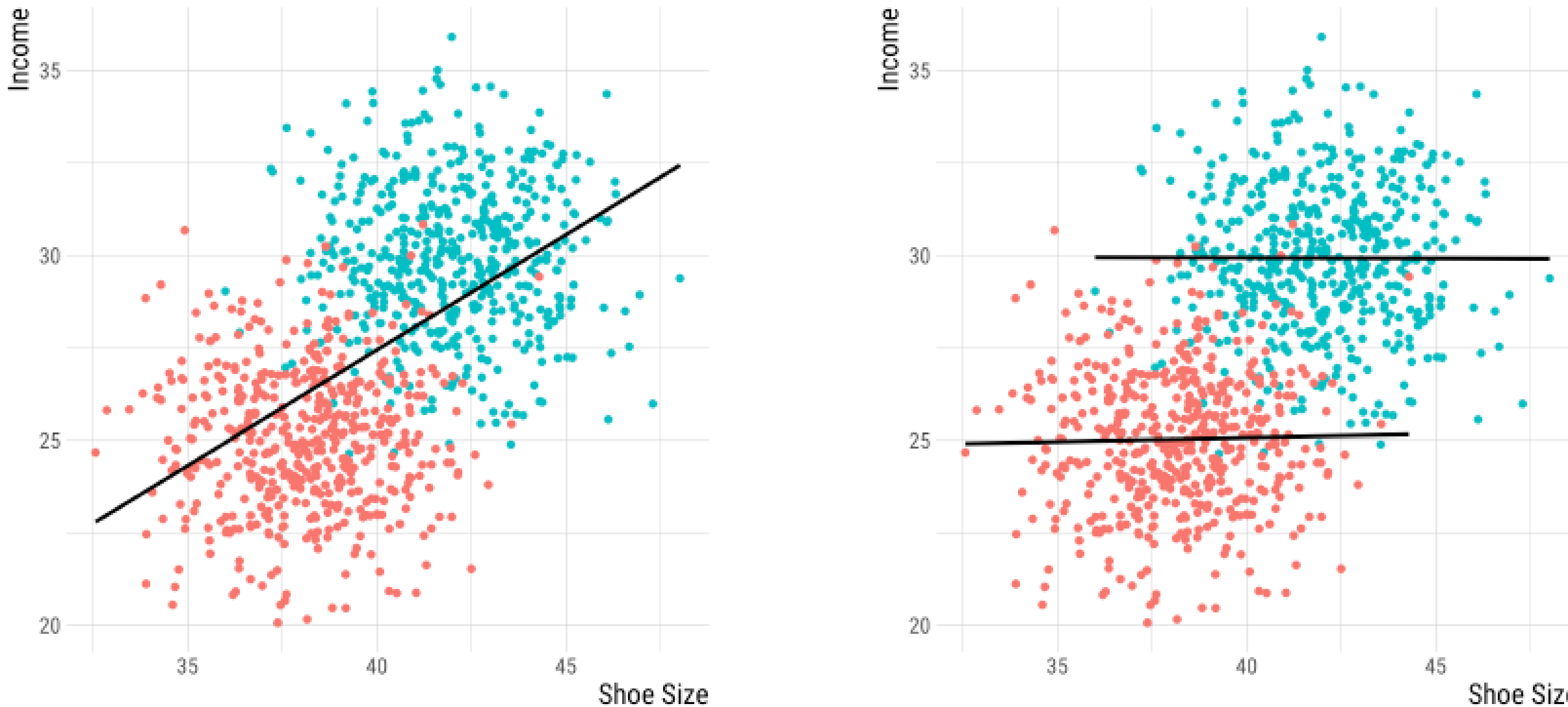
Conditioning on a collider (W03)





- *Y* SAT score, *D* motivation, *X* college admission (red: admitted, blue: rejected)
- These are simulated data

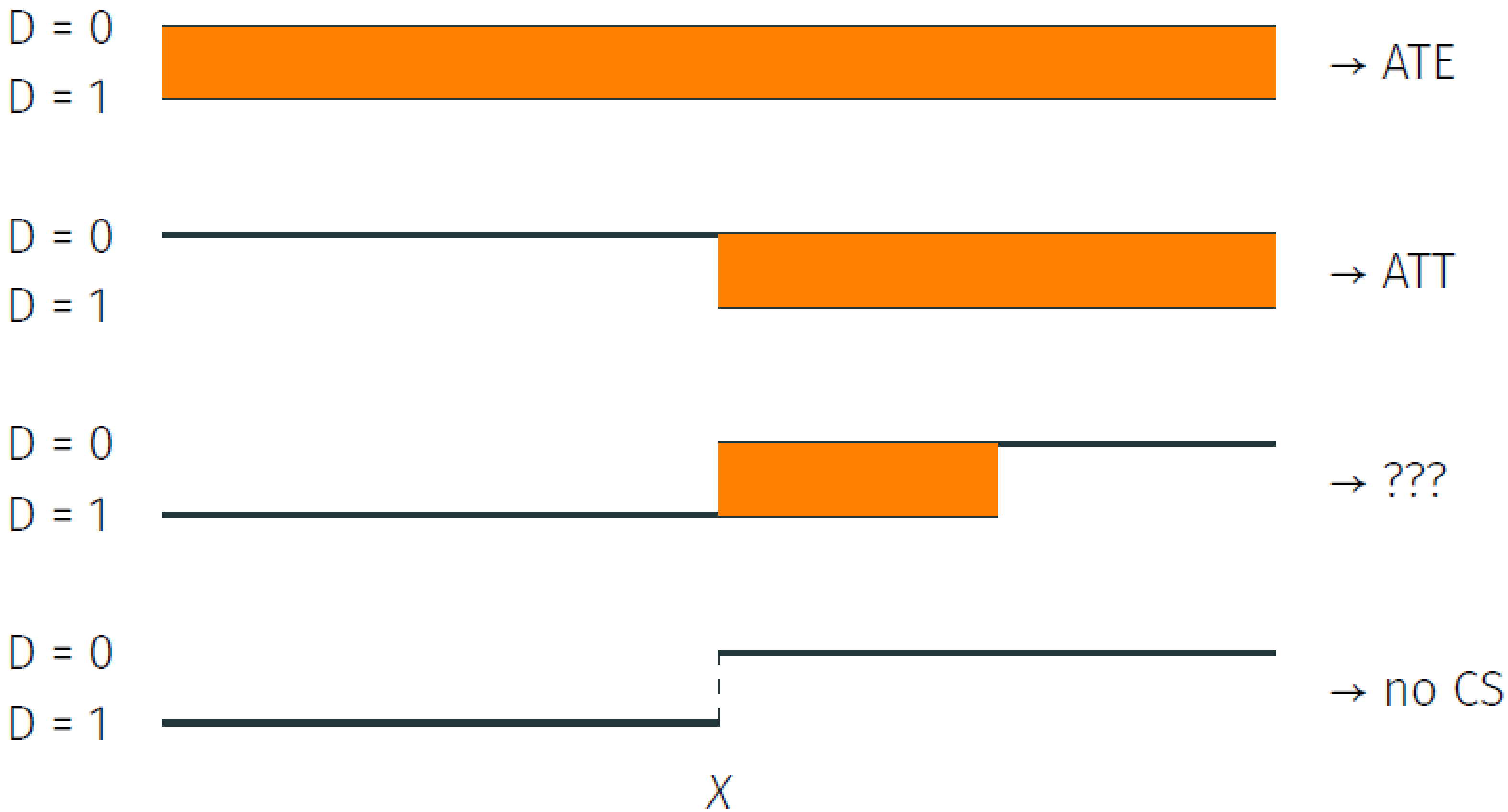
Conditioning on a confounder (W03)



- Y income, D shoe size, X gender (red: female, blue: male)
- These are simulated data

Common support in matching (W04)

Cases and interpretation



Matching (W04) – Catholic schooling and grades (from Simon's W04 script) – Balance tables

```
Call:
matchit(formula = catholic ~ race_white + w3income + p5hmage +
    p5numpla + w3momed_hsb, data = ecls_nomiss, method = "nearest")

Summary of balance for all data:
            Means Treated Means Control SD Control  Mean Diff  eQQ Med   eQQ Mean  eQQ Max
distance           0.1927        0.1379     0.0845     0.0549 5.67e-02     0.0548 7.60e-02
race_white         0.7411        0.5914     0.4916     0.1497 0.00e+00     0.1501 1.00e+00
w3income       82568.9357    55485.0210 43961.0872 27083.9146 2.50e+04 27069.1775 6.25e+04
p5hmage           39.5932       37.5658     6.5506     2.0274 2.00e+00     2.2544 7.00e+00
p5numpla           1.0917        1.1298     0.3910    -0.0380 0.00e+00     0.0399 2.00e+00
w3momed_hsb        0.2234        0.4609     0.4985    -0.2375 0.00e+00     0.2374 1.00e+00


Summary of balance for matched data:
            Means Treated Means Control SD Control Mean Diff eQQ Med  eQQ Mean  eQQ Max
distance           0.1927        0.1927     0.0846    0.0000       0    0.0000 3.30e-03
race_white         0.7411        0.7470     0.4349   -0.0059       0    0.0059 1.00e+00
w3income       82568.9357    81403.9926 46618.2406 1164.9430       0 1164.9430 6.25e+04
p5hmage           39.5932       39.5503     5.2243    0.0429       0    0.0873 1.00e+01
p5numpla           1.0917        1.0762     0.2970    0.0155       0    0.0200 2.00e+00
w3momed_hsb        0.2234        0.2152     0.4111    0.0081       0    0.0081 1.00e+00

Percent Balance Improvement:
            Mean Diff. eQQ Med eQQ Mean  eQQ Max
distance       99.9934     100  99.9689  95.6780
race_white     96.0477       0  96.0591   0.0000
w3income       95.6988     100  95.6964   0.0000
p5hmage        97.8841     100  96.1286 -42.8571
p5numpla       59.1653       0  50.0000   0.0000
w3momed_hsb    96.5746       0  96.5732   0.0000

Sample sizes:
          Control Treated
All          7915    1352
Matched      1352    1352
Unmatched    6563       0
Discarded       0       0
```

Dependent variable is not shown!

IV 2SLS estimation: ivreg output

From the Sesame Street example (assignment 4):
Effect of watching the show regularly on letter recognition

Y D Z

```
Call:
ivreg(formula = postlet ~ regular | encour, data = dat)

Residuals:
    Min      1Q  Median      3Q     Max 
-20.593  -9.593  -4.527  10.723  34.473 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   20.593      3.659   5.628 5.11e-08 ***
regular        7.934      4.606   1.723   0.0863 .  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1 

Residual standard error: 12.46 on 238 degrees of freedom
Multiple R-Squared: 0.1355,     Adjusted R-squared: 0.1318 
Wald test: 2.967 on 1 and 238 DF,  p-value: 0.08626 
```

2SLS estimate

First-stage: Regress D on Z, save fitted values D-hat

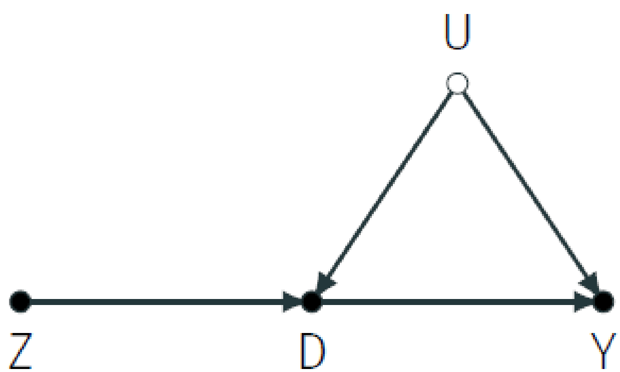
Second-stage: Regress Y on D-hat

Coefficient on regular = LATE

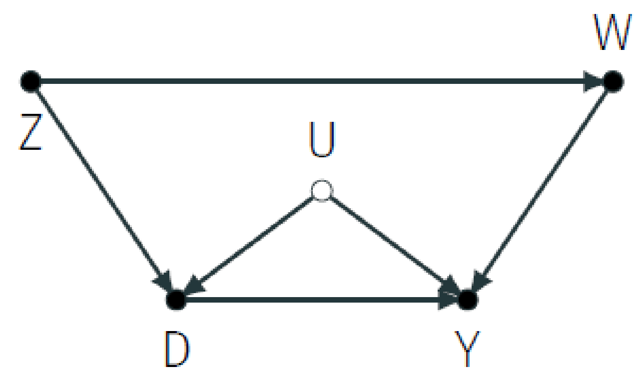
IV (W05)

Z as a...

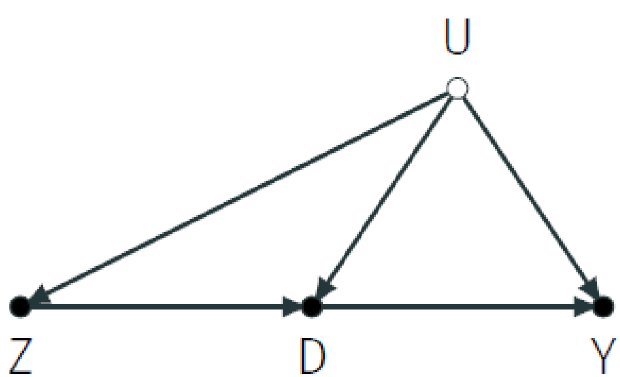
valid instrument for *D*



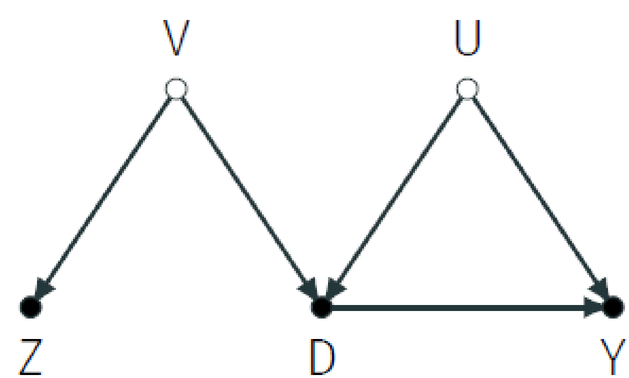
valid conditional instrument for *D*



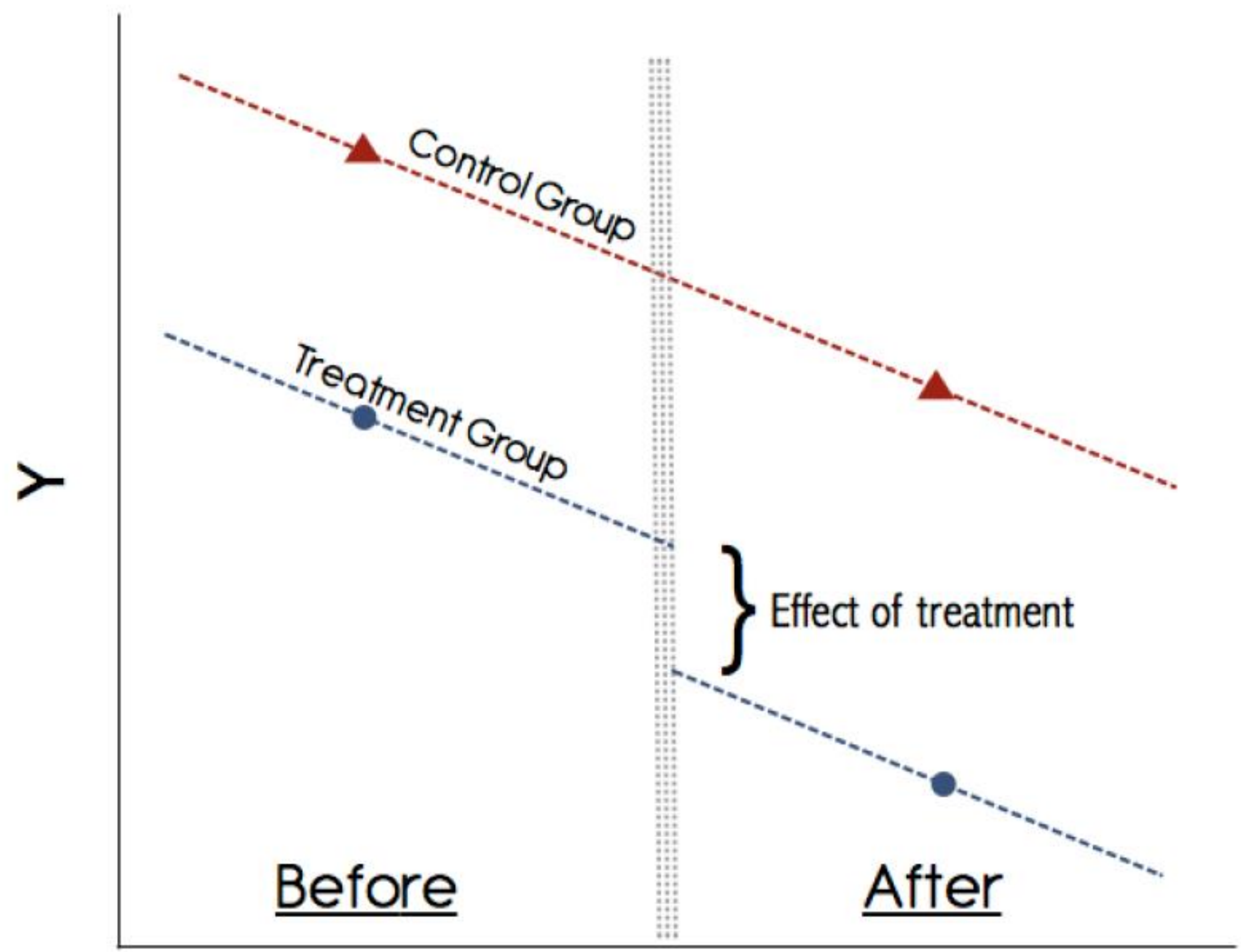
invalid instrument for *D*



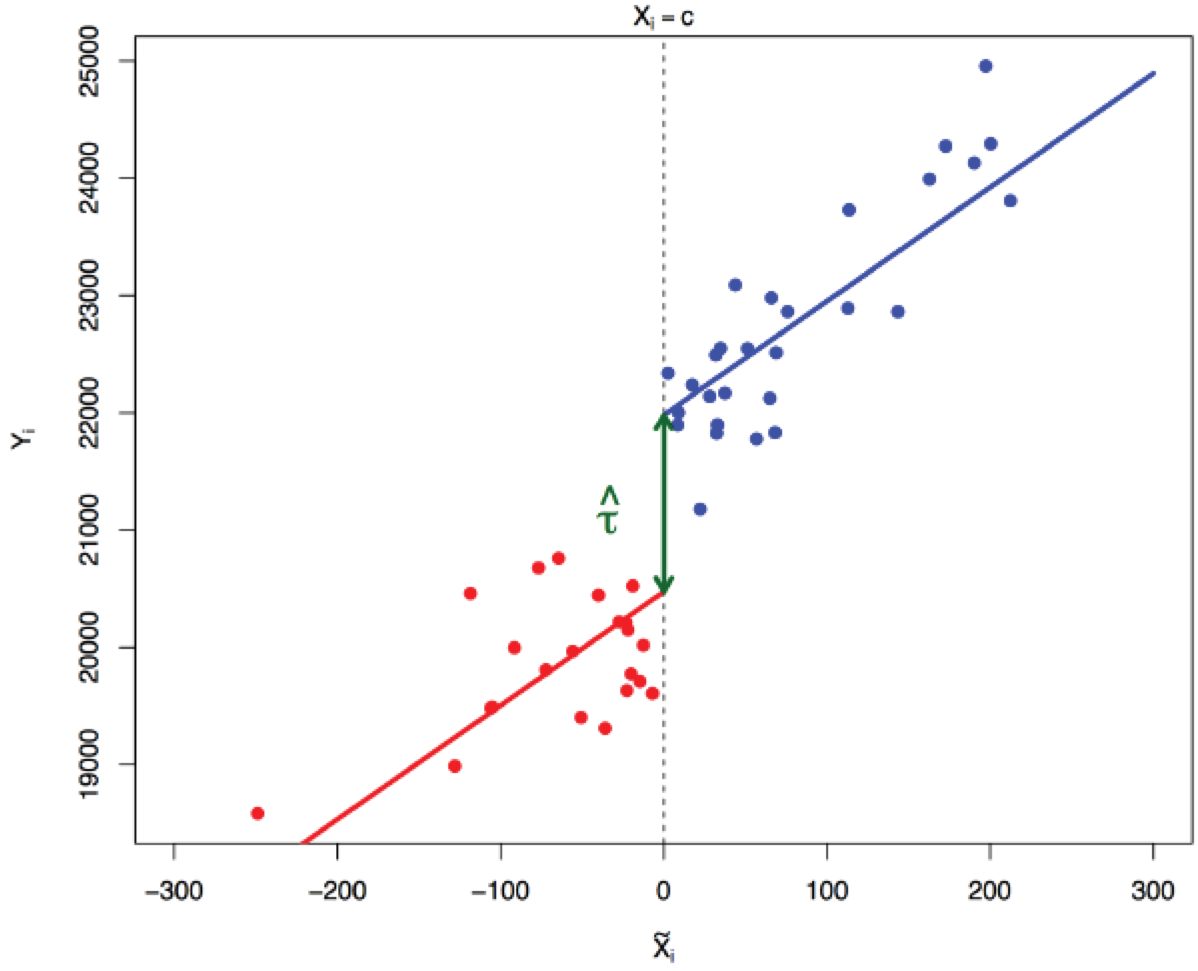
valid surrogate instrument for *D*



Diff-in-Diff (W06)

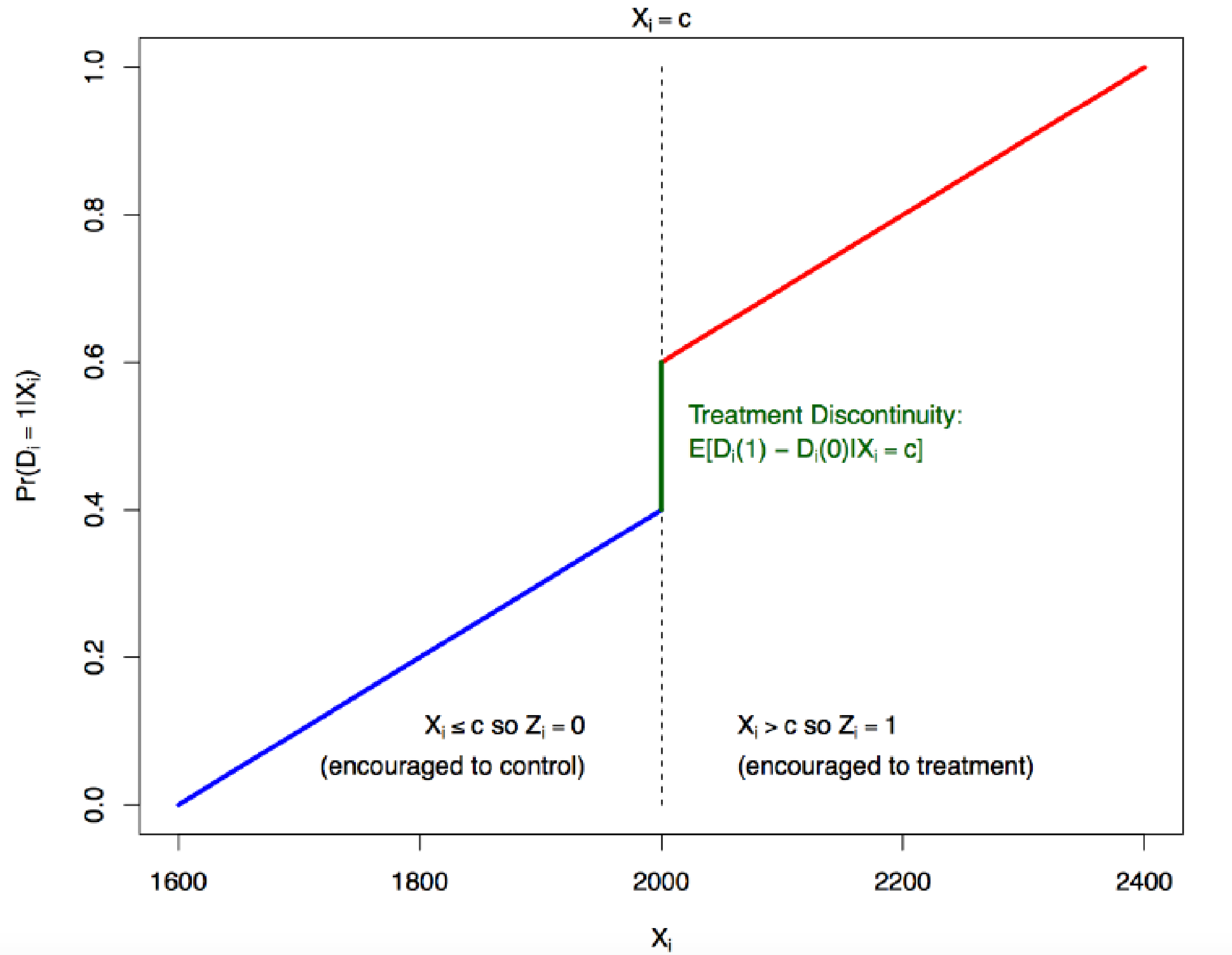


Sharp RD (W07)



$X_i = c$
25000
24000
23000
22000
21000
20000
19000
Y_i
$\hat{\tau}$
−300
−200
−100
0
100
200
300
$\tilde{X}_i$

Fuzzy RD (W07)



$X_i = c$
$Pr(D_i = 1|X_i)$
0.0
0.2
0.4
0.6
0.8
1.0
Treatment Discontinuity:
$E[D_i(1) - D_i(0)|X_i = c]$
$X_i \leq c$ so $Z_i = 0$
(encouraged to control)
$X_i > c$ so $Z_i = 1$
(encouraged to treatment)
1600
1800
2000
2200
2400
X_i

Mediators vs moderators (W08)

- Mediators are **intermediate variables** between the causal variable of interest and the outcome
- A moderator is a variable that affects the direction and/or strength of the relationship between the causal variable of interest and the outcome

Questions about application papers?

Matching (W04, slides 24-26)

Example: Untangling the Causal Effects of Sex on Judging

Application Paper!

Untangling the Causal Effects of Sex on Judging

Christina L. Boyd University at Buffalo, SUNY
Lee Epstein Northwestern University School of Law
Andrew D. Martin Washington University in St. Louis

We explore the role of sex in judging by addressing two questions of long-standing interest to political scientists: whether and in what ways male and female judges decide cases distinctly—"individual effects"—and whether and in what ways serving with a female judge causes males to behave differently—"panel effects." While we attend to the dominant theoretical accounts of why we might expect to observe either or both effects, we do not use the predominant statistical tools to assess them. Instead, we deploy a more appropriate methodology: semiparametric matching, which follows from a formal framework for causal inference. Applying matching methods to 13 areas of law, we observe consistent gender effects in only one—sex discrimination. For these disputes, the probability of a judge deciding in favor of the party alleging discrimination decreases by about 10 percentage points when the judge is a male. Likewise, when a woman serves on a panel with men, the men are significantly more likely to rule in favor of the rights litigant. These results are consistent with an informational account of gendered judging and are inconsistent with several others.

Example: Untangling the Causal Effects of Sex on Judging

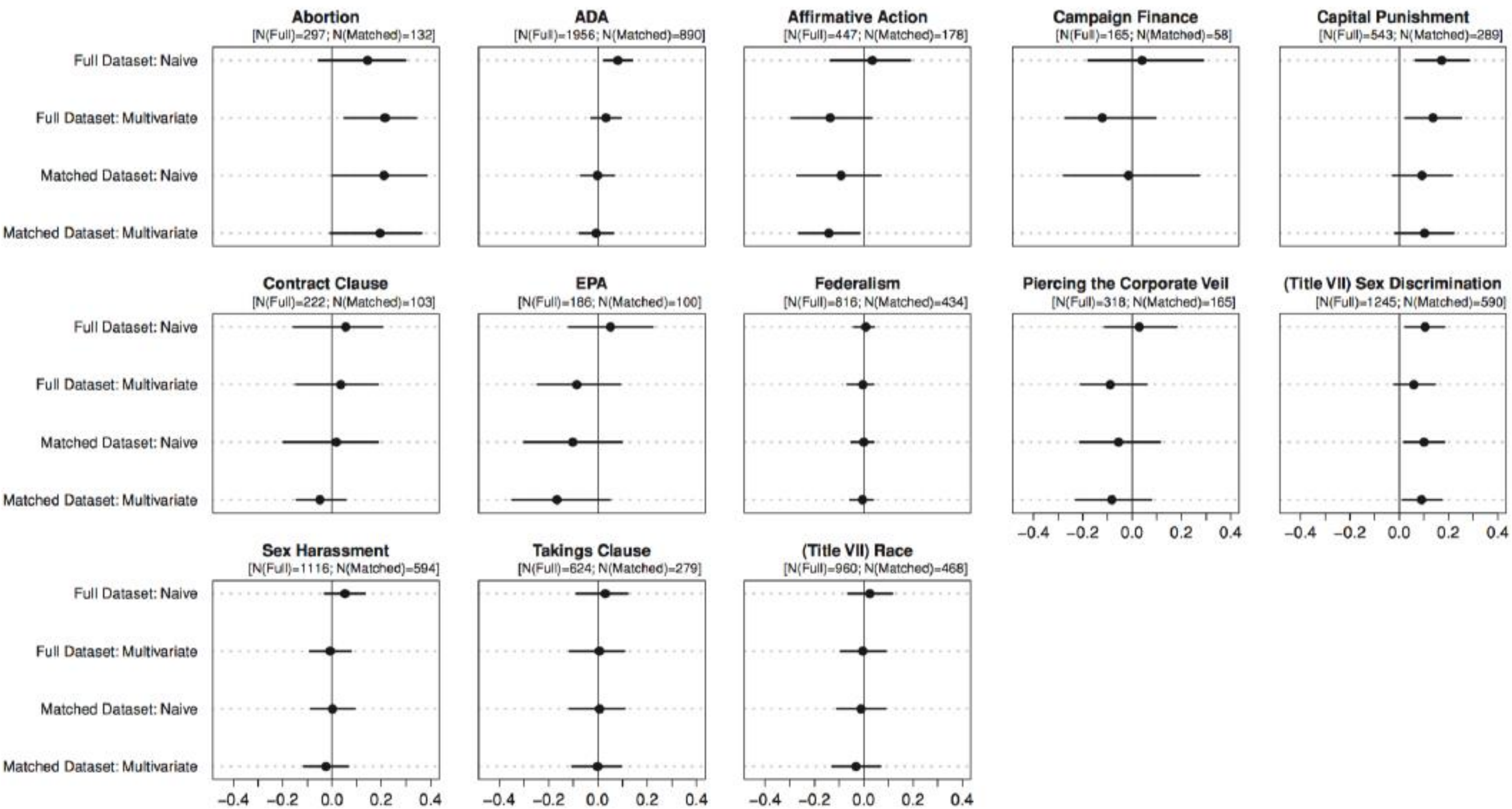
TABLE 2 **Matching Summary Statistics for the Individual Effects Analyses for ADA and Title VII Sex Discrimination Cases**

	ADA Cases						
	Full Data ($N = 1956$)				Matched Data ($N = 890$)		
Variable	**Mean Treated**	**Mean Control**	**eQQ Med**	**Percent Reduction**	**Mean Treated**	**Mean Control**	**eQQ Med**
Propensity Score	0.32	0.13	0.19	94.89	0.32	0.31	0.09
Minority Judge	0.09	0.11	0.00	·	0.09	0.12	0.00
Judicial Experience	0.47	0.47	0.00	·	0.47	0.48	0.00
Judicial Common Space	−0.17	0.06	0.17	98.04	−0.17	−0.17	0.06
Confirmation Year	1991.14	1985.17	5.00	92.60	1991.14	1990.70	2.00
	(Title VII) Sex Discrimination Cases						
	Full Data ($N = 1245$)				Matched Data ($N = 590$)		
Variable	**Mean Treated**	**Mean Control**	**eQQ Med**	**Percent Reduction**	**Mean Treated**	**Mean Control**	**eQQ Med**
Propensity Score	−1.13	−2.75	1.58	91.67	−1.13	−1.27	0.57
Minority Judge	0.12	0.09	0.00	30.39	0.12	0.14	0.00
Judicial Experience	0.45	0.45	0.00	·	0.45	0.43	0.00
Judicial Common Space	−0.12	0.10	0.16	81.48	−0.12	−0.08	0.11
Confirmation Year	1990.38	1984.58	6.00	98.12	1990.38	1990.27	2.00

The left portion of each table provides results for the full, unmatched data, while the right portion displays results after matching has taken place. eQQ med is the median difference in the empirical quantile-quantile plot (an eQQ med of zero is ideal).

Example: Untangling the Causal Effects of Sex on Judging

FIGURE 4 Dotplots of Average Treatment Effects (ATEs) for Individual Effects Across 13 Issue Areas



The lines represent 95% confidence intervals for the average treatment effect. For every issue area, the first two models are logistic regression models fit to each full, unbalanced dataset. The naive model includes only the judge's sex as a covariate. The other model includes the judge's sex and a number of controls, including ideology. The next two models show the ATE after nearest-neighbor matching with replacement on the estimated propensity score. The first is for a difference of proportions analysis. The second is for a logistic regression model with the judge's sex and a number of controls including ideology.

Example: Untangling the Causal Effects of Sex on Judging

Table A2 Logistic Regression Estimates for the Title VII Sex Discrimination Cases, Individual and Panel Effects

	Individual Effects				Panel Effects			
Covariates	**Full: Naive**	**Full: Multi-variate**	**Matched: Naive**	**Matched: Multi-variate**	**Full: Naive**	**Full: Multi-variate**	**Matched: Naive**	**Matched: Multi-variate**
(Intercept)	−0.68*	12.68	−0.66*	72.97*	−0.83*	3.94	−0.93*	7.59
	(0.06)	(12.78)	(0.10)	(22.22)	(0.08)	(13.59)	(0.09)	(15.11)
Treatment	0.44*	0.28	0.42*	0.46*	0.54*	0.65*	0.63*	0.72*
	(0.17)	(0.20)	(0.19)	(0.22)	(0.14)	(0.15)	(0.15)	(0.16)
Judge Ideology		−0.79*		−1.06*		−0.79*		−0.75*
		(0.21)		(0.31)		(0.23)		(0.26)
Year of Birth		−0.01		−0.04*		−0.00		−0.01
		(0.01)		(0.01)		(0.01)		(0.01)
Minority Judge		0.32		0.35		0.32		0.65*
		(0.21)		(0.27)		(0.23)		(0.30)
Lower Court Direction		1.08*		1.12*		1.10*		1.03*
		(0.14)		(0.24)		(0.15)		(0.18)
Circuit Ideology		−0.11		−0.26		−0.05		−0.03
		(0.30)		(0.40)		(0.33)		(0.36)
Female Maj. Opin. Writer		0.46*		0.51*				
		(0.18)		(0.23)				
Standard errors in parentheses; *p < 0.05								
N:	1245	1245	590	590	1075	1075	843	843
Log-Likelihood:	−797.42	−700.10	−338.49	−255.48	−673.83	−590.15	−508.98	−420.95

Average treatment effects reported in Figures 4 and 6 are derived from these estimates. Standard errors are in parentheses. To conserve space, estimates of year fixed effects are not reported. The naive models include only the treatment (for individual effects a female judge, for panel effects a mixed-sex panel) as a covariate. The other models include the treatment, ideology, and other reported covariates. Similar regression tables for the 12 other issue areas are reported in the online appendix.

Diff in Diff (W06, slides 33-37)

Application: Hitler's speeches and NSDAP success

Examining a Most Likely Case for Strong Campaign Effects: Hitler's Speeches and the Rise of the Nazi Party, 1927–1933

PETER SELB *University of Konstanz*
SIMON MUNZERT *Hertie School of Governance*

Hitler's rise to power amidst an unprecedented propaganda campaign initiated scholarly interest in campaign effects. To the surprise of many, empirical studies often found minimal effects. The predominant focus of early work was on U.S. elections, though. Nazi propaganda as the archetypal and, in many ways, most likely case for strong effects has rarely been studied. We collect extensive data about Hitler's speeches and gauge their impact on voter support at five national elections preceding the dictatorship. We use a semi-parametric difference-in-differences approach to estimate effects in the face of potential confounding due to the deliberate scheduling of events. Our findings suggest that Hitler's speeches, while rationally targeted, had a negligible impact on the Nazis' electoral fortunes. Only the 1932 presidential runoff, an election preceded by an extraordinarily short, intense, and one-sided campaign, yielded positive effects. This study questions the importance of charismatic leaders for the success of populist movements.

Application: Hitler's speeches and NSDAP success

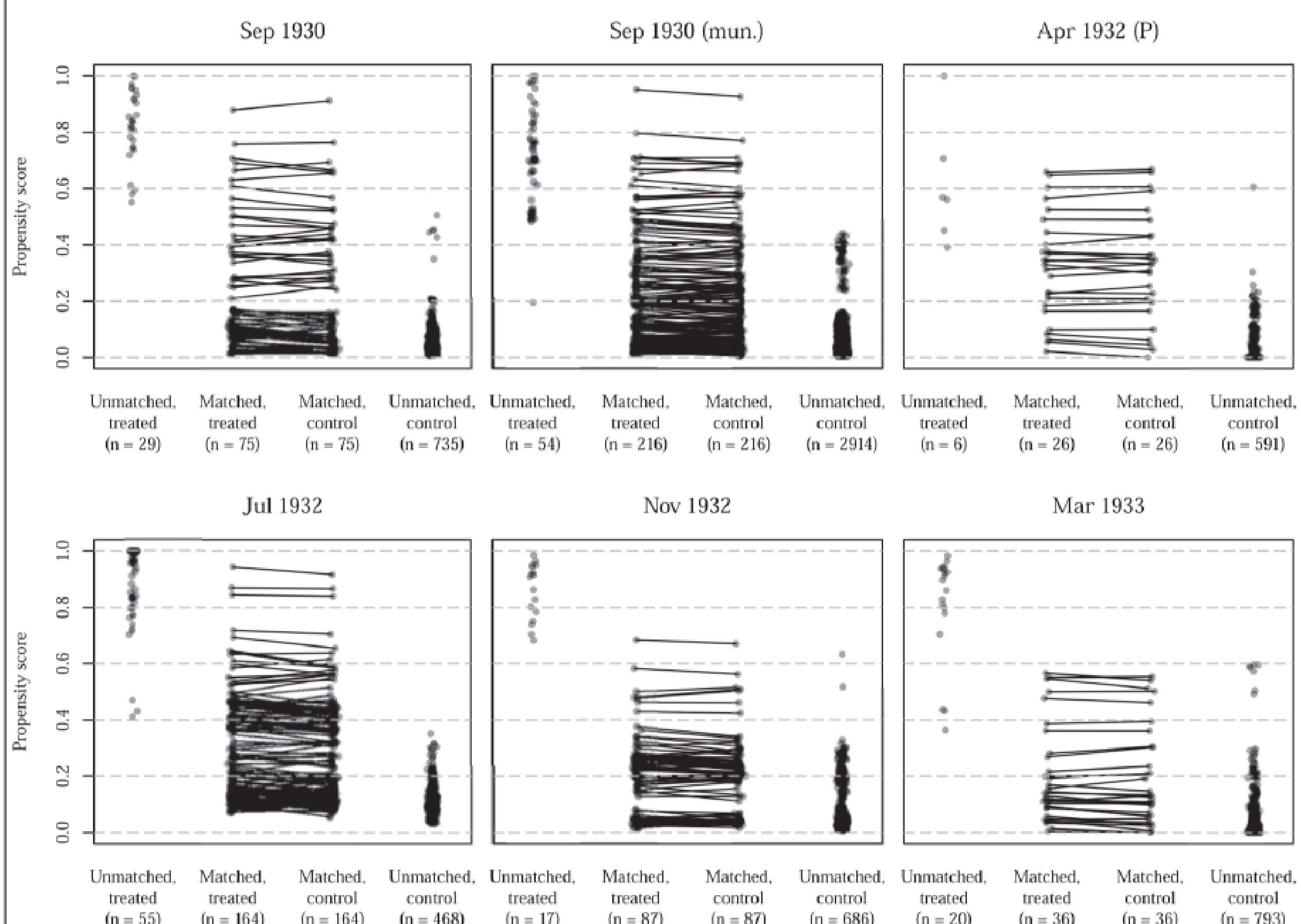
TABLE 1. Probit estimates of Hitler appearances by election. Standard errors in parentheses.

	Sep 1930	Sep 1930 (mun.)	Apr 1932 (P)	Jul 1932	Nov 1932	Mar 1933
Competitiveness 1	0.311	0.737***		−0.653*	−0.294	1.251*
	(0.291)	(0.153)		(0.386)	(0.424)	(0.748)
Competitiveness 2	0.668***	0.770***		0.333*	−0.256	0.094
	(0.226)	(0.125)		(0.179)	(0.209)	(0.310)
Organizational strength	−0.704	−0.616***	−0.166	−0.106	−0.257*	−0.240
	(0.513)	(0.239)	(0.260)	(0.190)	(0.155)	(0.173)
Distance to nearest airfield	−0.323	0.050	−1.943***	−0.334*	−0.367	−3.574***
	(0.282)	(0.164)	(0.612)	(0.190)	(0.235)	(0.780)
Number of eligibles	0.695***	0.736***	0.196	1.200***	0.593***	0.577**
	(0.164)	(0.103)	(0.216)	(0.270)	(0.214)	(0.240)
Previous NSDAP vote share	3.394	2.135**		0.579	0.989*	0.158
	(2.146)	(0.945)		(0.847)	(0.589)	(1.088)
Previous Hitler vote share			−5.544***			
			(1.823)			
Previous appearance	0.844***	0.886***	5.719	0.985***	0.967***	0.418**
	(0.185)	(0.086)	(240.510)	(0.129)	(0.129)	(0.199)
Goebbels appearance	1.077***	1.189***	7.513	0.708***	1.165***	1.618***
	(0.200)	(0.112)	(6,609.109)	(0.207)	(0.381)	(0.293)
(Intercept)	−2.269***	−2.715***	−4.750	−1.161***	−1.714***	−2.337***
	(0.249)	(0.140)	(240.510)	(0.328)	(0.425)	(0.750)
Mc-Fadden's Pseudo R2	0.31	0.27	0.51	0.24	0.23	0.44
Observations	1,000	3,864	685	1,000	953	953
Log Likelihood	−229.763	−712.633	−63.317	−398.459	−252.176	−118.828
Akaike Inf. Crit.	477.526	1,443.267	140.633	814.918	522.353	255.655

Note: *$p<0.1$; **$p<0.05$; ***$p<0.01$

Application: Hitler's speeches and NSDAP success

FIGURE 3. Predicted propensity scores by exposure and matching status. Lines indicate matched pairs.



Application: Hitler's speeches and NSDAP success

TABLE 2. Propensity score and covariate balance before and after matching. Mean differences on variables reported.

	Sep 1930			Sep 1930 (mun.)			Apr 1932 (P)		
Variable names	Before	After	% Impr.	Before	After	% Impr.	Before	After	% Impr.
Propensity score	0.32	0.01	97	0.24	0.01	97	0.35	0.00	100
Competitiveness 1	0.05	0.05	–8	0.10	0.05	48			
Competitiveness 2	0.16	0.09	44	0.16	0.03	84			
Organizational strength	0.18	–0.02	90	0.09	0.01	92	0.40	0.14	66
Distance to nearest airfield	–0.22	–0.09	59	–0.14	–0.03	77	–0.28	–0.03	90
Number of eligibles	0.78	–0.06	93	0.33	–0.01	98	0.85	0.26	69
Previous NSDAP vote share	0.01	0.00	78	0.01	0.00	75	–0.06	0.02	75
Previous appearance	0.27	–0.05	80	0.36	0.04	89	0.79	0.04	95
Goebbels appearance	0.41	0.03	93	0.34	–0.01	97	0.03	0.00	100

Application: Hitler's speeches and NSDAP success

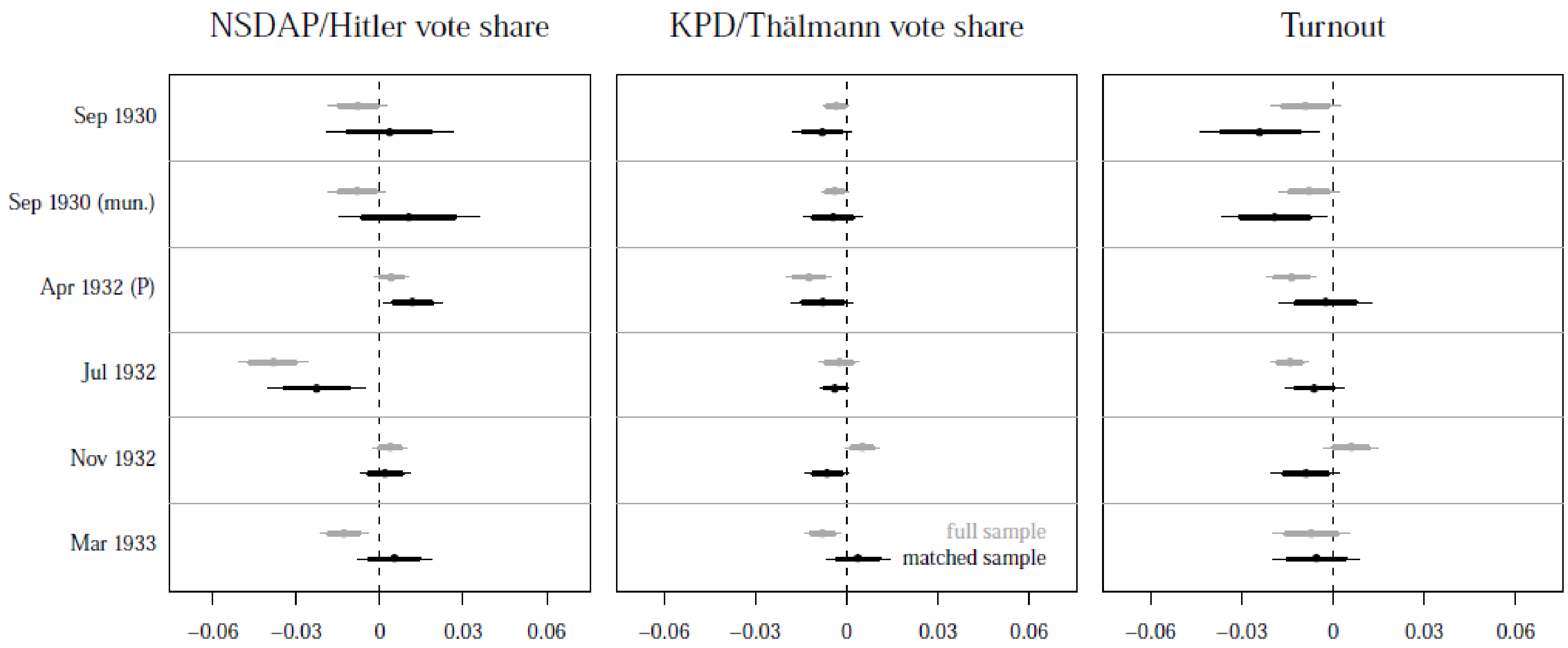
TABLE 2. Propensity score and covariate balance before and after matching. Mean differences on variables reported.

	Sep 1930			Sep 1930 (mun.)			Apr 1932 (P)		
Variable names	Before	After	% Impr.	Before	After	% Impr.	Before	After	% Impr.
Propensity score	0.32	0.01	97	0.24	0.01	97	0.35	0.00	100
Competitiveness 1	0.05	0.05	−8	0.10	0.05	48			
Competitiveness 2	0.16	0.09	44	0.16	0.03	84			
Organizational strength	0.18	−0.02	90	0.09	0.01	92	0.40	0.14	66
Distance to nearest airfield	−0.22	−0.09	59	−0.14	−0.03	77	−0.28	−0.03	90
Number of eligibles	0.78	−0.06	93	0.33	−0.01	98	0.85	0.26	69
Previous NSDAP vote share	0.01	0.00	78	0.01	0.00	75	−0.06	0.02	75
Previous appearance	0.27	−0.05	80	0.36	0.04	89	0.79	0.04	95
Goebbels appearance	0.41	0.03	93	0.34	−0.01	97	0.03	0.00	100

TABLE 3. Propensity score and covariate balance before and after matching, *continued*. Mean differences on variables reported.

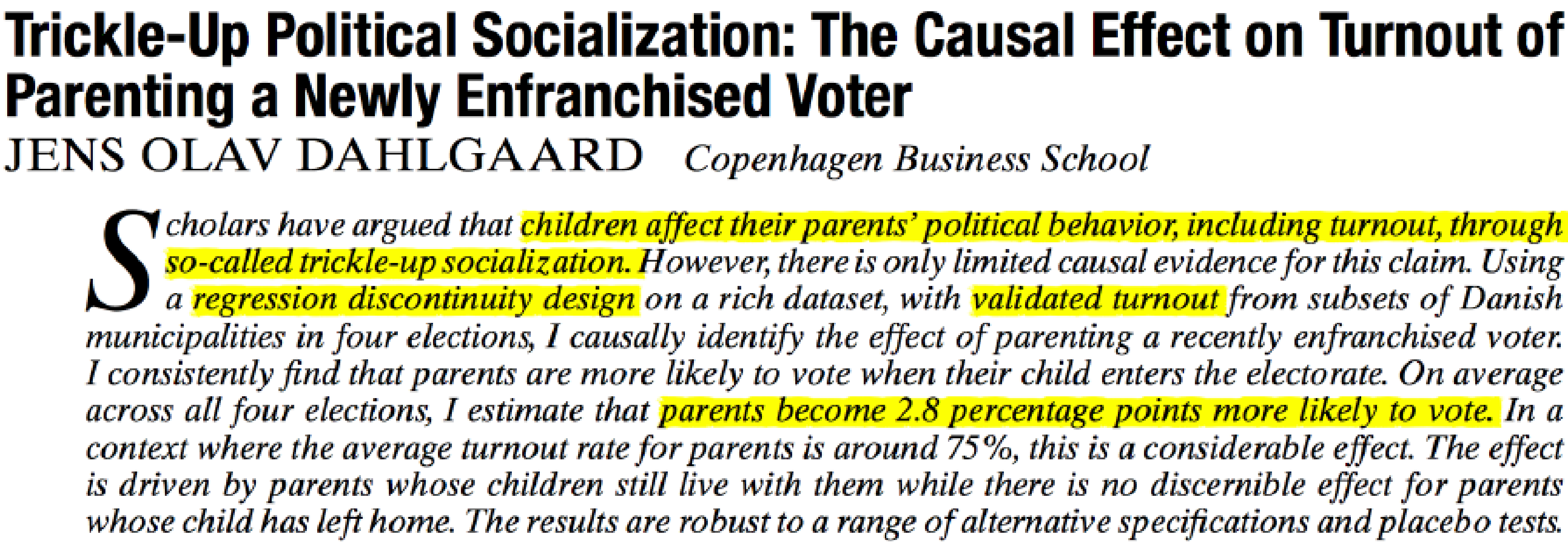
	July 1932			November 1932			Mar 1933		
Variable names	Before	After	% Impr.	Before	After	% Impr.	Before	After	% Impr.
Propensity score	0.28	0.01	96	0.23	0.00	99	0.37	0.01	98
Competitiveness 1	−0.02	0.00	94	0.00	0.03	−5479	−0.01	0.00	42
Competitiveness 2	0.03	−0.02	21	0.00	0.01	−248	0.02	−0.06	−234
Organizational strength	0.54	0.04	93	0.60	0.04	93	0.93	−0.11	88
Distance to nearest airfield	−0.16	−0.05	70	−0.19	−0.01	96	−0.32	−0.02	93
Number of eligibles	0.55	0.06	90	0.66	0.01	99	1.09	−0.14	87
Previous NSDAP vote share	0.00	0.00	92	0.00	0.01	−135	−0.03	0.01	76
Previous appearance	0.36	−0.05	85	0.51	−0.01	98	0.49	0.14	72
Goebbels appearance	0.20	0.04	82	0.13	0.01	91	0.36	0.03	92

Application: Hitler's speeches and NSDAP success



RDD (W07, slides 36-38)

Application: Trickle-up political socialization



Trickle-Up Political Socialization: The Causal Effect on Turnout of Parenting a Newly Enfranchised Voter

JENS OLAV DAHLGAARD *Copenhagen Business School*

Scholars have argued that children affect their parents' political behavior, including turnout, through so-called trickle-up socialization. However, there is only limited causal evidence for this claim. Using a regression discontinuity design on a rich dataset, with validated turnout from subsets of Danish municipalities in four elections, I causally identify the effect of parenting a recently enfranchised voter. I consistently find that parents are more likely to vote when their child enters the electorate. On average across all four elections, I estimate that parents become 2.8 percentage points more likely to vote. In a context where the average turnout rate for parents is around 75%, this is a considerable effect. The effect is driven by parents whose children still live with them while there is no discernible effect for parents whose child has left home. The results are robust to a range of alternative specifications and placebo tests.

Application: Trickle-up political socialization

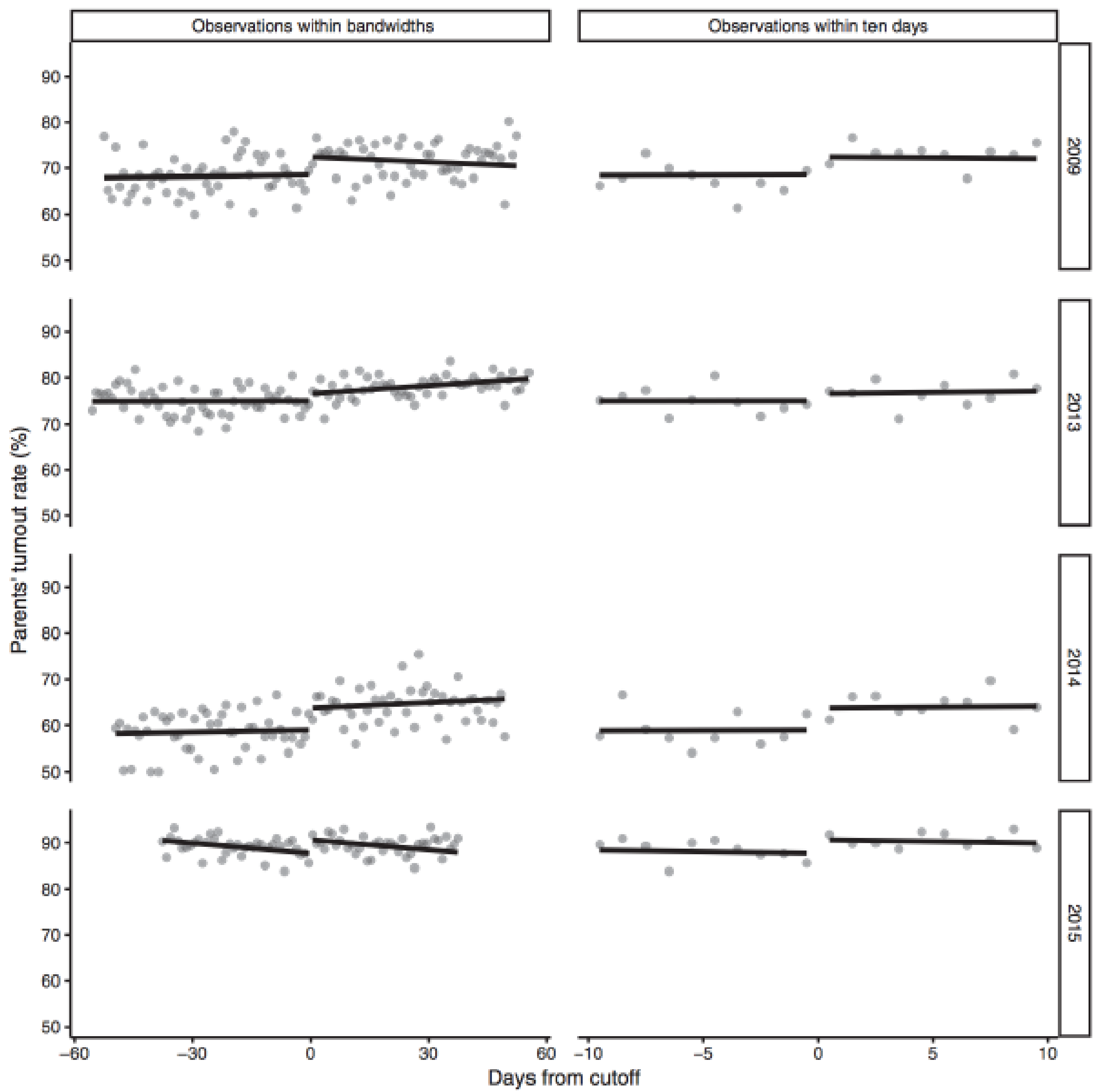
TABLE 1. Means within 6 Months of Cutoff

	2009		2013		2014		2015		Pooled	
Child is	Ineligible	Eligible	Ineligible	Eligible	Ineligible	Eligible	Ineligible	Eligible	Ineligible	Eligible
Parent voted (%)	69.04	72.04	75.81	78.60	58.36	62.87	88.72	89.65	74.48	77.31
Parent is native (%)[a]	86.23	86.85	87.50	88.61	92.34	92.01	92.81	92.22	89.54	89.81
Parent's age (Years)	43.69	44.11	44.36	44.80	44.60	45.09	44.78	45.15	44.39	44.81
Parent's income (1000 euro)	50.79	50.98	58.30	57.48	61.94	62.04	61.24	61.91	58.34	58.23
Parent is female (%)	51.16	51.08	51.09	51.08	50.78	51.06	51.04	51.01	51.03	51.06
Parent has bachelor's degree (%)	26.80	27.04	26.80	26.82	30.77	29.86	31.29	29.99	28.70	28.20
Child is female (%)	49.22	48.88	48.24	48.25	48.42	48.41	48.75	48.48	48.59	48.45
Child is oldest (%)	46.87	48.68	44.79	44.43	46.25	45.80	44.49	43.34	45.38	45.18
N	32,425	32,260	63,539	68,464	33,197	32,117	44,044	44,238	173,205	177,079

Note: N varies across elections as the proportion of voters for whom turnout was recorded varies. [a]The proportion of native Danes is higher in the 2014 and 2015 elections because of different requirements for which nationalities are eligible to vote.

Application: Trickle-up political socialization

FIGURE 1. Turnout over Child's Age in Days



Application: Trickle-up political socialization

TABLE 2. Effect on Turnout of Parenting an Eligible Voter

	All municipalities with data in each election				
	2009	2013	2014	2015	Precision weighted average
Estimate	3.83	1.54	4.80	2.95	2.80
St.err.	(1.54)	(0.95)	(1.60)	(0.97)	(0.58)
CI	[0.82, 6.84]	[−0.33, 3.41]	[1.67, 7.93]	[1.04, 4.86]	[1.66, 3.94]
Robust estimate	4.07	1.20	4.40	3.36	2.79
Robust st.err.	1.84	1.10	1.89	1.09	0.67
Robust CI	[0.46, 7.67]	[−0.95, 3.35]	[0.7, 8.1]	[1.22, 5.50]	[1.48, 4.09]
$\text{Turnout}_{\textit{parents of ineligible}}$	68.22	75.14	58.40	88.99	73.45
$N_{\textit{ineligible}}$	8,668	18,513	9,264	9,883	46,330
$N_{\textit{eligible}}$	8,659	19,902	8,963	9,517	47,043
Bandwidth	53	56	50	38	
$\text{Bandwidth}_{\textit{Bias − correction}}$	81	97	79	73	

DAGs

DAGs

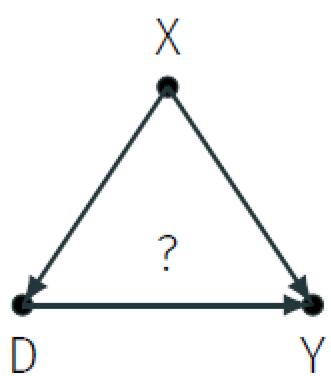
From “The book of why?”, Chapter 4

(a) In a chain junction, $A \to B \to C$, controlling for B prevents information about A from getting to C or vice versa.

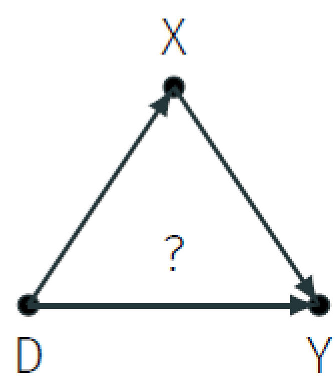
(b) Likewise, in a fork or confounding junction, $A \leftarrow B \to C$, controlling for B prevents information about A from getting to C or vice versa.

(c) Finally, in a collider, $A \to B \leftarrow C$, exactly the opposite rules hold. The variables A and C start out independent, so that information about A tells you nothing about C. But if you control for B, then information starts flowing through the “pipe,” due to the explain-away effect.

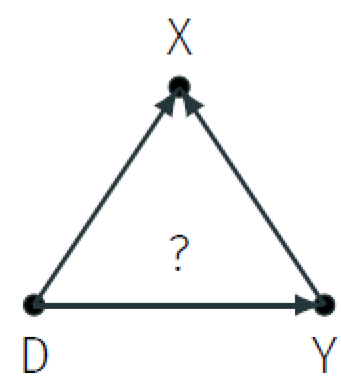
X is **confounder**.



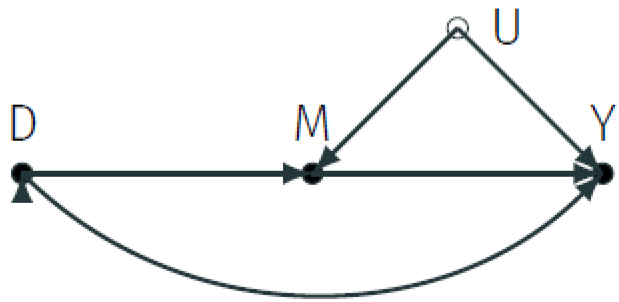
X is **mediator**.



X is **collider**.

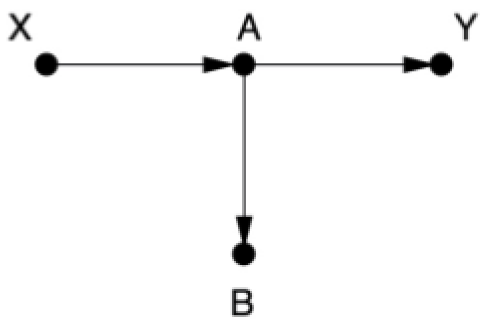


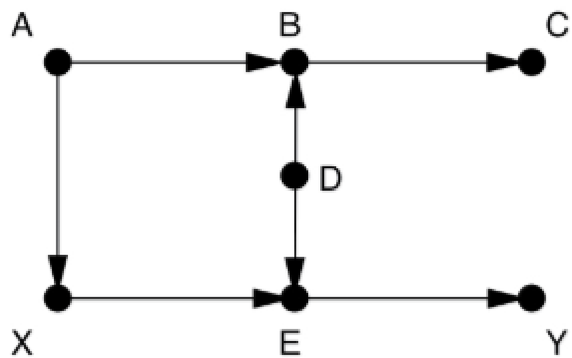
DAGs

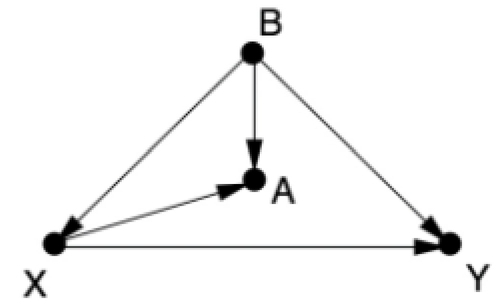


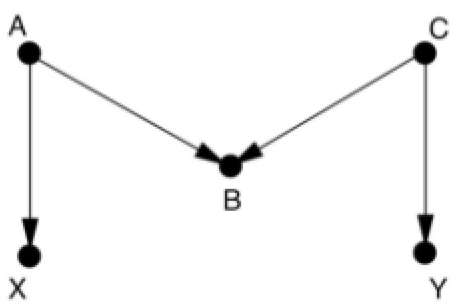
- **Are there back-door paths into D?**
  - No
- **What happens when I condition on M?**
  - I block a causal path 😱
  - I open up a non-causal path 😱
- **What kind of bias is introduced when controlling for M?**
  - Post-treatment bias

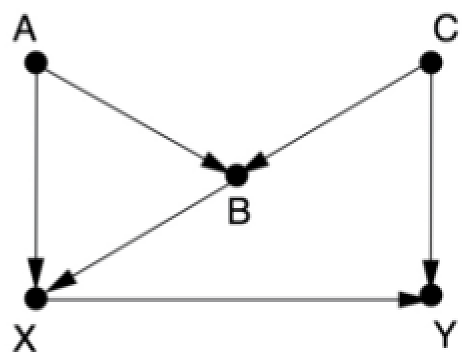
Let’s practice: identify the back-door criterion to deconfound the relationship between x and y











Notation (I)

$$ATE = E(Y^1) - E(Y^0)$$

$$NATE = E(Y^1|D=1) - E(Y^0|D=0)$$

$$\left.\begin{aligned} ATT &= E(Y^1|D=1) - E(Y^0|D=1) \\ ATC &= E(Y^1|D=0) - E(Y^0|D=0) \end{aligned}\right\} ATE = \pi ATT + (1-\pi) ATC$$

Ignorability/independence assumption

$$E(Y^1|D=1) = E(Y^1|D=0),\ E(Y^0|D=1) = E(Y^0|D=0) \quad \text{if } (Y^1, Y^0) \perp\!\!\!\perp D$$

Conditional Ignorability/independence assumption

$$E(Y^1|D=1,\mathbf{Z}) = E(Y^1|D=0,\mathbf{Z}),\ E(Y^0|D=1,\mathbf{Z}) = E(Y^0|D=0,\mathbf{Z})$$

In IV estimation: $\mathrm{ITT} = E(Y_i|Z_i=1) - E(Y_i|Z_i=0)$

$$\mathrm{LATE} = \frac{E(Y_i|Z_i=1) - E(Y_i|Z_i=0)}{E(D_i|Z_i=1) - E(D_i|Z_i=0)} = \frac{\mathrm{Cov}(Y_i, Z_i)}{\mathrm{Cov}(D_i, Z_i)}$$

$$\text{2SLS:} \begin{cases} D = \gamma_0 + \gamma_1 Z + v \\ \hat{D} = \hat{\gamma}_0 + \hat{\gamma}_1 Z \\ Y = \beta_0 + \beta^{2SLS} \hat{D} + e \end{cases}$$

Notation (II)

DiD: $\{E(Y_{i1}(1)|D_i^* = 1) - E(Y_{i1}(0)|D_i^* = 0)\} - \{E(Y_{i0}|D_i^* = 1) - E(Y_{i0}(0)|D_i^* = 0)\}$

Regression estimation of DiD: $Y_{it} = \beta_0 + \beta_1 D_i^* + \beta_2 P_t + \beta_{DD} D_i^* \times P_t + q_{it}$

D^*	$t = 0$	$t = 1$	Difference
1	$\beta_0 + \beta_1$	$\beta_0 + \beta_1 + \beta_2 + \beta_{DD}$	$\beta_2 + \beta_{DD}$
0	β_0	$\beta_0 + \beta_2$	β_2

FE – Time-demeaned equation $y_{it} - \bar{y}_i = \beta_1(x_{it} - \bar{x}_i) + \eta_{it} - \bar{\eta}_i, t = 1, 2, \ldots, T$

RD: $D_i = \begin{cases} 1, & \text{if } X_i > c \\ 0, & \text{if } X_i \leq c \end{cases}$

RD – LATE at threshold: $\tau_{SRD} = E(Y_i(1) - Y_i(0)|X_i = c)$