Building a Robot Judge: Data Science for Decision-Making

8. Bias and Discrimination

Q&A / Announcements

http://bit.ly/BRJ_Padlet8

Review: IV

- IV is a powerful tool to identify causal links, but relies on the quality of the instruments.
- Three dimensions:
  1. Relevance
     - Always report the first stage (F-test above 10)
  2. Exogeneity
     - Is conditional exogeneity (no Z-confounders) plausible?
     - Check if instrument is correlated with predetermined variables
  3. Exclusion restriction
     - Discuss the possible links between Z and Y
     - Specify group which is affected by the instrument (local average treatment effect)

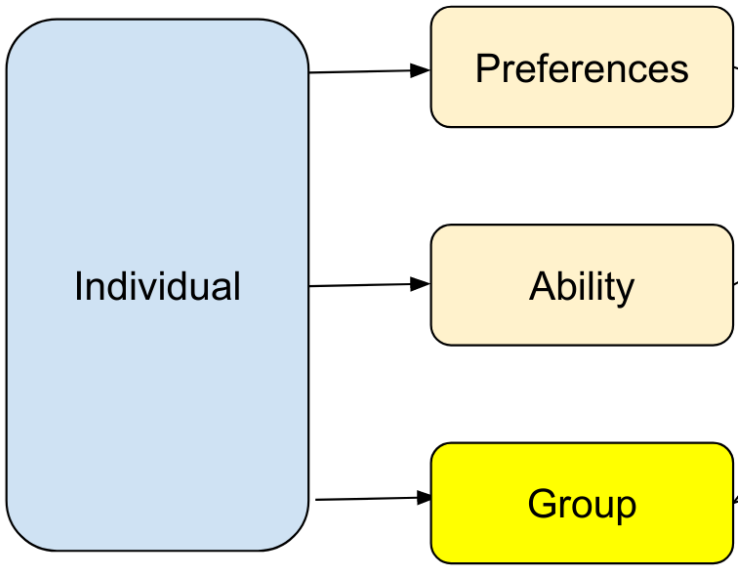
Recap: Adding Instruments to Causal Graphs

`http://bit.ly/BRJ-W7-A-graphs`

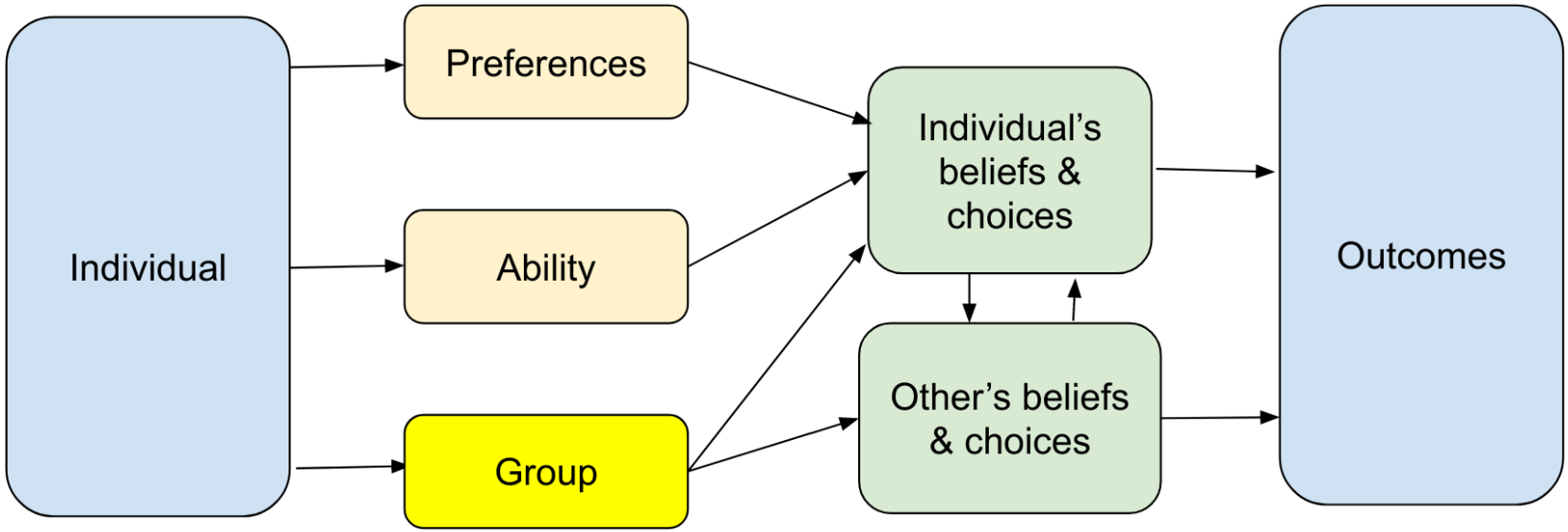
- Please rejoin the same breakout group # you were in last week.
- Continue discussing group member causal graphs
  - take turns defending/criticizing instrument
  - up to 5 minutes per group member
- When done, return to main zoom room, we will discuss some as a bigger group.

Motivation

- Systematic and persistent differences in labor-market and justice-system outcomes across groups – e.g. men/women, across racial/ethnic groups.
- Anecdotally, there are clear examples of prejudice or biased treatment.
- But disparate outcomes on average are also explained in part by differences in characteristics or choices across groups.
- If we want to reduce bias/prejudice, we have to distinguish it from other factors.



- There could be group differences in preferences and ability.
  - these could be correlated with group identity.



- Individual choices, especially around education and skills investments, depend on preferences, ability, and group identity.
- But they might also affect others' choices.
- *Note: This graph could be drawn in many other ways.*

Gender Wage Gap

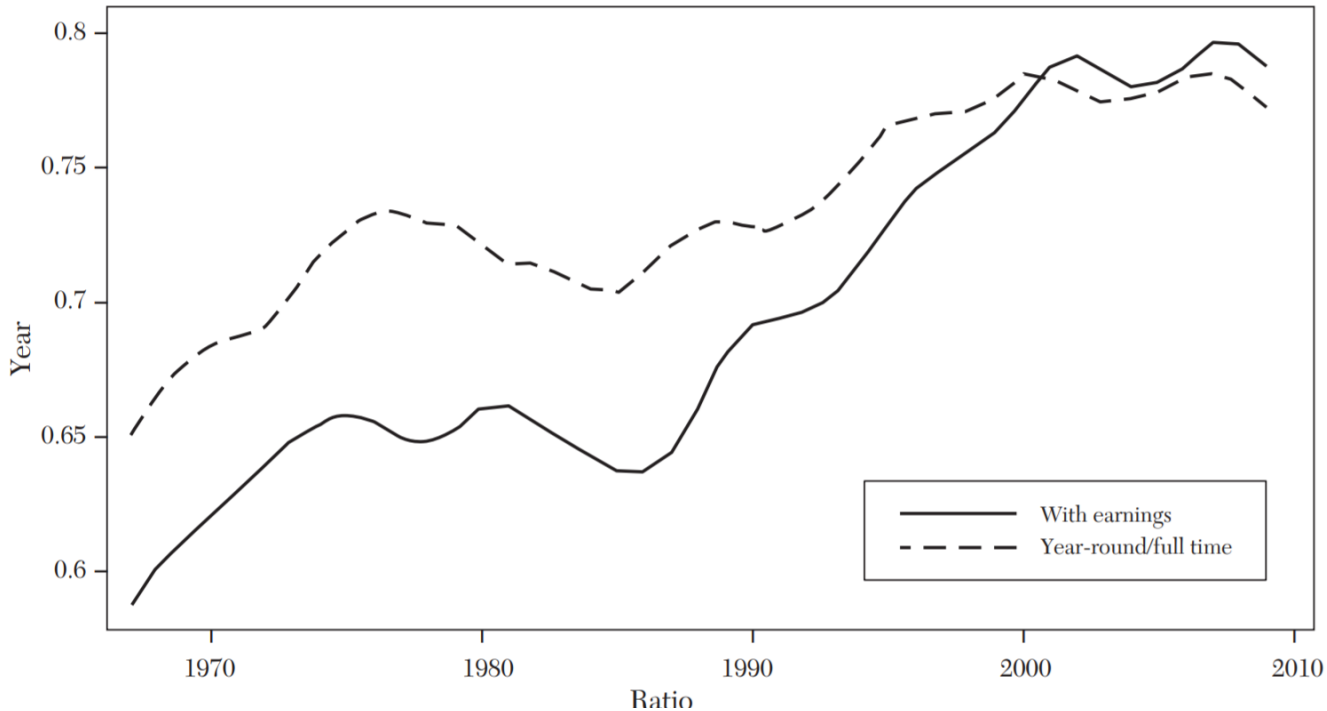
Non-discrimination reasons for the wage gap:

- women have different innate abilities
- women prefer less risky jobs, or shorter commute times, or fewer hours
- women take breaks from their career for childbearing, reducing skills/earnings.

Responses to discrimination:

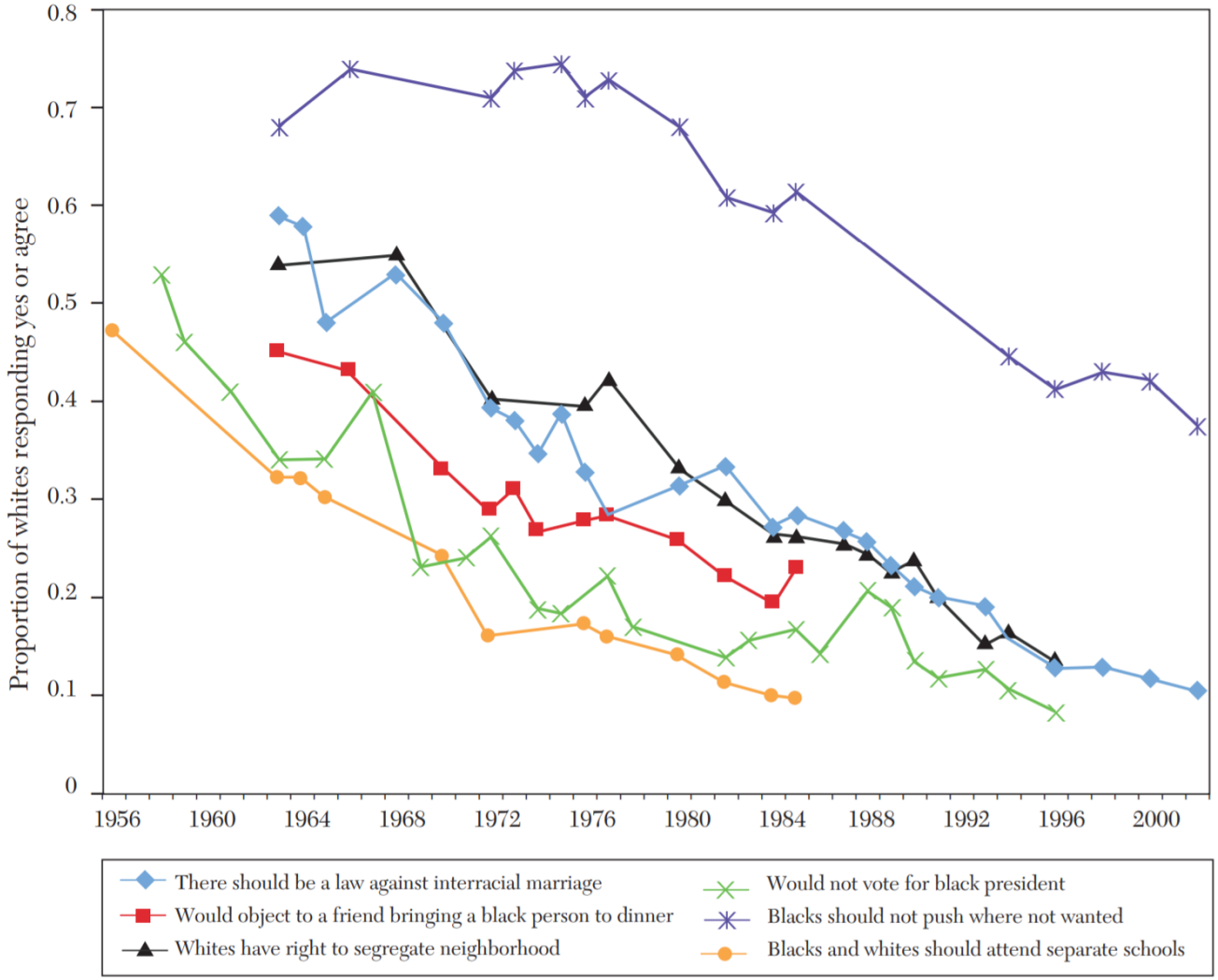
- women might stay out of the labor force to avoid discrimination

Black-to-White Ratio in Median Earnings, U.S. Workers, 1967-2010



- not all or even mainly discrimination.
- other important factors?
  - cognitive ability
  - education
  - labor force participation

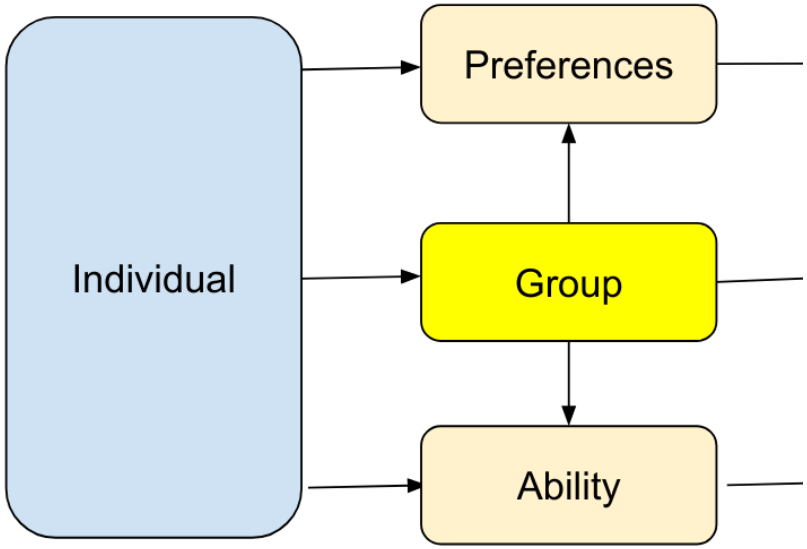
Trends in Measures of Self-Reported Prejudice, U.S. Respondents, 1956-2010



- caveat: whites could be more cautious because they think racist views are not acceptable
- but inter-racial marriage also increased significantly (8x since 1958).

Are preferences and ability really exogenous?

- Preferences/ability could be influenced by group identity, especially across generations.



Ability:

- disparities in prenatal/child health services
- disparities in school quality

Preferences:

- prejudice in generation $t \rightarrow$ demotivation in generation $t+1$

Is group identity exogenous?

- All group identity categories are subjective / socially constructed.
- Until recently, gender and race were considered pre-determined.
  - sex is consistent biological difference (but not 100%)
  - race/ethnicity is a relatively new concept
- religion usually (not always) considered a choice
- These are fluid concepts – we can abstract away from specific dimensions.

Defining Discrimination

- Arrow (1973):
  - Worker characteristics unrelated to productivity are valued on the market.
  - In law, can re-frame this as, "defendant characteristics unrelated to guilt/innocence are valued by the judicial process".
- What about:
  - more attractive person gets a modeling job
  - fire-fighters are mostly male
  - judge gives lenient sentence to defendant with young children

Discrimination: An Inclusive Definition

$$Y_i = \alpha G_i + X_i' \beta + \epsilon_i$$

- Y_i= outcome, e.g. wage or incarceration rate
- G_i= group membership
- X_i= other factors affecting outcome, e.g. ability/education

$$\alpha \neq 0 \leftrightarrow \text{discrimination}$$

Taste-Based Discrimination (Prejudice)

- "Taste for discrimination" (associated with Becker)
  - Firms willing to pay to associate with some persons, not others
  - E.g. act as if blacks more expensive to hire than they are
- Prejudice will reduce profits $\rightarrow$ in a competitive market, discriminating firms will be competed out.
  - could remain with other labor market frictions, e.g. imperfect competition
  - could remain in public sector (e.g. judicial decisions)

Statistical discrimination (stereotypes)

- Employers/judges have different information/beliefs about identity groups.
  - race/gender could in fact be correlated with productivity/criminality
- Different priors (stereotypes) about productivity/criminality.
  - could be self-confirming: employer/judge doesn't give the stereotyped group a chance to prove themselves.
  - another channel for self-confirmation: minority workers expect to be discriminated against, and therefore don't invest in education/skills.

Activity: Review, Four Types of Confounders

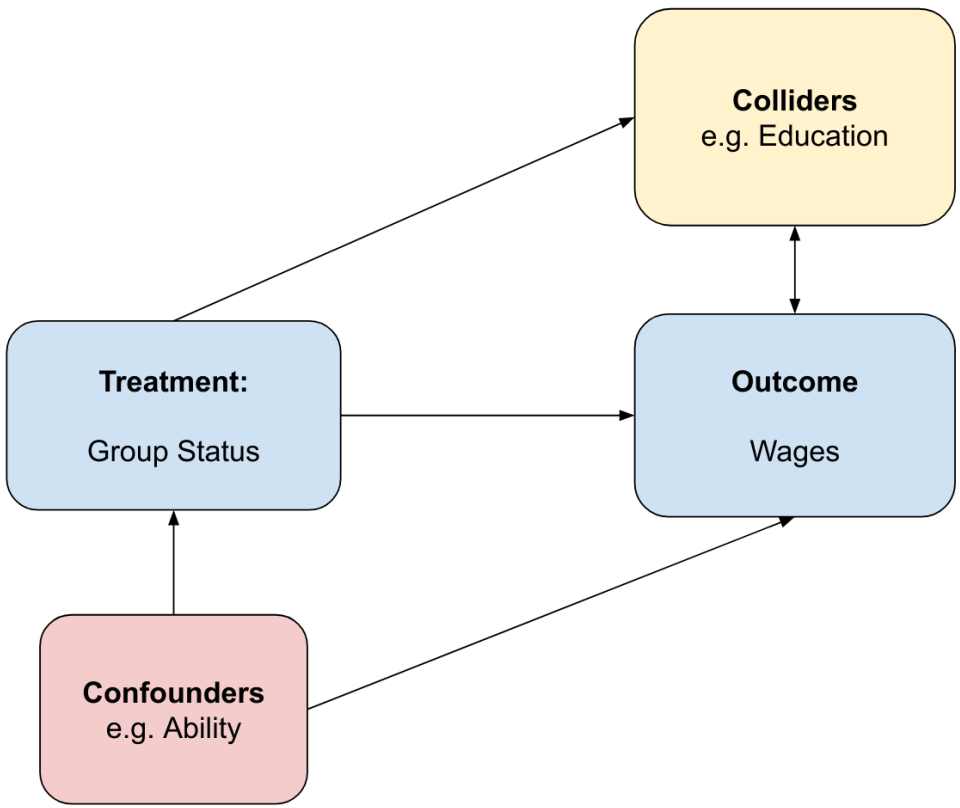
https://twitter.com/ASlavitt/status/1319870491063177216

https://padlet.com/eash44/p4h4614l70q4ixzd

Empirical Evidence

$$Y_i = \alpha G_i + \boldsymbol{X}_i'\beta + \epsilon_i$$

- Y_i= wage, G_i= group , $\boldsymbol{X}_i$= other factors
- $\alpha < 0$ often estimated for women/minorities



- The usual concern: (unobserved) confounders
- e.g. in one study, adding an ability test score (AFQT) explains 3/4 of racial wage gap (Neal and Johnson 1996).

Blind Auditions

Goldin and Rouse (2000)

- natural experiment: orchestras moved to blind auditions.
  - use fixed effects regression, comparing rate that women were hired before and after.
- positive effect: blind auditions helped women.

Resume Audit Study

Bertrand and Mullainathan (2004)

- 5,000 resumes sent to help-wanted ads in Boston and Chicago
- Randomized otherwise equivalent resumes to have African-American or White sounding names:
  - Emily Walsh or Greg Baker relative to Lakisha Washington or Jamal Jones
- Results:
  - 50% gap in callback rate for black-sounding names
- Caveats:
  - "Lakisha" or "Jamal" might signal other non-racial factors, e.g. socioeconomic status.
  - Fryer and Levitt (2004) find no long-term life outcome differences for people with more black-sounding names, adjusting for other background factors.

Jury Race in Criminal Trials

Anwar, Bayer, and Hjalmarsson (2012)

- Examine jury racial composition and trial outcomes in Florida, 2000-2010
- Exogenous treatment: day-to-day variation in composition of jury pool
  - Identification check: composition of jury pool uncorrelated with characteristics of the defendant and case.

Results

Anwar, Bayer, and Hjalmarsson (2012)

TABLE IV

REDUCED-FORM BENCHMARK REGRESSIONS

	(1)	(2)	(3)	(4)
Dependent variable	Any guilty conviction		Proportion guilty convictions	
Black defendant	0.150***	0.164***	0.156***	0.160***
	[0.056]	[0.058]	[0.055]	[0.057]
Any black in pool	0.069	0.105**	0.063	0.090*
	[0.048]	[0.051]	[0.047]	[0.050]
Black defendant * any black in pool	−0.168**	−0.166**	−0.174**	−0.155**
	[0.070]	[0.074]	[0.069]	[0.072]
Constant	0.656***	0.627***	0.600***	0.576***
	[0.039]	[0.041]	[0.038]	[0.040]
Includes controls for:				
Gender/age of pool	No	Yes	No	Yes
County dummy	No	Yes	No	Yes
Year of filing dummies	No	Yes	No	Yes
Observations	712	712	712	712
R-squared	0.01	0.07	0.01	0.08

Racial bias in vehicle searches

Knowles, Persico, and Todd (2001)

- Motivation: Black drivers are searched more often by police for drugs.
  - is this prejudice?
- Two reasons blacks get searched more:
  - Statistical discrimination: blacks are more likely to have drugs
  - Taste-based discrimination: police are prejudiced
- Can formally test in this context:
  - statistical discrimination $\rightarrow$ contraband discovery (successful search) rates will be the same for both groups.
  - prejudice $\rightarrow$ contraband discovery rates will be lower for black drivers, as threshold for search is lower.
- Empirical test:
  - data on 1500 traffic searches in Maryland, 1995-1999
  - contraband discovery rates are the same across races, consistent with statistical discrimination, but not taste-based discrimination

Discrimination as a self-fulfilling prophecy

Grover, Pallais, and Pariente (2017)

- In French grocery chain, implicit association test to store managers and ranked them as more or less implicitly biased toward ethnic minorities.
- When minority cashiers are (randomly) paired with biased managers, they are less productive (absent more often, scan items more slowly).
- With *unbiased* manageres, minority cashiers perform better than non-minority cashiers; performance is the same on average.
  - Consistent with statistical discrimination in hiring $\rightarrow$ minorities underperform when assigned to biased managers, so firm sets a higher hiring standard for minorities.

Practice: Paper on Bias in Employment Recognition

http://bit.ly/BRJ-W8-Activity

Regression Discontinuity Design (RDD)

- Added to Week 3 slides: Sharp RD
- Added to Week 7 slides: Fuzzy RD