Carnegie Mellon University
Heinz College
INFORMATION SYSTEMS • PUBLIC POLICY • MANAGEMENT

Nonparametric Assessment of Racial Disparities in Prosecutorial Peremptory Strikes: A Case Study in Mississippi

Zhiyu Guo, Edward H. Kennedy, Eli Ben-Michael

August 3, 2025

Carnegie Mellon University

Motivation

- **Legal Context:** Batson v. Kentucky (1986) prohibits race-based jury strikes

Motivation

- **Legal Context:** Batson v. Kentucky (1986) prohibits race-based jury strikes
- **Challenge:** Prosecutors can claim "race-neutral" reasons for strikes

Motivation

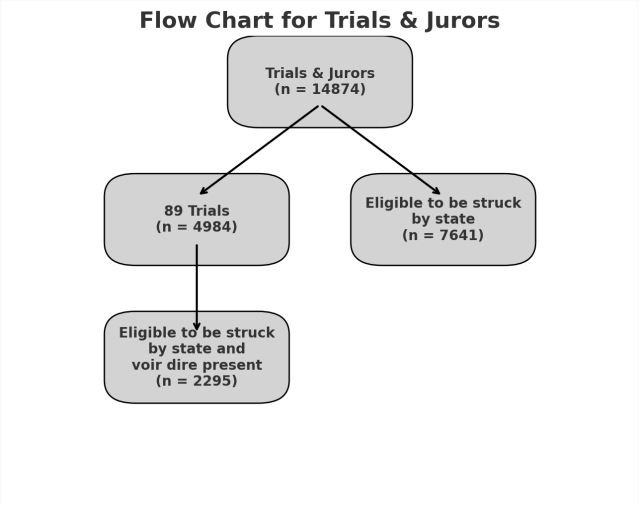
- **Legal Context:** Batson v. Kentucky (1986) prohibits race-based jury strikes
- **Challenge:** Prosecutors can claim "race-neutral" reasons for strikes
- **Statistical Question:** How do Black and white jurors of similar observable characteristics fare in prosecutorial peremptory strikes?

Motivation

- **Legal Context:** Batson v. Kentucky (1986) prohibits race-based jury strikes
- **Challenge:** Prosecutors can claim "race-neutral" reasons for strikes
- **Statistical Question:** How do Black and white jurors of similar observable characteristics fare in prosecutorial peremptory strikes?
- **Goal:** Develop flexible, nonparametric framework to quantify racial disparities controlling for high-dimensional covariates

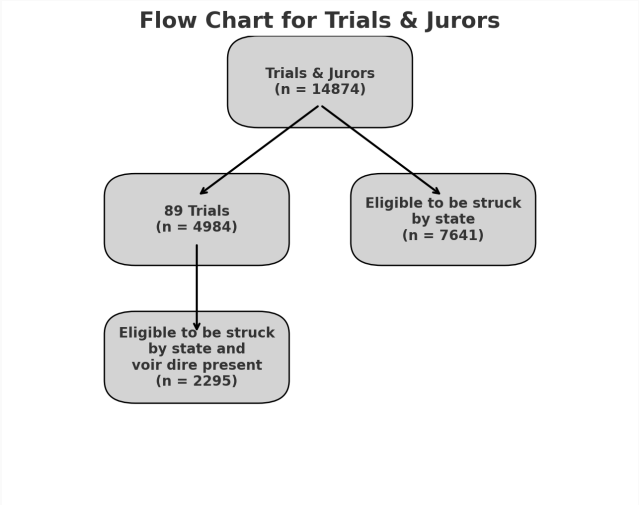
Data

Data



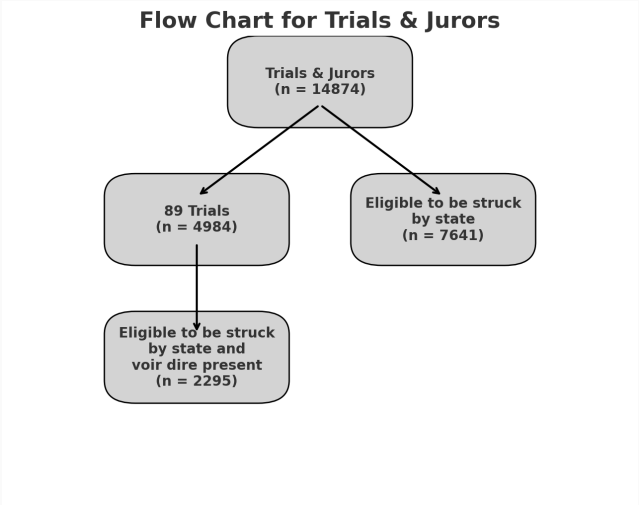
- 305 trials from 1992 to 2017, 14874 potential juror/venire members

Data



- 305 trials from 1992 to 2017, 14874 potential juror/venire members
- 89 trials with transcripts, 3546 voir dire records

Data



- 305 trials from 1992 to 2017, 14874 potential juror/venire members
- 89 trials with transcripts, 3546 voir dire records
- Focus: subsample that's eligible to be struck by state. There are 1541 (67.1%) white jurors, 741 (32.2%) Black jurors, and 13 with unknown race

Problem Setup

- $X \in \mathbb{R}^{120}$: gender, education level, marriage status, religious status, voir dire answers (all binary), judge, county, prosecutor(s), prosecutors more than 3 indicator, defendant race, crime type. Voir dire answers coded from transcripts, juror information from jury sheet, and trial information from court records.

Problem Setup

- $X \in \mathbb{R}^{120}$: gender, education level, marriage status, religious status, voir dire answers (all binary), judge, county, prosecutor(s), prosecutors more than 3 indicator, defendant race, crime type. Voir dire answers coded from transcripts, juror information from jury sheet, and trial information from court records.
- $A \in \{0, 1\}$: Juror race, use $A = 1$ to denote Black and $A = 0$ to denote white.

Problem Setup

- $X \in \mathbb{R}^{120}$: gender, education level, marriage status, religious status, voir dire answers (all binary), judge, county, prosecutor(s), prosecutors more than 3 indicator, defendant race, crime type. Voir dire answers coded from transcripts, juror information from jury sheet, and trial information from court records.
- $A \in \{0, 1\}$: Juror race, use $A = 1$ to denote Black and $A = 0$ to denote white.
- $Y \in \{0, 1\}$: Struck by state indicator.

Methods

Outline of Methods

- Target: Learning the conditional mean difference in strike rates for Black and white jurors: $\mathbb{E}[Y|A=1,X] - \mathbb{E}[Y|A=0,X]$

Outline of Methods

- Target: Learning the conditional mean difference in strike rates for Black and white jurors: $\mathbb{E}[Y|A=1,X] - \mathbb{E}[Y|A=0,X]$
- Summary measure: the average conditional difference in strike rates

Outline of Methods

- Target: Learning the conditional mean difference in strike rates for Black and white jurors: $\mathbb{E}[Y|A=1,X] - \mathbb{E}[Y|A=0,X]$
- Summary measure: the average conditional difference in strike rates
- Heterogeneity: the variance of conditional difference and variance-based variable importance measures

Outline of Methods

- Target: Learning the conditional mean difference in strike rates for Black and white jurors: $\mathbb{E}[Y|A=1,X] - \mathbb{E}[Y|A=0,X]$
- Summary measure: the average conditional difference in strike rates
- Heterogeneity: the variance of conditional difference and variance-based variable importance measures
- Robustness: sensitivity analysis to potential unmeasured/partially measured covariates

Average Controlled Difference

Average Controlled Difference

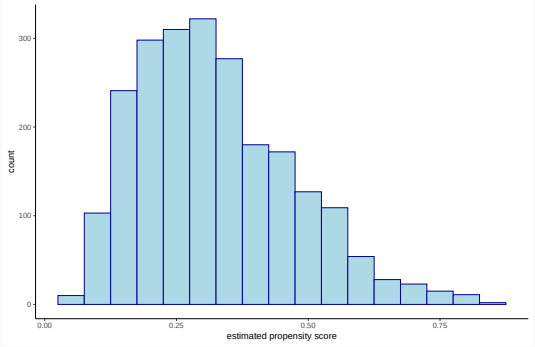
- Mean/ACD: $\mathbb{E}[\mathbb{E}[Y|A=1,X]-\mathbb{E}[Y|A=0,X]]$.

Average Controlled Difference

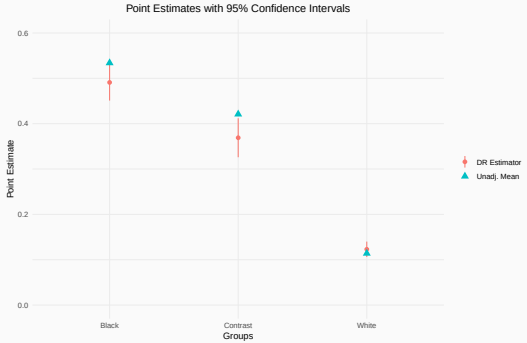
- Mean/ACD: $\mathbb{E}[\mathbb{E}[Y|A=1,X]-\mathbb{E}[Y|A=0,X]]$.
- Under positivity, can be estimated by a doubly robust estimator: $\hat{\psi}_{\mathsf{dr}}=\frac{1}{N}\sum_{i=1}^{N}\left[\hat{\mu}_1(X_i)+\frac{A_i(Y_i-\hat{\mu}_1(X_i))}{\hat{\pi}(X_i)}\right]-\left[\hat{\mu}_0(X_i)+\frac{(1-A_i)(Y_i-\hat{\mu}_0(X_i))}{1-\hat{\pi}(X_i)}\right]$

Average Controlled Difference

- Mean/ACD: $\mathbb{E}[\mathbb{E}[Y|A=1,X]-\mathbb{E}[Y|A=0,X]]$.
- Under positivity, can be estimated by a doubly robust estimator: $\hat{\psi}_{\mathsf{dr}} = \frac{1}{N}\sum_{i=1}^{N}\left[\hat{\mu}_1(X_i)+\frac{A_i(Y_i-\hat{\mu}_1(X_i))}{\hat{\pi}(X_i)}\right]-\left[\hat{\mu}_0(X_i)+\frac{(1-A_i)(Y_i-\hat{\mu}_0(X_i))}{1-\hat{\pi}(X_i)}\right]$



Main Results: Large Racial Disparity



- **ACD = 0.37** [0.31, 0.42] (37 percentage point difference)
- **Odds ratio: 6.91** - Black jurors nearly 7× more likely to be struck
- Highly significant, large effect size

Conditional Difference, Heterogeneity, and Variable Importance

Conditional Difference and Heterogeneity

- Motivation: We have a good estimate of the ACD, but is it uniform across jurors? If not, what covariates are associated with more/less conditional difference in strike rates?

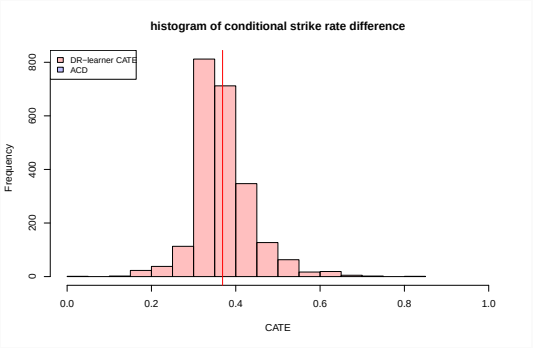
Conditional Difference and Heterogeneity

- Motivation: We have a good estimate of the ACD, but is it uniform across jurors? If not, what covariates are associated with more/less conditional difference in strike rates?
- We consider the conditional difference $\tau(X) = \mathbb{E}[Y|A = 1, X] - \mathbb{E}[Y|A = 0, X]$.

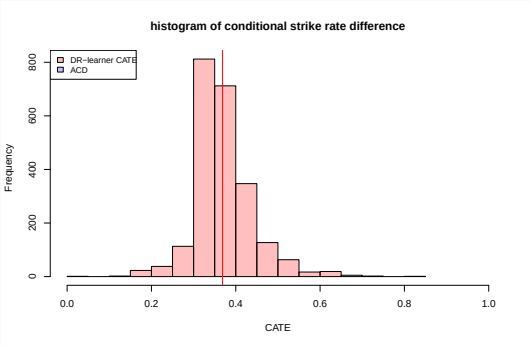
Conditional Difference and Heterogeneity

- Motivation: We have a good estimate of the ACD, but is it uniform across jurors? If not, what covariates are associated with more/less conditional difference in strike rates?
- We consider the conditional difference $\tau(X) = \mathbb{E}[Y|A = 1, X] - \mathbb{E}[Y|A = 0, X]$.
- Using the DR-learner, we observe an empirical variance of 0.005, which translates to 0.07 in SD scale.

Conditional difference and Heterogeneity

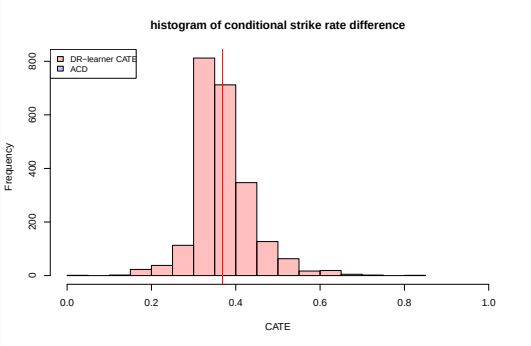


Conditional difference and Heterogeneity



- [Levy et al., 2018] derived an influence function based estimator for the variance:
  $\hat{Var}(\tau(X)) = \mathbb{P}_n\{2(\hat{\tau}(X) - \hat{\psi})(\frac{A}{\hat{\pi}(X)} - \frac{1-A}{1-\hat{\pi}(X)})(Y - \hat{\mu}_A(X)) + (\hat{\tau}(X) - \hat{\psi})^2\}$

Conditional difference and Heterogeneity



- [Levy et al., 2018] derived an influence function based estimator for the variance: $\hat{Var}(\tau(X)) = \mathbb{P}_n\{2(\hat{\tau}(X) - \hat{\psi})(\frac{A}{\hat{\pi}(X)} - \frac{1-A}{1-\hat{\pi}(X)})(Y - \hat{\mu}_A(X)) + (\hat{\tau}(X) - \hat{\psi})^2\}$
- The variance is estimated to be 0.05 [0.04, 0.05]. Using the delta method, we can also get the sqrt of the variance of 0.22 [0.19, 0.24] on the original ACD scale.

Variable Importance for Heterogeneity

Framework: [Hines et al., 2023] variance-based importance measures

$$\Theta_s = \text{Var}(\tau(X)) - \text{Var}(\tau_s(X)) = E[\text{Var}(\tau(X)|X_{-s})]$$

Variable Name	$\hat{\Theta}_s$	$\hat{\Theta}_s/\hat{\Theta}_{total}$
family/friend accused of crime	0.011 [0.008, 0.013]	0.221
family/friend in law enforcement	0.004 [0.003, 0.006]	0.085
knows the defendant	0.003 [0.002, 0.005]	0.072
widowed	0.003 [0.001, 0.005]	0.064
prosecutor 2: Kevin Horan	0.003 [0.001, 0.004]	0.052

Interpretation: Personal connections to crime explain $\sim 22\%$ of treatment effect variance

Variable Importance, Keep-One-In

Variable	$\hat{\Theta}_s$	$(1 - \hat{\Theta}_s)/\hat{\Theta}_{total}$
family/friend accused of crime	0.042 [0.031, 0.053]	0.127
knows the defendant	0.044 [0.034, 0.054]	0.080
has prior information on the case	0.044 [0.036, 0.053]	0.077
hesitation w/ death penalty	0.045 [0.035, 0.054]	0.071
prosecutor 2: Susan Denley	0.045 [0.036, 0.055]	0.062

Key takeaway: Personal connections to crime and knowing the defendant are consistently most important across both measures

Sensitivity Analysis: Unmeasured Covariates

Concern: Potential unmeasured covariates U (dialects, mannerisms, missing data)

Method: [Bonvini & Kennedy, 2022] proportion of unmeasured covariates

- Let $S \in \{0, 1\}$ indicate contamination, suppose $S \perp\!\!\!\perp Y \mid A, X$

Sensitivity Analysis: Unmeasured Covariates

Concern: Potential unmeasured covariates U (dialects, mannerisms, missing data)

Method: [Bonvini & Kennedy, 2022] proportion of unmeasured covariates

- Let $S \in \{0, 1\}$ indicate contamination, suppose $S \perp\!\!\!\perp Y \mid A, X$
- For fraction ϵ with $S = 0$: $\mathbb{E}[Y \mid A = 1, X] - \mathbb{E}[Y \mid A = 0, X] = \mathbb{E}[Y \mid A = 1, X, U] - \mathbb{E}[Y \mid A = 0, X, U]$

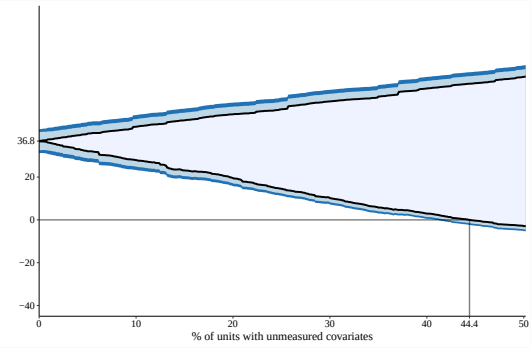
Sensitivity Analysis: Unmeasured Covariates

Concern: Potential unmeasured covariates U (dialects, mannerisms, missing data)

Method: [Bonvini & Kennedy, 2022] proportion of unmeasured covariates

- Let $S \in \{0, 1\}$ indicate contamination, suppose $S \perp\!\!\!\perp Y \mid A, X$
- For fraction ϵ with $S = 0$: $\mathbb{E}[Y \mid A = 1, X] - \mathbb{E}[Y \mid A = 0, X] = \mathbb{E}[Y \mid A = 1, X, U] - \mathbb{E}[Y \mid A = 0, X, U]$
- Otherwise: unmeasured covariates can alter conditional expectations

Robustness to Large Proportion of Unmeasured Covariates



Result: 44.4% [41.1%, 47.7%] of the sample would need to have unmeasured covariates that substantially alter their expected strike rates to invalidate the results - highly robust finding

Conclusion

Statistical Evidence:

- Strong, robust evidence of 37% racial disparity
- "Race-neutral" factors systematically disadvantage Black jurors
- Personal crime connections associated with additional disparities

Policy Response from Other Places:

- **Arizona:** Abolished peremptory strikes (2022)
- **WA, CT, CA:** Limited peremptory strikes
- **International:** UK (1988), Canada (2019) abolished

Contribution: Rigorous statistical framework supporting audits and legal reforms

Future Statistical Directions

- **Text Analysis:** NLP methods for transcript-derived covariates, potential for bias detection
- **Network Effects:** Prosecutor behavior correlation, spillover effects
- **Dynamic Models:** Evolution of discrimination patterns over time

Questions?