Estimating Racial Disparities when Race is Not Observed

Kosuke Imai

Harvard University

Bridging Prediction and Intervention Problems in Social Systems
Banff International Research Station for Mathematical Innovation and Discovery
June 4, 2024

Joint work with Cory McCartan (Penn State), Robin Fisher (Treasury), Jacob Goldin (Chicago Law), and Daniel E. Ho (Stanford Law)

Motivation

- Importance of racial disparity estimation in many fields: public health, employment, voting, criminal justice, taxation, housing, lending, and internet technology
- But, often individual race is not available
  - law may prohibit collection of information about race (e.g., Equal Credit Opportunity Act)
  - agencies and companies may not wish to collect such information
- How should we estimate racial disparities when race is not observed?
  - Standard methods use BISG (Bayesian Improved Surname Geocoding)
  - But, it has been shown that they are likely to yield biased estimates
- Can we improve the standard methods and eliminate their bias?
- Executive Order 13985: Advancing Racial Equity and Support for Underserved Communities through the Federal Government

Motivating Application: Racial Disparity in US Tax System

- Brown (2022) *The whiteness of wealth: How the tax system impoverishes Black Americans and how we can fix it*
- Racial disparity estimation is important, but IRS does not collect individual race information
- Census Bureau cannot share the individual data (Title 13)
- Home Mortgage Interest Deduction (HMID)
  - Brown describes HMID as "little more than the twenty-first-century version of redlining" and concludes it "must be repealed"
  - HMID does not encourage home ownership but increases housing price
  - 90% of taxpayers take standard deduction and does not itemize HMID
- We analyze a random 10% sample of the individual tax returns (1040s) from 2019, leading to a total of 17 million observations

Setup

- Data
  - Y_i: outcome of interest (categorical)
  - R_i: (unobserved) race
  - S_i: surname
  - G_i: residence location
  - W_i: covariates of interest
  - X_i: other Census variables (optional)
- Census data
  - $\mathbb{P}(G_i = g, R_i = r, X_i = x)$
  - $\mathbb{P}(R_i = r, S_i = s)$ for frequently occurring surnames
- Racial disparity estimands
  - $\mathbb{P}(Y_i = y \mid R_i = r) - \mathbb{P}(Y_i = y \mid R_i = r')$ for $r \neq r'$
  - $\mathbb{P}(Y_i = y \mid R_i = r, W_i = w) - \mathbb{P}(Y_i = y \mid R_i = r', W_i = w)$
- Regression estimands
  - short regression: $\mathbb{P}(Y_i = y \mid R_i = r)$
  - long regression: $\mathbb{P}(Y_i = y \mid R_i = r, X_i = x)$

Standard Estimation Methods

1. Predict race via BISG (or its variant)
   - Assumption: $G_i \perp\!\!\!\perp S_i \mid R_i$
   - Bayes rule:

$$\begin{aligned}
\hat{P}_{ir} &= \mathbb{P}(R_i = r \mid G_i = g, S_i = s) \\
&= \frac{\mathbb{P}(S_i = s \mid G_i = g, R_i = r)\,\mathbb{P}(G_i = g, R_i = r)}{\sum_{r'} \mathbb{P}(S_i = s \mid G_i = g, R_i = r')\,\mathbb{P}(G_i = g, R_i = r')} \\
&= \frac{\mathbb{P}(S_i = s \mid R_i = r)\,\mathbb{P}(G_i = g, R_i = r)}{\sum_{r'} \mathbb{P}(S_i = s \mid R_i = r')\,\mathbb{P}(G_i = g, R_i = r')}
\end{aligned}$$

   - With covariates: $\{G_i, X_i\} \perp\!\!\!\perp S_i \mid R_i$
2. Estimate racial disparities $\mu_{Y|R}(y \mid r) = \mathbb{P}(Y_i = y \mid R_i = r)$
   - weighting:

$$\hat{\mu}^{\text{wtd}}_{Y|R}(y \mid r) = \frac{\sum_i \mathbf{1}\{Y_i = y\}\hat{P}_{ir}}{\sum_i \hat{P}_{ir}}$$

   - thresholding: use the racial group with the largest probability as imputed race

Good Race Prediction Can Bias Racial Disparity Estimates

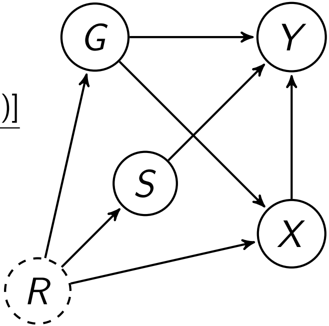
- Bias of the weighting estimator (Chen *et al.* 2019)

$$\hat{\mu}^{\text{wtd}}_{Y|R}(y \mid r) - \mathbb{P}(Y_i = y \mid R_i = r)$$
$$= -\frac{\mathbb{E}[\text{Cov}(\mathbf{1}\{Y_i = y\}, \mathbf{1}\{R_i = r\} \mid G_i, X_i, S_i)]}{\mathbb{P}(R_i = r)}$$

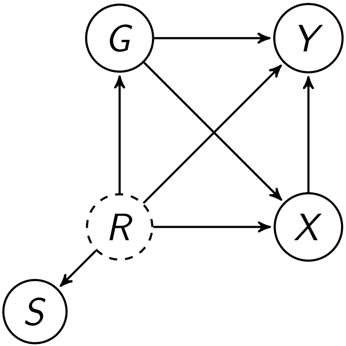
- Required assumption:

$$Y_i \perp\!\!\!\perp R_i \mid G_i, S_i, X_i$$

- Problem: race affects many aspects of the society



New Identification Strategy



- Required assumption:

$$Y_i \perp\!\!\!\perp S_i \mid G_i, R_i, X_i$$

- Surname as a proxy for race
- Race can directly or indirectly affects the outcome
- Example: anonymous application screening
- Potential violations:
  - name-based discrimination
  - coarse racial categories

Surname as a High-dimensional Instrument

- Identification (Kuroki and Pearl, 2014)

$$\overbrace{\mathbb{P}(Y_i = y \mid G_i = g, X_i = x, S_i = s)}^{\text{observed data}}$$
$$= \sum_{r \in \mathcal{R}} \underbrace{\mathbb{P}(Y_i = y \mid R_i = r, G_i = g, X_i = x)}_{\text{unknown parameters}} \underbrace{\mathbb{P}(R_i = r \mid G_i = g, X_i = x, S_i = s)}_{\text{BISG probability}}$$

  - $(|\mathcal{Y}| - 1) \times |\mathcal{G}| \times |\mathcal{X}| \times |\mathcal{S}|$ equations
  - $(|\mathcal{Y}| - 1) \times |\mathcal{G}| \times |\mathcal{X}| \times |\mathcal{R}|$ unknown parameters

- OLS estimator (see also Fong and Tyler, 2021):

$$\hat{\boldsymbol{\mu}}^{(\text{ols})}_{Y|RGX}(y \mid \cdot, g, x) = (\hat{\mathbf{P}}^{\top}_{\mathcal{I}(xg)} \hat{\mathbf{P}}_{\mathcal{I}(xg)})^{-1} \hat{\mathbf{P}}_{\mathcal{I}(xg)} \mathbb{1}\{\mathbf{Y}_{\mathcal{I}(xg)} = y\},$$

  - compute this for each (g, x), and aggregate them using $\mathbb{P}(G_i = g, X_i = x \mid R_i = r)$
  - unbiased estimate of $\mathbb{P}(Y_i = y \mid R_i = r)$
  - ignores the fact that $\mathbb{P}(Y_i = y \mid R_i = r, G_i = g, X_i = x)$ is probability

BIRDiE (Bayesian Instrumental Regression for Disparity Estimation)

- Flexible and scalable probabilistic model that integrates BISG
- Posterior:

$$\pi(\Theta, \mathbf{R} \mid \mathbf{Y}, \mathbf{G}, \mathbf{X}, \mathbf{S}) \propto \pi(\Theta) \prod_{i=1}^{N} \underbrace{\pi(Y_i \mid R_i, G_i, X_i, \Theta)}_{\text{complete-data model}} \underbrace{\pi(R_i \mid G_i, X_i, S_i)}_{\text{BISG prob. } \hat{P}_{ir}}$$

- EM algorithm: updated race probabilities
- Models:
  1. Complete-pooling:

$$Y_i \mid R_i, G_i, X_i, \Theta \sim \mathrm{Cat}_{\mathcal{Y}}(\boldsymbol{\theta}_{R_i}), \quad \boldsymbol{\theta}_r \overset{iid}{\sim} \mathrm{Dir}(\boldsymbol{\alpha})$$

  2. Saturated (no pooling):

$$Y_i \mid R_i, G_i, X_i, \Theta \sim \mathrm{Cat}_{\mathcal{Y}}(\boldsymbol{\theta}_{R_i G_i X_i}), \quad \boldsymbol{\theta}_{rgx} \overset{iid}{\sim} \mathrm{Dir}(\boldsymbol{\alpha})$$

  3. Partial pooling (mixed effects): $\mathbf{W}$ group-level covariates, $\mathbf{Z} = (X, G)$

$$Y_i \mid R_i, G_i, X_i, \Theta \sim \mathrm{Cat}_{\mathcal{Y}}(g^{-1}(\boldsymbol{\mu}_{rgx})), \quad \mu_{rgxy} = \mathbf{W}\boldsymbol{\beta}_{ry} + \mathbf{Z}\mathbf{u}_{ry}$$

$$\mathbf{u}_{ry} \mid \phi_{ry} \sim \mathcal{N}\big(0, \Sigma(\boldsymbol{\phi}_{ry})\big), \quad \boldsymbol{\beta}_{ry} \overset{iid}{\sim} f_\beta, \quad \boldsymbol{\phi}_{ry} \overset{iid}{\sim} f_\phi$$

Sensitivity Analysis

- Potential violation of the key identifying assumption
  - name-based discrimination
  - racial category is too coarse
- Suppose we can have information about finer ethnic groups

$$f : \mathcal{S} \to \mathbb{R}^d, \ d \ll |\mathcal{S}|$$

  - $f(\text{Imai}) = \text{Japanese}$, $f(\text{McCartan}) = \text{Irish}$, etc.
  - Assume instead

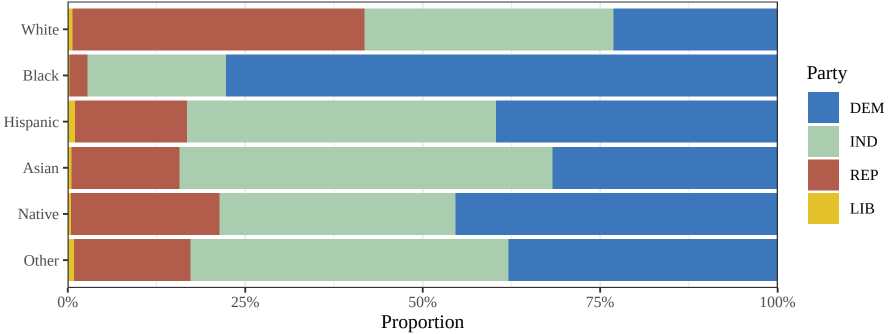
$$Y_i \perp\!\!\!\perp S_i \mid f(S_i), R_i, G_i, X_i$$

- 1930 Census provides 22 groups
  - Anglosphere and Black surname (third-or-more generation Whites and Blacks): Smith, Williams, Brown, ...
  - First wave European immigration (German, Nordic, and Irish): Burns, Olson, Wagner, ...
  - East Asian (Chinese, Japanese, Korean), South Asian (Indian, Southwest Asian), Southeast Asian and Pacific (Vietnamese, Filipino)
  - Non-Cuban Hispanic (Mexican, Latin American), Cuban

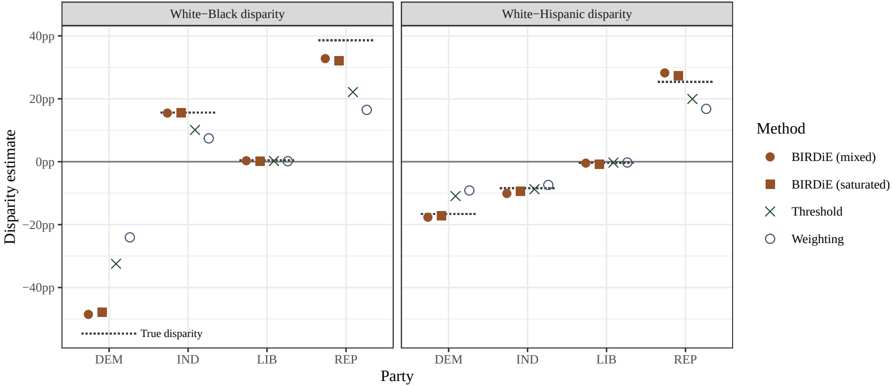
Empirical Validation

- 2022 North Carolina voter file: 5.8M voters with self-reported race
- Subset 1M voters $\rightsquigarrow$ negligible sampling uncertainty

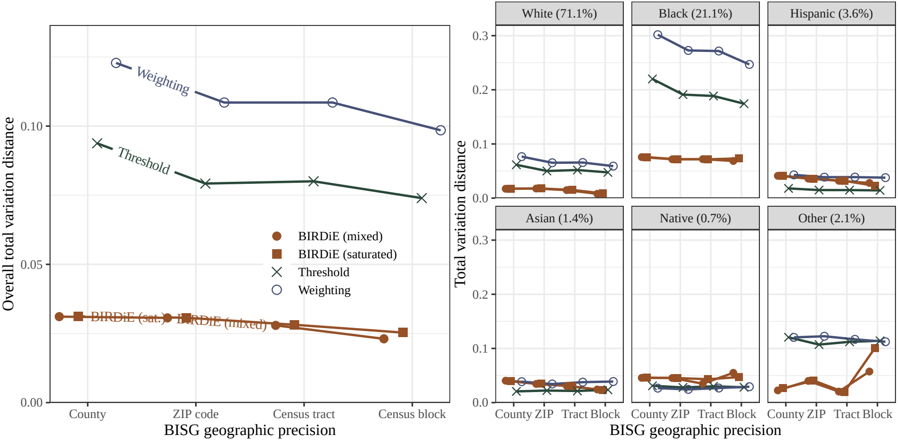
- Focus on party registration



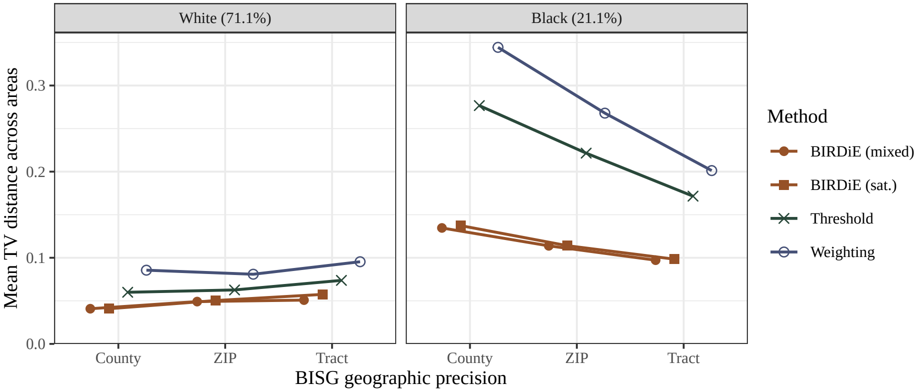
Estimates of Racial Disparity in Party Registration



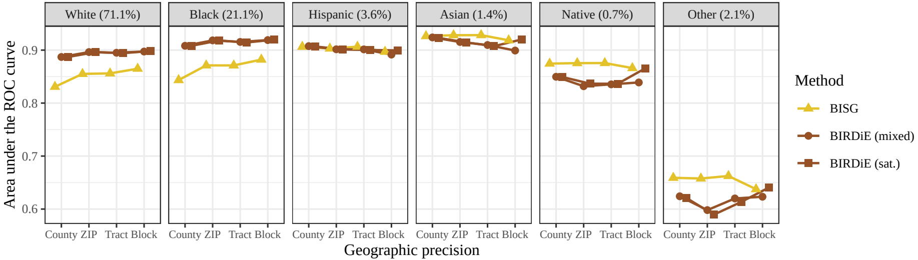
Total Variation Distance



Small Area Estimation

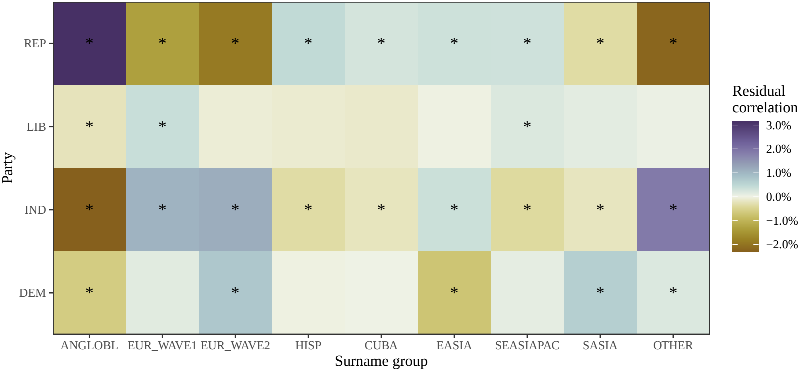


Improved Race Probabilities



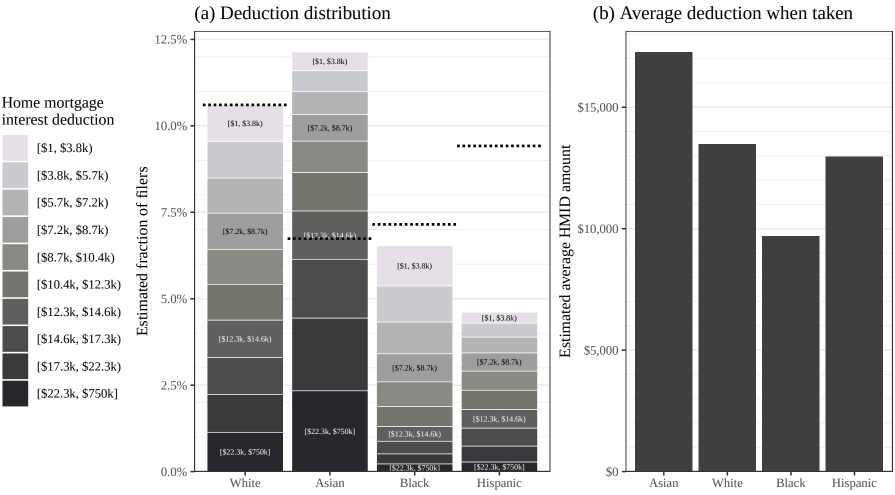
Robustness Analysis

- Surname groups from 1930 Census
- Added 3,000 Asian surnames to account for more recent immigration
- Correlation between BIRDiE residuals and nine surname groups



- Including these in BIRDiE does not substantially alter the estimates

Racial Disparity in HMID



Concluding Remarks

- BIRDiE
  - new identification assumption
  - flexible modeling with scalable estimation
  - improved BISG race probabilities
  - sensitivity analysis
- Future work
  - additional empirical validations: understanding bias
  - better use of auxiliary information in sensitivity analysis
  - make BIRDiE more robust to small bias in BISG probabilities

The paper is available at

https://imai.fas.harvard.edu/research/birdie.html

The software is available at

https://corymccartan.com/birdie/