Estimating Racial Disparities when Race is Not Observed

Kosuke Imai

Harvard University

Annual Meeting of the American Political Science Association
September 2, 2023

Joint work with Cory McCartan, Jacob Goldin, and Daniel E. Ho

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?

The Setup

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

Standard Estimation Methods

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

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

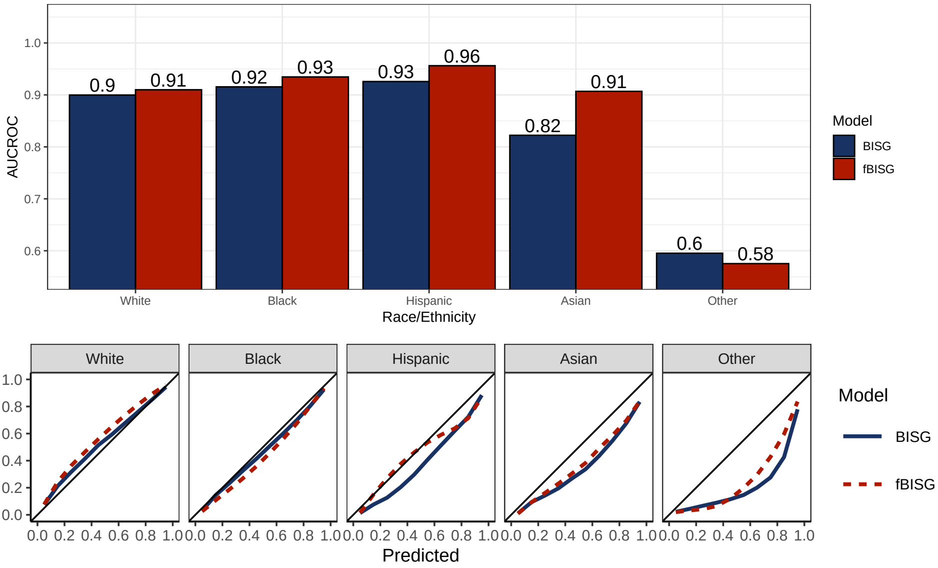
   - With covariates: $(G_i, X_i) \perp\!\!\!\perp S_i \mid R_i$
   - WRU software package (Imai and Kahna 2016)

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

BISG Prediction Works Reasonably Well (Imai et al. 2022. *Sci. Adv.*)



Good Race Prediction Can Bias Racial Disparity Estimates

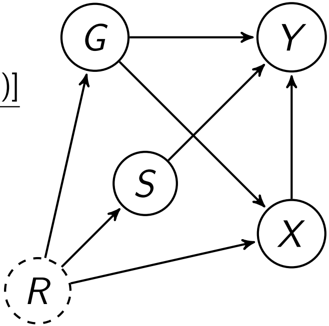
- Bias of the weighted estimator (Chen *et al.* 2019)

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

  - bias tends to be large for minority groups
  - racial disparity tends to be underestimated

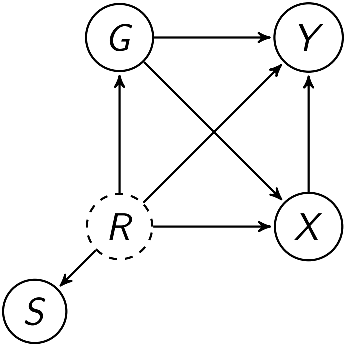
- Required assumption:

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



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
- Potential violations:
    - name-based discrimination
    - coarse racial categories
    - but conditional on true race
- Anonymous application

Surname as a High-dimensional Instrument

- Identification (see also 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 and x, and aggregate
  - 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:

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

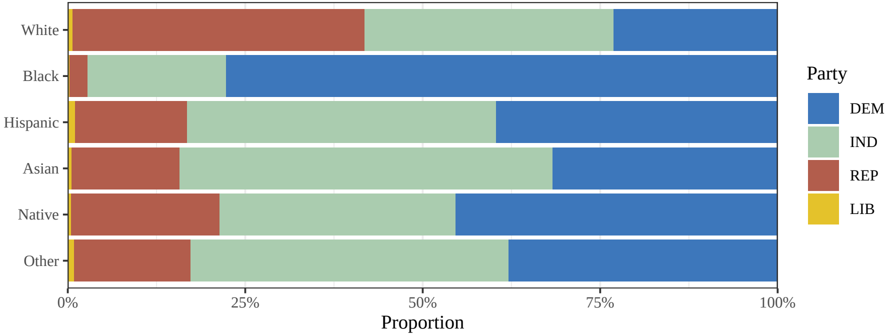
- Models (Categorical $\sim$ Dirichlet)
    1. Complete-pooling
    2. Saturated (no pooling)
    3. Partial pooling (mixed effects)

- Computation:
    1. Small samples: full MCMC (e.g., via Stan)
    2. Large samples: EM algorithm

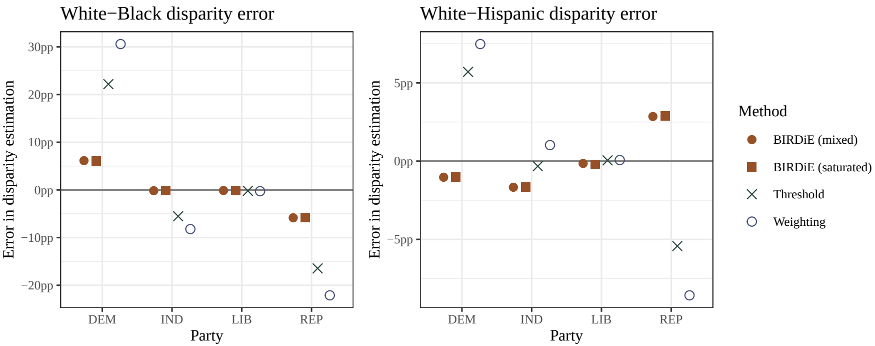
Empirical Validation

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

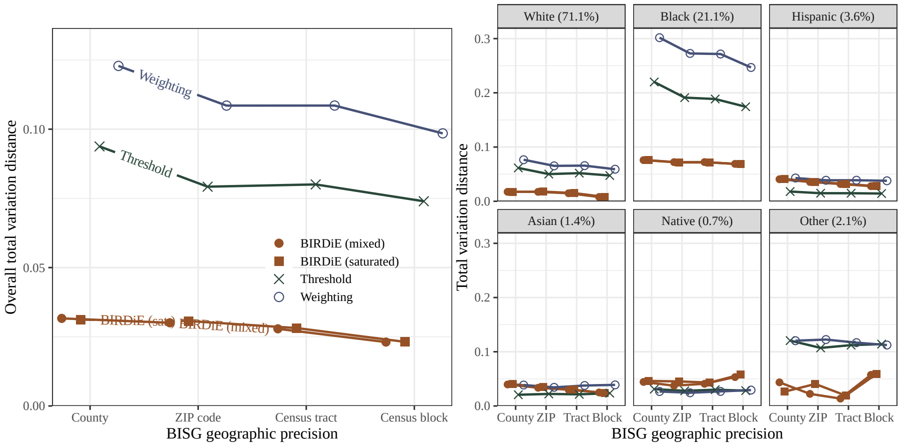
- Focus on party registration



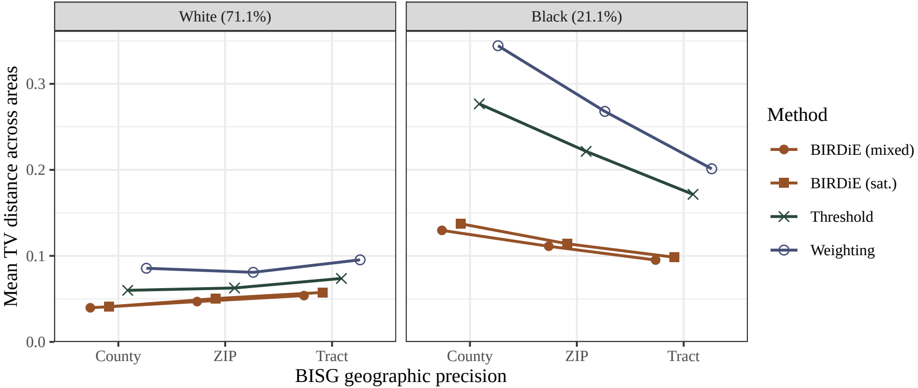
Estimates of Racial Disparity in Party Registration



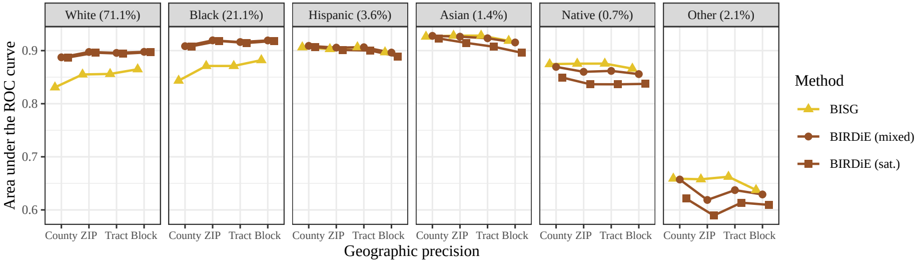
Total Variation Distance



Small Area Estimation



Improved Race Probabilities



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
  - more general approach to data combination and record linkage

The paper is available at

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

The software is available at

https://corymccartan.com/birdie/