Estimating Racial Disparities when Race is Not Observed

Kosuke Imai

Harvard University

Interdisciplinary Seminar on Social Science Methodology
Interdisciplinary Seminar in Quantitative Methods,
University of Michigan
March 10, 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 prohibits 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

- Sec. 4. Identifying Methods to Assess Equity. (a) The Director of the Office of Management and Budget (OMB) shall, in partnership with the heads of agencies, study methods for assessing whether agency policies and actions create or exacerbate barriers to full and equal participation by all eligible individuals. The study should aim to identify the best methods, consistent with applicable law, to assist agencies in assessing equity with respect to race, ethnicity, religion, income, geography, gender identity, sexual orientation, and disability.
- Sec. 5. Conducting an Equity Assessment in Federal Agencies. The head of each agency, or designee, shall, in consultation with the Director of OMB, select certain of the agency's programs and policies for a review that will assess whether underserved communities and their members face systemic barriers in accessing benefits and opportunities available pursuant to those policies and programs.

Overview of the Talk

1. Existing methods are likely to be biased
   - BISG predictions are typically accurate and well calibrated
   - Still, estimates of racial disparities based on them can be biased
   - This is because race affects many aspects of our society

2. BIRDiE (Bayesian Instrumental Regression for Disparity Estimation)
   - New and more credible identification assumption
   - Flexible model allows for various racial disparity estimands
   - Sensitivity analysis for potential violation of the assumption
   - Open-source software package birdie available

3. Empirical validation
   - North Carolina voter file where self-reported race is observed
   - Estimates of racial differences in party registration
   - BIRDiE yields much smaller bias than the standard methods
   - Results are robust to potential violation of assumptions

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
  - $\mathbb{P}(Y_i = y \mid R_i = r)$: short regression
  - $\mathbb{P}(Y_i = y \mid R_i = r, X_i = x)$: long regression
- 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:

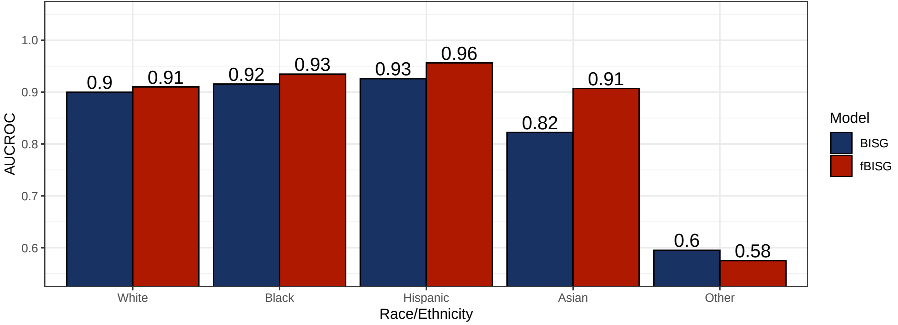
$$
\begin{aligned}
\hat{P}_{ir} &= \mathbb{P}(R_i = r \mid G_i = g, S_i = s) \\
&= \frac{\mathbb{P}(S_i = s \mid G_i = g, R_i = r)\,\mathbb{P}(G_i = g, R_i = r)}{\sum_{r'} \mathbb{P}(S_i = s \mid G_i = g, R_i = r')\,\mathbb{P}(G_i = g, R_i = r')} \\
&= \frac{\mathbb{P}(S_i = s \mid R_i = r)\,\mathbb{P}(G_i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

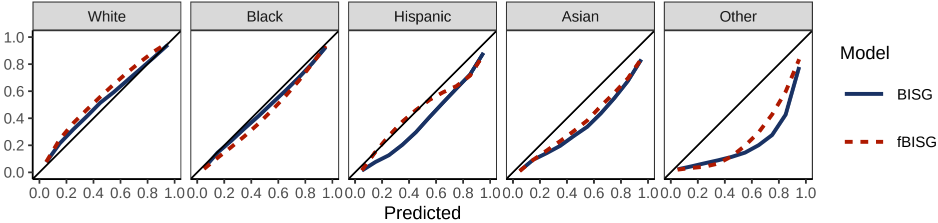
   - With covariates: $(G_i, X_i) \perp\!\!\!\perp S_i \mid R_i$
   - WRU software package (Imai and Kahna 2016)
2. Estimate racial disparities $\mu_{Y|R}(y \mid r) = \mathbb{P}(Y_i = y \mid R_i = r)$
   - weighting:

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

   - 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

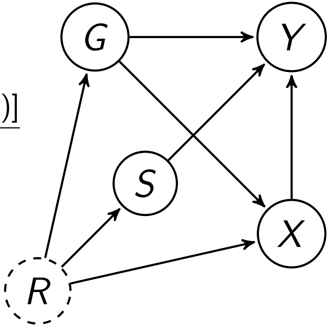
- Bias of the weighted estimator (Chen *et al.* 2019)

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

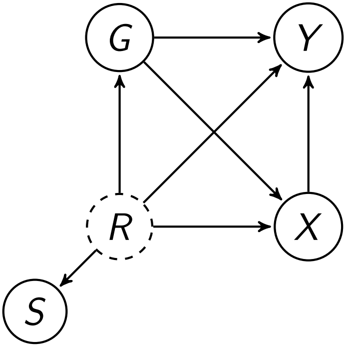
- Required assumption:

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

- 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
- Potential violations:
  - name-based discrimination
  - coarse racial categories
- Anonymous application

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 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:
  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, $\mathbf{Z} = (X, G)$

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

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

Computation

1. Small samples: direct inference

$$\pi(\Theta \mid \mathbf{Y}, \mathbf{G}, \mathbf{X}, \mathbf{S}) \propto \pi(\Theta) \prod_{i=1}^{N} \sum_{r \in \mathcal{R}} \pi(Y_i \mid r, G_i, X_i, \Theta) \hat{P}_{ir}$$

   - low-dimensional parameter space, MCMC is applicable (e.g., Stan)
   - but it is not scalable

2. Large samples: EM algorithm
   - E-step: update race probability (improvement upon BISG prob.)

$$\tilde{P}^{(t)}_{ir|Y} = \frac{\pi(Y_i \mid r, G_i, , X_i, \Theta^{(t)}) \hat{P}_{ir}}{\sum_{r' \in \mathcal{R}} \pi(Y_i \mid r', G_i, , X_i, \Theta^{(t)}) \hat{P}_{ir'}}$$

   - M-step: maximize each (y, x, g) group separately

$$\log \pi(\Theta^{(t+1)}) + \sum_{r \in \mathcal{R}} \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} \sum_{g \in \mathcal{G}} \log \pi(y \mid r, g, x, \Theta^{(t+1)}) \left(\sum_{i \in \mathcal{I}(yxg)} \tilde{P}^{(t)}_{ir|Y} \right)$$

Additional Explanatory Variables

- Long regression: $\mathbb{P}(Y_i = y \mid R_i = r, W_i = w)$ where W_i is not part of (X_i, G_i)

- Two strategies:
  1. Joint modeling: $\mathbb{P}(Y_i, W_i \mid R_i)$
  2. Iterative modeling: fit $\mathbb{P}(W_i \mid R_i)$ first and then use the updated race probability to fit $\mathbb{P}(Y_i \mid W_i, R_i)$

- Both approaches require:

$$W_i \perp\!\!\!\perp S_i \mid R_i, G_i, X_i \quad \text{and} \quad Y_i \perp\!\!\!\perp S_i \mid W_i, R_i, G_i, X_i$$

  or equivalently

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

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

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

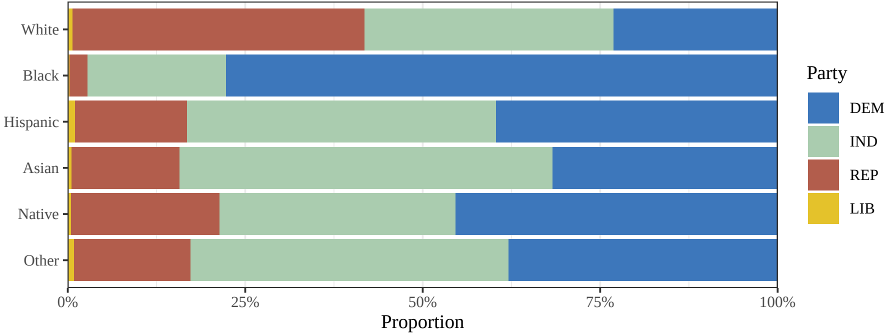
  - $f(\text{Imai}) = \text{Japanese}$, $f(\text{McCartan}) = \text{Irish}$, etc.
  - Assume instead

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

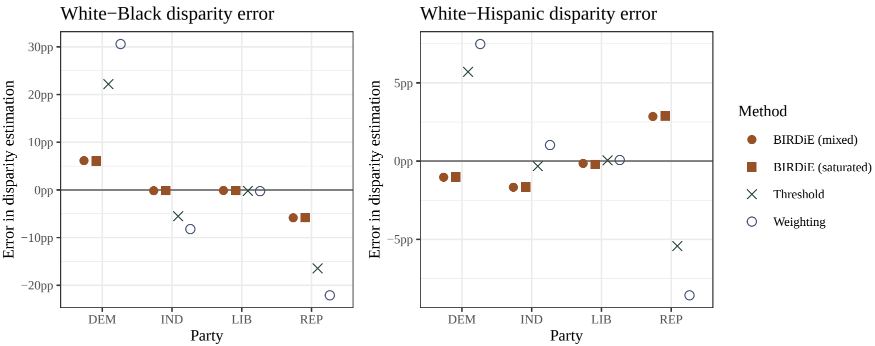
- 1930 Census provides 22 groups
  - Anglosphere and Black surname (third-or-more generation Whites and Blacks): Smith, Williams, Brown, ...
  - First wave European immigration (German, Nordic, and Irish): Burns, Olson, Wagner, ...
  - East Asian (Chinese, Japanese, Korean), South Asian (Indian, Southwest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

Empirical Validation

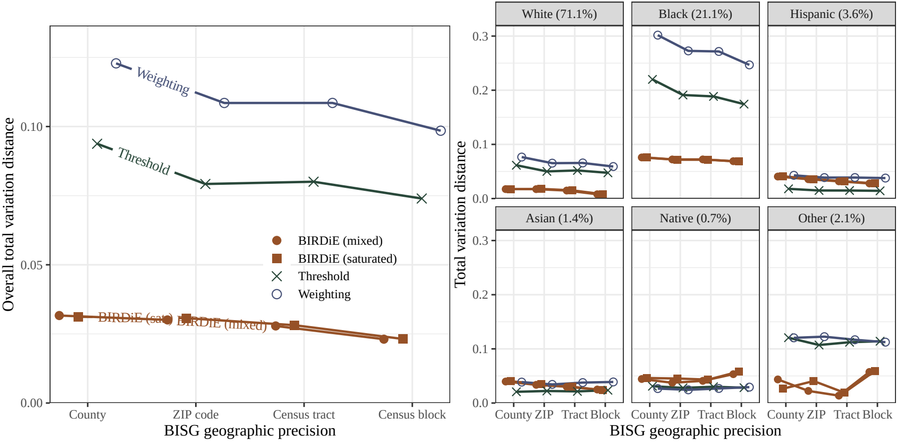
- 2022 North Carolina voter file: 5.8 millon voters with self-reported race
- Subset 1 million voters $\rightsquigarrow$ negligible sampling uncertainty
- Focus on party registration



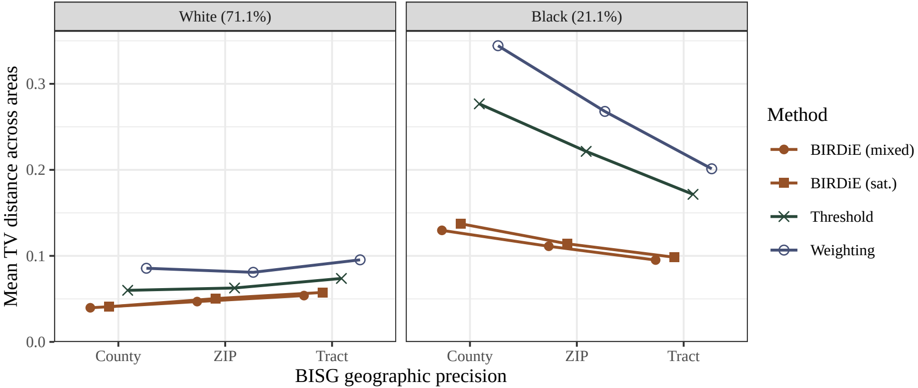
Estimates of Racial Disparity in Party Registration



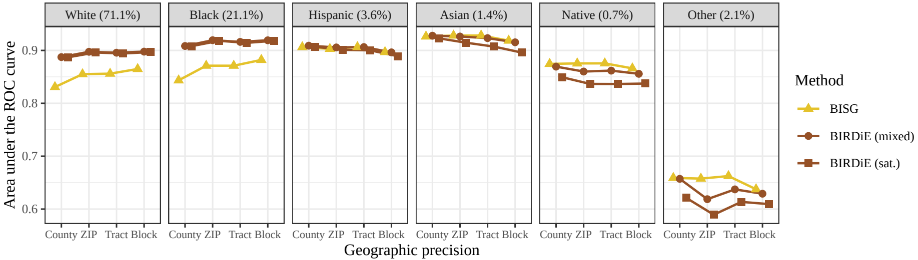
Total Variation Distance



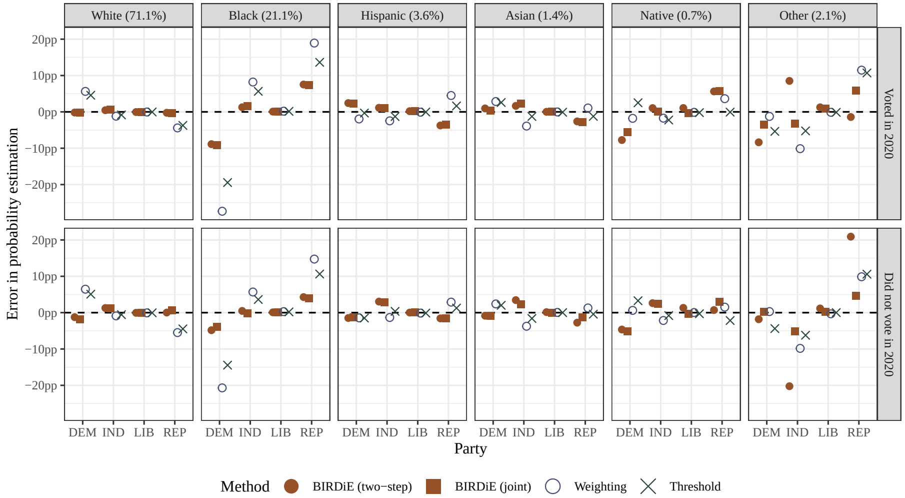
Small Area Estimation



Improved Race Probabilities

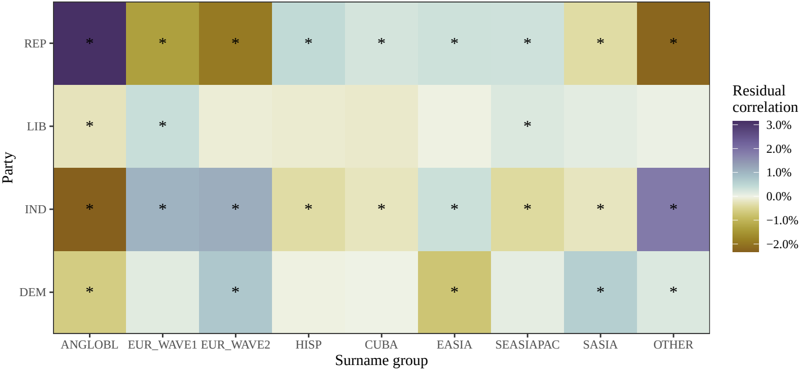


Estimates Conditional on an Additional Variable



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

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/