

Supplementary Material - Appendix SA2

Evaluation Results Based on Voter Registration Data for North Carolina

This appendix includes the data description and evaluation results based on voter registration data for North Carolina. The voter registration data is publicly available for free from the North Carolina State Board of Elections (<http://dl.ncsbe.gov/index.html?prefix=data>), and, in raw form, includes all active and inactive voters in the state who are currently on record (as of the date of the data download, which for this paper is September 28, 2017). The voter sample used for evaluation is obtained from the raw data by first selecting all active voters with valid race/ethnicity who were born on or after 1925 (4,450,135 voters), and then from those keep only the voters with available name and geographic information. These selection criteria result in a validation sample with 3,603,031 observations.

Table 1. Match Rates for Names and Geography

Match rate for first names	0.892
Match rate for surnames	0.941
Match rate for geography ²	0.961
Match rate for matches on both surname and geography	0.905
Match rate for matches on all three features	0.810
<i>N obs</i> ¹	<i>4,450,135</i>

Note 1: Includes all individuals with valid race/ethnicity who were active voters as of September 28, 2017 and were born on or after 1925.

Note 2: Voters who were matched with a geographical unit without any people are considered failed matches and thus are excluded from the numerator of the match rate.

Table 2: Self-Reported Race/Ethnicity Composition for the Validation Sample

Race/Ethnicity	Nobs	Percent
Hispanic	86,723	2.4%
NH Black	615,949	17.1%
NH White	2,847,336	79.0%
NH API	19,291	0.5%
NH AIAN	23,607	0.7%
NH Multiracial	10,455	0.3%
<i>Total</i>	<i>3,603,361</i>	<i>100.0%</i>

NOTE: The validation sample includes all active voters with valid race/ethnicity who were born on or after 1925, and have non-missing name and geographic demographic information.

Table 3: Distribution of Loans by Race/Ethnicity: Proxies vs. Self-Report

Method	Hispanic	NH Black	NH White	NH API	Weighted Average Overall Deviation from Self-Report	% Diff. in Average Deviation BIFSG - BISG	Relative Efficiency of BIFSG vs. BISG
Self-Report	2.4%	17.1%	79.0%	0.5%	(0)		
BIFSG	2.6%	21.0%	73.8%	0.6%	4.8%	-22.0%	64.6%
BISG	3.0%	21.8%	72.3%	0.8%	6.2%		

NOTE: Differences in prevalences between BIFSG and BISG are statistically significant at the 5 percent level for all racial/ethnic groups, as indicated by both pairwise t-tests and non-parametric sign tests.

Table 4. Correlation between Proxy Probability and Self-Reported Race/Ethnicity

Method	Correlation Coefficients					% Diff. in Average Correlation BIFSG - BISG	Relative Efficiency of BIFSG vs. BISG
	Hispanic	NH Black	NH White	NH API	Weighted Average		
BIFSG	0.761	0.668	0.681	0.741	0.678	5.3%	10.8%
BISG	0.725	0.635	0.646	0.695	0.644		

NOTE: Differences between the BIFSG and BISG correlations are statistically significant at the 5 percent level for all racial/ethnic groups, as indicated by a comparison of the 95 percent confidence intervals for Fisher's Z transformation of the correlation coefficients.

Table 5. Accuracy Metrics for Alternative Classification Methods

Classification Method	Proxy Method	Race/ Ethnicity	Total Applications in Group		% Unclassified ¹	False Negative Rate ² (%)	False Positive Rate ³ (%)
			Based on Self-Reporting	Estimated by Proxy Method			
MAP	BIFSG	Hispanic	86,723	83,737	0	27.8	25.2
	BISG	Hispanic	86,723	90,880	0	26.8	30.1
	BIFSG	NH Black	615,949	654,455	0	30.5	34.6
	BISG	NH Black	615,949	648,664	0	34.1	37.4
	BIFSG	NH White	2,847,336	2,831,229	0	8.3	7.8
	BISG	NH White	2,847,336	2,825,521	0	9.2	8.5
	BIFSG	NH API	19,291	16,810	0	33.6	23.8
	BISG	NH API	19,291	18,972	0	33.8	32.7
80% Threshold	BIFSG	Hispanic	86,723	60,969	26.3	19.6	15.7
	BISG	Hispanic	86,723	54,935	36.5	22.3	22.1
	BIFSG	NH Black	615,949	321,956	45.5	20.2	16.8
	BISG	NH Black	615,949	230,181	58.8	22.6	14.6
	BIFSG	NH White	2,847,336	2,256,512	21.8	2.6	3.9
	BISG	NH White	2,847,336	2,059,445	28.8	2.1	3.7
	BIFSG	NH API	19,291	12,412	25.4	24.8	12.7
	BISG	NH API	19,291	13,072	26.9	24.3	18.3

Note 1: % Unclassified represents the number of individuals that self-report belonging to a particular group but cannot be classified by the proxy method, as percent of the total population that self-reports belonging to that group.

Note 2: False negatives in a particular group are individuals who self-report belonging to that group, but whom the proxy method categorizes into another group. The false negative rate is the ratio of the number of false negatives in that group to the total population that self-reports belonging to that group.

Note 3: False positives in a particular group are individuals whom the proxy method assigns to that group, when in fact they belong to another group. The false positive rate is computed as the ratio of the number of false positives categorized in that group to the total population that the proxy classifies in that group.