Regulatory Arbitrage or Random Errors? Implications of Race Prediction Algorithms in Fair Lending Analysis

Daniel L. Greenwald	Sabrina T. Howell	Cangyuan Li	Emmanuel Yimfor
New York University	New York University	New York University	University of Michigan

Outline

Motivation

- Race proxies used in high-stakes contexts where race not observed
  - Regulators, firms, administrators, researchers
- Algorithms predict race using racial distribution of names and locations
  - Thought to have large error rates, esp. for Black Americans
  - Errors could be correlated with socioeconomic characteristics
- We study this issue in lending context
  - Self-reported race collected only for home mortgage applications
  - Regulators (CFPB, Fed, etc.) use predicted race to assess lender compliance with fair lending laws for many loan products
    - Auto
    - Personal
    - Student
    - Small Business
- **Do algorithmic race measures bias fair lending assessments? What are implications for lender incentives?**

What We Do: Setup

- Simple model of lending under different regulatory environments
  - Assume lenders would approve lower share of one group (B) than another (A) in absence of regulation
  - Regulator seeking to constrain this difference cannot observe race, uses noisy algorithm
  - Lenders tilt approval policies toward people with high algorithm-predicted prob in Group B
    → Regulation less efficient at reducing between-race inequality
    → Within-group inequality affected if errors correlated with socioeconomic chars
- Measures of Race
  - Focus on standard Baysian-Improved Surname Geocoding (BISG) algorithm
  - Compare to novel image-based race measure
  - Benchmark against self-identified race
- Setting: Small business lending
  - Applications to marketplace fintech lender
  - PPP loans

Key Findings

- BISG poorly predicts whether individual is Black
  - More false classifications than correct ones
  - Errors related to socioeconomic advantage
- Fair lending evaluations assess whether lender approves similar share of applicants in protected groups as control groups
  - Large variation across lenders in difference in approval rates between image-based race and BISG-based race
    $\rightarrow$ Could lead to faulty compliance decisions
  - Horse-race: image-based race predicts approval, BISG does not
- In counterfactual, policy shift from predicted to actual race
  $\rightarrow$ Reduces between-race inequality: Reallocates to Black
  $\rightarrow$ Increases within-race inequality: Reallocates to advantaged

Contribute to 3 Strands of Literature

- (1) Racial disparities in access to financial services
  - Mostly focused on residential mortgages and consumer credit markets
    Tootell, 1996; Bayer et al., 2018; Bhutta and Hizmo, 2021; Dobbie et al., 2020; Giacoletti et al., 2021; Begley and Purnanandam, 2021; Blattner and Nelson, 2021
  - Role of different lenders and especially new fintechs in serving minority and underserved populations
    Buchak et al., 2018; Tang, 2019; Fuster et al., 2019; Balyuk et al., 2020; Erel and Liebersohn, 2020; Berg et al., 2020; D'Acunto et al., 2020; Fairlie and Fossen, 2021; Bartlett et al., 2021; Chernenko and Scharfstein, 2021; Howell et al. 2022
  - We are first to examine how lender disparities in serving different groups depends on way race is measured
- (2) Racial disparities in entrepreneurship and beyond
  - Blanchflower et al. (2003), Robb and Robinson (2018), Asiedu et al. (2012), Bellucci et al. (2013), Fairlie et al. (2022), Arnold et al. (2018), Knowles et al. (2001), Anwar and Fang (2006), Charles and Guryan (2008), Price and Wolfers (2010)
  - We are first to address lender compliance with fair lending laws for small business loans

Contribute to 3 Strands of Literature

- (3) Methodology of identifying race
  - Relevant to research and policy that require measures of race, especially contexts where self-identified data are unavailable
    Dimmock et al., 2018; Pool et al., 2015; Egan et al., 2022; Frame et al., 2022; Ambrose et al., 2021; Jiang et al., 2021; Howell et al., 2022
  - Join new literature using image-based analysis (Athey et al. 2022)
  - We offer guidance on addressing bias from correlation between socioeconomic chars and algorithm errors

Outline

Simple Model: No Regulation

- Consider a lender who lends to two groups, A and B.
- Value of lending to individual i of type $j \in \{A, B\}$ is sum of group-specific mean and idiosyncratic shock:

$$v_{i,j} = \mu_j + \varepsilon_i, \qquad \varepsilon_i \sim U[\varepsilon^{min}, \varepsilon^{max}]$$

- With **no regulation ("NR")**, optimally approve borrower if $\varepsilon_i > -\mu_j$
- Probability of approval with no regulation:

$$\pi_{i,j}^{NR} = \text{const} + \gamma_1(\mu_B - \mu_A)\mathbb{I}_{j=B}$$

- We assume $\mu_A > \mu_B$ so that in absence of regulation lenders would provide fewer loans to Group B (motive for regulation).

Simple Model: Regulation Based on Actual Race

- Now imagine regulator wants to reduce gap in approval rates across groups, and can observe **actual race ("AR")**.
  - Constraint: gap between Group A and B approval rates $\leq \kappa$.

- Optimal lender policy is to approve borrower if $\varepsilon_i > \bar{\varepsilon}_j^{AR}$ (notation: λ^{AR} is multiplier on constraint, s_j is population share)

$$\bar{\varepsilon}_A^{AR} = -\mu_A \underbrace{+ \frac{\lambda^{AR}}{s_A}}_{\text{approval} \downarrow}, \qquad \bar{\varepsilon}_B^{AR} = -\mu_B \underbrace{- \frac{\lambda^{AR}}{s_B}}_{\text{approval} \uparrow}$$

Approval rate:

$$\pi_{i,j}^{AR} = \text{const} + \gamma_1 \left[\underbrace{(\mu_B - \mu_A)}_{<0} + \underbrace{\lambda^{AR}(s_A^{-1} + s_B^{-1})}_{>0} \right] \mathbb{I}_{j=B}$$

Simple Model: Regulation Based on Predicted Race

- Now assume regulator wants to close gap in approval rates but can only observe **predicted race ("PR")** from an algorithm (e.g., BISG).
  - Constraint: **predicted gap** between Group A and B approval rates $\leq \kappa$.
  - Let q denote predicted probability that borrower is in Group B.
- Optimal lender policy is to approve borrower if $\varepsilon_i > \bar{\varepsilon}_j^{PR}(q)$ for

$$\bar{\varepsilon}_j^{PR}(q) = -\mu_j - \lambda^{PR}\left[\frac{q}{s_B} - \frac{1-q}{s_A}\right]$$

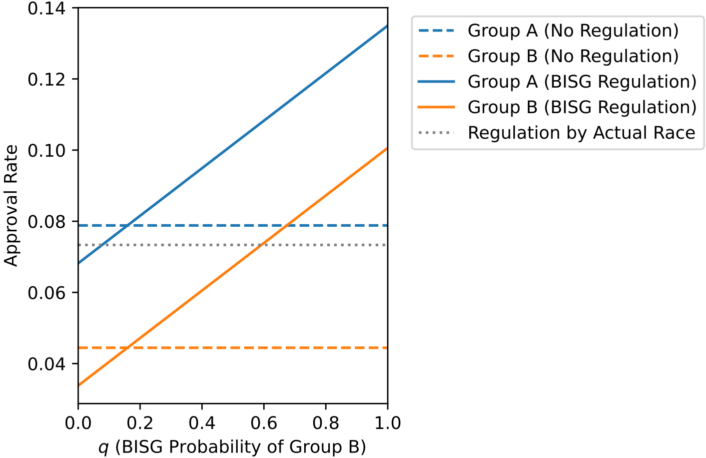
- Probability of approval:

$$\pi_i^{PR} = \text{const} + \underbrace{\gamma_1(\mu_B - \mu_A)\mathbb{I}_{j=B}}_{\text{original term}} + \underbrace{\gamma_1\lambda^{PR}\left(s_B^{-1} + s_A^{-1}\right)q_i}_{\text{effect of constraint}}$$

- More approval for borrowers with higher predicted probability q, but gap between members of Groups A and B unchanged conditional on q.

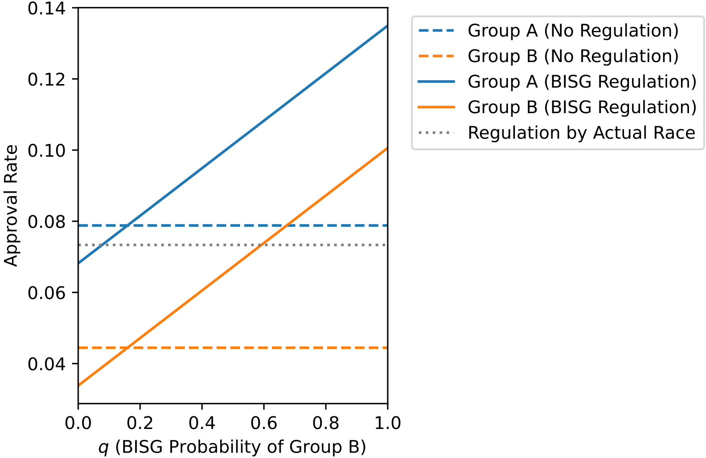
Simple Model: Numerical Example

- Below: approval rates by regulatory regime and q.
- With no regulation, large and constant gap between Groups A and B (dashed lines).



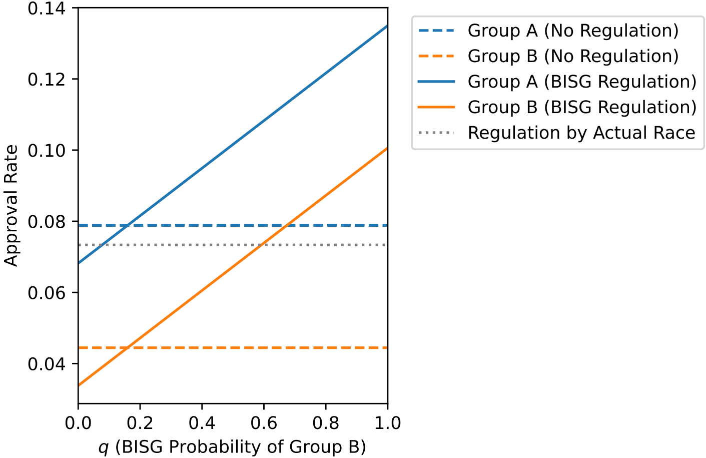
Simple Model: Numerical Example

- Dotted line: constraint based on **actual race (AR)** equalizes approval rates must be equal across groups.



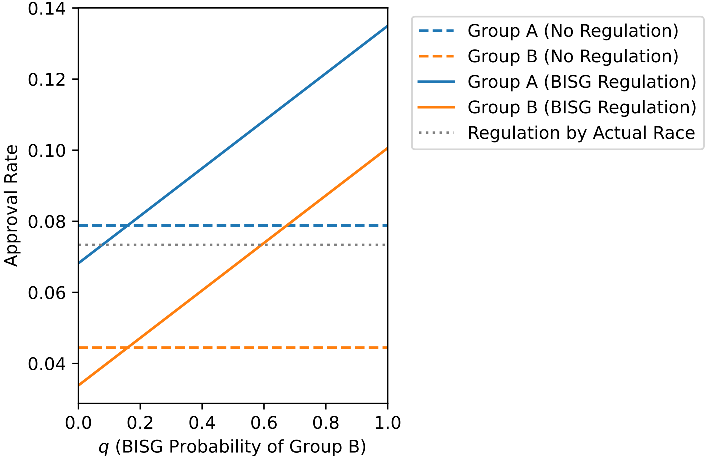
Simple Model: Numerical Example

- Solid lines: constraint based on **predicted race (PR)**.
- Tilts lending toward high q borrowers (who relax constraint), but gap is equally large conditional on q.



Simple Model: Numerical Example

- PR policy still somewhat effective at reducing gap because Group B has higher q on average.
- But substantial gap in actual approvals remains even when $\kappa = 0$.



Outline

Setting

- Focus on small business lending
  - Extensive evidence of racial disparities in credit
  - Regulators particularly focused on compliance with fair lending laws
  - Contribute to debate on Dodd-Frank Section 1071: Require small business lenders to collect & report information about race
- Employ two sources of data
  - (1) Lendio: Loan applications and funded loans from online small business loan marketplace Details
    - Enable us to observe lender approval decisions in a real-world context
  - (2) Paycheck Protection Program Loans: Govt-guaranteed, forgivable loans to small businesses during COVID-19 Details
    - Include self-identified measures of race in a real-world, non-mortgage lending context
  - Neither sample representative of U.S. small businesses or their lenders, but provide real-world comparisons of measures of race & concrete examples of error rate implications

LinkedIn and Geography-based Covariates

- Zip-level covariates from the 2019 American Community Survey (ACS)
- Racial animus measures from
  - Implicit Association Test (IAT); Xu et al., 2014
  - Nationscape survey which asks how favorably White respondents rate Black Americans; Bursztyn et al., 2021
  - Dissimilarity Index (differences in distributions of White and Black residents across city tracts); Massey and Denton, 1988
  - Isolation Index (probability of a Black resident sharing same city tract with another Black resident); Massey and Denton, 1988
- Individual-level education data from LinkedIn
  - Able to parse out degree by identifying keywords such as “Bachelor”, “Bachelor’s of Science”, “Master’s”, and so on from the text

▸ Lendio Covariate Sum Stats ▸ PPP Covariate Sum Stats

Outline

Self-Identified Race

- Race that an individual reports for themselves
- Person's self-ID race may differ from how they are perceived
  - e.g., you self-ID as White but loan officer perceives you as Black
  - Find many such cases in our data based on clerical review of images

Bayesian Improved Surname Geocoding (BISG)

- BISG combines two measures of race:
  - Geography-based: Assigns probability of individual's race based on proportion of individuals in a given location who are of same race
    - We use zip code, which is standard
  - Surname-based: Assigns probability of individual's race based on frequency distribution of names within population
- Issues:
  - Few names have strong correlation with Black
    - Among 10 most common last names for Black Americans (12% of Black pop), only one is majority Black
    - Some names strongly corr with Black (e.g. 90% of people surnamed "Washington" are Black), but compose small share of Black pop
- We use business owner name and address
  - Business address and residential address may differ
    - Match to real estate data (Infutor) to obtain home zip code
- Standard Python library calculates BISG race
  - Returns percent chance person is Hispanic, White, Black, Asian, Pacific Islander/Alaska Native, or Multiracial
  - Use either continuous measure or randomly assign with weight

Image-based Race

- Inferred from an individual's appearance

1. Obtain images from LinkedIn
   - Only use those where we can find the company name on profile
2. Use pre-trained classifier (VGG-Face via DeepFace)
3. Train random forest model on dataset of $\approx 170,000$ images of entrepreneurs
4. Apply clerical review to model output
   - ML-based classifications achieves accuracy of 91%
5. Classify each applicant as Black or not-Black

Suppose 25% of *marginal* applicants Black (Image or Self-ID)

If lenders observe race, a 60% approval rate of all marginal applicants might yield:

We get a different *picture* when we use BISG as a proxy

Why? Sorting on BISG (regulatory motive) changes the composition of marginal applicants that are approved

Martin Brown
BISG score: 74

Claudette Hudson
BISG score: 67

Britt Wagner
BISG score: 0.001

Jay Thomas
BISG score: 0.001

Image-based race is *similar*, not equal, to Self-ID; Image-based race approximates Self-ID, but is closer to perceived race

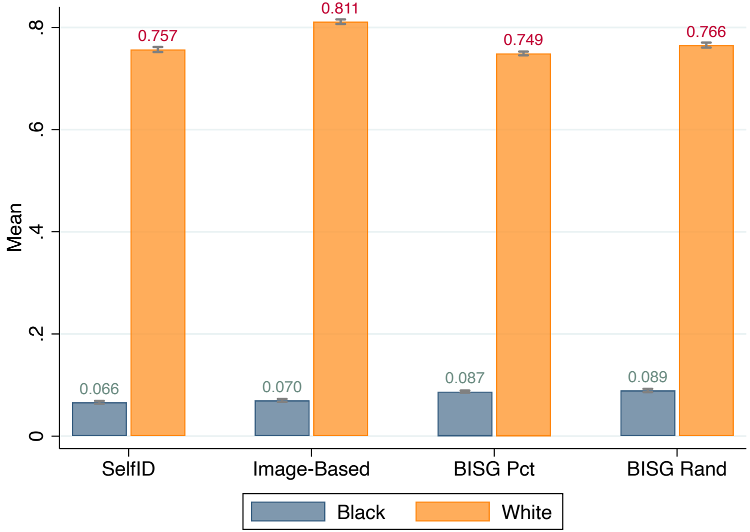
Marcy Ybarra
Self-ID: Black
Image: Hispanic

Daniel Bailey
Self-ID: Black
Image: White

Jessica Williams
Self-ID: Black
Image: Asian

Mary Reed
Self-ID: Black
Image: White

Race Shares Across Measures



- PPP, Unique borrower level (very similar in Lendio)

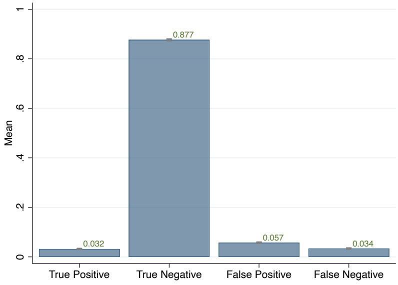
Correlations Between Race Variables (PPP)

	Black (SelfID)	Black (Image)
Black (Image)	0.87***	1.00
BISG Black Percent	0.54***	0.56***
N = 28,990		

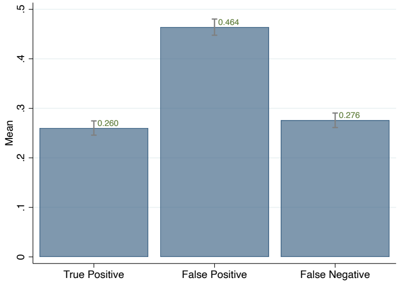
	Black (SelfID)	Black (Image)
Black (Image)	0.87***	1.00
Black (BISG)	0.37***	0.38***
N = 28,990		

	Black (SelfID)	Black (Image)
Black (Image)	0.87***	1.00
Black (Geography)	0.19***	0.21***
N = 28,994		

BISG Error Rates Relative to SelfID Race (PPP)



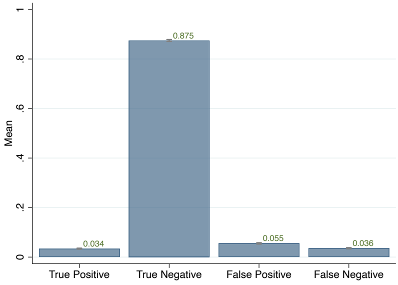
Baseline = SelfID (PPP)



Baseline = SelfID (PPP, Within Black)

- **True Positive**: Person is Black and is predicted as Black
- **True Negative**: Person is non-Black and is predicted as non-Black
- **False Positive**: Person is non-Black and is predicted as Black
- **False Negative**: Person is Black and is predicted as non-Black

BISG Error Rates Relative to Image-Based Race (PPP)



Baseline = Image (PPP)

Baseline = Image (PPP, Within Black)

- Similar in Lendio

Takeaway

- Race measures deviate from one another
- BISG performs much worse than image-based race when self-identified race is the benchmark
- BISG predicts more false positives and false negatives than true positives when classifying Black

Outline

Hypothesis

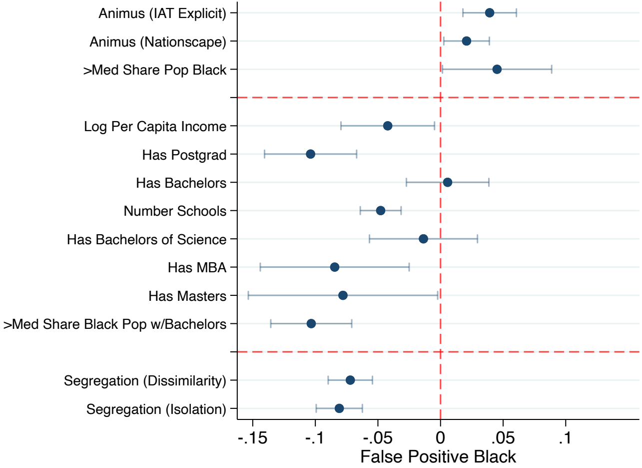
- If errors in BISG-predicted race are random $\rightarrow$ noise
  - Problematic because will increase the challenges of ascertaining compliance with fair lending standards
  - Make research estimates of disparate impact less precise
- More concerning: If errors are systematically related to characteristics relevant for underwriting decisions
  - BISG's reliance on location and name may lead errors to be correlated with socioeconomic advantage
  - We expect when BISG judges an individual as Black who is not Black (a false positive), that person may have systematically different characteristics associated with higher underwriting risk
  - Expect reverse for false negatives; associated with lower underwriting risk

Analysis

- Exclude true negatives: Focus attention on narrow sample that at least one measure judged to be Black
- (1) Compare false positives (image-based not Black, BISG Black) to other 2 groups who are image-based Black (true positive + false negative)
- (2) Compare false negatives (image-based Black, BISG not Black) to other 2 groups who are BISG Black (true positive + false positive)
- Run series of regressions that are simple correlations, e.g. projecting an indicator for false positive Black on a characteristic of the borrower

Socioeconomic Characteristics and False Positive BISG Error

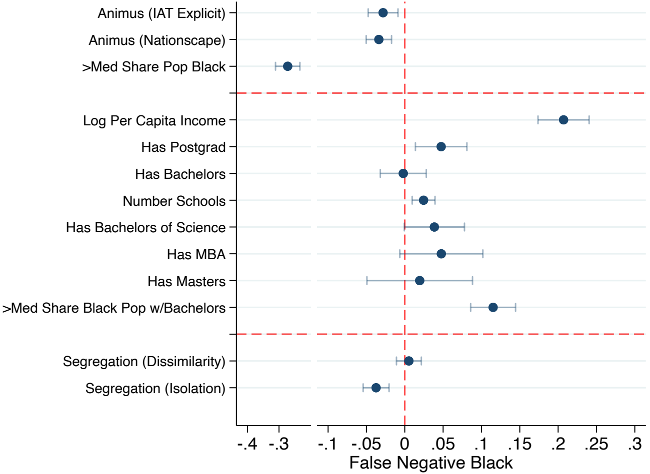
- Predict False Positive in PPP sample



- Similar results in SelfID and Lendio samples

Socioeconomic Characteristics and False Negative BISG Error

- Predict False Negative in PPP sample



- Similar results in ▸ SelfID and Lendio samples

Takeaway

- Geographies where BISG tends to make errors—predicting people to be Black when they are not—are also areas with particularly strong historic systematic disadvantage for Black borrowers
- Individual-specific higher education, which is strongly related to wealth formation and is likely highly related to borrower risk (e.g., Crissey 2009), associated with BISG being much less likely to predict Black when not Black
- $\rightarrow$ False positives are relatively more socioeconomically disadvantaged

Outline

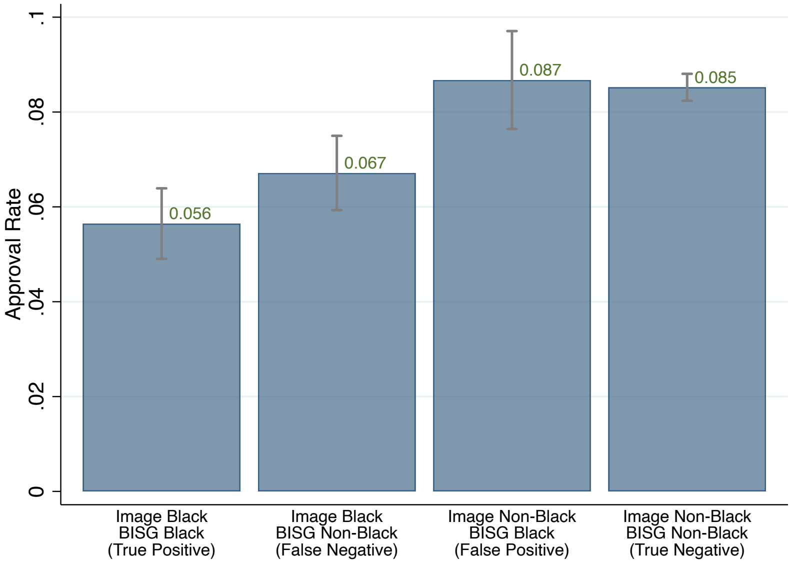
Regulatory Context

- Key element of complying with fair lending rules is disparate treatment and disparate impact analyses: Is lender serving protected groups (e.g., Black) in a similar way as the majority group (e.g., White)?
  - Full analysis requires information on applicant risk
- Comparing approval rates across groups is important first step
  - If lender can show approves similar share of applicants in protected groups as control groups, regulators will not typically look further for evidence of discriminatory conduct
- E.g. U.S. Interagency Fair Lending Examination Procedures
  - Apply to 5 federal agencies including Federal Reserve and FDIC)
  - First indicators of potential disparate treatment in underwriting is "Substantial disparities among the approval/denial rates for applicants by monitored prohibited basis characteristic."
- We focus on disparities in approval rates, important dimension of compliance evaluation

Lender-Level Summary Statistics, Lendio Approval Statistics by Race

	N	Mean	Median	SD
Share of Applicants by Race:				
Share Apps from Black (Image)	101	0.120	0.105	0.104
Share Apps from White (Image)	101	0.740	0.753	0.162
Share Apps from Black (BISG)	101	0.116	0.109	0.068
Share Apps from White (BISG)	101	0.691	0.687	0.121
Approval Rate Among Applicants of Race:				
Approval Rate Black (Image)	101	0.076	0.000	0.195
Approval Rate White (Image)	101	0.091	0.045	0.125
Approval Rate Black (BISG)	101	0.090	0.024	0.198
Approval Rate White (BISG)	101	0.100	0.043	0.148
Loan Rate Among Borrowers of Race:				
Share Loans to Black (Image)	101	0.097	0.000	0.195
Share Loans to White (Image)	101	0.587	0.742	0.383
Share Loans to Black (BISG)	101	0.081	0.047	0.119
Share Loans to White (BISG)	101	0.539	0.667	0.346
Difference in Rates by Race Measure (Image Less BISG):				
Diff Approval Rate Black	101	-0.003	0.000	0.062
Diff Approval Rate White	101	-0.002	0.000	0.080
Diff Loan Rate Black	101	0.016	0.000	0.135
Diff Loan Rate White	101	0.027	0.042	0.181

Lendio - Approval Rate by Group



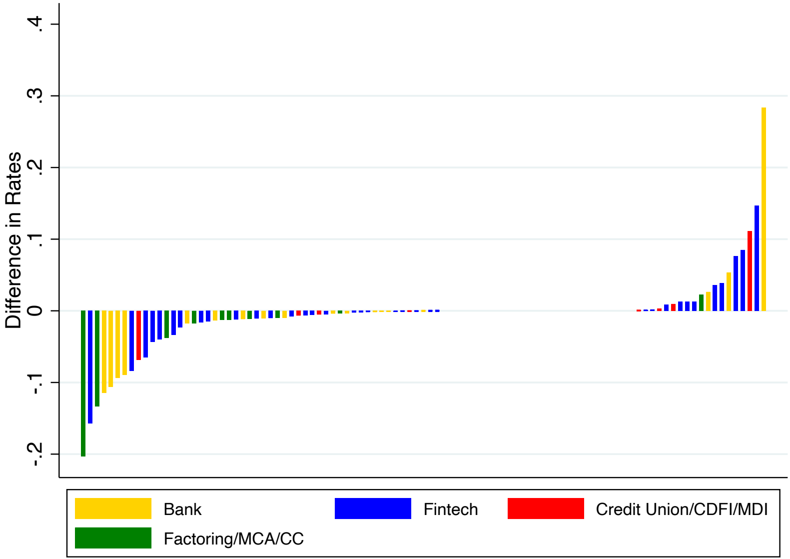
Differences in Approval Rates

- At lender level, construct measure for difference in approval rates using image-based race vs. BISG-based race

$$\Delta_{\text{Share Black Appr}} = \bar{\pi}_B - \bar{\pi}_B^{BISG} = \frac{\#\,\text{Image Black Approved}}{\#\,\text{Image Black Applicants}} - \frac{\#\,\text{BISG Black Approved}}{\#\,\text{BISG Black Applicants}}$$

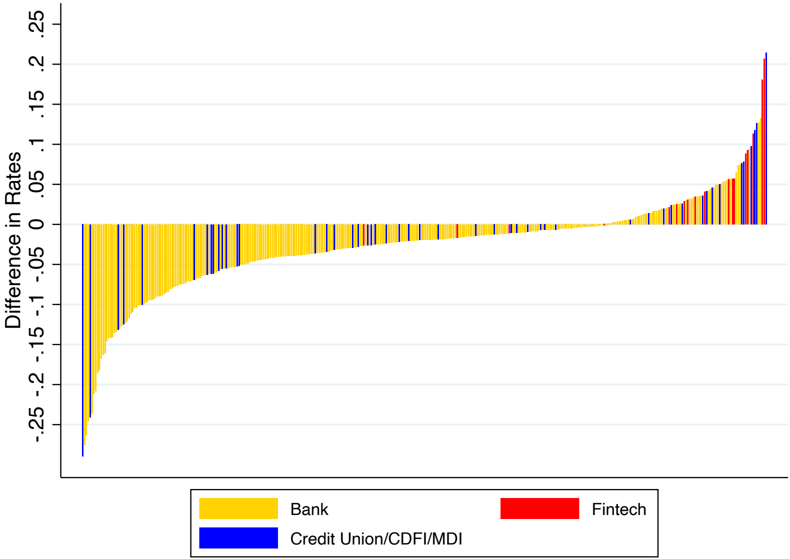
- When $\Delta_{\text{Share Black Appr}}$ is $+$, lender serving the Black pop at a higher rate than they appear to be with BISG (either more false neg or less false pos)
- Since false positive is correlated with disadvantaged socioeconomic status, a lender who serves more advantaged Black borrowers—perhaps because the lender is "cream skimming" or because of demand-side factors—will have a higher difference
- When $\Delta_{\text{Share Black Appr}}$ is $-$, BISG errors make lender appear more compliant with fair lending laws than they actually are

Lender-level $\Delta_{\text{Share Black Appr}}$ (Lendio)



- Large variation across lenders

Lender-level $\Delta_{\text{Share Black Appr}}$ (PPP)



- Based on share loans to Black (rather than approval rate)

Suggestive ordering by lender type

- $\Delta_{\text{Share Black Appr}}$ more often negative for banks and factoring/MCA/CC factoring, and more often positive for fintech lenders
  - MCA, factoring, business CC products are long-standing and pre-fintech, very high interest rates
- Banks and other conventional small business lenders that typically rely on soft information for underwriting (Petersen and Rajan, 1994; Berger and Black, 2011) on the negative side
- While fintechs, which are most automated and arms-length (Howell et al. (2022) and Balyuk et al. (2020)) on the positive side.
- Caveat: sample of lenders is far from representative of small business lenders in the U.S.

Relationship between Lender Type and Differences in Lending Rates Across Race Measures (Lendio, PPP)

	Lendio (Share Approved)			PPP (Share Loans)		
	$\Delta > 0$	Δ	$\Delta >$ 75 Pctile	$\Delta > 0$	Δ	$\Delta >$ 75 Pctile
	(1)	(2)	(3)	(4)	(5)	(6)
Fintech	0.14	0.02*	0.17*	0.31***	0.04**	0.38***
	(0.10)	(0.01)	(0.09)	(0.10)	(0.01)	(0.10)
Factoring/MCA/CC	0.01	-0.02	-0.06			
	(0.15)	(0.02)	(0.14)			
Large Bank				-0.07	0.01	-0.04
				(0.10)	(0.02)	(0.11)
Medium Bank				-0.02	-0.01**	-0.03
				(0.04)	(0.01)	(0.05)
Credit Union/CDFI				0.21***	0.01	0.18**
				(0.07)	(0.01)	(0.07)
MDI				0.20**	0.02	0.17
				(0.10)	(0.01)	(0.11)
Observations	92	92	92	438	438	438
R-squared	0.027	0.069	0.051	0.058	0.041	0.058
Y-mean	0.250	-0.004	0.228	0.221	-0.027	0.249

Implications for Policy: One Possible Interpretation

- Fintechs tend to be less regulated and have much higher costs of capital: Looking to cream-skim
- Banks appear more compliant using BISG-based race predictions than they would using image-based race: Benefit from BISG-based fair lending evaluation
  - Consistent with this, vociferously lobbying against rule requiring lenders to collect self-reported race data in small business lending

Crux of Compliance Evaluation: Approval Disparities

- Explore at the application level whether the measures of race predict loan approval using Lendio data

$$\mathbb{1}(\text{Approved}_{i,l}) = \alpha_l + \alpha_t + \beta \mathbb{1}(\text{Black}_i) + \mathbf{X}_i \delta + \varepsilon_{il}$$

- FE for lender (α_l) and application year (α_t)
- Control for log amount of funding sought and in some models socioecon chars
- Drop applications where cannot ascertain whether application approved vs. rejected

How Image and BISG Race Measures Predict Loan Approval

Dependent Variable:	Approved					
	(1)	(2)	(3)	(4)	(5)	(6)
Black (Image)	-0.017***		-0.015***	-0.014***		
	(0.005)		(0.005)	(0.005)		
Black (BISG)		-0.011**	-0.005	-0.004		
		(0.005)	(0.005)	(0.005)		
True Positive Black (BISG)					-0.021***	-0.020***
					(0.006)	(0.007)
False Positive Black (BISG)					-0.004	-0.002
					(0.006)	(0.006)
False Negative Black (BISG)					-0.013**	-0.013**
					(0.006)	(0.006)
Observations	49,375	49,375	49,375	49,375	49,375	49,375
Application Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Lender FE	Yes	Yes	Yes	Yes	Yes	Yes
Log Amt Sought	Yes	Yes	Yes	Yes	Yes	Yes
Socioecon Controls	No	No	No	Yes	No	Yes
P-value			0.109	0.096	0.014	0.018
R-squared	0.074	0.074	0.074	0.075	0.074	0.075
Y-mean	0.08	0.08	0.08	0.08	0.08	0.08

- Col 1-2: Both indicators negatively predict approvals, but image-based race has 50% larger impact

How Image and BISG Race Measures Predict Loan Approval

Dependent Variable:	Approved					
	(1)	(2)	(3)	(4)	(5)	(6)
Black (Image)	-0.017***		-0.015***	-0.014***		
	(0.005)		(0.005)	(0.005)		
Black (BISG)		-0.011**	-0.005	-0.004		
		(0.005)	(0.005)	(0.005)		
True Positive Black (BISG)					-0.021***	-0.020***
					(0.006)	(0.007)
False Positive Black (BISG)					-0.004	-0.002
					(0.006)	(0.006)
False Negative Black (BISG)					-0.013**	-0.013**
					(0.006)	(0.006)
Observations	49,375	49,375	49,375	49,375	49,375	49,375
Application Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Lender FE	Yes	Yes	Yes	Yes	Yes	Yes
Log Amt Sought	Yes	Yes	Yes	Yes	Yes	Yes
Socioecon Controls	No	No	No	Yes	No	Yes
P-value			0.109	0.096	0.014	0.018
R-squared	0.074	0.074	0.074	0.075	0.074	0.075
Y-mean	0.08	0.08	0.08	0.08	0.08	0.08

- Col 3-4: Predictive power of BISG indicator subsumed by the image-based indicator, provides no independent variation

How Image and BISG Race Measures Predict Loan Approval

Dependent Variable:	Approved					
	(1)	(2)	(3)	(4)	(5)	(6)
Black (Image)	-0.017***		-0.015***	-0.014***		
	(0.005)		(0.005)	(0.005)		
Black (BISG)		-0.011**	-0.005	-0.004		
		(0.005)	(0.005)	(0.005)		
True Positive Black (BISG)					-0.021***	-0.020***
					(0.006)	(0.007)
False Positive Black (BISG)					-0.004	-0.002
					(0.006)	(0.006)
False Negative Black (BISG)					-0.013**	-0.013**
					(0.006)	(0.006)
Observations	49,375	49,375	49,375	49,375	49,375	49,375
Application Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Lender FE	Yes	Yes	Yes	Yes	Yes	Yes
Log Amt Sought	Yes	Yes	Yes	Yes	Yes	Yes
Socioecon Controls	No	No	No	Yes	No	Yes
P-value			0.109	0.096	0.014	0.018
R-squared	0.074	0.074	0.074	0.075	0.074	0.075
Y-mean	0.08	0.08	0.08	0.08	0.08	0.08

- Col 5-6: Disaggregate BISG indicator, omitted group is true negatives
- Large negative coeff for true positive, but zero for false positive
- Since BISG indicator mixes these 2 groups (of similar size)
    - Has much lower predictive power
    - And misses negative impact of false negative

Counterfactual Exercise

Characteristic	(1) BISG Weight ↓	(2) Image Weight ↑	(3) Net Change
Panel A: Applicant Level			
BISG False Positive Black	-0.211	-0.105	-0.316
BISG False Negative Black	-0.131	0.783	0.652
Image Race = Black	-0.585	1.473	0.888
Panel B: Geographic Level			
Log Per Capita Income	0.326	-0.207	0.119
Share Pop Black	-0.348	0.289	-0.058
Share Pop w/Bachelors	0.078	-0.048	0.030

- Row 1: Reducing weight in approval regression on BISG-Black or increasing weight on image-Black reduces share of false positives who are approved

Counterfactual Exercise

Characteristic	(1) BISG Weight ↓	(2) Image Weight ↑	(3) Net Change
Panel A: Applicant Level			
BISG False Positive Black	-0.211	-0.105	-0.316
BISG False Negative Black	-0.131	0.783	0.652
Image Race = Black	-0.585	1.473	0.888
Panel B: Geographic Level			
Log Per Capita Income	0.326	-0.207	0.119
Share Pop Black	-0.348	0.289	-0.058
Share Pop w/Bachelors	0.078	-0.048	0.030

- Row 2: Reducing weight on BISG-Black actually decreases approval rate for false negatives (they still have above-average BISG-Black score), but combined, the policies increase Black borrower approval rate (col 3)

Counterfactual Exercise

Characteristic	(1) BISG Weight ↓	(2) Image Weight ↑	(3) Net Change
Panel A: Applicant Level			
BISG False Positive Black	-0.211	-0.105	-0.316
BISG False Negative Black	-0.131	0.783	0.652
Image Race = Black	-0.585	1.473	0.888
Panel B: Geographic Level			
Log Per Capita Income	0.326	-0.207	0.119
Share Pop Black	-0.348	0.289	-0.058
Share Pop w/Bachelors	0.078	-0.048	0.030

- Row 3: Reducing weight on BISG-Black actually decreases approval rate for image-Black (because image-based race and BISG race correlated), but since correlation highly imperfect, more than undone by larger increase in Black
- $\rightarrow$ Net effect of regulatory change strongly positive

Counterfactual Exercise

Characteristic	(1) BISG Weight ↓	(2) Image Weight ↑	(3) Net Change
Panel A: Applicant Level			
BISG False Positive Black	-0.211	-0.105	-0.316
BISG False Negative Black	-0.131	0.783	0.652
Image Race = Black	-0.585	1.473	0.888
Panel B: Geographic Level			
Log Per Capita Income	0.326	-0.207	0.119
Share Pop Black	-0.348	0.289	-0.058
Share Pop w/Bachelors	0.078	-0.048	0.030

- Rows 4-6: Regulatory change leads to more lending to higher income, more educated areas, less to high-pop-Black areas, which are traditionally underserved
- $\rightarrow$ May increase within-race inequality

Takeaway

- Predictive power of the BISG indicator on approvals is attenuated by its classification errors, and is largely subsumed by our image-based measure.
- If lender serves pop with high false negative rate, and these truly Black individuals are less likely to get loans (as is the case in our sample on average)
    - Then lender will appear to be approving a larger share of Black applicants, and be more compliant than are in reality
    - May create distortionary incentives for lender
- In extension, add first name
    - First name widely used to test for discrimination in correspondence audit studies
    - Algorithm performs somewhat better
        - But driven by stronger predictive power of false positives
        - May reflect false positives being more associated with lower socioeconomic status when first name included (i.e., first name is more "Black")

Outline

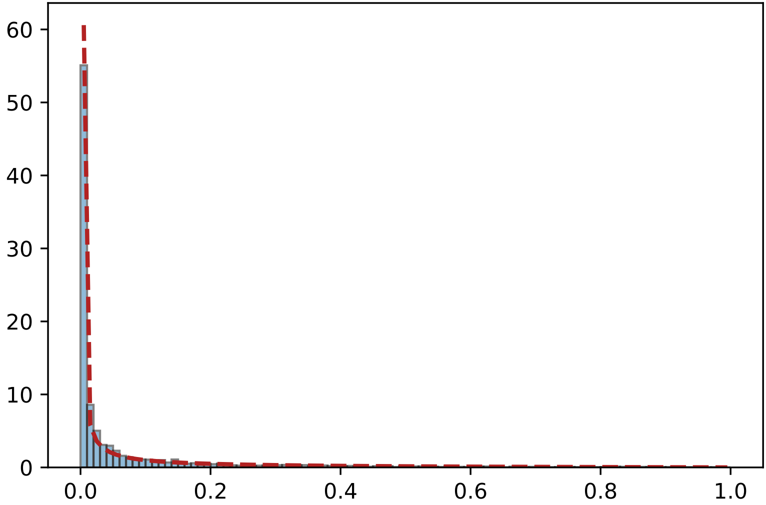
Concluding Discussion

- “Folk knowledge” that widely used race prediction algorithms based on demographic characteristics of name and location perform poorly, especially for predicting Black
- If errors correlated with socioeconomic characteristics that are, in turn, related to loan profitability
  - $\rightarrow$ Link between apparent compliance with fair lending laws and the measure of race (image, self-reported, algorithmically predicted)
- Has important implications for policy
  - Whether particular lenders are sanctioned
  - Whether new fintech lenders given freer rein if can serve truly Black applicants at higher rates than traditional lenders

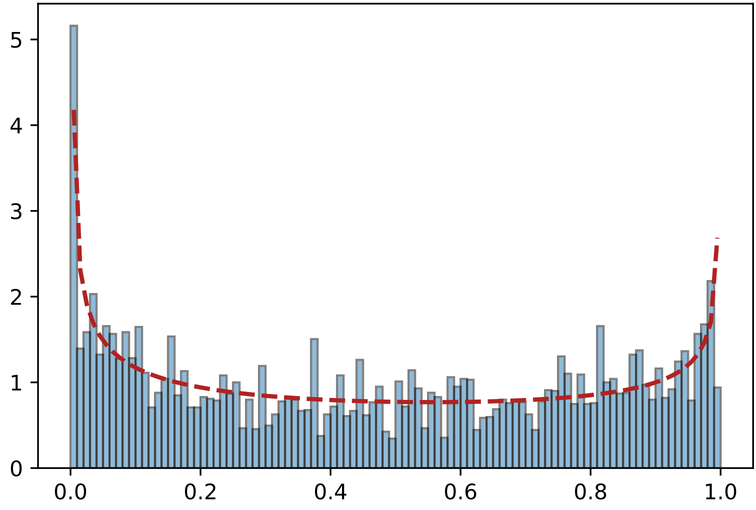
Concluding Discussion

- We offer first systematic documentation of socioeconomic implications errors in race prediction algorithms, focusing on standard (BISG)
- Develop measure of perceived race using applicant images
- Image-based Black race negatively predicts loan approval much more strongly than BISG-based Black race
  - Reflects lower chance of approval among individuals who are false negative Black
- BISG errors will generate substantial differences in compliance evaluations depending on the type of borrower a lender serves
- Regulators, researchers, practitioners should consider objective before choosing method
  - E.g. if aim is to identify people who are Black and also relatively disadvantaged within Black pop, BISG works fairly well
  - E.g. if aim is to focus on discrimination on the basis of skin tone and facial features alone, BISG has important shortcomings

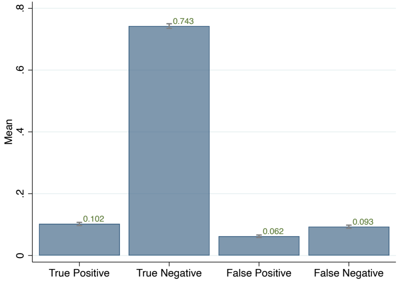
BISG Densities by Actual Race, Non-Black



BISG Densities by Actual Race, Black



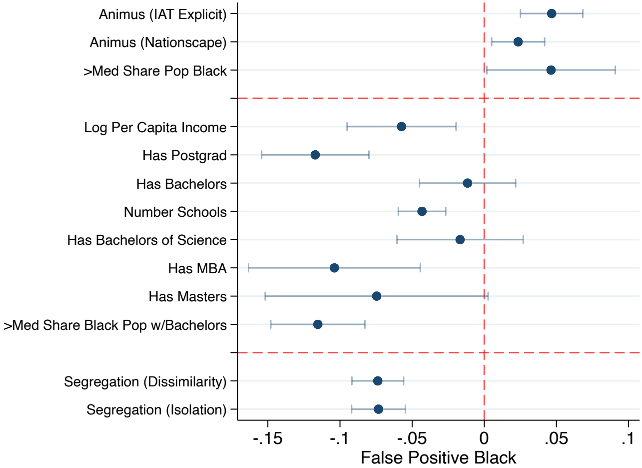
BISG Error Rates (Unique Borrower Level), Lendio Image



Baseline = Image (Lendio)

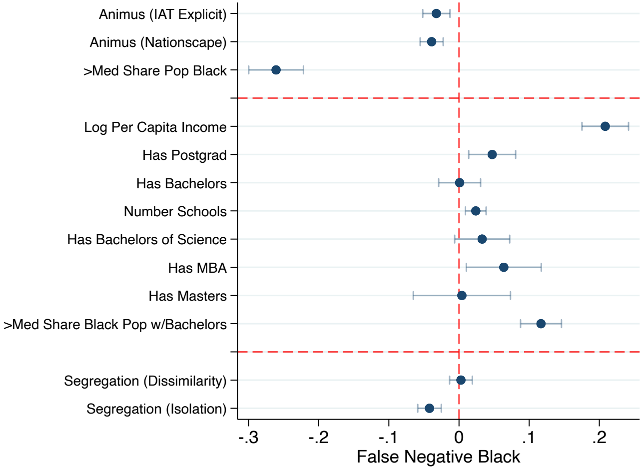
Baseline = Image (Lendio, Within Black)

Correlation Between Socioeconomic Covariates and BISG Errors Relative to Self-Identified Race (PPP), False Positive



Back

Correlation Between Socioeconomic Covariates and BISG Errors Relative to Self-Identified Race (PPP), False Negative



Back

Lendio Loan Applications

- Lendio is an online loan marketplace for small businesses
  - Firms submit one application to Lendio, who forwards to lenders
  - Lenders decide to make an offer, which the borrower can accept / reject
- We use data from 2017-2019
- 674,203 applications from 160,942 unique firms
  - After BISG: 139,759 firms
  - After image-based: 11,566 firms, 49,401 applications Lendio Sum Stats
- We do not observe if loan was not funded because lender rejected, or applicant rejected an offer
  - Lendio only forwards application to additional lender if rejected
  - Identify rejected as application not funded and sent elsewhere subsequently

Back

Paycheck Protection Program Loans

- PPP data from April 3, 2020 to May 31, 2021
- No application data, only loans that were actually made
- Begin with 11.8 million PPP loans
  - Restrict to 4,775,702 loans made before Feb 24, 2021 (when rules were changed to prioritize lending to small and minority-owned firms)
  - Restrict to 933,645 loans where borrower self-reported race
  - Restrict to 867,151 loans with "valid" person names
  - 27,861 loans with BISG and image-based race

▸ PPP Sum Stats ▸ Back

Lender Classification

- Banks
- Credit Unions
- CDFIs/MDIs: Community Development Financial Institutions, nonprofits, and Minority Depository Institutions, as classified by the FDIC
- Factoring/MCA/CC (Lendio-only): Factoring, Merchant Cash Advance and business credit card lenders
  - High-cost alternatives to bank loans for small businesses
  - Factoring: Selling accounts receivable to lender
  - MCA: Loan repayment is percentage of sales
- Fintech: Lenders designated as such by the SBA, online lenders founded since 2005, recieved VC investment, or originate primarily for fintech partners / platforms

Loan Application and Lender Summary Statistics (Lendio)

Panel A: Application-Level Data

	N	Mean	Median	SD
Loan Approval:				
Amount Sought	47,504	104,014	50,000	372,628
Amount Funded	3,875	52,031	26,000	98,213
Approved	47,504	0.082	0.000	0.274
Rejected	47,504	0.918	1.000	0.274
Share Lender Type:				
Bank	47,504	0.243	0.000	0.429
Fintech	47,504	0.486	0.000	0.500
Credit Union/CDFI	47,504	0.137	0.000	0.343
MDI	47,504	0.001	0.000	0.022
Factoring/MCA/CC	47,504	0.134	0.000	0.341

Panel B: Unique Applicant-Level Data

	N	Mean	Median	SD
Loan Approval:				
Amount Sought	11,190	99,732	49,999	520,159
Amount Funded	2,891	51,818	27,500	73,033
Approved	11,190	0.157	0.000	0.330
Rejected	11,190	0.843	1.000	0.330
Share Lender Type:				
Bank	11,190	0.316	0.214	0.346
Fintech	11,190	0.425	0.463	0.351
Credit Union/CDFI	11,190	0.158	0.000	0.262
MDI	11,190	0.001	0.000	0.019
Factoring/MCA/CC	11,190	0.101	0.000	0.189

Panel C: Unique Lender-Level Data

	N	Mean	Median	SD
Loan Variables:				
Number Loans	101	438.087	39.000	957.677
Amount Funded	101	53,549	32,812	48,522
Share Lender Type:				
Bank	103	0.311	0.000	0.465
Fintech	103	0.456	0.000	0.501
Credit Union/CDFI	103	0.087	0.000	0.284
MDI	103	0.019	0.000	0.139
Factoring/MCA/CC	103	0.107	0.000	0.310

Back

Loan Application and Lender Summary Statistics (Lendio), Application-Level Data

	N	Mean	Median	SD
Loan Approval:				
Amount Sought	49,401	102,451	50,000	185,662
Amount Funded	4,012	52,100	25,390	97,739
Approved	49,401	0.081	0.000	0.273
Rejected	49,401	0.919	1.000	0.273
Share Lender Type:				
Bank	49,401	0.240	0.000	0.427
Fintech	49,401	0.486	0.000	0.500
Credit Union/CDFI	49,401	0.138	0.000	0.345
MDI	49,401	0.000	0.000	0.021
Factoring/MCA/CC	49,401	0.135	0.000	0.342

Loan Application and Lender Summary Statistics (Lendio), Unique Applicant-Level Data

	N	Mean	Median	SD
Loan Approval:				
Amount Sought	11,566	96,815	49,999	225,195
Amount Funded	2,995	51,467	27,000	73,367
Approved	11,566	0.157	0.000	0.330
Rejected	11,566	0.843	1.000	0.330
Share Lender Type:				
Bank	11,566	0.311	0.200	0.343
Fintech	11,566	0.427	0.467	0.351
Credit Union/CDFI	11,566	0.160	0.000	0.262
MDI	11,566	0.000	0.000	0.016
Factoring/MCA/CC	11,566	0.102	0.000	0.191

Loan Application and Lender Summary Statistics (Lendio), Unique Lender-Level Data

	N	Mean	Median	SD
Loan Variables:				
Number Loans	101	446.653	40.000	965.231
Amount Funded	101	164,956	105,000	154,838
Share Lender Type:				
Bank	101	0.307	0.000	0.464
Fintech	101	0.465	0.000	0.501
Credit Union/CDFI	101	0.089	0.000	0.286
MDI	101	0.020	0.000	0.140
Factoring/MCA/CC	101	0.109	0.000	0.313

Loan and Lender Summary Statistics (PPP)

Panel A: Unique Borrower-Level Data

	N	Mean	Median	SD
Loan Approval:				
Number Loans	22,618	614.636	231.000	761.995
Loan Amt	22,618	138001.568	38,461.000	384265.622
Share Lender Type:				
Large Bank	22,618	0.401	0.000	0.490
Medium Bank	22,618	0.280	0.000	0.449
Small Bank	22,618	0.144	0.000	0.352
Fintech	22,618	0.103	0.000	0.304
Credit Union/CDFI	22,618	0.041	0.000	0.199
MDI	22,618	0.030	0.000	0.170

Panel B: Unique Lender-Level Data

	N	Mean	Median	SD
Loan Variables:				
Number Loans	369	61.295	20.000	184.416
Loan Amt	369	166730.113	44,166.648	497478.855
Share Lender Type:				
Large Bank	369	0.046	0.000	0.210
Medium Bank	369	0.423	0.000	0.495
Small Bank	369	0.344	0.000	0.476
Fintech	369	0.049	0.000	0.216
Credit Union/CDFI	369	0.098	0.000	0.297
MDI	369	0.041	0.000	0.198

Back

Loan and Lender Summary Statistics (PPP), Unique Borrower-Level Data

	N	Mean	Median	SD
Loan Approval:				
Number Loans	27,861	682.036	287.000	830.806
Loan Amt	27,861	124177.394	30,000.000	364681.422
Share Lender Type:				
Large Bank	27,861	0.365	0.000	0.481
Medium Bank	27,861	0.266	0.000	0.442
Small Bank	27,861	0.145	0.000	0.352
Fintech	27,861	0.138	0.000	0.345
Credit Union/CDFI	27,861	0.056	0.000	0.230
MDI	27,861	0.030	0.000	0.170

Loan and Lender Summary Statistics (PPP), Unique Lender-Level Data

	N	Mean	Median	SD
Loan Variables:				
Number Loans	439	63.465	19.000	198.361
Loan Amt	439	156578.898	38,200.000	378820.353
Share Lender Type:				
Large Bank	439	0.039	0.000	0.193
Medium Bank	439	0.408	0.000	0.492
Small Bank	439	0.355	0.000	0.479
Fintech	439	0.046	0.000	0.209
Credit Union/CDFI	439	0.112	0.000	0.315
MDI	439	0.041	0.000	0.199

Correlations Between Race Variables (PPP), Part 2

	Black (SelfID)	Black (Image)
Black (Image)	0.87***	1.00
Black (Surname)	0.18***	0.19***
N = 29,002		

	Black (SelfID)	Black (Image)
Black (Image)	0.87***	1.00
Black (First-name+Surname)	0.25***	0.27***
N = 26,444		

	Black (SelfID)	Black (Image)
Black (Image)	0.87***	1.00
Black (BIFSG)	0.41***	0.43***
N = 26,427		

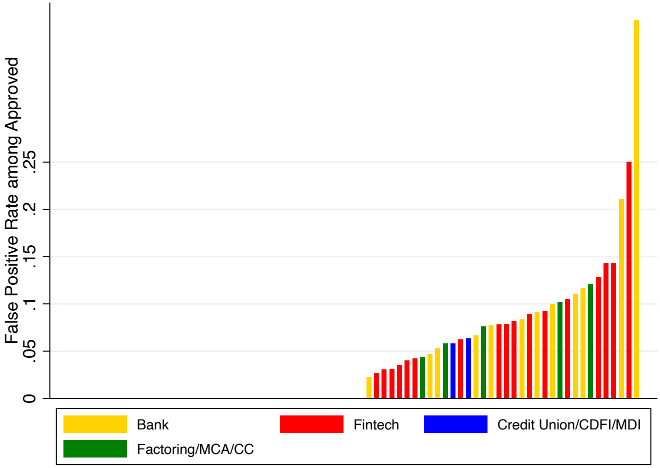
Lender-Level Summary Statistics, PPP Statistics by Race

	N	Mean	Median	SD
Loan Rate Among Borrowers of Race:				
Share Loans to Black (Image)	439	0.052	0.029	0.078
Share Loans to White (Image)	439	0.861	0.903	0.137
Share Loans to Black (BISG)	439	0.080	0.050	0.085
Share Loans to White (BISG)	439	0.771	0.795	0.168
Difference in Rates by Race Measure (Image Less BISG):				
Diff Loan Rate Black	439	-0.028	-0.018	0.066
Diff Loan Rate White	439	0.090	0.080	0.097

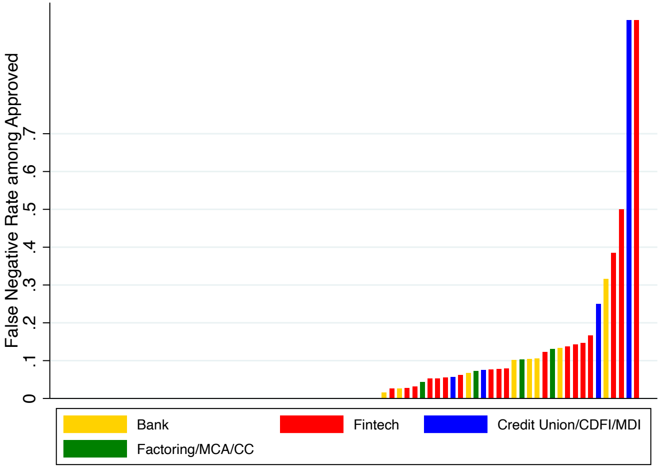
How Image and Geography-based Race Measures Predict Loan Approval

Dependent Variable:	Approved							
	Geography				Surname			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Black (Image)	-0.017***	-0.015***			-0.017***	-0.015***		
	(0.005)	(0.005)			(0.005)	(0.005)		
Black (Algorithm)	0.001	0.005			0.001	0.001		
	(0.006)	(0.007)			(0.006)	(0.006)		
True Positive Black			-0.012	-0.005			-0.026***	-0.024***
			(0.008)	(0.010)			(0.006)	(0.007)
False Positive Black			-0.001	0.002			0.007	0.007
			(0.008)	(0.008)			(0.007)	(0.007)
False Negative Black			-0.019***	-0.017***			-0.012**	-0.010*
			(0.005)	(0.005)			(0.006)	(0.006)
Observations	49,375	49,375	49,375	49,375	49,375	49,375	49,375	49,375
Application Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lender FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Log Amt Sought	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Socioecon Controls	No	Yes	No	Yes	No	Yes	No	Yes
P-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R-squared	0.074	0.075	0.074	0.075	0.074	0.075	0.074	0.075
Y-mean	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08

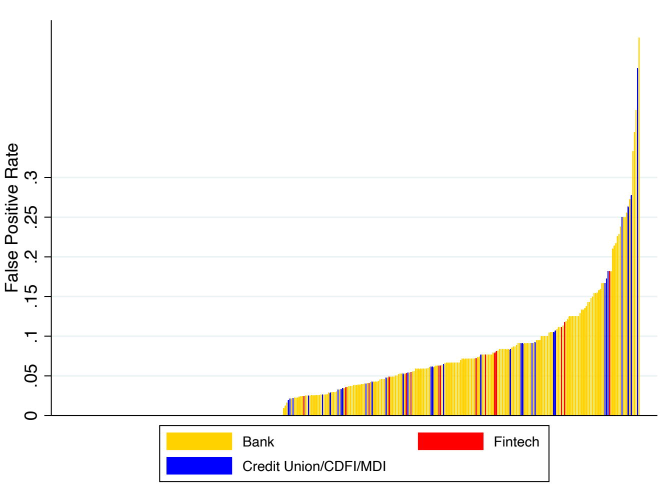
Lender-level average false positive rate (Lendio)



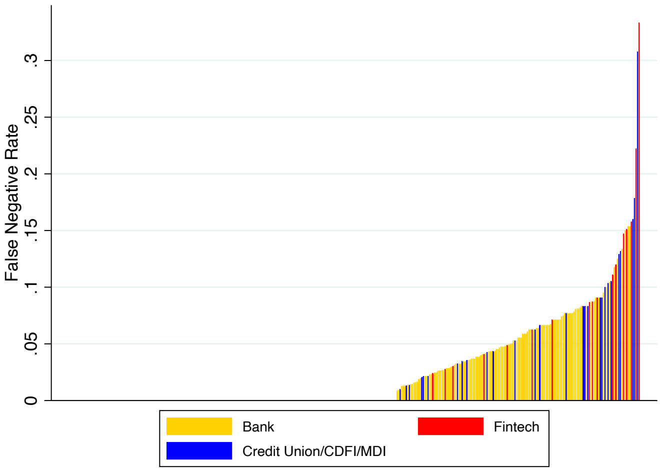
Lender-level Average False Negative Rate (Lendio)



Lender-level Average False Positive Rate (PPP)



Lender-level Average False Negative Rate (PPP)



Applicant Covariate Summary Statistics (Lendio, One-per-applicant Level)

	N	Mean	Median	SD
Covariates (Geographic Level):				
Log Per Capita Income	13,172	10.53	10.49	0.42
Animus (IAT Explicit)	13,172	-0.08	-0.06	0.76
Animus (Nationscape)	13,172	-0.10	0.03	0.95
Segregation (Dissimilarity)	13,172	-0.10	-0.02	0.91
Segregation (Isolation)	13,172	-0.10	-0.11	0.95
Share Pop Black	13,172	0.14	0.06	0.18
Share Black Pop w/Bachelors	13,172	0.21	0.18	0.15
Covariates (Geographic Level) Within Image + BISG Black Population:				
Log Per Capita Income	3,390	10.39	10.36	0.40
Animus (IAT Explicit)	3,390	0.01	-0.00	0.73
Animus (Nationscape)	3,390	0.08	0.11	0.82
Segregation (Dissimilarity)	3,390	0.00	0.14	0.92
Segregation (Isolation)	3,390	0.35	0.37	0.83
Share Pop Black	3,390	0.29	0.22	0.25
Share Black Pop w/Bachelors	3,390	0.18	0.16	0.12
Covariates (Applicant Level, From LinkedIn):				
Has Bachelors	13,172	0.45	0.00	0.50
Has Postgrad	13,172	0.15	0.00	0.35
Number Schools	13,172	1.72	1.83	0.86
Has Bachelors of Science	13,172	0.11	0.00	0.31
Has Masters	13,172	0.03	0.00	0.17
Has MBA	13,172	0.05	0.00	0.22
Covariates (Applicant Level, From LinkedIn) Within Image + BISG Black Population:				
Has Bachelors	3,390	0.43	0.00	0.49
Has Postgrad	3,390	0.15	0.00	0.36
Number Schools	3,390	1.75	1.83	0.88
Has Bachelors of Science	3,390	0.09	0.00	0.29
Has Masters	3,390	0.03	0.00	0.18
Has MBA	3,390	0.05	0.00	0.22

Back

PPP - Borrower Covariate Variable Summary Statistics (One-per-applicant Level)

	N	Mean	Median	SD
Covariates (Geographic Level):				
Log Per Capita Income	26,427	10.62	10.59	0.41
Animus (IAT Explicit)	26,427	-0.19	-0.15	0.75
Animus (Nationscape)	26,427	-0.23	-0.22	1.00
Segregation (Dissimilarity)	26,427	-0.02	0.02	0.91
Segregation (Isolation)	26,427	-0.26	-0.21	0.98
Share Pop Black	26,427	0.09	0.04	0.14
Share Pop w/Bachelors	26,427	0.30	0.28	0.14
Share Black Pop w/Bachelors	26,427	0.21	0.19	0.16
Covariates (Geographic Level) Within Image + BISG Black Population:				
Log Per Capita Income	3,629	10.46	10.43	0.43
Animus (IAT Explicit)	3,629	-0.05	-0.08	0.76
Animus (Nationscape)	3,629	0.03	0.11	0.89
Segregation (Dissimilarity)	3,629	0.07	0.15	0.91
Segregation (Isolation)	3,629	0.33	0.33	0.87
Share Pop Black	3,629	0.27	0.19	0.25
Share Pop w/Bachelors	3,629	0.26	0.23	0.15
Share Black Pop w/Bachelors	3,629	0.19	0.17	0.12
Covariates (Applicant Level, From LinkedIn):				
Has Bachelors	26,427	0.59	1.00	0.49
Number Schools	26,427	1.82	2.00	0.95
Has Bachelors of Science	26,427	0.17	0.00	0.38
Has Masters	26,427	0.04	0.00	0.20
Has MBA	26,427	0.07	0.00	0.25
Covariates (Applicant Level, From LinkedIn) Within Self Identified + BISG Black Population:				
Has Bachelors	3,578	0.60	1.00	0.49
Number Schools	3,578	1.87	2.00	0.99
Has Bachelors of Science	3,578	0.17	0.00	0.37
Has Masters	3,578	0.05	0.00	0.21
Has MBA	3,578	0.08	0.00	0.27

Back

Correlations Between Race Variables (Lendio), Part 1

	Black (Image)
BISG Black Percent	0.65***
N = 62,151	

	Black (Image)
Black (BISG)	0.48***
N = 62,151	

	Black (Image)
Black (Geography)	0.25***
N = 62,156	

Correlations Between Race Variables (Lendio), Part 2

	Black (Image)
Black (Surname)	0.27***
N = 62,174	

	Black (Image)
Black (First-name+Surname)	0.36***
N = 55,362	

	Black (Image)
Black (BIFSG)	0.52***
N = 55,338	

PPP - Regressions with Image-based False Positive Black

Dependent Variable:	False Positive Black												
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Animus (IAT Explicit)	0.03***												
	(0.01)												
Animus (Nationscape)		0.03***											
		(0.01)											
Above Median Share Pop Black			0.11***										
			(0.02)										
Log Per Capita Income				-0.07***									
				(0.02)									
Has Postgrad					-0.08***								
					(0.02)								
Has Bachelors						0.01							
						(0.01)							
Number Schools							-0.03***						
							(0.01)						
Has Bachelors of Science								-0.01					
								(0.02)					
Has MBA									-0.07***				
									(0.03)				
Has Masters										-0.06*			
										(0.03)			
Above Median Share Black Pop w/Bachelors											-0.12***		
											(0.01)		
Segregation (Dissimilarity)												-0.06***	
												(0.01)	
Segregation (Isolation)													-0.05***
													(0.01)
Observations	4,882	4,882	4,880	4,882	4,882	4,882	4,882	4,882	4,882	4,882	4,850	4,882	4,882
R-squared	0.003	0.002	0.007	0.004	0.005	0.000	0.004	0.000	0.002	0.001	0.014	0.011	0.007
Y-mean	0.461	0.461	0.461	0.461	0.461	0.461	0.461	0.461	0.461	0.461	0.461	0.461	0.461

PPP - Regressions with Image-based False Negative Black

Dependent Variable:	False Positive Black												
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Animus (IAT Explicit)	-0.02**												
	(0.01)												
Animus (Nationscape)		-0.03***											
		(0.01)											
Above Median Share Pop Black			-0.33***										
			(0.02)										
Log Per Capita Income				0.21***									
				(0.01)									
Has Postgrad					0.03**								
					(0.01)								
Has Bachelors						-0.01							
						(0.01)							
Number Schools							0.02***						
							(0.01)						
Has Bachelors of Science								0.03*					
								(0.02)					
Has MBA									0.04				
									(0.02)				
Has Masters										0.02			
										(0.03)			
Above Median Share Black Pop w/Bachelors											0.11***		
											(0.01)		
Segregation (Dissimilarity)												-0.01	
												(0.01)	
Segregation (Isolation)													-0.06***
													(0.01)
Observations	4,882	4,882	4,880	4,882	4,882	4,882	4,882	4,882	4,882	4,882	4,850	4,882	4,882
R-squared	0.001	0.005	0.074	0.038	0.001	0.000	0.002	0.001	0.000	0.000	0.016	0.001	0.015
Y-mean	0.278	0.278	0.278	0.278	0.278	0.278	0.278	0.278	0.278	0.278	0.278	0.278	0.278

PPP - Regressions with SelfID-based False Positive Black

Dependent Variable:	False Positive Black												
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Animus (IAT Explicit)	0.03***												
	(0.01)												
Animus (Nationscape)		0.01											
		(0.01)											
Above Median Share Pop Black			0.00										
			(0.02)										
Log Per Capita Income				0.01									
				(0.02)									
Has Postgrad					-0.09***								
					(0.02)								
Has Bachelors						-0.02							
						(0.01)							
Number Schools							-0.05***						
							(0.01)						
Has Bachelors of Science								-0.03					
								(0.02)					
Has MBA									-0.04*				
									(0.03)				
Has Masters										-0.09***			
										(0.03)			
Above Median Share Black Pop w/Bachelors											-0.07***		
											(0.01)		
Segregation (Dissimilarity)												-0.07***	
												(0.01)	
Segregation (Isolation)													-0.09***
													(0.01)
Observations	4,973	4,973	4,972	4,973	4,973	4,973	4,973	4,973	4,973	4,973	4,966	4,973	4,973
R-squared	0.002	0.000	0.000	0.000	0.007	0.000	0.009	0.000	0.001	0.002	0.005	0.019	0.025
Y-mean	0.400	0.400	0.400	0.400	0.400	0.400	0.400	0.400	0.400	0.400	0.400	0.400	0.400

PPP - Regressions with SelfID-based False Negative Black

Dependent Variable:	False Positive Black												
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Animus (IAT Explicit)	-0.02**												
	(0.01)												
Animus (Nationscape)		-0.03***											
		(0.01)											
Above Median Share Pop Black			-0.26***										
			(0.02)										
Log Per Capita Income				0.20***									
				(0.01)									
Has Postgrad					0.04***								
					(0.01)								
Has Bachelors						0.01							
						(0.01)							
Number Schools							0.02***						
							(0.01)						
Has Bachelors of Science								0.05***					
								(0.02)					
Has MBA									0.04*				
									(0.02)				
Has Masters										0.02			
										(0.03)			
Above Median Share Black Pop w/Bachelors											0.12***		
											(0.01)		
Segregation (Dissimilarity)												-0.01	
												(0.01)	
Segregation (Isolation)													-0.05***
													(0.01)
Observations	4,973	4,973	4,972	4,973	4,973	4,973	4,973	4,973	4,973	4,973	4,966	4,973	4,973
R-squared	0.001	0.003	0.038	0.036	0.001	0.000	0.002	0.001	0.001	0.000	0.017	0.000	0.008
Y-mean	0.291	0.291	0.291	0.291	0.291	0.291	0.291	0.291	0.291	0.291	0.291	0.291	0.291

Lendio - Regressions with Image-based False Positive Black

Dependent Variable:	False Positive Black												
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Animus (IAT Explicit)	0.01												
	(0.01)												
Animus (Nationscape)		0.00											
		(0.01)											
Above Median Share Pop Black			0.06***										
			(0.02)										
Log Per Capita Income				-0.04*									
				(0.02)									
Has Postgrad					-0.04**								
					(0.02)								
Has Bachelors						-0.00							
						(0.02)							
Number Schools							-0.03***						
							(0.01)						
Has Bachelors of Science								0.02					
								(0.03)					
Has MBA									-0.06				
									(0.04)				
Has Masters										-0.10**			
										(0.05)			
Above Median Share Black Pop w/Bachelors											-0.07***		
											(0.02)		
Segregation (Dissimilarity)												-0.06***	
												(0.01)	
Segregation (Isolation)													-0.07***
													(0.01)
Observations	3,232	3,232	3,232	3,232	3,232	3,232	3,232	3,232	3,232	3,232	3,227	3,232	3,232
R-squared	0.000	0.000	0.002	0.001	0.001	0.000	0.004	0.000	0.001	0.002	0.005	0.017	0.016
Y-mean	0.303	0.303	0.303	0.303	0.303	0.303	0.303	0.303	0.303	0.303	0.303	0.303	0.303

Lendio - Regressions with Image-based False Negative Black

Dependent Variable:	False Positive Black												
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Animus (IAT Explicit)	-0.01												
	(0.01)												
Animus (Nationscape)		-0.02*											
		(0.01)											
Above Median Share Pop Black			-0.37***										
			(0.02)										
Log Per Capita Income				0.28***									
				(0.02)									
Has Postgrad					-0.00								
					(0.02)								
Has Bachelors						-0.01							
						(0.02)							
Number Schools							0.02*						
							(0.01)						
Has Bachelors of Science								0.00					
								(0.03)					
Has MBA									-0.01				
									(0.04)				
Has Masters										-0.01			
										(0.05)			
Above Median Share Black Pop w/Bachelors											0.12***		
											(0.02)		
Segregation (Dissimilarity)												-0.01	
												(0.01)	
Segregation (Isolation)													-0.08***
													(0.01)
Observations	3,232	3,232	3,232	3,232	3,232	3,232	3,232	3,232	3,232	3,232	3,227	3,232	3,232
R-squared	0.001	0.001	0.085	0.056	0.000	0.000	0.001	0.000	0.000	0.000	0.017	0.001	0.019
Y-mean	0.330	0.330	0.330	0.330	0.330	0.330	0.330	0.330	0.330	0.330	0.330	0.330	0.330

Lendio - Correlations Between Application and Approvals by Race

	Share Apps from Black (Image)
Share Apps from Black (Image)	1.00
Approval Rate Black (Image)	0.43***

	Share Apps from White (Image)
Share Apps from White (Image)	1.00
Approval Rate White (Image)	0.13

	Share Apps from Black (BISG)
Share Apps from Black (BISG)	1.00
Approval Rate Black (BISG)	0.06

	Share Apps from White (BISG)
Share Apps from White (BISG)	1.00
Approval Rate White (BISG)	-0.10

How Image and Firstname+Surname-based Race Measures Predict Loan Approval

Dependent Variable:	Approved			
	(1)	(2)	(3)	(4)
Black (Image)	-0.014***	-0.012**		
	(0.005)	(0.006)		
Black (Firstname+Surname)	-0.011*	-0.011*		
	(0.006)	(0.006)		
True Positive Black (Firstname+Surname)			-0.027***	-0.025***
			(0.007)	(0.008)
False Positive Black (Firstname+Surname)			-0.008	-0.008
			(0.008)	(0.007)
False Negative Black (Firstname+Surname)			-0.013**	-0.011*
			(0.006)	(0.006)
Observations	43,943	43,943	43,943	43,943
Application Year FE	Yes	Yes	Yes	Yes
Lender FE	Yes	Yes	Yes	Yes
Log Amt Sought	Yes	Yes	Yes	Yes
Socioecon Controls	No	Yes	No	Yes
P-value	0.000	0.000	0.000	0.000
R-squared	0.074	0.076	0.074	0.076
Y-mean	0.08	0.08	0.08	0.08

How Image and BISFG-based Race Measures Predict Loan Approval

Dependent Variable:	Approved			
	(1)	(2)	(3)	(4)
Black (Image)	-0.012**	-0.012**		
	(0.006)	(0.006)		
Black (BIFSG)	-0.011*	-0.010		
	(0.006)	(0.007)		
True Positive Black (BIFSG)			-0.023***	-0.022***
			(0.006)	(0.007)
False Positive Black (BIFSG)			-0.011	-0.010
			(0.009)	(0.009)
False Negative Black (BIFSG)			-0.012*	-0.012*
			(0.007)	(0.007)
Observations	43,923	43,923	43,923	43,923
Application Year FE	Yes	Yes	Yes	Yes
Lender FE	Yes	Yes	Yes	Yes
Log Amt Sought	Yes	Yes	Yes	Yes
Socioecon Controls	No	Yes	No	Yes
P-value	0.000	0.000	0.000	0.000
R-squared	0.074	0.076	0.074	0.076
Y-mean	0.08	0.08	0.08	0.08

Predictive Power of All Race Measures on Loan Approval

Dependent Variable:	Approved					
	(1)	(2)	(3)	(4)	(5)	(6)
Black (Image)	-0.017***					
	(0.005)					
Black (BISG)		-0.011**				
		(0.005)				
Black (Geography)			-0.002			
			(0.006)			
Black (Surname)				-0.003		
				(0.005)		
Black (Firstname+Surname)					-0.016***	
					(0.005)	
Black (BIFSG)						-0.017***
						(0.006)
Observations	49,375	49,375	49,375	49,375	43,943	43,923
Application Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Lender FE	Yes	Yes	Yes	Yes	Yes	Yes
Log Amt Sought	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.074	0.074	0.074	0.074	0.074	0.074
Y-mean	0.08	0.08	0.08	0.08	0.08	0.08