NBER WORKING PAPER SERIES

REGULATORY ARBITRAGE OR RANDOM ERRORS? IMPLICATIONS OF RACE PREDICTION ALGORITHMS IN FAIR LENDING ANALYSIS

Daniel Greenwald
Sabrina T. Howell
Cangyuan Li
Emmanuel Yimfor

Working Paper 31646
http://www.nber.org/papers/w31646

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2023

Howell and Yimfor's work on this project is funded by the Alfred P. Sloan Foundation (G-2022-19451). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Regulatory Arbitrage or Random Errors? Implications of Race Prediction Algorithms in Fair Lending Analysis
Daniel Greenwald, Sabrina T. Howell, Cangyuan Li, and Emmanuel Yimfor
NBER Working Paper No. 31646
August 2023
JEL No. C81,G21,G23,G28,J15

ABSTRACT

When race is not directly observed, regulators and analysts commonly predict it using algorithms based on last name and address. In small business lending—where regulators assess fair lending law compliance using the Bayesian Improved Surname Geocoding (BISG) algorithm—we document large prediction errors among Black Americans. The errors bias measured racial disparities in loan approval rates downward by 43%, with greater bias for traditional vs. fintech lenders. Regulation using self-identified race would increase lending to Black borrowers, but also shift lending toward affluent areas because errors correlate with socioeconomics. Overall, using race proxies in policymaking and research presents challenges.

Daniel Greenwald
NYU Stern School of Business
KMC 9-93
United States
New York, New 10012
dlg340@stern.nyu.edu

Sabrina T. Howell
NYU Stern School of Business
KMC 9-93
44 West 4th Street
New York, NY 10012
and NBER
showell@stern.nyu.edu

Cangyuan Li
NYU Stern School of Business
cl4220@stern.nyu.edu

Emmanuel Yimfor
Columbia Business School
Department of Finance
Columbia University
665 W 130th St.
New York, NY 10027
eay2121@columbia.edu

There are many high-stakes contexts in which regulators, firms, administrators, and researchers do not directly observe race and instead rely on proxies.[1] One important setting is lending, where regulators use race prediction algorithms to assess compliance with fair lending laws for auto, personal, and small business loans, among other products.[2] How accurate are these proxies? How might their prediction errors influence the distribution of lending? And which borrowers and lenders stand to benefit if the regulatory regime switches from proxies to more direct data on race?

We study these questions in the setting of small business lending, where regulators are especially attentive to compliance with fair lending laws because information asymmetry and financial constraints make small business owners more vulnerable to discrimination.[3] Small business lenders do not collect and report applicant race, unlike home mortgage lenders. This setting thus allows us to study an environment in which the use of proxies is central to measuring regulatory compliance and enables us to contribute to an active policy debate about whether or not to require small business lenders to collect self-identified race data.[4] We focus our analysis on Black Americans because they have historically faced discrimination in credit markets—a key motivation for fair lending laws—and because name-based race prediction algorithms may be particularly problematic for them.[5]

In this context, we study the frequency of prediction errors, their impact on underwriting policy both in aggregate and across lenders, and how these impacts vary with socioeconomic characteristics of borrowers. To obtain prediction errors when self-identified race is not available, we construct a novel measure of race using images ("image-based race"), which best aligns with visual perception of race. Image-based race is relevant to many contexts where discrimination on the basis of visual perception is a concern and could be applied in many additional settings.

[1] For example, in the absence of data on race for healthcare patients, healthcare plans and administrators use racial prediction algorithms to assess differences across groups in disease incidence and in the quality of care (Fremont et al., 2016). Economists studying issues of race also often use prediction algorithms. Examples include Pool et al. (2015); Dimmock et al. (2018); Ambrose et al. (2021); Egan et al. (2022); Frame et al. (2022); and Howell et al. (2022).

[2] See, for example, prohibitions on discrimination in the Equal Credit Opportunity Act (ECOA) (CFPB Methodology). Also, this list of products excludes mortgages, for which self-identified race is collected under the Home Mortgage Disclosure Act.

[3] The U.S. Interagency Fair Lending Examination Procedures, which apply to five federal agencies, note that when it comes to commercial loans, "Although ECOA prohibits discrimination in all commercial credit activities of a covered institution, the agencies recognize that small businesses (sole proprietorships, partnerships, and small, closely-held corporations) may have less experience in borrowing. Small businesses may have fewer borrowing options, which may make them more vulnerable to discrimination. Therefore, in implementing these procedures, examinations should generally be focused on small business credit."

[4] See here. More than 10 years ago, the Dodd-Frank Act directed the CFPB to adopt regulations on this matter (referred to as "1071" due to the section of the act), but due to stiff opposition from the banking community (see here and here), they did not do so. In late March 2023, a final rule was issued requiring data collection (see here), but banks and members of Congress are against it (see here).

[5] Largely due the legacy of slavery, there is large overlap between surnames of Black and White Americans, reducing the statistical informativeness of surnames. Name-based proxies typically perform better when identifying individuals with Hispanic or Asian backgrounds.

To build intuition and predictions for our empirical analysis, we provide a simple model of lending under different regulatory environments. We assume that the average benefit lenders receive from serving one racial group ("Group B") is lower than the other ("Group A"), leading to lower approval rates for Group B in the absence of regulation.[6] In our model, regulators aim to limit the disparity in approval rates between the two groups. However, the regulator cannot observe actual race and instead uses a noisy algorithm that predicts race. In order to maintain compliance by reducing the measured gap between groups, lenders tilt their approval policy for both racial groups to increase lending to borrowers with a high *algorithm-predicted* probability of being in Group B. However, among borrowers with a given algorithm-predicted probability of being in Group B, the gap in approval rates between *actual* members of Groups A and B remains just as large as without regulation. As a result, algorithm-based regulation is only partially effective at closing approval gaps across groups, which remain higher than algorithm-based gaps perceived by the regulator. Moreover, algorithm-based regulation distorts lending *within* each racial group across borrowers with different algorithm-based scores, which may change the distribution of lending with respect to socioeconomic covariates.

To empirically examine prediction errors and test these hypotheses, we employ two sources of data on small business lending. The first is a dataset of loan applications and funded loans between 2017 and 2019 from Lendio, an online loan marketplace for small businesses; these data enable us to observe lender approval decisions in a real-world context. The second data source is the Paycheck Protection Program, which provided government-guaranteed, forgivable loans to small businesses during the COVID-19 pandemic. Notably, while PPP data condition on receiving a loan and represent an unusual underwriting environment due to government guarantees, they include *self-identified* measures of race in a real-world, nonmortgage lending context, allowing us to measure various predicted race errors against this "gold standard" benchmark. While neither of these samples is perfectly representative of U.S. small businesses or their lenders, they provide real-world laboratories to study prediction errors from different measures of race. We augment these sources with information on the socioeconomic characteristics of each applicant's ZIP code, as well as education data sourced from their public LinkedIn profiles.

For our main analysis, we construct two measures of race. The first is the standard algorithmic prediction based on name and location (Bayesian Improved Surname Geocoding, or BISG), which is widely used by regulators, researchers, and practitioners. Our second measure is image-based race, which approximates how an individual is usually perceived in the U.S. and is closer to self-identified race than BISG (an assumption we test below). To obtain an image of firm owners' faces, we match them to LinkedIn profiles, requiring that the firm in our loan data appears on the LinkedIn

[6]We remain agnostic on the origin of these preferences, which could be the result of taste-based discrimination, statistical discrimination, or unconscious bias among loan officers.

profile. Next, we use an image classifier to obtain facial embeddings (distinctive facial features) for each image. Using a separate dataset of firm founder images for which we have race, we train a random forest model to predict race (with 91% accuracy) using facial embeddings and apply this model to our main datasets to classify applicants as Black or non-Black. Finally, we conduct clerical reviews of the output to mitigate the prediction errors from the model. After the filters and matching, we observe image- and BISG-based race for about 12,000 unique applicants in the Lendio data and 28,000 unique borrowers in the PPP data.[7] Because we obtained images by matching names to a public source (LinkedIn), our procedure requires only basic information on the applicant and firm as inputs, and it could be broadly applied in other contexts.

For the purposes of our analysis, we denote an applicant who is Black and correctly identified as such by BISG as a *true positive* and denote an applicant who is non-Black and correctly identified as such by BISG as a *true negative*. Conversely, we denote an applicant who is Black but inaccurately classified by BISG as non-Black as a *false negative* and an applicant who is non-Black but inaccurately classified by BISG as Black as a *false positive*. We use a shorthand of calling either image-based or self-identified Black race the "true" race to compare with BISG-based race, but we view these measures as capturing different dimensions of race in the U.S. and do not consider any measure to be the "truth" in a fundamental sense.

We first evaluate image- and BISG-based race measures against self-identified race in the PPP data. We find that BISG has a high error rate when classifying Black borrowers, with more than twice as many errors (false positives or false negatives) as true positives. The image-based measure performs better, with a correlation of 0.87 between image-based and self-identified indicators for being Black, compared to a correlation of just 0.52 for BISG-based and self-identified indicators for being Black. We note that the imperfect correlation between image-based and self-identified race does not necessarily reflect errors; image-based race may be more relevant for discrimination when it differs from self-identified race because it more accurately reflects how an individual is perceived by others, which may differ from how they perceive themselves.

To understand which borrowers stand to benefit from regulators using a direct measure of race rather than a proxy, we explore how these prediction errors vary with socioeconomic characteristics. We find that the geographies where BISG tends to make false positive errors (predicting people to be Black when they are not) are also areas with particularly strong historical systematic disadvantage for Black borrowers, including lower per capita income, more racial animus, higher Black population shares, and less geographic segregation. Turning to individual characteristics, we find that more educated individuals are less likely to generate false positive errors. These patterns are reversed for false negatives. The results are robust to using self-identified

[7]The 12,000 applicants correspond to about 50,000 applications since Lendio sends applications to multiple lenders, and firms also sometimes apply on multiple occasions.

race as "true" race (only available in the PPP sample), as well as across the PPP and Lendio samples using image-based race as "true" race.

Having established the presence of large prediction errors using BISG, we next analyze the impact of these errors on measured racial disparities in approvals in our Lendio data. Our analysis is motivated by the process for conducting fair lending evaluations across a range of federal U.S. regulators, which begins by assessing whether the lender approves a similar share of applicants in protected groups as in the majority group.[8] Our theory predicts that when regulators evaluate compliance using BISG, lenders are incentivized to prioritize non-Black borrowers with high BISG scores (false positives) over Black borrowers with low BISG scores (false negatives), which biases the regulator-measured approval gap down relative to the true one.

Testing this hypothesis in the data, we find that the gap in approval rates between non-Black and Black applicants is 1.3pp when classifying applicants using BISG, but is nearly twice as large (2.3pp) when classifying applicants using our image-based measure. This implies that a regulator who can only observe the BISG-based measure would substantially underestimate racial disparities. In regressions, we find that this difference in the predictive power of BISG-based race and image-based race on approvals is robust to various specifications and controls. Our empirical analysis is thus consistent with lenders responding to regulatory incentives as predicted by theory, creating an illusion of better compliance with fair lending laws, and reducing the ability of regulation to narrow true approval gaps between racial groups.

The above results highlight average approval gaps in our overall sample. Since regulators focus on lender-level evidence of discrimination, we construct a lender-level measure for the difference in approval rates using image-based race vs. BISG-based race, which we call $\Delta_{\text{Share Black Appr}}$. When this difference is positive, the lender is serving the actual (image-based) Black population at a higher rate than they appear to be with BISG; that is, the lender is serving more Black borrowers classified as non-Black by BISG (false negatives). Since there is a positive correlation between false negatives and high socioeconomic status (e.g., for a Black borrower with a racially ambiguous last name living in a "White" neighborhood), a lender that serves more advantaged Black borrowers will have a higher difference, potentially representing a response to "cream-skimming" incentives. In contrast, when $\Delta_{\text{Share Black Appr}}$ is more negative, the lender appears more consistent with fair lending laws than they actually are, potentially representing a response to compliance incentives. We find substantial variation across lenders, suggesting that errors in predicting the race of individual borrowers translate into large errors in evaluating compliance at the lender level.

In both Lendio and PPP data (where we use loan shares to Black borrowers rather than approval

[8]The U.S. Interagency Fair Lending Examination Procedures, which apply to five federal agencies including the Federal Reserve Board and the Federal Deposit Insurance Corporation, detail how approval rates should be used by fair lending examiners. Note that in the absence of information about risk, the expectation is not zero difference. However, a wider difference merits a closer investigation by the regulator.

shares), we find that $\Delta_{\text{Share Black Appr}}$ is more negative for banks and other conventional lenders that typically rely on soft information for underwriting (Petersen and Rajan, 1994; Berger and Black, 2011), while it is more positive for fintechs, which are more automated and arms-length (Balyuk et al., 2020; Howell et al., 2022). There are many possible reasons for this difference, but one interpretation is that fintechs—which tend to be more lightly regulated—have weaker incentives to improve perceived compliance using BISG-based measures. These results imply that moving from measures based on BISG to measures based on self-identified race for regulation would have impacts that vary widely across both lenders and lender types, with conventional lenders being particularly affected.

Last, we estimate a counterfactual in which regulators move from evaluating compliance with fair lending laws using BISG-based race to using actual race (proxied here by image-based race). Our analysis reveals that this policy change would increase the share of loans to borrowers who are actually Black, reducing discrimination based on skin color. However, this shift would also reallocate loans toward areas with higher incomes, fewer Black households, and higher levels of education. Thus, it could inadvertently increase within-race or geographic inequality in lending, even though it reduces between-race inequality by increasing lending to actual Black borrowers.

Related Literature. This paper contributes to several strands of literature. First, we build on research about racial disparities in access to financial services, which has mostly focused on residential mortgages and consumer credit markets (Tootell, 1996; Bayer et al., 2018; Begley and Purnanandam, 2021; Bhutta and Hizmo, 2021; Blattner and Nelson, 2021; Dobbie et al., 2021; Giacoletti et al., 2021). Also related is work on the role of different lenders, especially the role of emerging fintech firms and traditional banks, in serving minority groups and underserved populations (Buchak et al., 2018; Tang, 2019; Fuster et al., 2019; Berg et al., 2020; D'Acunto et al., 2020; Erel and Liebersohn, 2020; Bartlett et al., 2022). Other work studies how removing names from applications for employment or loans affects outcomes (Bartik and Nelson, 2021; Kabir and Ruan, 2023). More broadly, we join the literature on bias against Black Americans across a wide range of settings, including Knowles et al. (2001), Anwar and Fang (2006), Charles and Guryan (2008), Price and Wolfers (2010), and Arnold et al. (2018). To our knowledge, this paper is the first to examine how disparities in serving different groups—in our case, disparities across lenders—depends on the way race is measured.

We join a small literature on racial disparities in entrepreneurship and small business finance specifically. Blanchflower et al. (2003) find racial differences in access to small business credit, while Robb and Robinson (2018) do not find such differences. Other work on the role of race in small business lending includes Fairlie and Robb (2007), Asiedu et al. (2012), Bellucci et al. (2013), and Fairlie et al. (2022). A recent literature has focused specifically on racial disparities in the PPP (Erel and Liebersohn, 2020; Chernenko and Scharfstein, 2021; Fairlie and Fossen, 2021; Howell et al., 2022). To our knowledge, this project is the first effort to focus explicitly on compliance with fair lending laws in small business finance.

Finally, we contribute to work on the methodologies used in identifying race. We create a measure of race based on images and compare its effectiveness to other established approaches, using self-identified race as a benchmark. This is relevant to research and policy that require measures of race, especially contexts where self-identified data are unavailable (Pool et al., 2015; Dimmock et al., 2018; Ambrose et al., 2021; Jiang et al., 2021; Egan et al., 2022; Frame et al., 2022). We also join a new literature with image-based analysis; for example, Athey et al. (2022) show that microloan applicants who smile in their online profile photograph are more successful in obtaining credit. Our results provide guidance on best practices for researchers and regulators, such as the need to address bias arising from the correlation between socioeconomic characteristics and errors in proxies for race.

Overview. The rest of the paper is organized as follows. Section 1 presents the algorithmic, image-based, and self-reported measures of race we will use in our analysis. Section 2 constructs a simple model to show how the measure used by regulators influences lending. Section 3 documents our data sources. Section 4 compares our measures of race to construct classification errors. Section 5 shows that these errors are not random but vary with socioeconomic characteristics. Section 6 shows how these errors influence measured approval rates by race. Section 7 studies the link between prediction errors and approval by lender. Section 8 presents our empirical counterfactual exercise from a switch to self-reported race regulation. Section 9 concludes.

1 Measures of Race

This section describes the measures of race we use, highlighting their strengths and limitations and detailing our novel methodology for inferring race based on an individual's image. Importantly, we do not believe there is a single absolute truth when it comes to measuring race, and thus the context of how race is used in a particular decision-making or research process should inform which measure is best suited to the application.

Self-Identified Race. Self-identified race refers to the race that an individual reports for themselves. Self-identified race and ethnicity for loan applicants are typically collected for home mortgages and used for assessing compliance with fair lending laws in mortgage underwriting. While self-identified race can be seen as the "gold standard" in terms of minimizing measurement error, we note that an individual's self-identified race may differ from how others perceive them—a crucial distinction as many economic questions revolve around whether agents are treated differently because they are *perceived* to be of a particular race.

Bayesian Improved Surname Geocoding (BISG). BISG is the standard method used by regulators to predict race in the absence of data on self-identified race. BISG combines two measures of race based on geography and surnames. The geography-based measure assigns the probability of an individual's race based on the racial composition of the specific geography. The geography is most often a ZIP code, but can be a census block, census tract, county, or state. The surname-based measure uses the frequency distribution of names within a population to predict race. However, this method poses a practical challenge as many Black Americans have racially ambiguous surnames. For example, while the most common last names among Black people—Williams, Johnson, Smith, Jones, Brown, Jackson, Davis, Thomas, Harris, and Robinson—account for approximately 12% of all Black Americans, people identified as Black in the U.S. census comprise a minority of individuals with these names, except in one case (53% of people with the name Jackson).[9]

To generate BISG probabilities, we first identify the owner's name and their location (discussed in Section 3), and then employ the Surgeo library in Python (see Section B.1 for details). The algorithm's output are probabilities that an individual is Hispanic, White, Black, Asian, Pacific Islander/Alaska Native, or Multiracial. We transform this output by summing the probabilities that the applicant is non-Black and retaining two columns, one for the probability that the applicant is Black and the other for the probability that the applicant is non-Black. We also use the probabilities to randomly assign an applicant as Black or non-Black (BISG_{RAND}).[10] For instance, if an applicant has a 20 percent probability of being Black, then there is a 20 percent chance that we will assign them the Black label. For analyses with individual-level data, we use BISG_{RAND}. For lender-level analysis, we use the raw probability of being Black as determined by the BISG algorithm.

Economists studying race sometimes use Bayesian Improved Firstname and Surname Geocoding (BIFSG), which also employs the first name (e.g., Pool et al., 2015; Dimmock et al., 2018; Ambrose et al., 2021; Egan et al., 2022; Frame et al., 2022). While the BISG method is defined when an individual has a surname that is shared by more than 100 or more people (the 2010 U.S. Census surname database contains race and ethnicity percentages of 151,671 unique surnames covering 89.9% of U.S. population), the BIFSG method is only defined for people with common first names. The most common source has just 4,250 first names.[11] In fact, of the most common distinctive Black names in Fryer and Levitt (2004)—Deshawn, Tyrone, Reginald, Shanice, Precious, Kiara, and Deja—only two out of seven are among the candidate names, meaning that the others have no BIFSG prediction. Our main analysis employs BISG to avoid the restriction to common first names.

[9] See the 2010 Decennial Census.
[10] We set a seed for this random assignment to ensure replicability.
[11] See here.

Image-Based Race. Image-based race is inferred from an individual's appearance. We obtain images by matching individuals in our data to LinkedIn and then downloading their complete LinkedIn profiles.[12] To mitigate errors from associating an image with the wrong applicant, we retain only those observations where the company name listed on the applicant's LinkedIn profile matches the borrowing company's name on the application, and the employment start date precedes the application date.

We use a pretrained image classifier to obtain facial embeddings, which are loadings of an image on different facial attributes. Specifically, we use the VGG-Face classifier, which is wrapped in the *DeepFace* Python package developed by Serengil and Ozpinar (2020). With these embeddings, we train a random forest model on a dataset consisting of around 170,000 images of founders of venture-backed startups. The model achieved 91% accuracy in a hold-out sample. We then apply the model to the facial embeddings in our sample to obtain a preliminary classification. It is worth noting that automated face recognition is not infallible and can result in false positives and false negatives, particularly for Black applicants photographed in very bright lighting or White applicants photographed in settings with very little lighting.[13] To address such potential errors, we conducted manual reviews of all images using the applicant's LinkedIn profile information when the image alone proved insufficient for classification. We classify each applicant as either Black or non-Black.

2 Theory

This section offers a simple model to fix ideas about how different measures of race could systematically bias disparate impact assessments in lending, which in turn could create incentives to distort lending in order to game the compliance system. We first present the structure and equilibrium of the model under various regulatory assumptions, then provide a simple numerical example to display its properties.

Lending Technology. Consider a lender who lends to two groups, A and B. The value of lending to an individual i of type $j \in \{A, B\}$ is assumed to be

$$v_{i,j} = \mu_j - \varepsilon_i. \tag{1}$$

The term μ_j represents a group-specific benefit to the lender of making a loan to an individual of Group j, which could represent actual average differences in profitability, confounds between race and other variables that influence the value of lending, or a subjective preference for lending to individuals of a certain race. We order A and B so that $\mu_A > \mu_B$, implying that in the absence of

[12] When an image is not available on LinkedIn, we also draw from other websites such as Facebook and Twitter.

[13] At this stage, we also screen out typically South Asian names using a name classifier, as South Asians are often mistakenly classified by the facial recognition software as Black.

regulation, lenders would provide fewer loans to members of Group B. The term ε_i is an idiosyncratic type, which creates variation in the cost of lending to different borrowers within a group. For simplicity, and to motivate the linear probability regressions we will run later, we assume the uniform distribution $\varepsilon_{i,j} \sim U[\varepsilon^{min}, \varepsilon^{max}]$, implying that the CDF of $\varepsilon_{i,j}$ is equal to

$$F_\varepsilon(\varepsilon_{i,j}) = \underbrace{\left(-\frac{\varepsilon^{min}}{\varepsilon^{max} - \varepsilon^{min}}\right)}_{\gamma_0} + \underbrace{\left(\frac{1}{\varepsilon^{max} - \varepsilon^{min}}\right)}_{\gamma_1} \varepsilon_{i,j}, \tag{2}$$

where for ease of notation we define γ_0 and γ_1 so that $F_\varepsilon(\varepsilon_{i,j}) = \gamma_0 + \gamma_1 \varepsilon_{i,j}$.

No Regulation Equilibrium. It is clear from equation (1) that the optimal policy for the lender is to approve loans to all borrowers of Group j with $\varepsilon < \bar{\varepsilon}_j$, for some threshold value $\bar{\varepsilon}_j$. In the absence of a regulatory constraint, the lender therefore chooses $\{\bar{\varepsilon}_A, \bar{\varepsilon}_B\}$ to maximize

$$V = \sum_{j \in \{A,B\}} s_j \int^{\bar{\varepsilon}_j} (\mu_j - \varepsilon)\, dF_\varepsilon(\varepsilon), \tag{3}$$

where s_j is the share of the population in Group j. The first order condition for this problem is

$$\bar{\varepsilon}_j^{NR} = \mu_j, \tag{4}$$

where the superscript NR stands for "no regulation." Assuming that $\mu_A > \mu_B$, this implies that $\bar{\varepsilon}_A > \bar{\varepsilon}_B$ and therefore that the probability of being approved, equal to $F_\varepsilon(\bar{\varepsilon}_j)$, is higher for Group A than Group B. Formally, if we define π_i^{NR} to be the probability of approval for individual i under no regulation, we obtain

$$\pi_i^{NR} = \text{const} + \underbrace{\gamma_1(\mu_B - \mu_A)}_{<0} B_i, \tag{5}$$

where, in a slight abuse of notation, B_i is an indicator for being in Group B.

Regulatory Constraint on Actual Race. Depending on the drivers of these disparities in lending, a regulator may wish to require lenders to allocate credit more equally across groups. We first consider the case where the regulator can observe the actual group (race) to which each applicant belongs and sets a constraint that the share of loans approved for members of Group B can be lower than that of Group A by no more than some amount $\kappa \geq 0$:

$$F_\varepsilon(\bar{\varepsilon}_A) - F_\varepsilon(\bar{\varepsilon}_B) \leq \kappa. \tag{6}$$

The lender's problem now is to maximize (3) subject to (6). The optimal policy is to approve loans for borrowers of type j with $\varepsilon_{i,j} < \bar{\varepsilon}_j^{AR}$, where

$$\bar{\varepsilon}_A^{AR} = \mu_A - \frac{\lambda^{AR}}{s_A}, \tag{7}$$

$$\bar{\varepsilon}_B^{AR} = \mu_B + \frac{\lambda^{AR}}{s_B}. \tag{8}$$

Where λ^{AR} is the multiplier on constraint (6) and the superscript AR stands for "actual race." These expressions nest (3) if $\lambda^{AR} = 0$, which occurs when the regulatory constraint is slack. When the constraint binds ($\lambda^{AR} > 0$), the effect is to lower the cutoff for Group A (reducing approvals) and to raise the cutoff for Group B (increasing approvals). The approval rate for individual i is

$$\pi_i^{AR} = \text{const} + \gamma_1 \left[\underbrace{(\mu_B - \mu_A)}_{<0} + \underbrace{\lambda^{AR}(s_A^{-1} + s_B^{-1})}_{>0} \right] B_i. \tag{9}$$

Relative to (5), the regulation increases the probability of approval for applicants in Group B, and decreases the probability of approval for applicants in Group A.

Regulatory Constraint on Predicted Race. We now consider the most empirically realistic case, in which regulators cannot observe actual race, but instead use a predictive algorithm (BISG). In this case, we assume that the lender can either still observe actual race, or some covariates that influence μ_j and are correlated with race but not captured by BISG, such as income or education.

For this scenario, we guess and verify that the optimal policy is to approve all borrowers with $\varepsilon < \bar{\varepsilon}_j(q)$ for some thresholds $\bar{\varepsilon}_j(q)$ that depend on Group j and each borrower's BISG-predicted probability of being in Group B, denoted q. Since regulators use BISG to predict the amount of approved loans L_j to Group j, they estimate

$$\hat{L}_A = \sum_{j \in \{A,B\}} s_j \int (1-q) F_\varepsilon(\bar{\varepsilon}_j(q))\, dF_{q,j}(q), \tag{10}$$

$$\hat{L}_B = \sum_{j \in \{A,B\}} s_j \int q F_\varepsilon(\bar{\varepsilon}_j(q))\, dF_{q,j}(q), \tag{11}$$

where $F_{q,j}$ is the CDF of the distribution of q for Group j, and the hats indicate that these are the BISG-predicted, rather than actual, values.

The lender now requires that the BISG-predicted share of loans approved for members of Group B can be lower than that of Group A by no more than some amount $\kappa \geq 0$. Since the estimated approval rate for Group j using BISG probabilities is $\hat{L}_j/s_j$, this constraint is

$$\frac{\hat{L}_A}{s_A} - \frac{\hat{L}_B}{s_B} \leq \kappa, \tag{12}$$

which we show in Appendix C can be written as

$$\sum_{j\in\{A,B\}} s_j \int \left[\frac{1-q}{s_A} - \frac{q}{s_B}\right] F_\varepsilon(\bar{\varepsilon}_j(q))\, dF_{q,j}(q) \leq \kappa. \tag{13}$$

The lender now maximizes

$$V = \sum_{j\in\{A,B\}} s_j \int \int^{\bar{\varepsilon}_j(q)} (\mu_j - \varepsilon)\ dF_\varepsilon(\varepsilon)\, dF_{q,j} \tag{14}$$

subject to (13). The optimal policy is to approve loans for borrowers of Group j and BISG probability q if $\varepsilon_{i,j} < \bar{\varepsilon}_j^{PR}(q)$ for

$$\bar{\varepsilon}_j^{PR}(q) = \mu_j + \lambda^{PR}\left[\frac{q}{s_B} - \frac{1-q}{s_A}\right], \tag{15}$$

where the superscript PR stands for "predicted race." The intuition for (15) is that the regulator predicts that this applicant is in Group B with probability q, which loosens the constraint, but is in Group A with probability $1-q$, which tightens the constraint. In the case that the BISG algorithm is perfect, and q is always zero or one, then (15) becomes (7) and (8) for members of Groups A and B, respectively, nesting our earlier results.

Computing the probability of approval, we obtain

$$\pi_i^{PR} = \text{const} + \gamma_1(\mu_B - \mu_A)B_i + \gamma_1\lambda^{PR}\left(s_B^{-1} + s_A^{-1}\right) q_i. \tag{16}$$

Importantly, the loading on B_i is exactly the same as in the no regulation equilibrium (equation (5)), meaning that the gap between groups A and B remains unchanged conditional on q. Intuitively, this holds because lenders only receive credit for fair lending based on q, rather than the actual group j.

The effect on approvals of moving from regulation based on predicted (BISG) race to regulation based on actual (self-identified) race is the difference between (9) and (16), which is equal to

$$\pi_i^{AR} - \pi_i^{PR} = \text{const} + \gamma_1\lambda^{AR}(s_A^{-1} + s_B^{-1})B_i - \gamma_1\lambda^{PR}\left(s_B^{-1} + s_A^{-1}\right) q_i. \tag{17}$$

Equation (17) shows that such a change in policy would reduce the coefficient on q (the BISG probability of Group B) in an approval rate regression, while increasing the coefficient on the indicator for actually being in Group B. We will use this equation to motivate our empirical design in Section 8.

Approval Rate Decomposition. A primary focus of regulators is on gaps in approval rates between racial groups. To provide intuition on the statistical sources of bias in these measures, we decompose the difference in the actual average approval rate for Black applicants (denoted $\bar{\pi}_B$) and the average predicted approval rate for Black applicants using BISG (denoted $\bar{\pi}_B^{BISG}$). These can be written as

$$\bar{\pi}_B = \frac{P(Approved, B)}{P(B)} = \frac{E[\mathbf{1}_i \times B_i]}{E[B_i]}, \tag{18}$$

$$\bar{\pi}_B^{BISG} = \frac{P^{PR}(Approved, B)}{P^{PR}(B)} = \frac{E[\mathbf{1}_i \times q_i]}{E[q_i]}, \tag{19}$$

where $\mathbf{1}_i$ is an indicator for being approved, q_i is the BISG probability of an applicant being Black, B_i is an indicator for the applicant's actual race being Black, and P^{PR} is the perceived probability using BISG.

If we assume that the BISG algorithm is unbiased on average (i.e., it uses the correct unconditional probability of an applicant being Black), we have $E[q_i] = E[B_i] = s_B$. Combining this assumption with equations (18) and (19) we obtain

$$\begin{aligned}
\bar{\pi}_B - \bar{\pi}_B^{BISG} &= s_B^{-1} E\Big[\mathbf{1}_i(B_i - q_i)\Big] \\
&= s_B^{-1}\Big\{\text{Cov}\left(\mathbf{1}_i, (B_i - q_i)\right) + E\left[\mathbf{1}_i\right] E\left[B_i - q_i\right]\Big\} \\
&= s_B^{-1}\text{Cov}\Big(\mathbf{1}_i, \underbrace{(B_i - q_i)}_{\text{BISG error}}\Big),
\end{aligned}$$

where the third step follows from the assumption that BISG is unbiased ($E[B_i - q_i] = 0$). This expression implies that when the BISG algorithm is unbiased on average, the difference between the perceived and actual approval rates depends on the covariance of approval with $B_i - q_i$, which represents the error between the true indicator for being Black and the BISG probability of being Black. If we define, for example, $(B_i - q_i)_+ \equiv \max(B_i - q_i, 0)$, then we can apply the identity

$$B_i - q_i = \underbrace{(B_i - q_i)_+}_{\text{false negative error}} - \underbrace{(q_i - B_i)_+}_{\text{false positive error}},$$

which cleanly separates the error into false negative and false positive components.[14] Substituting

[14] To see this, note that for a "pure" false negative ($B_i = 1, q_i = 0$) we would have $B_i - q_i = 1$ and so $(B_i - q_i)_+ = 1, (q_i - B_i)_+ = 0$. Conversely, for a "pure" false positive ($B_i = 0$, $q_i = 1$) we would have $B_i - q_i = -1$ and so $(B_i - q_i)_+ = 0, (q_i - B_i)_+ = 1$.

back into our earlier expression, we obtain

$$\bar{\pi}_B - \bar{\pi}_B^{BISG} = s_B^{-1} \left\{ \mathrm{Cov}\Big(\mathbf{1}_i, \underbrace{(B_i - q_i)_+}_{\text{false neg.}}\Big) - \mathrm{Cov}\Big(\mathbf{1}_i, \underbrace{(q_i - B_i)_+}_{\text{false pos.}}\Big) \right\}. \tag{20}$$

Summing up, equation (20) shows that the actual approval rate will be lower than the perceived approval rate when approval covaries negatively with the probability of being a false negative and when approval covaries positively with the probability of being a false positive.

Numerical Example. To close the theory section, we provide a numerical example to illustrate the mechanisms at work. We first parameterize the model. We map Group A to non-Black borrowers and Group B to Black borrowers. Accordingly, we set the population shares of Group A and Group B to 83.96% and 16.04% respectively, matching the image-based shares of loan applications by non-Black and Black borrowers in our Lendio data (which are introduced in detail below in Section 3). For the distributions $F_{q,j}$, we choose beta distributions, which are appropriate for distributions of probabilities q as they are bounded between 0 and 1. Beta distributions are parameterized by two shape parameters: α and β. For Groups A (non-Black) and B (Black), we choose these parameters to match the mean and variance of the BISG probabilities for these groups, where we assign applicants to groups based on our image-based measure of race. The resulting beta distributions fit the empirical distributions well, as shown in Figure 1.

For simplicity, we set $\gamma_0 = 0$ and $\gamma_1 = 1$, so that $\varepsilon \sim U[0, 1]$. To calibrate the μ parameters, we set up a predicted race equilibrium that corresponds as closely as possible to the data. Since the approval rate gap constraint in the model is binding in the predicted race equilibrium, we set $\kappa = 1.32\%$ so that we exactly reproduce the approval rate gap in the data between non-Black and Black applicants as predicted by BISG (8.33% for BISG non-Black applicants, 7.01% for BISG Black applicants). We then set $\mu_A = 9.02\%$ and $\mu_B = 4.58\%$ so that the approval rates by actual race in this predicted race equilibrium exactly match the approval rates by image-based race in the data (8.53% for image non-Black applicants, 6.20% for image Black applicants).[15]

With the model calibrated, from this point on we consider a policy experiment that sets $\kappa = 0$ to simulate hypothetical regulation that aims to equalize the actual or predicted approval rates of the two groups. The approval rates implied by the model can be seen in Figure 2. The dashed lines show the approval rates under the no regulation scenario (π^{NR}). These lines are perfectly flat due to our assumption (made to better illustrate the mechanism) that the fundamental value of lending to a borrower does not depend directly on q conditional on a borrower's actual Group j. In this case, the approval rate for Group B is 4.58%, substantially lower than the approval rate of 9.02% for Group

[15] These approval shares differ from μ_A and μ_B because of the lender's incentives to reduce the regulator's perceived approval gap according to BISG.

A. However, a regulator using BISG would predict that the approval rates are 5.32% for Group B vs. 8.88% for Group A, meaning that the actual gap is 24.8% larger than the BISG-implied one. This downward bias when measuring the approval gap using BISG occurs even though q is close to unbiased and approval policies do not depend on BISG conditional on actual race. Instead, the BISG-based classification overpredicts the approval rate for Group B because the false positive borrowers mistakenly added to Group B have a much higher approval rate than the false negative borrowers mistakenly removed from Group B, even though these groups are similar in size.

From this baseline, we can turn to the solid lines, which show the equilibrium under regulation using predicted race (π^{PR}). In this case, the approval rates for both groups now show a strong upward tilt with respect to q. Under the new policy, lending to high-q borrowers (of either group) helps the lender meet its regulatory constraint, increasing the value of loans to these borrowers and generating this slope. At the same time, the gap between approval rates for the two groups conditional on q is unchanged from the no regulation baseline.

Under the predicted race policy, the regulator (using BISG to predict race) believes that the gap in approval rates is completely eliminated, with both groups being approved at rate 8.07%. However, the approval gap using actual race is not eliminated, with approval rates for Group A and B of 8.24% and 7.15%, respectively. Thus, the Predicted Race policy is only partially effective, leaving 24.6% of the initial 4.44pp gap from the no regulation equilibrium unclosed. To the extent that the predicted race regulation manages to be effective, it does so despite the difference between the two groups' approval rates (the vertical distance between the two solid lines in Figure 2) remaining exactly the same as in the no regulation equilibrium conditional on q (equal to $\mu_A - \mu_B$ at all points). Instead, this occurs because the *densities* of q are not the same across groups, with the q distribution for Group B having much more mass to the right (where approval rates are elevated compared to the no regulation baseline) than to the left (where approval rates are depressed compared to the no regulation baseline), while the Group A distribution has much more mass in the low-q region to the left. As a result, Group B sees a greater increase in approvals despite no change in the gap between π_A^{PR} and π_B^{PR} conditional on q.

Third, we consider a regulatory regime in which the lender requires the true approval rates (based on actual race) to be the same across groups. The approval rates under this equilibrium (π^{AR}) are identical across groups and are plotted as a single gray dotted line in Figure 2. As discussed above, this change of regime increases the loading of approval rate on an indicator for being in Group B, shown by the large average shift upward of the orange line and average shift downward of the blue line. However, this change also reduces the loading of approval rates on q, which removes the positive slope in the predicted race equilibrium and returns the slope of the approval rate with respect to q to zero.

We summarize the differences between these regulatory regimes in Figure 3, which shows the approval rates by actual and BISG-predicted group. The figure shows that the predicted race policy has outcomes very similar to the actual race policy for both true positive and true negative applicants—borrowers in Groups B and A, respectively, who are correctly classified as such by BISG. In particular, approval rates in the predicted race equilibrium are slightly higher than in the actual race equilibrium for true positive borrowers and slightly lower than in the actual race equilibrium for true negative borrowers. However, the two policies deviate widely for borrowers incorrectly classified by BISG. In particular, the predicted race equilibrium exhibits approval rates for false positive applicants well above those of the actual race equilibrium. This occurs because lenders receive regulatory credit for lending to these applicants in the predicted race equilibrium (since the regulator believes they are in Group B, even though they are actually in Group A) but do not in the actual race equilibrium where they are known by the regulator to be in Group A. Similarly, lending to false negative applicants (in Group B but predicted to be in Group A) is markedly lower in the predicted race equilibrium, since lenders do not receive regulatory credit for these approvals.

In summary, BISG-based regulation can be partially successful at increasing lending to Group B since BISG is a somewhat informative signal, so that Group B on average has higher values for q. However, because the predicted race regulation incentivizes lending more to false positive applicants and less to false negative applicants, it does not completely close the approval rate gap between groups. Thus, moving to an actual race regime based on self-identified race would reduce the true between-race approval gap. To the extent that actual race and BISG-predicted race (encoded in q above) covary differently with other socioeconomic variables (i.e., false negative and true positive borrowers differ systematically), moving to actual race regulation may also influence these socioeconomic measures among approved borrowers.

3 Data Sources

We use two primary sources of data on applicants and borrowers for small business loans. In this section, we describe them as well as the supplementary sources we draw from.

Lendio Loan Applications. We use basic data on loan applications and funded loans from Lendio, an online loan marketplace for small businesses. These data offer a rare chance to observe lenders' approval decisions in a real-world market context. Firms submit a single application, and Lendio forwards that application to one or more lenders on its platform. Those lenders decide whether to make offers to the borrower, who then decides whether to take up the loan. We employ Lendio data from 2017–2019. We identify the applicant as the primary contact on the loan application.[16]

[16] To avoid spurious results from small samples, we exclude lenders who received 10 or fewer applications. This allows us to construct more reliable lender-specific approval probabilities.

The raw Lendio dataset has 674,203 applications from 160,942 unique firms, with each application sent to about four lenders on average. Of the 160,942 firms, the BISG algorithm produces a race prediction for 139,759. We obtain image-based race for 11,190. This relatively small number reflects our strict approach to minimize false positives. We require that: i) the borrowing company's name corresponds with at least one experience entry on the applicant's LinkedIn profile; ii) the applicant must have commenced their tenure at the company prior to the application date; and iii) the applicant must not have terminated their tenure at the company before the application date. While the small sample in our context could pose a challenge for applications such as regulatory supervision, image-based race may be useful when images that are certain to correspond to the individual are more accessible for a representative population. Our approach allows us to study the limitations of using proxies for race as substitutes for self-identified race. Summary statistics about the data used in analysis are in Table 1. Focusing on the application level, the average (median) application seeks just over \$100,000 (\$50,000) in funding, but the approved amounts are much lower, at around \$52,000 (\$25,000). The average approval rate is 8.1%.

Unfortunately, we do not directly observe whether a loan was not funded because the lender formally rejected the application or made an offer that was not taken up. However, based on our understanding of Lendio's process, we are able to compute an implied measure of approval or rejection.[17] The key is that Lendio will typically only forward an application to an additional lender when it is rejected. Thus, if a particular loan application is not funded and we observe that Lendio subsequently forwards that application to further lenders, we can safely infer that the application was rejected in the first round. To build intuition, we present a practical example:

1. Lendio receives an application on June 1.
2. Lendio sends this application to two lenders on June 2, but neither provide funding.
3. On June 16, Lendio fowards the same application to two new lenders. Based on information from Lendio, we can infer that the lenders from June 2 rejected the application.
4. Suppose one lender from the June 16 group approves the loan. In our data, we identify this lender as making an approval decision.
5. However, we exclude the nonfunding lender from June 16 in our analysis. We cannot confirm whether the borrower was rejected or rejected the lender's offer.

Paycheck Protection Program Loans. The second main source of data, which allows us a rare chance to observe self-identified race in a nonmortgage lending context, is from the Paycheck Protection Program (PPP), which was established by the CARES Act in March 2020 to help small

[17]We thank the Lendio staff, including Katherine Chandler and Brock Blake, for their helpful insights.

businesses struggling during the COVID-19 pandemic. With more than $800 billion in loans, it is one of the largest single public finance programs in U.S. history. To facilitate the speedy disbursal of PPP funds, the federal government outsourced the origination of PPP loans to private lenders. PPP loans were federally guaranteed, uncollateralized, and forgiveable if used for eligible expenses (in particular, payroll).

We begin with public administrative data from the Small Business Administration (SBA) on 11.8 million PPP loans made between April 3, 2020, and May 31, 2021.[18] Unfortunately, no data on PPP applications are available. We first restrict the sample to 4,775,702 "first draw" loans made before February 24, 2021, when program rules were changed to explicitly prioritize lending to small firms and minority-owned businesses. Further, we only consider the 933,645 loans for which borrowers voluntarily reported their race. It is important to note the potential bias in self-reporting race. Our analysis does not aim to provide representative figures for the U.S. population or small business owners. Instead, we seek to show real-world comparisons between measures of race within these selected samples, highlighting how and in what ways they can differ.

Of 933,645 loans with self-identified owner race, 255,355 are from borrowers with identifiable personal names. For the others, we use the first executive officer listed in the state business registration, as supplied by analytics firm Middesk.[19] As with our Lendio dataset, we then implement the BISG and image-based race classifications. The supplemented PPP data results in 867,151 borrowers with "valid" person names. Out of these, the BISG algorithm produces a racial classification for 702,080. From this subset, we manage to assign image-based race for 33,661. Again, the match rate reflects the stringent matching criteria we described above. We further filter the data to include only lenders with more than 10 loans, resulting in a sample of 22,618 borrowers. Summary statistics about these loans and lenders are in Table 2. The average (median) loan is $138,000 ($39,000). These are roughly similar to the amounts in the much larger sample of PPP loans used in Howell et al. (2022).

Lender Classification In both the Lendio and PPP data, we divide lenders into the following mutually exclusive groups, roughly following the approach in Howell et al. (2022):

1. Large, medium, and small banks;[20]

2. Credit unions, community development financial institutions (CDFIs) and minority depository institutions (MDIs);[21]

[18] These data are publicly available here: https://www.sba.gov/funding-programs/loans/covid-19-relief-options/paycheck-protection-program/ppp-data.

[19] These data on current firm officers as of July 2021 are drawn from secretary of state registrations. The owner is identified as the first individual listed as owner or principal under "business contacts" in secretary of state filings.

[20] We define these gruops by dividing banks into three equal groups according to assets.

[21] We identify credit unions based on the lender name (i.e., having "credit union" or "CU" at the end of the name). We identify CDFIs and MDIs using the FDIC classification.

3. Factoring, merchant cash advance (MCA), and business credit card (CC) lenders: These are longstanding alternatives to bank loans for small businesses, which typically charge very high interest rates. Factoring involves selling accounts receivable to the lender. MCAs are loan agreements where repayment is a percentage of sales. They appear only in the Lendio data since they are not SBA-approved lenders and SBA approval was required to participate in the PPP.

4. Fintechs: These include all lenders officially designated as fintechs by the SBA, plus online lenders who originate primarily for or via fintech partners or platforms, online lenders founded since 2005, and online lenders that received venture capital investment.

There are 369 unique lenders in the PPP data, of which 20 are fintechs. There are 101 unique lenders in the Lendio data, of which 47 are fintechs. The chance of a fintech (bank) loan among unique borrowers in Lendio is 42.5% (31.6%). Table 2 shows that the chance of getting a fintech (bank) loan in our PPP data is 10.3% (82.5%).

Geography-Based Covariates. We also collect data on socioeconomic characteristics of the firms in our data, which are summarized in Appendix Tables A.1 and A.2. Since BISG is based in large part on geographic variation, we are interested in how classification errors vary with geographic characteristics. To this end, we collect data from the U.S. Census Bureau's American Community Survey on ZIP code–level income and demographics, using data from 2019. We focus on two demographic variables: the share of the population that is Black and the share of the Black population with a bachelor's degree. The latter variable represents a proxy for a relatively advantaged (i.e. lower loan risk) Black population.[22]

We also collect two measures of anti-Black racial animus. The first comes from the implicit association test (IAT), which assesses implicit bias against Black individuals and is commonly used by researchers (Xu et al., 2014). The second measure follows Bursztyn et al. (2021) and is based on how favorably White respondents rate Black Americans as a group in the Nationscape survey (Tausanovitch and Vavreck, 2020). These measures shed light on how lenders' racial preferences might affect the error rate of proxies for race.

Last, we collect two measures of local residential segregation (Massey and Denton, 1988). The dissimilarity index captures differences in the distributions of White and Black residents across city tracts. The isolation index estimates the probability of a Black resident sharing the same city tract with another Black resident. These segregation variables are connected to racial preferences, but also enhance the precision of the geographic component of race. Both the animus and segregation variables are demeaned and divided by the standard deviation for ease of interpretation.

[22]Borrower education is widely known to predict loan performance, so much so that there have been concerns among policymakers that using education data in underwriting could disadvantage protected groups who typically have less access to education (see here). To highlight how education is relevant for lending, see here.

LinkedIn Profile Covariates. To understand how prediction errors vary with borrower-specific characteristics, we use information about education from the LinkedIn profiles. These data are reported in a standardized way on LinkedIn, so we can identify whether a person has a bachelor's, bachelor's of science, master's, and MBA.[23] We expect that the latter three variables are associated with better career options, greater wealth, more financial sophistication, and lower risk from the lender's perspective. These data are also summarized in Appendix Tables A.1 and A.2.

Infutor Data. While our data provide the address of each applicant's business, the BISG algorithm requires each applicant's residential address as an input. We obtain this by merging the applicant data with Infutor's CRD4 dataset, which contains the residential address history of most U.S. adults. We match each applicant to the closest residential address that has an inhabitant with the same first and last name at the time of the application in the CRD4 data. To compute distance, we first calculate the latitude and longitude of each city as the averages of the corresponding variables across all addresses in the Infutor data with the same city and state. We then compute distance between each city in our application data (the city of the borrowing firm) and each residential address from their latitudes and longitudes using the Haversine formula. In cases where we cannot match an applicant to a resident within 100 km, we use the business address ZIP code in place of a residential ZIP code. The results are very similar using only the business location.

4 Race Prediction Outputs: Comparing Measures of Race

This section compares BISG- and image-based measures of race and benchmarks them against self-identified race in the PPP data. We focus on the PPP data as this has a larger sample and includes self-identified race (recall we only use those borrowers who self-identify race to build the PPP sample). Essentially all of the findings also apply to Lendio, and we point to the parallel Lendio statistics in footnotes.

We first compare the measures. Note that self-identified race and image-based race should not be expected to capture the same concept of race, in that how one self-identifies is not always how one is perceived. Nonetheless, a higher correlation with self-identified race should indicate a better measure. Figure A.6 shows that across all applicants in the PPP sample, 6.6% report being Black (SelfID) while 76% report being White. Using image-based race, these statistics are 7% and 81%, respectively. Using BISG, they are about 9% and 75%.[24]

Correlation rates between the variables are in Table 3. In the top two panels, we present the correlation between indicators for being self-identified Black, image-based Black, and the continuous and indicator measures of BISG Black (recall we construct the indicator using random assignment with a weight corresponding to the continuous probability). When we use two binary

[23]We also collected MD, PhD, and JD degrees, but these were excluded due to the small sample sizes.
[24]The corresponding figure for Lendio is Appendix Figure A.1.

variables, the correlation between self-identified and image-based race is 87%, while the correlation between self-identified and BISG race is just 37%. In the subsequent rows, we present the same correlations for predictions that use alternative inputs. The two inputs to BISG are geography and surname. We can see that using each of these individually to predict self-identified Black performs very poorly, but that geography does a bit better than name.[25]

In the final panel, we use a version of BISG that takes into account first name (BISFG). This performs better than BISG, with a correlation of 41% with self-identified race. However, we focus on BISG in much of the analysis because the CFPB and other U.S. agencies use BISG to evaluate fair lending compliance, and we wish to speak directly to the implications of this standard. Also, the sample is larger because many first names do not have a race distribution, which means that in practice, BISFG is less useful.

Having established that image-based race is better correlated with self-identified race, we turn to BISG errors, which we describe with the terms *false positive* and *false negative*. These take either image-based or self-identified race as the "truth," but note that we mean "truth" and use terminology such as "is Black" in only a statistical sense. The reader should keep in mind that this is a simplification and no single measure is the truth for every individual. With this in mind, we use the term "false positive" to mean that a person is not Black according to our chosen baseline measure, but BISG identifies them as Black. Likewise, "false negative" implies that the person is Black according to our chosen baseline measure, but BISG fails to recognize them as such.

In the full PPP sample, Figure 5 Panels A and C show that using either self-identified or image-based race as true race, about 88% of the sample are true negatives. To analyze the socioeconomic predictors of BISG errors, we concentrate on the subsample that is either true Black or BISG-predicted Black (true positives, false positives, and false negatives). In this population, there are about 3,630 unique applicants. Figure 5 Panel B (D) shows that when "actual race" is based on self-identified (image-based) race, the true positive rate is 26% (27%), the false positive rate is 46% (44%), and the false negative rate is 28% (29%). Thus, BISG is more likely to misclassify Black applicants than to correctly classify them. Once again, the results are very similar using self-identified and image-based race as the "truth," which validates the image-based measure. In Appendix A.7, we show that the same patterns hold using image-based race in the Lendio data.

In sum, these statistics document that race measures differ substantially, with BISG performing much worse than image-based race when self-identified race is the benchmark. In fact, the BISG algorithm produces more false positives and false negatives than true positives when categorizing Black applicants. These errors make evaluating compliance with fair lending standards more challenging and reduce the precision of estimates of disparate impact.

[25] The corresponding rates for Lendio are in Appendix Table A.3.

5 Race Prediction Errors Are Not Random

Building on our previous finding that the BISG algorithm has a high error rate, we now study whether those errors are systematically related to applicant characteristics.

We focus attention on the narrow sample where at least one race measure classified the applicant as Black. This allows us to make two comparisons. The first is between the false positives and the individuals who *are* Black according to the image-based measure (comprising the true positive and false negative groups). The second comparison is between the false negatives and individuals who *are* Black according to BISG (the combined true positive and false positive groups). We exclude the true negatives from this analysis because they would otherwise dominate the sample, preventing us from effectively highlighting the differences of interest between the other groups. Within this sample, we correlate being false positive and false negative Black with socioeconomic characteristics of the borrower using both the PPP and Lendio data. As the PPP dataset is considerably larger and more representative (since a wide range of firms applied to PPP), these data are our primary focus. We present results for both image-based and self-identified race measures within the PPP data, which helps to further validate the image-based measure.

To quantify these effects, we run a series of regressions where we regress being either false positive or false negative Black on each of a set of socioeconomic characteristics in our PPP data. Figure 6 displays the regression results, where blue dots represent the coefficient on false negative and orange dots represent the coefficient on false positive. Panel A uses our image-based measure as "true" race for classification, while Panel B uses self-reported race. For both panels, we restrict to our "within Black" sample that excludes true negatives (correctly classified non-Black applicants). Our results are highly similar across these two classification types, so we report statistics from Panel A (image-based race) unless otherwise noted. All regressions are additionally reported in tabular form in Appendix Table A.5.

We begin with demographic variables at the ZIP level. Because BISG combines a surname probability, which does not depend on geography, and ZIP-level race shares, these demographic variables will summarize how BISG errors vary across geography. The first row of Figure 6 represents the regression results using the independent variable of whether an area's share of Black residents is above or below the national median. Since BISG is increasing with the Black share of the population by construction, BISG will predict that all applicants are more likely to be Black in areas with high Black share. These higher BISG probabilities of being Black in turn create more false positive errors (when the applicant is actually non-Black) and fewer false negative errors (when the applicant is actually Black). This intuition is confirmed by Figure 6 Panel A (image-based race), which shows that residing in a location with an above-median Black share of the population increases the chance of being a false positive by 5pp (12% of the mean) and decreases the chance of being a false negative by 27pp (94% of the mean).[26]

Next, while the Black share of the area's population influences which type of errors we observe, the total number of prediction errors is related to the racial diversity within the ZIP code. Areas with more heterogeneous populations will have more misclassified applicants, while a hypothetical area with no heterogeneity (a single racial group only) would be perfectly classified by BISG. Correspondingly, the next two rows show that measures of segregation predict lower levels of both types of error.[27]

We consider other socioeconomic characteristics at the geographic (ZIP) level in the following two rows of Figure 6. While BISG is mechanically determined by demographics, how these characteristics line up with demographics provides important context for the impacts of regulatory policy. The results show that local income and the share of the local Black population with a bachelor's degree have more false negative errors and fewer false positive errors, with particularly strong effects on the share of false negatives. These result stem from a confound with race: because areas with higher incomes and education levels tend to have a lower Black population on average, BISG tends to produce more false negative and fewer false positive errors in these areas. We also consider measures of racial animus at the ZIP level, which we find to be associated with more false positive errors and fewer false negative errors, meaning that Black borrowers are less likely to be misclassified in areas with higher racial animus.

The last six variables of Figure 6 comprise a set of individual-level education indicators obtained from applicants' LinkedIn profiles. These results align with our previous results at the geographic level, albeit with more noise. For all of the postgraduate education outcomes as well as for the bachelor's of science, we see that higher education predicts more false negative errors and fewer false positive errors. The effect of having any postgraduate degree is particularly strong, reducing the chance of being a false positive by 10pp (over 22% of the mean) and increasing the change of being a false negative by 5pp (17% of the mean) in Panel A (image-based race). There is no predictive power for having a bachelor's degree, which may reflect the fact that this is less informative for small business owners who are also on LinkedIn, a sample selected on a higher likelihood of having a bachelor's. Overall, the correlations confirm at the individual level that highly educated Black borrowers are particularly likely to be misclassified as non-Black by BISG.

These results have important effects for regulatory policy. In light of our findings in Section 2, BISG-based regulation incentivizes lending toward borrowers with false positive errors and away from borrowers with false negative errors. Applied to our findings above, this means that BISG-based regulation encourages more lending to borrowers in disadvantaged areas: those with a high Black share of the population, lower incomes, higher racial animus, and where Black residents are

[26]See Tables A.4 and A.5 for precise regression coefficients.

[27]The lone exception is the dissimilarity measure of segregation in the False Negative model, which cannot be distinguished from zero.

less educated. At the individual level, we theorize that BISG-based regulation incentivizes lending away from highly educated Black borrowers and toward less educated non-Black borrowers. These potential socioeconomic effects of regulation at both the geographic and individual level will be important considerations for policymakers.

6 Approval Rate Analysis

In the previous section, we showed the existence of large prediction errors and demonstrated that they covary with socioeconomic characteristics at the geographic and individual level. In this section, we study how these errors aggregate to bias measured disparities in approval rates across racial groups, which are frequently used as inputs in evaluating compliance with fair lending laws.

A central part of complying with fair lending rules is disparate treatment and disparate impact analyses, where the compliance officer or regulator asks whether the lender is serving protected groups (e.g., Black individuals) in a similar way to the majority group (e.g., White individuals). Note that a full analysis requires information on the risk level of applicants, which we do not observe. However, comparing approval rates across groups is an important first step; if a lender can show that they approve a similar share of applicants in protected groups as control groups, then government regulators will not typically look further for evidence of discriminatory conduct. The U.S. Interagency Fair Lending Examination Procedures, which apply to five federal agencies including the Federal Reserve Board and the Federal Deposit Insurance Corporation, detail how approval rates should be used by fair lending examiners. The first indicators of disparate treatment in underwriting are "substantial disparities among the approval/denial rates for applicants by monitored prohibited basis characteristic." In order to determine whether a detailed investigation is necessary, the procedures mandate that "after calculating denial rates between the control and prohibited basis groups for the underwriting centers, examiners should select the centers with the highest fair lending risk." Therefore, in our analysis, we focus on disparities in approval rates as an important dimension of compliance evaluation.

In the remainder of this section, we first measure approval rates at the aggregate level, showing that the use of BISG biases implied approval rates for Black borrowers up and would lead regulators to perceive a smaller gap between Black and non-Black (or Black and White) borrowers than actually exists using our image-based measure. We add rigor to these results in a formal regression setting to establish the statistical significance of our results and robustness in the presence of controls and fixed effects. Last, we extend our analysis by incorporating a variant of BISG that adds information on first name (BIFSG).

Results: Approval Rates. We now measure implied approval rates under our various measure of race in the Lendio sample, which are displayed in Table 4. We find that BISG errors have a large effect on measured approval rates. Using our image-based measure, applications by Black applicants are approved at a 6.2% rate, while applications by non-Black applicants are approved at a 8.5% rate, for a difference of 2.3pp. But when classifying race using BISG, the corresponding approval rates are 7.0% for Black applicants and 8.3% for non-Black applicants, implying a difference of only 1.3pp. As a result, using BISG in place of a more accurate measure would lead regulators to understate the true gap in approval rates by over 43%. Repeating this exercise to compare Black and White applicants specifically, we see that moving from image-based to BISG-based race reduced the measured approval gap between White and Black applicants from 2.5pp to 1.4pp, corresponding to an even larger 44% decrease. In summary, approval rate disparities between Black applicants and other applicants appear dramatically smaller when predicting race using BISG instead of our image-based measure.

We break down these approval rates by BISG error type in Figure 7. The lowest approval rate, at 5.6%, is for true positive Black, where both BISG and image measures agree that the business owner is Black. The next-lowest rate is for false negatives at 6.7%, corresponding to borrowers for whom image-based race is Black but BISG predicts non-Black. The approval rate is significantly higher for false positives, where BISG (and thus the regulator) predicts an applicant is Black but they are not actually Black, at 8.7%. Finally, the approval rate is 8.5% for true negatives. As derived in Section 2, the higher approval rates for false positives compared to false negatives drives the bias in measured approval rates. This gap between false positives and negatives aligns closely with the predictions of the model, since under BISG-based regulation a lender does not get "credit" for lending to false negative borrowers under the fair lending evaluation, but will for lending to false positive borrowers.

Having documented differences in approval rates across lenders when employing different race measures, we now turn to the initial phase of a compliance evaluation, which looks for differences in approval rates across protected and control groups. Specifically, as the disparity in the approval rates across Black and White individuals narrows, the likelihood of regulators initiating an investigation decreases. Therefore, we explore whether different race measures at the application level predict different loan approval gaps. We show that BISG errors lead the BISG-based Black measure to have poor predictive power.

While the results above document differences in approval rates across lenders when employing different measures of race, it is possible that these differences could be explained by confounding, or could simply be due to noise in an insufficiently large sample. To address this, we use our Lendio data to formally estimate variants of the regression model

$$\mathbb{1}(\text{Approved}_{i,l}) = \alpha_l + \alpha_t + \beta\mathbb{1}(\text{Black}_i) + \mathbf{X}_i\delta + \varepsilon_{il}, \tag{21}$$

where i denotes an applicant and l denotes a lender. We include fixed effects for the lender (α_l) and for the year of application (α_t) and control for the log amount of funding sought by the applicant. In some models, we further include the full set of socioeconomic characteristics seen in Figure 6.

The results are in Table 5. Columns 1 and 2 regress approval on a single indicator for being Black, either our image-based measure (Column 1) or the BISG-predicted measure (Column 2). These regressions echo the kind of analysis a regulator might perform to evaluate disparate lending outcomes across races. Comparing these columns shows that, while both indicators negatively predict approvals, the indicator on image-based race is larger in magnitude, with a point estimate of -1.8pp, approximately 64% larger than the corresponding coefficient on the BISG-based measure (-1.1pp). The estimated effect of the image-based measure is also more statistically significant, with a t-statistic of -3.6, compared to -2.2 for the BISG-based measure. Thus, a regulator using BISG would perceive approval gaps to be both smaller and more likely to be explained by noise than a regulator using an image-based measure.

We next incorporate both indicators on the right hand side of the regression, with the results displayed in Column 3. These estimates show that the predictive power of the BISG indicator is largely subsumed by the image-based indicator, failing to provide independent variation. Specifically, the point estimate on the image-based indicator remains virtually unchanged from Column 1 at -1.6pp, whereas the point estimate on the BISG indicator loses its statistical significance. Column 4 displays estimates of this regression with socioeconomic controls and finds that the results are essentially unchanged. These results imply that the BISG-based race measure is not only weaker than the image-based measure in terms of predicting approvals, but is essentially redundant for fair lending evaluations, with very little additional predictive information added compared to the image-based measure alone.

Finally, we disaggregate our results by splitting the positive or negative values of the BISG indicator into subcategories depending on whether the classification aligns with our image-based measure or not, keeping the true negatives as the omitted group. We see a very large negative coefficient of -2.3pp for the true positive category where both image and BISG classify borrowers as Black. However, the coefficient on the false positive group, which BISG incorrectly classifies as Black, is virtually zero. By averaging these two groups, which are of similar size, the overall BISG indicator loses predictive power. Furthermore, the BISG indicator overlooks the strong negative coefficient of -1.4pp for the false negative group of those BISG incorrectly identified as non-Black. These results support the conclusion that classification errors vitiate BISG's predictive power, and cause it to be subsumed by the image-based measure when they are both included.

Adding First Name. In an extension, we add the first name to last name and geography. In the social sciences, the most rigorous evidence of racial discrimination is from correspondence audit studies in which first names are used to signal race (Butler and Broockman, 2011; Milkman et al., 2012; Bartoš et al., 2016; Giulietti et al., 2019).[28] Distinctively Black first names may be related to parental identity and socioeconomic status (Fryer and Levitt, 2004; Gaddis, 2017; Kreisman and Smith, 2023). As far as we know, first name is not used by regulators in assessing fair lending.

In Appendix Table A.11, we evaluate the BIFSG algorithm, which extends BISG by using both first and last name (see Voicu (2018)). This approach has the downside of more missing observations (see Section 1). BIFSG has better predictive power over approvals than BISG (columns 1–2), though image-based race continues to outperform.[29] The improved power with first names is mostly due to a larger negative coefficient on false positives. Therefore, the improved performance of BIFSG is driven by how its errors correlate with approval rates; this could reflect false positives being more strongly associated with lower socioeconomic status when first name is included (i.e., the first name is more "Black"), consequently decreasing the likelihood of loan approval. Indeed, we observe that measures of disadvantage, such as low local per capita income, more strongly predict being false positive Black using BISFG. In sum, while the inclusion of first names in the race prediction algorithm improves its performance slightly, the improvement does not reflect the impact of true positives but rather an increase in false positives associated with lower socioeconomic status.

Overall, this section shows that in our sample, image-based race predicts loan approval better than BISG-based race. Disaggregating the BISG errors reveals that BISG's poor performance is largely due to false negative Black individuals being less likely to get approved. This is relevant for evaluating lender compliance with fair lending laws. Suppose a lender primarily serves a demographic with a high false negative rate. If these individuals, who are indeed Black, are less likely to secure loans, as is the case in our sample, the lender will seem, under a BISG regime, to approve a larger share of Black applicants than they really do, creating an illusion of better compliance with fair lending laws. BISG errors could also induce distortionary incentives to cater to a demographic with a higher false positive rate in order to maintain compliance. These errors and potential for manipulation may undermine the intent of the law if they make true Black applicants less able to secure credit.

7 Lender-Level Analysis

Having established the impact of prediction errors on the aggregate population, we now refine our scope to study how these effects vary at the lender level. The key point to keep in mind is that if two lenders serve different socioeconomic groups, the bias in BISG may lead fair lending regulators

[28]A prominent example where the title conveys the process is Bertrand and Mullainathan (2004): "Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination."

[29]First names alone are not especially predictive (Appendix Table A.12). In Appendix Table A.13, we predict approval using each race measure separately. Here we see that algorithms which include the first-name (columns 5–6) are more predictive than BISG or its components (columns 2–4).

to arrive at different compliance conclusions even if the two lenders are actually lending to Black-owned firms at the same rates. We construct a measure for the difference in approval rates at the lender level using image-based race vs. BISG-based race as:

$$\begin{aligned}\Delta_{\text{Share Black Appr}} &= \bar{\pi}_B - \bar{\pi}_B^{BISG} \\ &= \frac{\#\,\text{Image Black Approved}}{\#\,\text{Image Black Applicants}} - \frac{\#\,\text{BISG Black Approved}}{\#\,\text{BISG Black Applicants}}. \end{aligned} \tag{22}$$

The first term on the right hand side of equation (22) is the number of Black applicants who are approved scaled by the total number of Black applicants, using the image-based indicator for being Black. The second term is defined analogously, except we use the continuous percent Black term produced by the BISG algorithm.[30]

From equation (20) we know that $\Delta_{\text{Share Black Appr}}$ increases with the false negative rate among approved borrowers and decreases with the false positive rate.[31] Since there is a positive correlation between false negatives and advantaged socioeconomic status, a lender serving more privileged Black borrowers would have a higher $\Delta_{\text{Share Black Appr}}$. In contrast, lenders serving a higher proportion of false positive borrowers would appear more compliant with fair lending laws than they actually are, with a lower $\Delta_{\text{Share Black Appr}}$.[32]

Lender-Level Results: Lendio Sample. We summarize $\Delta_{\text{Share Black Appr}}$ in our Lendio sample in Table 7 Panel A (bottom section, “Difference in Rates by Race Measure”). On average, it is close to zero, yet there is large variation. To explore this variation, Figure 8 plots $\Delta_{\text{Share Black Appr}}$ for each lender. The graph shows large variation across lenders, with some having large negative and others large positive $\Delta_{\text{Share Black Appr}}$. Furthermore, there seems to be some suggestive ordering by lender type, with $\Delta_{\text{Share Black Appr}}$ being more frequently negative for banks and factoring/MCA/CC and more commonly positive for fintech lenders. Note that merchant cash advances, factoring, and business credit card products are long-standing and predate fintechs; they are typically associated with very high interest rates. However, this sample of lenders is far from representative of small business lenders in the U.S.; for example, Lendio’s client base skews substantially towards fintechs. Nevertheless, it is noteworthy that banks and other conventional small business lenders lean towards the negative side while fintechs lean towards the positive side. Banks typically rely on soft information for underwriting (Petersen and Rajan, 1994; Berger and Black, 2011), while

[30] Specifically, we sum the probabilities that an applicant is Black according to BISG within each lender’s portfolio, and divide the sum by the sum of probabilities than each applicant within the lender’s portfolio falls into all the different racial categories.

[31] We present the lender-specific false positive and false negative rates in Appendix Figures A.2, A.3, A.4, and A.5.

[32] To put it differently, if $\Delta_{\text{Share Black Appr}}$ is positive, then $\frac{\#\,\text{Image Black Approved}}{\#\,\text{Image Black Applicants}} > \frac{\#\,\text{BISG Black Approved}}{\#\,\text{BISG Black Applicants}}$. This implies that the lender is serving the Black population at a higher rate than BISG makes it appear. Conversely, if $\Delta_{\text{Share Black Appr}}$ is negative, then the lender is not serving as high a share of Black applicants as it appears.

fintechs and large banks are more automated and arms-length than small banks (Howell et al., 2022; Balyuk et al., 2020).

Table 8 assesses whether certain lender types are associated with levels of $\Delta_{\text{Share Black Appr}}$. To do this, we run a simple regression to measure the links between $\Delta_{\text{Share Black Appr}}$ and three lender types—banks, fintechs, and factoring/MCA/CC—with banks as the reference category. We consider the following dependent variables: an indicator for $\Delta_{\text{Share Black Appr}}$ being positive, the continuous value of $\Delta_{\text{Share Black Appr}}$, and an indicator for $\Delta_{\text{Share Black Appr}}$ being above its 75th percentile. As shown in columns 1–3, fintechs are more likely to be at the top of the distribution, although the results are noisy.

Lender-Level Results: PPP Sample. The PPP data has the advantage of many more lenders, which are largely representative of the universe of U.S. small business lenders. While the PPP data do not have anything analagous to an approval or rejection rate, we can take a different approach to defining disparate impact by measuring the share of loans to Black-owned firms:

$$\Delta_{\text{Share Loans Black}} = \frac{\#\,\text{Image Black Borrowers}}{\#\,\text{All Borrowers}} - \frac{\#\,\text{BISG Black Borrowers}}{\#\,\text{All Borrowers}}. \tag{23}$$

We again take the difference between the share using image-based race and the share using BISG race. The lender-specific differences for the 368 unique lenders in our PPP analysis sample are shown in Figure 9. Factoring/MCA/CC lenders are absent in the PPP data as they were never SBA-approved to participate. As in the Lendio data, fintechs are more frequently on the positive (right) side of the distribution, while banks are more commonly concentrated on the left. We assess whether there are significant differences in Table 8 columns 4–6, using six lender categories. Since we have many lenders, we consider small, medium, and large banks separately, with small banks as the omitted category. Here, we see a more precise result, with fintechs being much more likely to have a high $\Delta_{\text{Share Loans Black}}$ compared to small banks. For example, they are 64pp more likely to have a positive $\Delta_{\text{Share Loans Black}}$, which is 256% of the mean (column 4). Credit unions, CDFIs, and minority depository institutions are also somewhat more likely to have higher $\Delta_{\text{Share Loans Black}}$.

The above results have nuanced implications for policy, though further research is needed to determine whether these findings can be generalized to a broader, more representative sample. The results suggest that some banks might benefit from a BISG-based fair lending evaluation. Interestingly, banks have persistently opposed the Dodd-Frank rule suggesting that lenders and regulators collect self-identified race data in small business lending.[33] The results are also consistent with fintechs serving Black applicants with higher socioeconomic status. This highlights intricate equity implications of shifting from BISG-based race measures to ones more reflective of

[33] The Dodd-Frank Act required the CFPB to adopt regulations on this matter (referred to as "1071" due to the Section of the Act). There has been stiff opposition from the small banking community (see here and here).

how individuals are typically perceived by others (image-based race) or self-identified race. Since false negatives are associated with higher socioeconomic status, if lenders adjust lending towards more false positives and fewer false negatives, the net effect could lead to lending to individuals of lower socioeconomic status, independent of race.

8 Counterfactual Exercise

In Section 2 we examined a regulatory environment in which lenders are constrained by how much their approval rates may differ across groups. In that setting, shifting from a regulatory environment based on BISG-predicted race to one based on actual race would lead to a linear reduction in approvals for BISG-predicted Black ("BISG-Black") applicants and a linear increase in approvals for actual Black applicants. In this final exercise, we study the potential impact of such a counterfactual policy change, where we proxy for actual Black race with image-based Black race ("image-Black"). Since we do not observe the environment without regulatory constraints, we cannot directly identify the size of the λ multipliers. As a result, the exact size of this shift is not clear. Instead, our results allow us to describe the *direction* of this change, without taking a stand on the magnitude.

We begin with our approval rate regression, equation (21), and consider changing the coefficients on the BISG-Black and image-Black variables while holding the overall approval rate fixed. Specifically, we can compute counterfactual approval rates as follows:

$$Approved_i^{BISG\downarrow} = Approved_i - h \times (BISG_i - \overline{BISG}), \tag{24}$$

$$Approved_i^{Image\uparrow} = Approved_i + h \times (Black_i - \overline{Black}), \tag{25}$$

$$Approved_i^{Both} = Approved_i + h \times \Big[(Black_i - \overline{Black}) - (BISG_i - \overline{BISG})\Big]. \tag{26}$$

Equation (24) computes a counterfactual approval rate where the weight on BISG-Black probability has been lowered by an arbitrary small constant h. To focus on changes in the tilt toward or away from different types of applicants, rather than changes in the overall level of borrowing, we remove the mean of the BISG-Black variable before multiplying by h so that the average approval rate is unchanged between equations (24) and (21). In a second counterfactual experiment, we apply a symmetric procedure, increasing the weight on the image-Black variable to create the counterfactual approval rate in equation (25). Finally, we simultaneously apply both an increase in the weight on the image-Black measure and a decrease in the weight on the BISG-Black measure in Equation (26). This last experiment most closely approximates the directional shift from the BISG-based regulatory environment to one based on actual race.

We study the effect of these changes on the characteristics of the approved population. For each characteristic Z and each counterfactual scenario C, we compute the weighted average of that characteristic among approved applications as

$$ApprovedShare_Z^C = \frac{\sum_i Z_i Approved_i^C}{\sum_i Approved_i^C}. \tag{27}$$

The directional change in that variable under the counterfactual scenario is

$$dApprovedShare_Z^C = \frac{ApprovedShare_Z^C - ApprovedShare_Z}{h} = \frac{\text{Cov}(Z_i, X_{ij})}{\overline{Approved}}, \tag{28}$$

where $ApprovedShare_Z$ is from equation (27) using the actual approval rate $Approved_i$ in place of $Approved_i^C$, $\text{Cov}(Z_i, X_{ij})$ is the sample covariance of Z_i with X_{ij}, $\overline{Approved}$ is the sample mean approval rate, and X_{ij} is the policy variable we are adjusting, which is either $BISG_i, Black_i$, or $(Black_i - BISG_i)$. The variable $dApprovedShare_Z^C$ thus represents the derivative of the share of the approved population with characteristic Z as we move in the direction of policy counterfactual C. We derive the second equality in Appendix C.

The results are displayed in Table 9, where columns (1), (2), and (3) correspond to the counterfactuals lowering the weight on BISG-Black (equation (24)), increasing the weight on image-Black (equation (25)), and applying both changes at once (equation (26)), respectively. The first two rows of Panel A consider correct classifications: true positives and true negatives. Since BISG-Black and image-Black are positively correlated, reducing the weight on BISG reduces the share of loans for true positives and increases the share to true negatives. Increasing the weight on image-Black has the reverse effects. The measurement error in BISG leads to stronger effects for correctly classified borrowers when the weight on image-Black increases.

The third and fourth rows show that the effects on false positive and false negative applicants vary widely across the experiments. Both components of the policy reduce false positives, since these borrowers have high values of BISG-Black but are not Black according to the image-based measure. Interestingly, while the image-based policy increases the share of loans going to false negatives, reducing the weight on BISG-Black actually *decreases* lending to this population. The reason is that the typical false negative borrower still has an above-average BISG-Black score (i.e., compared to the true negative non-Black population), implying that these borrowers still see their approval rates decline when lending to high BISG-Black borrowers falls. Combined, however, the two policies together increase the share of lending going to Black borrowers falsely classified as non-Black by BISG.

The fifth row of Panel A displays the impact on lending by image-based race. Because image-based race and BISG-predicted race are correlated, reducing the weight on the BISG-predicted Black probability also reduces lending to borrowers identified as Black by image. Because this correlation is imperfect, however, this effect is more than undone by the larger increase in lending to Black borrowers from directly increasing weight on the image-based Black measure in Column (2). As a result, the combined effect of the shift in regulatory regime is a net increase in lending to image-based Black borrowers.

In Panel B of Table 9, we analyze how these counterfactuals affect socioeconomic characteristics at the ZIP code level. We begin with per capita income in an applicant's ZIP code. Since this variable is negatively correlated with both the BISG-Black and image-Black measures, it increases in column 1 and decreases in column 2. However, the negative correlation is stronger for the BISG measure because BISG, by construction, loads more heavily on geography. As a result, when applying both policies at the same time, the effect of changing the weight on BISG-Black dominates, increasing lending to higher-income areas. In this sense, the change in regulatory environment may increase inequality in lending between high-income and low-income areas, despite potentially decreasing inequality in lending between Black and non-Black borrowers.

The remaining rows of Panel B show a similar pattern. Both the BISG-Black and image-Black measures are positively correlated with the Black share of the local population, while both the BISG-Black and image-Black measures are negatively correlated with the share of the local population with a bachelor's degree. However, in both cases the correlation with the geographic characteristic is stronger for the BISG-based variable than the image-based variable. As a result, tilting lending away from BISG-Black borrowers and toward image-Black borrowers decreases the share of lending going to more Black areas and increases the share of lending going to more educated areas.

9 Conclusion and Policy Discussion

There is "folk knowledge" among practitioners and researchers that the widely used race prediction algorithms based on name and location perform poorly. Although these demography-based measures might predict race reasonably well for some groups, they can have large error rates when names have ambiguous cultural origins or when the population in specific locations is diverse. Recognizing the discrimination historically faced by Black Americans in credit markets and the consequent interest of regulators, compliance officers, and researchers in this group, we concentrate on analyzing the accuracy of these proxies in predicting whether a small business borrower is Black. If errors in these algorithms are correlated with socioeconomic characteristics that are related to loan profitability, this could influence apparent compliance with fair lending laws based on the employed measure of race—whether it's image-based, self-identified, or BISG. This has important implications for policy, including potential sanctions for certain lenders and giving more latitude to new fintech lenders, particularly if they serve Black applicants at higher rates than conventional lenders.

Despite anecdotal reports, there is, to our knowledge, no comprehensive documentation of the potential consequences of race prediction algorithm errors in a nonmortgage context.

Understanding the performance, comparative efficiencies, and correlations with socioeconomic traits of these algorithms is, therefore, a unique contribution.

In this paper, we introduce an image-based measure of perceived race, which we show better correlates with self-identified race than BISG. We then show that the large errors in BISG yield more combined false positives (not being Black when BISG predicts Black) and false negatives (being Black when BISG predicts not-Black) than true positives. These errors are systematically related to measures of socioeconomic advantage; false positive Black individuals tend to be more disadvantaged, while false negative Black individuals are generally more advantaged. Using data on loan approvals, we show that image-based Black race is a stronger negative predictor of loan approval than BISG-based Black race, reflecting lower approval rates for false negative Black applicants. Our theoretical framework shows how BISG-based fair lending compliance evaluations could incentivize lenders to manipulate their performance by adjusting their lending rates to individuals whom BISG misclassifies.

Our findings imply that regulators, researchers, and practitioners should consider their specific objectives before selecting a method for measuring race. For instance, if a regulator's goal is to identify individuals who are Black and relatively disadvantaged within the Black community, our results indicate that BISG may serve quite well. However, if the goal is to identify Black individuals who experience discrimination based solely on skin tone and facial features, our findings expose significant limitations in BISG, suggesting that self-identified or image-based data might be more suitable. Our results have real-world implications beyond lending to domains such as university admissions, healthcare, and research.

References

Ambrose, Brent W, James N Conklin, and Luis A Lopez, "Does borrower and broker race affect the cost of mortgage credit?," *Review of Financial Studies*, 2021, *34* (2), 790–826.

Anwar, Shamena and Hanming Fang, "An alternative test of racial prejudice in motor vehicle searches: Theory and evidence," *American Economic Review*, 2006, *96* (1), 127–151.

Arnold, David, Will Dobbie, and Crystal S Yang, "Racial bias in bail decisions," *Quarterly Journal of Economics*, 2018, *133* (4), 1885–1932.

Asiedu, Elizabeth, James A Freeman, and Akwasi Nti-Addae, "Access to credit by small businesses: How relevant are race, ethnicity, and gender?," *American Economic Review*, 2012, *102* (3), 532–37.

Athey, Susan, Dean Karlan, Emil Palikot, and Yuan Yuan, "Smiles in profiles: Improving fairness and efficiency using estimates of user preferences in online marketplaces," Technical Report, National Bureau of Economic Research 2022.

Balyuk, Tetyana, Allen N Berger, and John Hackney, "What is fueling fintech lending? The role of banking market structure," *The Role of Banking Market Structure (June 23, 2020)*, 2020.

Bartik, Alexander and Scott Nelson, "Deleting a signal: Evidence from pre-employment credit checks," *Review of Economics and Statistics (Forthcoming)*, 2021.

Bartlett, Robert, Adair Morse, Richard Stanton, and Nancy Wallace, "Consumer-lending discrimination in the fintech era," *Journal of Financial Economics*, 2022, *143* (1), 30–56.

Bartoš, Vojtěch, Michal Bauer, Julie Chytilová, and Filip Matějka, "Attention discrimination: Theory and field experiments with monitoring information acquisition," *American Economic Review*, 2016, *106* (6), 1437–75.

Bayer, Patrick, Fernando Ferreira, and Stephen L Ross, "What Drives Racial and Ethnic Differences in High-Cost Mortgages? The Role of High-Risk Lenders," *Review of Financial Studies*, 2018, *31* (1), 175–205.

Begley, Taylor A and Amiyatosh Purnanandam, "Color and credit: Race, regulation, and the quality of financial services," *Journal of Financial Economics*, 2021, *141* (1), 48–65.

Bellucci, Andrea, Alexander Borisov, and Alberto Zazzaro, "Do banks price discriminate spatially? Evidence from small business lending in local credit markets," *Journal of Banking & Finance*, 2013, *37* (11), 4183–4197.

Berg, Tobias, Valentin Burg, Ana Gombović, and Manju Puri, "On the rise of fintechs: Credit scoring using digital footprints," *Review of Financial Studies*, 2020, *33* (7), 2845–2897.

Berger, Allen N and Lamont K Black, "Bank size, lending technologies, and small business finance," *Journal of Banking & Finance*, 2011, *35* (3), 724–735.

Bertrand, Marianne and Sendhil Mullainathan, "Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination," *American Economic Review*, 2004, *94* (4), 991–1013.

Bhutta, Neil and Aurel Hizmo, "Do minorities pay more for mortgages?," *Review of Financial Studies*, 2021, *34* (2), 763–789.

Blanchflower, David G, Phillip B Levine, and David J Zimmerman, "Discrimination in the small-business credit market," *Review of Economics and Statistics*, 2003, *85* (4), 930–943.

Blattner, Laura and Scott Nelson, "How Costly is Noise? Data and Disparities in Consumer Credit," *arXiv preprint arXiv:2105.07554*, 2021.

Buchak, Greg, Gregor Matvos, Tomasz Piskorski, and Amit Seru, "Fintech, regulatory arbitrage, and the rise of shadow banks," *Journal of Financial Economics*, 2018, *130* (3), 453–483.

Bursztyn, Leonardo, Thomas Chaney, Tarek Alexander Hassan, and Aakaash Rao, "The Immigrant Next Door: Exposure, Prejudice, and Altruism," Technical Report, National Bureau of Economic Research 2021.

Butler, Daniel M and David E Broockman, "Do politicians racially discriminate against constituents? A field experiment on state legislators," *American Journal of Political Science*, 2011, *55* (3), 463–477.

Charles, Kerwin Kofi and Jonathan Guryan, "Prejudice and wages: An empirical assessment of Becker's The Economics of Discrimination," *Journal of Political Economy*, 2008, *116* (5), 773–809.

Chernenko, Sergey and David S. Scharfstein, "Racial Disparities in the Paycheck Protection Program," Technical Report, Available at SSRN: https://ssrn.com/abstract=3907575 or http://dx.doi.org/10.2139/ssrn.3907575 2021.

D'Acunto, Francesco, Pulak Ghosh, Rajiv Jain, and Alberto G Rossi, "How costly are cultural biases?," *Available at SSRN 3736117*, 2020.

Dimmock, Stephen G, William C Gerken, and Nathaniel P Graham, "Is fraud contagious? Coworker influence on misconduct by financial advisors," *The Journal of Finance*, 2018, *73* (3), 1417–1450.

Dobbie, Will, Andres Liberman, Daniel Paravisini, and Vikram S Pathania, "Measuring bias in consumer lending," *Review of Economic Studies*, 2021, *88* (6), 2799–2832.

Egan, Mark, Gregor Matvos, and Amit Seru, "When Harry fired Sally: The double standard in punishing misconduct," *Journal of Political Economy*, 2022, *130* (5), 1184–1248.

Erel, Isil and Jack Liebersohn, "Does fintech substitute for banks? Evidence from the Paycheck Protection Program," Technical Report, National Bureau of Economic Research 2020.

Fairlie, Robert, Alicia Robb, and David T Robinson, "Black and White: Access to capital among minority-owned start-ups," *Management Science*, 2022, *68* (4), 2377–2400.

Fairlie, Robert W and Alicia M Robb, "Why are Black-owned businesses less successful than White-owned businesses? The role of families, inheritances, and business human capital," *Journal of Labor Economics*, 2007, *25* (2), 289–323.

_ **and Frank Fossen**, "Did the $660 billion Paycheck Protection Program and $220 Bbillion Economic Injury Disaster Loan Program get disbursed to minority communities in the early stages of COVID-19?," Technical Report, National Bureau of Economic Research 2021.

Frame, W Scott, Ruidi Huang, Erik J Mayer, and Adi Sunderam, "The impact of minority representation at mortgage lenders," Technical Report, National Bureau of Economic Research 2022.

Fremont, Allen, Joel S Weissman, Emily Hoch, and Marc N Elliott, "When race/ethnicity data are lacking: Using advanced indirect estimation methods to measure disparities," *Rand Health Quarterly*, 2016, *6* (1).

Fryer, Roland G and Steven D Levitt, "The causes and consequences of distinctively Black names," *Quarterly Journal of Economics*, 2004, *119* (3), 767–805.

Fuster, Andreas, Matthew Plosser, Philipp Schnabl, and James Vickery, "The role of technology in mortgage lending," *Review of Financial Studies*, 2019, *32* (5), 1854–1899.

Gaddis, S Michael, "How Black are Lakisha and Jamal? Racial perceptions from names used in correspondence audit studies," *Sociological Science*, 2017, *4*, 469–489.

Giacoletti, Marco, Rawley Heimer, and Edison G Yu, "Using high-frequency evaluations to estimate discrimination: Evidence from mortgage loan officers," *Available at SSRN 3795547*, 2021.

Giulietti, Corrado, Mirco Tonin, and Michael Vlassopoulos, "Racial discrimination in local public services: A field experiment in the United States," *Journal of the European Economic Association*, 2019, *17* (1), 165–204.

Howell, Sabrina T, Theresa Kuchler, David Snitkof, Johannes Stroebel, and Jun Wong, "Lender automation and racial disparities in credit access," *Journal of Finance (Forthcoming)*, 2022.

Jiang, Erica Xuewei, Yeonjoon Lee, and Will Shuo Liu, "Disparities in consumer credit: The role of loan officers in the FinTech era," *Available at SSRN*, 2021.

Kabir, Poorya and Tianyue Ruan, "Reducing Racial Disparities in Consumer Credit: Evidence from Anonymous Loan Applications," *Available at SSRN 4364852*, 2023.

Knowles, John, Nicola Persico, and Petra Todd, "Racial bias in motor vehicle searches: Theory and evidence," *Journal of Political Economy*, 2001, *109* (1), 203–229.

Kreisman, Daniel and Jonathan Smith, "Distinctively Black names and educational outcomes," *Journal of Political Economy*, 2023, *131* (4), 877–897.

Massey, Douglas S and Nancy A Denton, "The dimensions of residential segregation," *Social Forces*, 1988, *67* (2), 281–315.

Milkman, Katherine L, Modupe Akinola, and Dolly Chugh, "Temporal distance and discrimination: An audit study in academia," *Psychological Science*, 2012, *23* (7), 710–717.

Petersen, Mitchell A and Raghuram G Rajan, "The benefits of lending relationships: Evidence from small business data," *Journal of Finance*, 1994, *49* (1), 3–37.

Phillips, P Jonathon, Amy N Yates, Ying Hu, Carina A Hahn, Eilidh Noyes, Kelsey Jackson, Jacqueline G Cavazos, Géraldine Jeckeln, Rajeev Ranjan, Swami Sankaranarayanan et al., "Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms," *Proceedings of the National Academy of Sciences*, 2018, *115* (24), 6171–6176.

Pool, Veronika K, Noah Stoffman, and Scott E Yonker, "The people in your neighborhood: Social interactions and mutual fund portfolios," *Journal of Finance*, 2015, *70* (6), 2679–2732.

Price, Joseph and Justin Wolfers, "Racial discrimination among NBA referees," *Quarterly journal of economics*, 2010, *125* (4), 1859–1887.

Robb, Alicia and David T Robinson, "Testing for racial bias in business credit scores," *Small Business Economics*, 2018, *50* (3), 429–443.

Serengil, Sefik Ilkin and Alper Ozpinar, "Lightface: A hybrid deep face recognition framework," in "2020 Innovations in Intelligent Systems and Applications Conference (ASYU)" IEEE 2020, pp. 1–5.

Tang, Huan, "Peer-to-peer lenders versus banks: Substitutes or complements?," *Review of Financial Studies*, 2019, *32* (5), 1900–1938.

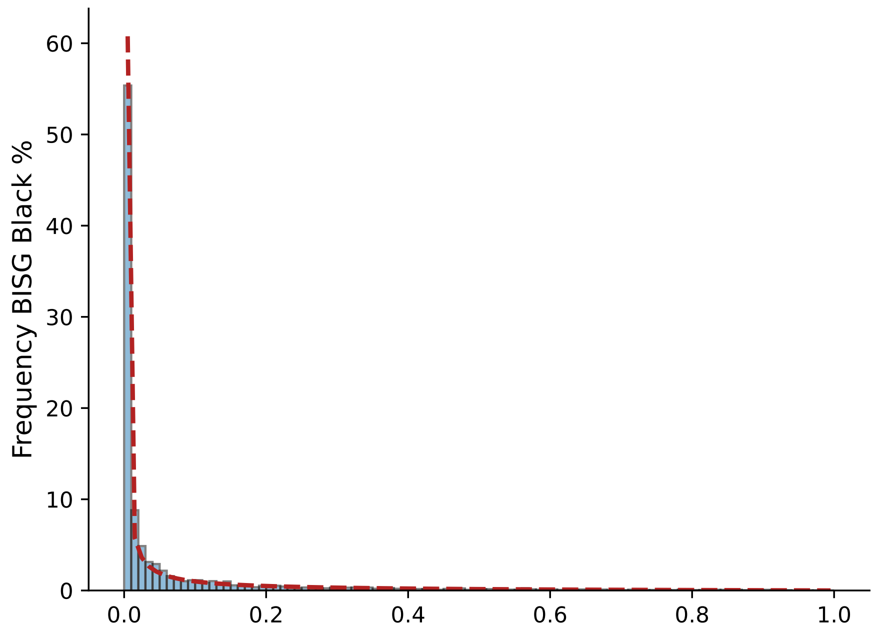
Tausanovitch, Chris and Lynn Vavreck, "Democracy Fund + UCLA Nationscape, Oct 10–17, 2019 (version 20200814)," Technical Report 2020.

Tootell, Geoffrey MB, "Redlining in Boston: Do mortgage lenders discriminate against neighborhoods?," *Quarterly Journal of Economics*, 1996, *111* (4), 1049–1079.

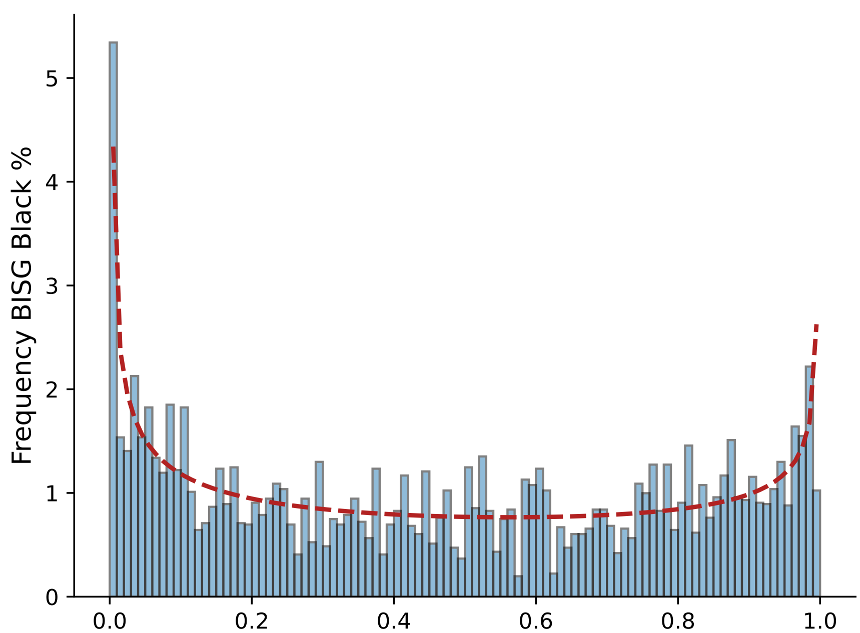
Voicu, Ioan, "Using first name information to improve race and ethnicity classification," *Statistics and Public Policy*, 2018, *5* (1), 1–13.

Xu, Kaiyuan, Brian Nosek, and Anthony Greenwald, "Psychology data from the race implicit association test on the project implicit demo website," *Journal of Open Psychology Data*, 2014, *2* (1).

Figure 1: BISG Densities by Actual Race



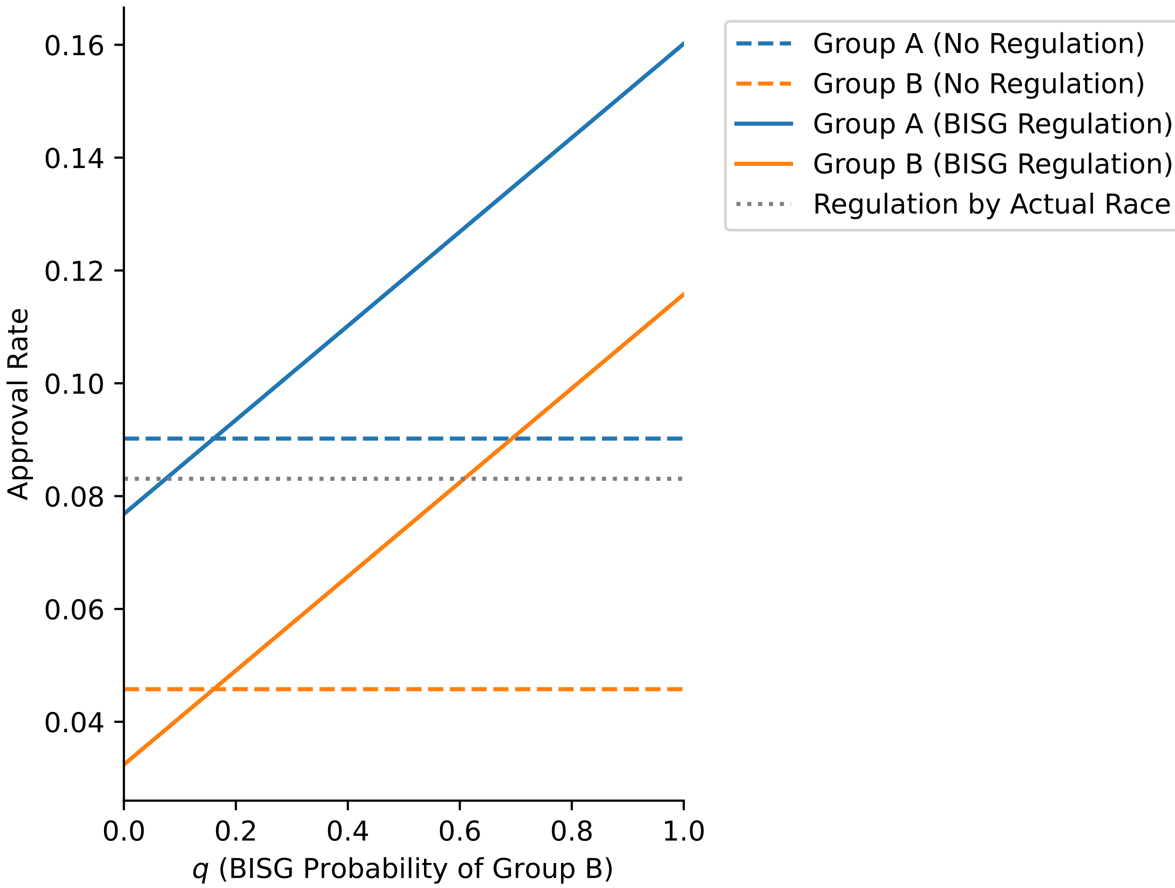
(A) Non-Black



(B) Black

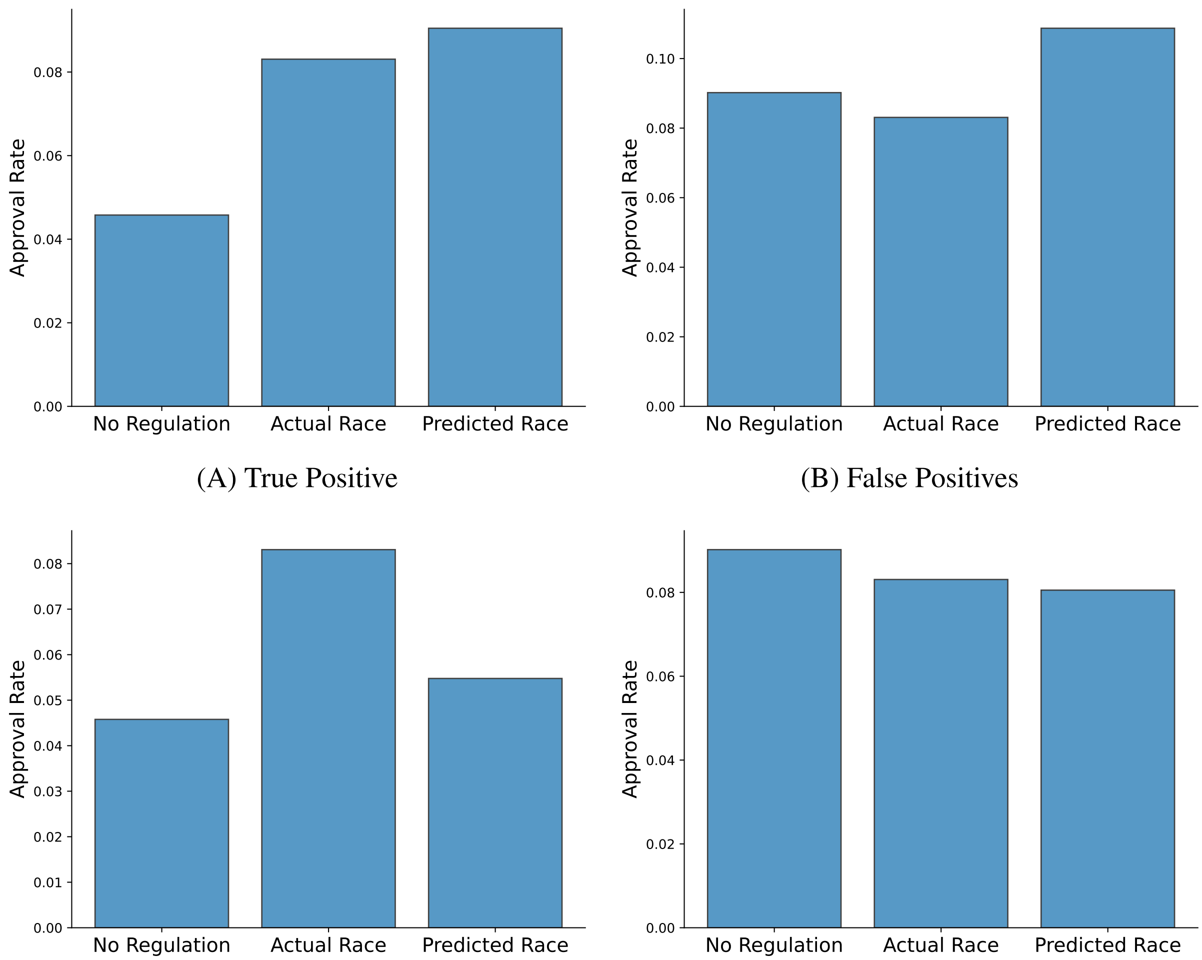
Note: The figures above display the empirical and fitted distributions of the probability of being Black according to the BISG algorithm. Panels (A) and (B) display borrowers identified to be non-Black and Black, respectively by our image-based measure. For each group, blue bars display our data's empirical distribution in one-percent bins, while the red dashed line displays the fitted beta distribution used by the model.

Figure 2: Numerical Example: Lender Approval Policy



Note: This figure displays the model's approval rates. The blue dashed line labeled π_A^{NR} displays the approval rate for Group A under the no regulation equilibrium, while the blue solid line labeled π_A^{PR} displays the approval rate for Group A under the predicted race regulation equilibrium. The orange dashed line labeled π_B^{NR} and orange solid line labeled π_B^{PR} display the same objects for Group B. The gray dashed line labeled $\pi_A^{AR} = \pi_B^{AR}$ displays the approval rate for both Groups A and B under the actual race equilibrium, as these turn out to be identical in this case.

Figure 3: Model Numerical Example: Approvals by Classification Type



(A) True Positive

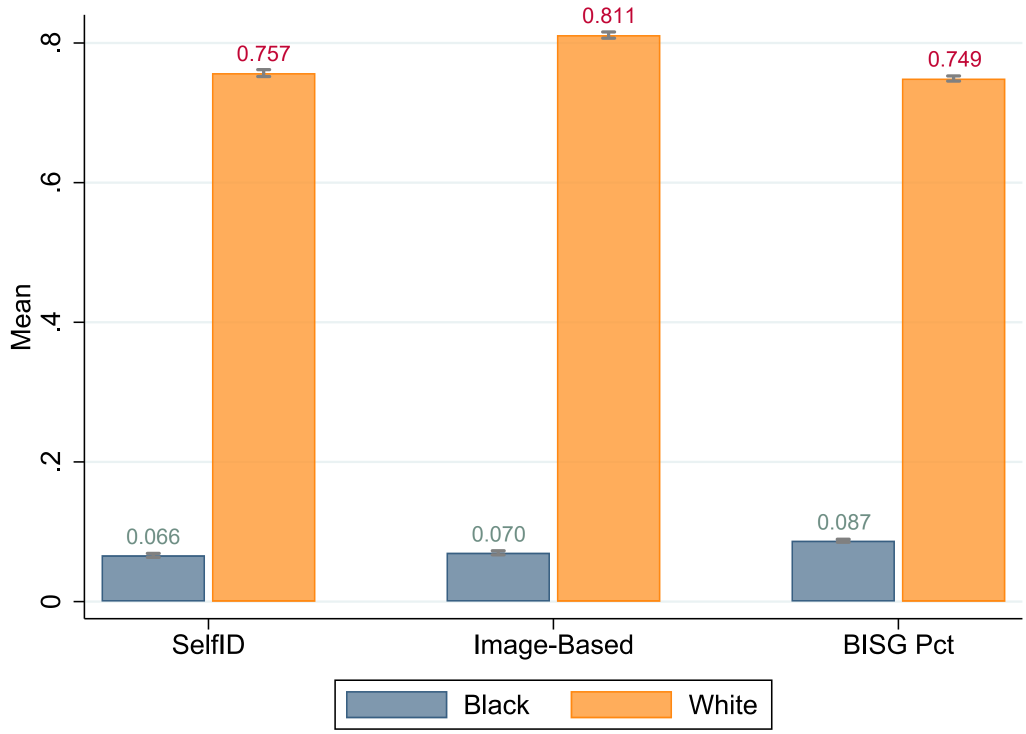
(B) False Positives

(C) False Negative

(D) True Negative

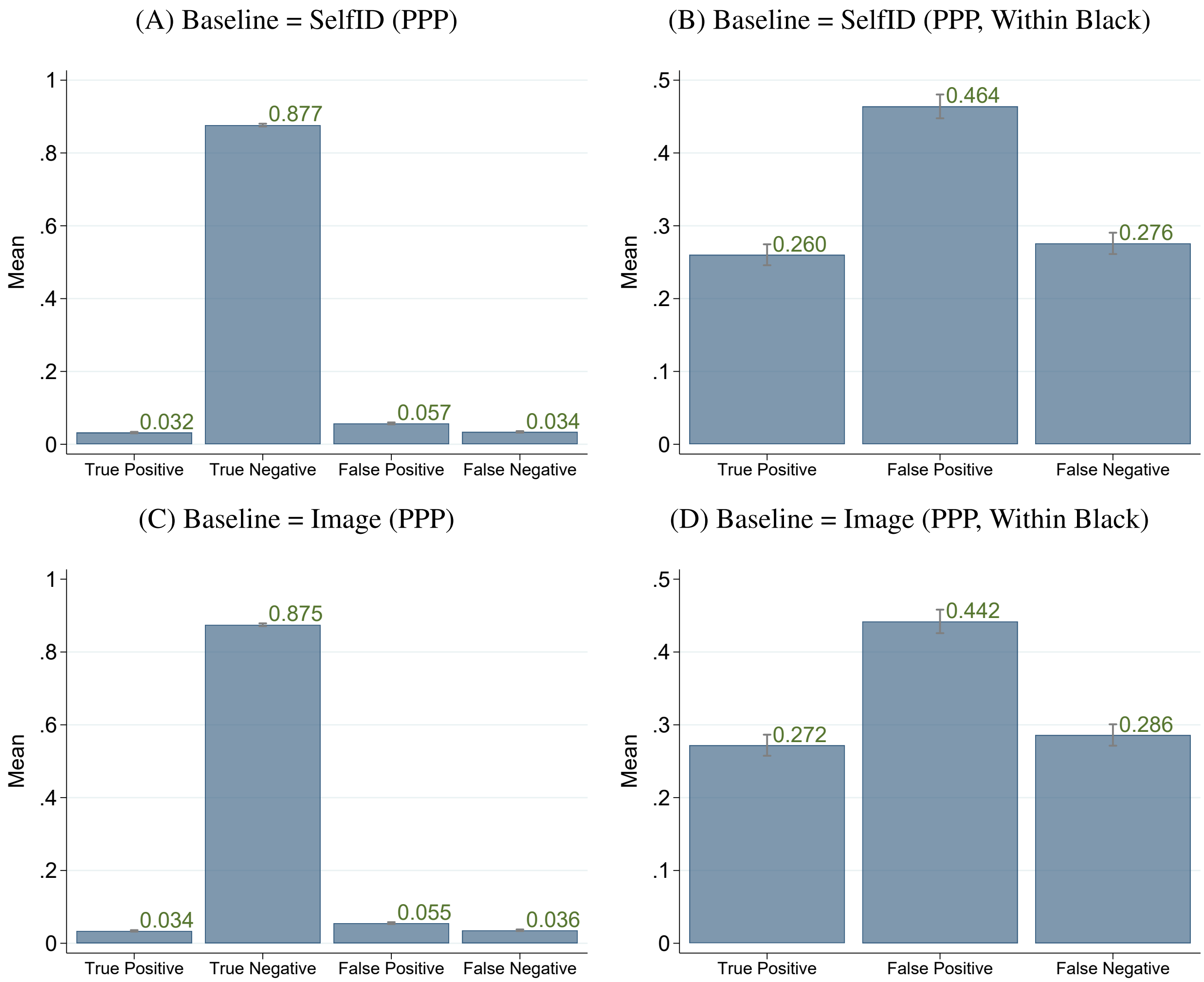
Note: This figure reports approval rates by classification group in the model's numerical example. Each panel displays the approval rates for a given classification of applicants over all three computed equilibria: no regulation, actual race, and predicted race. For the panel classifications, "True Positive" refers to applicants in Group B correctly classified by BISG to be in Group B, "False Postive" refers to applicants in Group A falsely classified by BISG to be in Group B, "False Negative" refers to applicants in Group B falsely classified by BISG to be in Group A, and "True Negative" refers to applicants in Group A correctly classified by BISG to be in Group A. For organization, the top and bottom rows of the figure correspond to applicants in Groups B and A, respectively, while the left and right columns correspond to applicants predicted by BISG as being in Groups B and A, respectively.

Figure 4: Borrower Race Variable Statistics (PPP, Unique Borrower–Level)



Note: This figure reports the mean of each race variable on the subsample of observations for which BISG is able to predict race, $N = 35,072$. The terms SelfID and image-based are binary variables. For example, 76% of our sample self-identifies as White. The continuous variable BISG Pct is defined as the actual percent chance the borrower is Black / White according to the BISG algorithm.

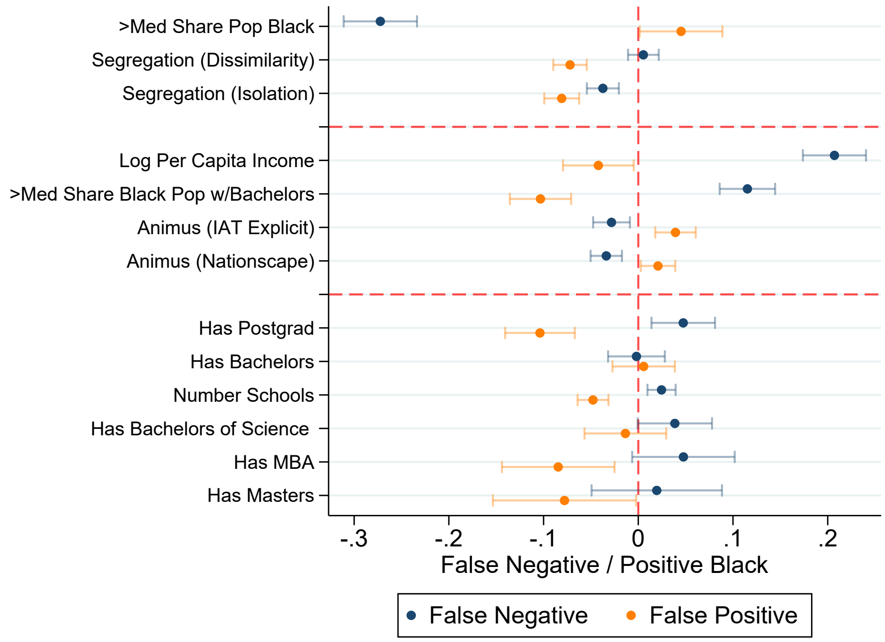
Figure 5: BISG Error Rates (Unique Borrower–Level)



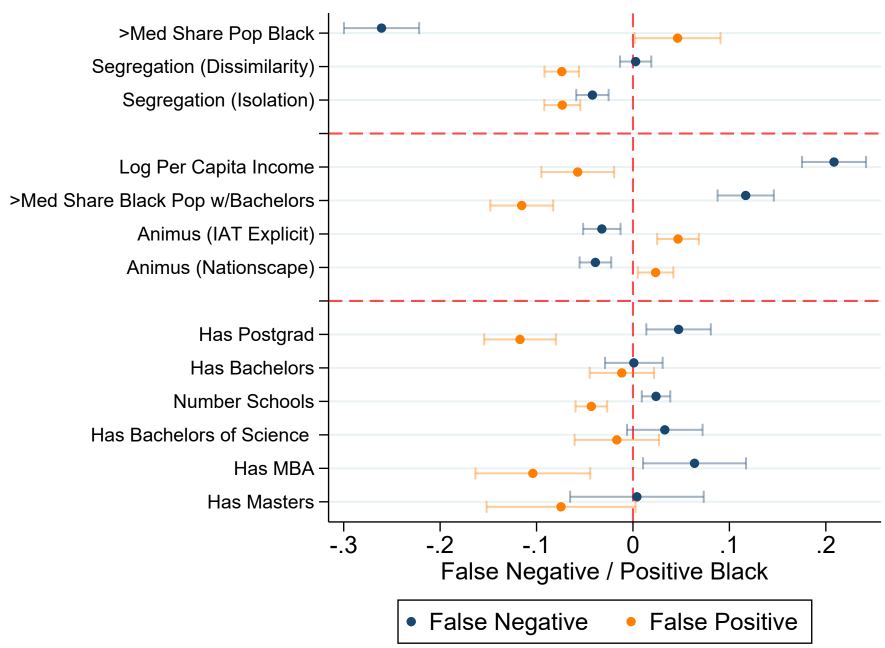
Note: This figure presents mean error types for observations where BISG can predict race. Each error type is defined relative to a benchmark "true race." For example, True Positive (SelfID) means that the business owner self-identifies as Black and is categorized by BISG as Black. True Positive (image-based) means both image-based and BISG measures classify the owner as Black. Panels A and B, containing all error types, have 35,072 observations. Panels C and D only include the "Within Black" sample, which is the subsample where the borrower is classified as Black by either the algorithm or by the "true race" measure. Panel C has 4,973 observations. Panel D has 4,882 observations.

Figure 6: Effect of Socioeconomic Covariates on BISG Errors (PPP)

(A) True Classification = Image-Based Race

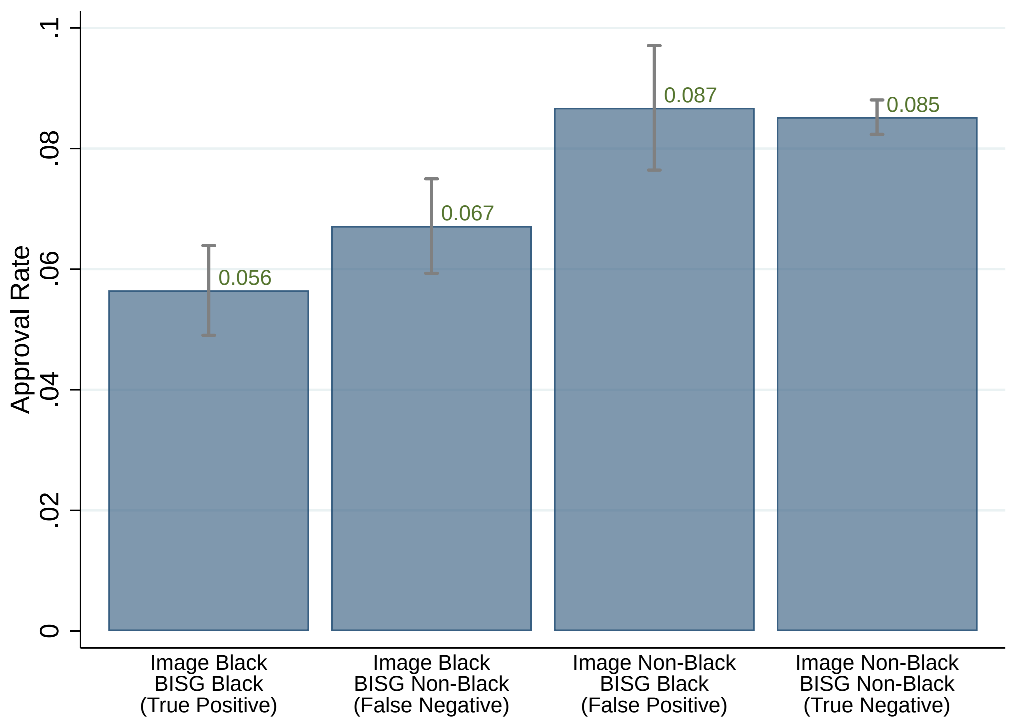


(B) True Classification = Self-Reported Race



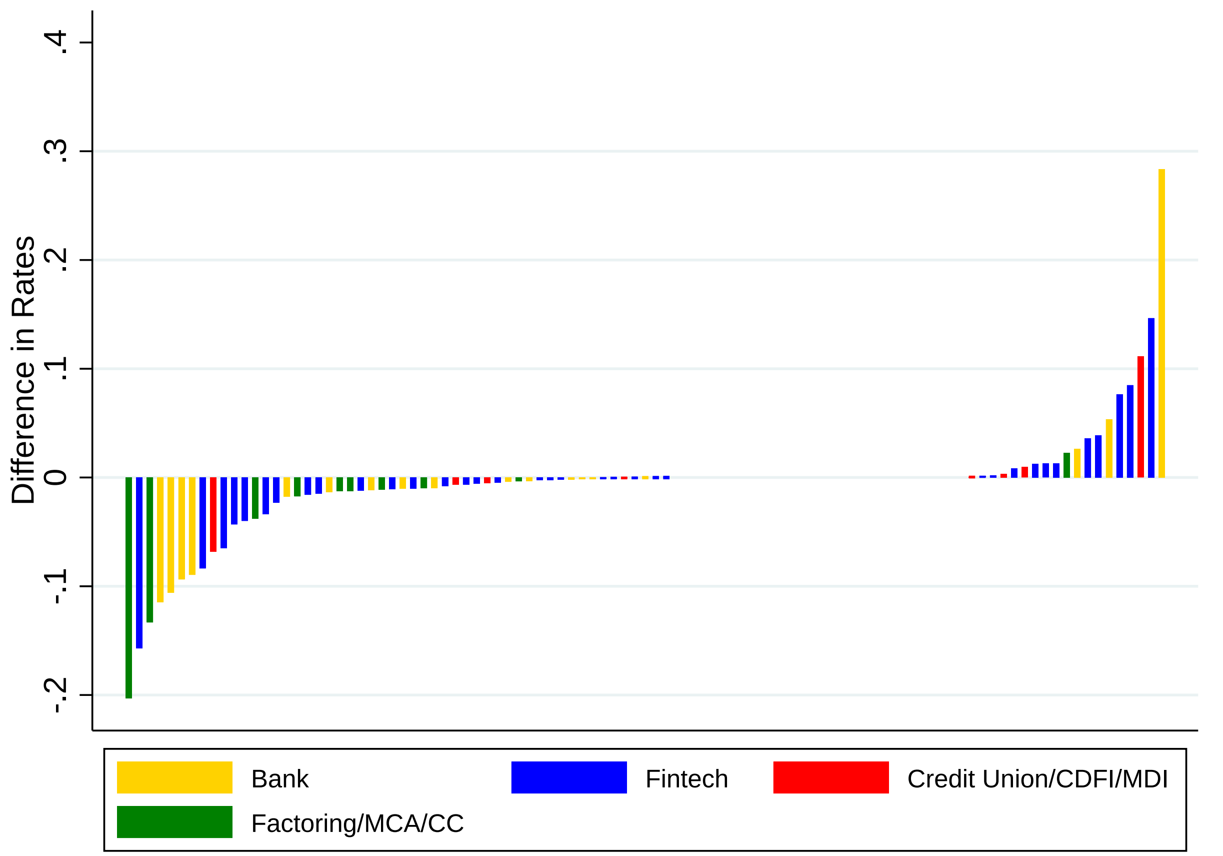
Note: This figure shows estimates of a set of regressions, each of either False Negative Black or False Positive Black on one of the socioeconomic characteristics in each row of the y-axis, using the "within Black" sample. The indicator variable False Positive Black is equal to one if BISG categorizes the borrower as Black, but our "correct" measure (either image-based or self-reported race) does not. The indicator variable False Negative Black is equal to one if BISG categorizes the borrower as non-Black, but the baseline measure categorizes them as Black. Panel A's sample includes borrowers classified as Black either by the algorithm or by the image-based measure ($N = 4,882$). Panel B's sample comprises borrowers who either self-identified as Black on their PPP loan application or were algorithmically classified as Black; $N = 4,973$.

Figure 7: Lendio Approval Rate by Group



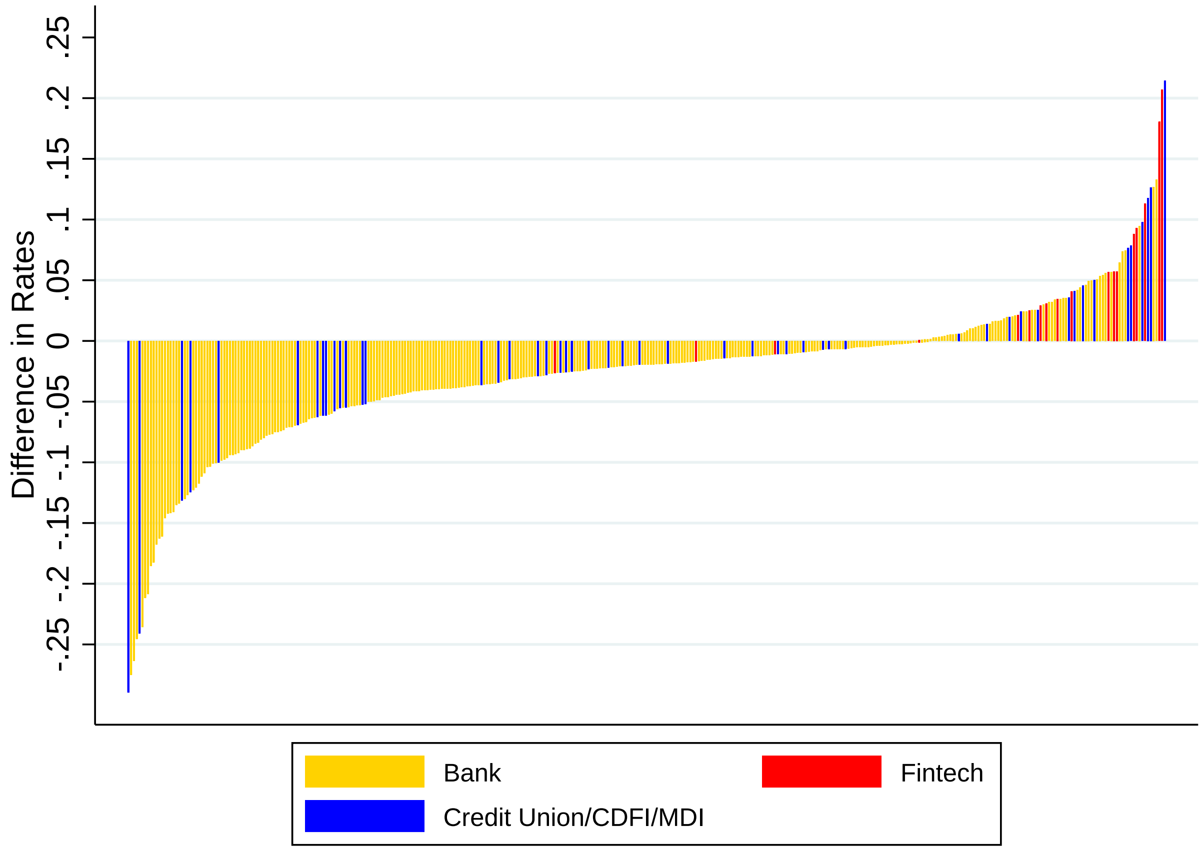
Note: This figure plots the approval rate by error type, focusing on a subset of 47,481 applicants for whom we could calculate job tenure and confidently determine loan rejection. We are able to calculate job tenure if the firm in the Lendio application matched a firm listed on the applicant's LinkedIn profile. We are confident the applicant was rejected if, for a given application date, their application was not approved by any lender.

Figure 8: Lender-Level Difference in Loan Approval Rates by Race Measure (Lendio)



Note: This figure plots, for each lender, the $\Delta_{\text{ShareApprovedBlack}}$. This is the difference in the approval rates among Black applicants between image-based and BISG measures, corresponding to $\pi_{i,j}^{AR} - \pi_{i,j}^{PR}$ in the model. Each bar represents $\Delta_{\text{ShareApprovedBlack}}$ for one of the 101 unique lenders in the Lendio analysis data. The bars are colored according to the lender type.

Figure 9: Lender-Level Difference in Lending Rates by Race Measure (PPP)



Note: This figure plots, for each lender, the $\Delta_{\text{ShareLoansBlack}}$. This is the difference in the lending rates to Black firm owners between image-based and BISG measures, corresponding to $\pi^{AR}_{i,j} - \pi^{PR}_{i,j}$ in the model. Each bar represents $\Delta_{\text{ShareLoansBlack}}$ for one of the 369 unique lenders in the PPP analysis data. The bars are colored according to the lender type.

Table 1: 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	103	438.087	39.000	957.677
Amount Funded	103	1,957,476	253,498	4,182,994
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

Note: This table reports loan application summary statistics focusing on a subset of applicants for whom we could calculate job tenure and confidently determine loan rejection. We are able to calculate job tenure if the firm in the Lendio application matched a firm listed on the applicant's LinkedIn profile. We are confident the applicant was rejected if, for a given application date, their application was not approved by any lender.

Table 2: 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

Note: This table reports loan summary statistics at the borrower (Panel A) and lender (Panel B) levels.

Table 3: Correlations between Race Variables (PPP)

	Black (SelfID)	Black (Image)
Black (Image)	0.87***	1.00
Black (BISG)	0.37***	0.38***
N = 28,990		
Black (Image)	0.87***	1.00
Black (BISG)	0.37***	0.38***
N = 28,990		
Black (Image)	0.87***	1.00
Black (Geography)	0.19***	0.21***
N = 28,994		
Black (Image)	0.87***	1.00
Black (Surname)	0.18***	0.19***
N = 29,002		
Black (Image)	0.87***	1.00
Black (Firstname+Surname)	0.25***	0.27***
N = 26,444		
Black (Image)	0.87***	1.00
Black (BIFSG)	0.41***	0.43***
N = 26,427		

Note: This table shows correlation coefficients between race variables. BISG Black Percent is the continuous probability of being Black from BISG, while all other variables are indicators. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 4: Lendio Application-Level Summary Statistics

	N	Mean	Median	SD
Applicants by Race:				
Apps from Black (Image)	47,504	0.160	0.000	0.367
Apps from White (Image)	47,504	0.700	1.000	0.458
Apps from Non-Black (Image)	47,504	0.840	1.000	0.367
Apps from Black (BISG)	47,504	0.138	0.000	0.345
Apps from White (BISG)	47,504	0.705	1.000	0.456
Apps from Non-Black (BISG)	47,504	0.862	1.000	0.345
Approvals Among Applicants of Race:				
Approved Black (Image)	7,618	0.062	0.000	0.241
Approved White (Image)	33,232	0.087	0.000	0.282
Approved Non-Black (Image)	39,886	0.085	0.000	0.279
Approved Black (BISG)	6,560	0.070	0.000	0.255
Approved White (BISG)	33,514	0.084	0.000	0.277
Approved Non-Black (BISG)	40,944	0.083	0.000	0.277
Loans Among Approved Borrowers:				
Approval Share Black (Image)	3,875	0.122	0.000	0.327
Approval Share White (Image)	3,875	0.746	1.000	0.435
Approval Share Non-Black (Image)	3,875	0.878	1.000	0.327
Approval Share Black (BISG)	3,875	0.118	0.000	0.323
Approval Share White (BISG)	3,875	0.726	1.000	0.446
Approval Share Non-Black (BISG)	3,875	0.882	1.000	0.323
Difference in Loans (Image Less BISG):				
Diff Loans Black	3,875	0.004	0.000	0.363
Diff Loans White	3,875	0.020	0.000	0.440
Diff Loans Non-Black	3,875	-0.004	0.000	0.363

Note: This table presents loan application summary statistics for the Lendio sample. The unit of observation is a loan application. The top group includes all applications. The second panel assesses the rate of approval among applicants predicted to be of a particular race. The third panel reports the share of approvals going to borrowers predicted to be of a particular race. The fourth panel reports the difference in loan shares between the different measures of race in the third panel.

Table 5: How Image and BISG Race Measures Predict Loan Approval

Dependent Variable:	Approved					
	(1)	(2)	(3)	(4)	(5)	(6)
Black (Image)	-0.018***		-0.016***	-0.016***		
	(0.005)		(0.005)	(0.005)		
Black (BISG)		-0.011**	-0.004	-0.002		
		(0.005)	(0.005)	(0.006)		
True Positive Black (BISG)					-0.023***	-0.021***
					(0.006)	(0.007)
False Positive Black (BISG)					-0.000	0.001
					(0.007)	(0.007)
False Negative Black (BISG)					-0.014**	-0.013**
					(0.006)	(0.006)
Observations	47,481	47,481	47,481	47,481	47,481	47,481
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			0.061	0.047	0.003	0.004
R-squared	0.073	0.073	0.073	0.075	0.073	0.075
Y-mean	0.08	0.08	0.08	0.08	0.08	0.08

Note: Columns (1) and (2) in this table provide estimates based on Equation (21), where the key independent variables indicate whether image-based race and BISG-based race classify the applicant as Black. In columns (3) and (4), we report the p-value on a one-tailed t-test testing whether the coefficient on image-based race is larger than the coefficient on BISG-based race. In columns (5) and (6), we report a similar P-value for whether the True Positive coefficient is significantly larger than the False Positive coefficient. Columns (3) and (4) decompose prediction errors by providing estimates based on Equation (21), using four indicators that describe how image-based race aligns with BISG-based race: True Positive, False Positive, False Negative, and True Negative Black. The base group True Negative is omitted. For instance, True Positive indicates that the business owner is classified as Black by both image-based and BISG-based measures. Standard errors are double-clustered by lender and borrower; ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 6: 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)	-0.018***	-0.017***			-0.019***	-0.018***		
	(0.005)	(0.005)			(0.005)	(0.005)		
Black (Algorithm)	0.001	0.003			0.004	0.004		
	(0.006)	(0.007)			(0.006)	(0.006)		
True Positive Black			-0.015*	-0.010			-0.023***	-0.022***
			(0.008)	(0.010)			(0.006)	(0.007)
False Positive Black			-0.001	0.002			0.010	0.011
			(0.008)	(0.009)			(0.008)	(0.008)
False Negative Black			-0.019***	-0.018***			-0.014***	-0.013**
			(0.005)	(0.005)			(0.005)	(0.006)
Observations	47,481	47,481	47,481	47,481	47,481	47,481	47,481	47,481
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	0.073	0.075	0.073	0.075	0.073	0.075	0.073	0.075
Y-mean	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08

Note: Columns (1) and (2) in this table provide estimates based on Equation 21, where the key independent variables indicate whether image-based race and zip-based race classify the applicant as Black. Columns (5) and (6) do the same using surname-based race. In Columns (1) and (2), as well as (5) and (6), we report the P-value on a one-tailed t-test testing whether the coefficient on image-based race is larger than the coefficient on zip or surname-based race, respectively. In Columns (3), (4), (7), and (8), we report a similar P-value on a one-tailed t-test for whether the True Positive coefficient is significantly larger than the False Positive coefficient. Columns (3), (4), (7), and (8) decompose prediction errors by estimating variants of Equation 21, using four indicators that describe how image-based race aligns with zip / surname-based race: True Positive, False Positive, False Negative, and True Negative Black. True Negative is the base group and is omitted. For instance, True Positive indicates that the business owner is classified as Black by both image-based and zip/surname-based measures. Standard errors are double-clustered by lender and borrower.***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 7: Lender-Level Summary Statistics (Lendio, one-per-lender)

Panel A: Lendio Approval Statistics by Race

	N	Mean	Median	SD
Share of Applicants by Race:				
Share Apps from Black (Image)	103	0.136	0.118	0.116
Share Apps from White (Image)	103	0.725	0.727	0.159
Share Apps from Black (BISG)	103	0.127	0.114	0.093
Share Apps from White (BISG)	103	0.699	0.696	0.139
Approval Rate Among Applicants of Race:				
Approval Rate Black (Image)	103	0.069	0.000	0.197
Approval Rate White (Image)	103	0.098	0.043	0.155
Approval Rate Black (BISG)	103	0.097	0.021	0.219
Approval Rate White (BISG)	103	0.104	0.044	0.165
Loan Rate Among Borrowers of Race:				
Share Loans to Black (Image)	103	0.091	0.000	0.188
Share Loans to White (Image)	103	0.557	0.714	0.391
Share Loans to Black (BISG)	103	0.074	0.043	0.108
Share Loans to White (BISG)	103	0.551	0.678	0.353
Difference in Rates by Race Measure (Image Less BISG):				
Diff Approval Rate Black	103	-0.006	-0.000	0.053
Diff Approval Rate White	103	-0.006	0.000	0.070
Diff Loan Rate Black	103	0.016	0.000	0.131
Diff Loan Rate White	103	0.006	0.020	0.164

Panel B: PPP Approval Statistics by Race

	N	Mean	Median	SD
Loan Rate Among Borrowers of Race:				
Share Loans to Black (Image)	369	0.053	0.025	0.080
Share Loans to White (Image)	369	0.849	0.889	0.145
Share Loans to Black (BISG)	369	0.080	0.054	0.076
Share Loans to White (BISG)	369	0.778	0.798	0.152
Difference in Rates by Race Measure (Image Less BISG):				
Diff Loan Rate Black	369	-0.027	-0.019	0.066
Diff Loan Rate White	369	0.071	0.059	0.090

Note: This table shows lender-level approval statistics by race. “Share Apps” is the proportion of applications from applicants identified as a particular race out of total applications received by that lender. “Approval Rate” is the fraction of applications approved from applicants of a specified race. “Share Loans” is the ratio of approved loans from borrowers of a certain race to the total number of approved loans. “Diff Approval Rate” is the difference between “Approval Rate” measured by image-based race and “Approval Rate” measured by BISG. Similarly, “Diff Loan Rate” is the difference between “Share Loans” measured by image-based race and “Share Loans” measured using BISG.

Table 8: 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	0.15	-0.00	0.12	0.59***	0.08***	0.64***
	(0.10)	(0.01)	(0.08)	(0.10)	(0.01)	(0.10)
Factoring/MCA/CC	0.02	-0.04*	-0.02			
	(0.15)	(0.02)	(0.13)			
Large Bank				0.22**	0.02	0.21**
				(0.10)	(0.02)	(0.11)
Medium Bank				-0.02	-0.01*	-0.00
				(0.05)	(0.01)	(0.05)
Credit Union/CDFI				0.15**	0.01	0.15*
				(0.08)	(0.01)	(0.08)
MDI				0.01	0.00	0.00
				(0.11)	(0.02)	(0.11)
Observations	94	94	94	368	368	368
R-squared	0.030	0.045	0.029	0.110	0.109	0.113
Y-mean	0.245	-0.006	0.170	0.234	-0.026	0.250

Note: Columns 1-3 show estimates of the association between lender type and percentiles of $\Delta_{\text{ShareApprovedBlack}}$, the difference in approval rate of Black applicants based on image-based race versus BISG-based race. We exclude Credit Unions and CDFIs due to their small representation. The omitted group is small banks. Columns 4-6 report estimates of how lender type is associated with percentiles of $\Delta_{\text{ShareLoansBlack}}$, the difference in the share of Black borrowers as determined by image-based race versus BISG-based race. Here too, the omitted group is small banks. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 9: Counterfactual Exercise

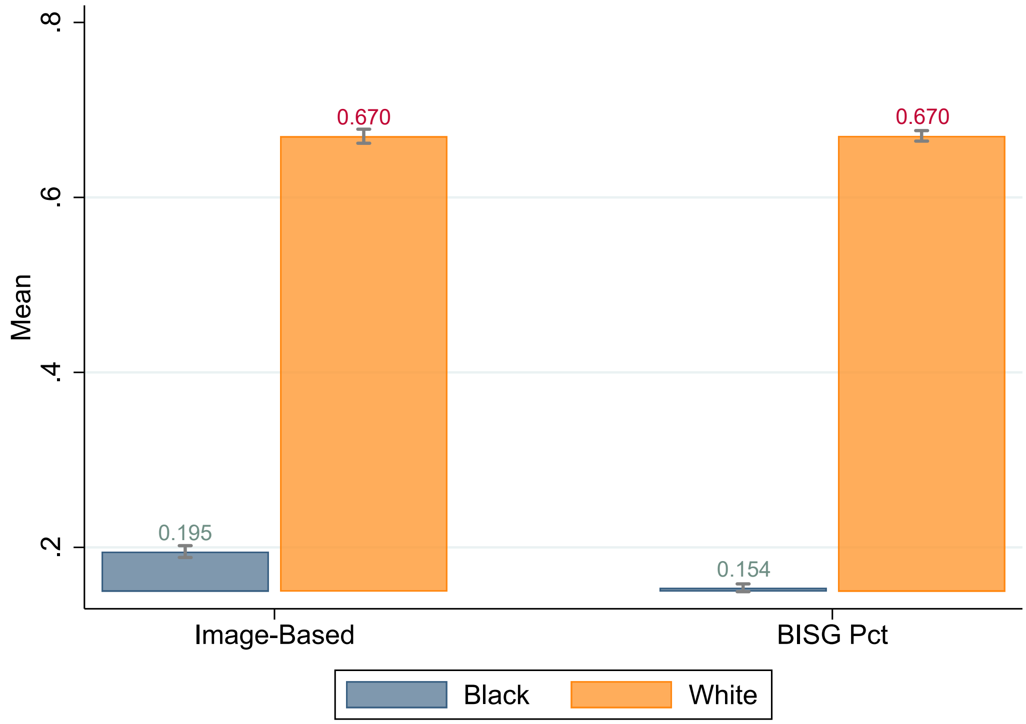
Characteristic	(1) BISG Weight ↓	(2) Image Weight ↑	(3) Net Change
Panel A: BISG + Image Classification Types			
BISG True Positive Black	-0.590	0.877	0.286
BISG True Negative Black	0.949	-1.728	-0.779
BISG False Positive Black	-0.198	-0.125	-0.323
BISG False Negative Black	-0.161	0.977	0.816
Image Black	-0.751	1.853	1.102
Panel B: Geographic Covariates			
Log Per Capita Income	0.395	-0.273	0.122
Share Pop Black	-0.412	0.369	-0.044
Share Pop w/ Bachelors	0.096	-0.061	0.034

Note: This table displays the results from the counterfactual exercise in Section 8. Specifically, each cell displays the derivative of the share of approved applications going to borrowers with a given characteristic with respect to a marginal decrease in the weight in our approval regression (21) on BISG (Column 1), a marginal increase in the weight on being classified as Black by our image-based measure (Column 2), and both marginal changes simultaneously (Column 3). See Section 8 for further details.

Appendix

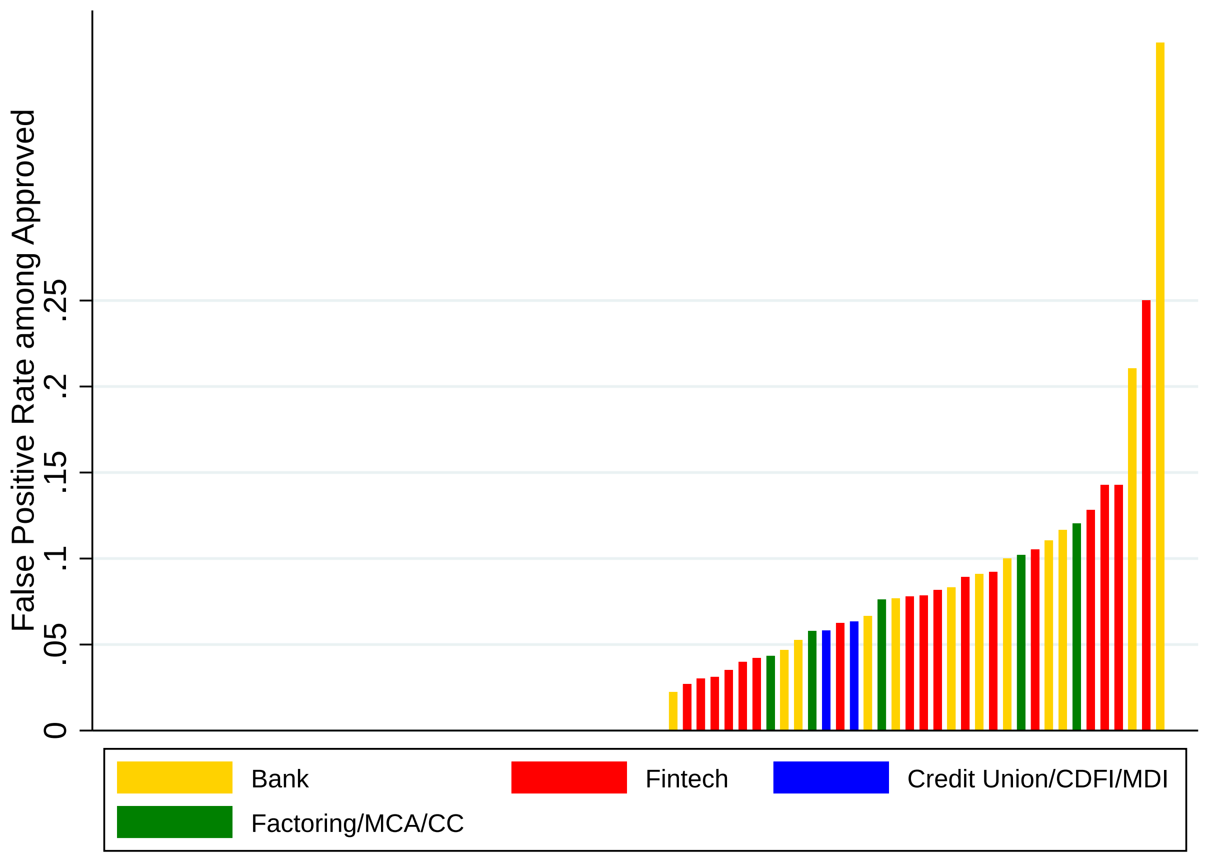
(For Online Publication)

Figure A.1: Borrower Race Variable Statistics (Lendio, Unique Applicant-Level)



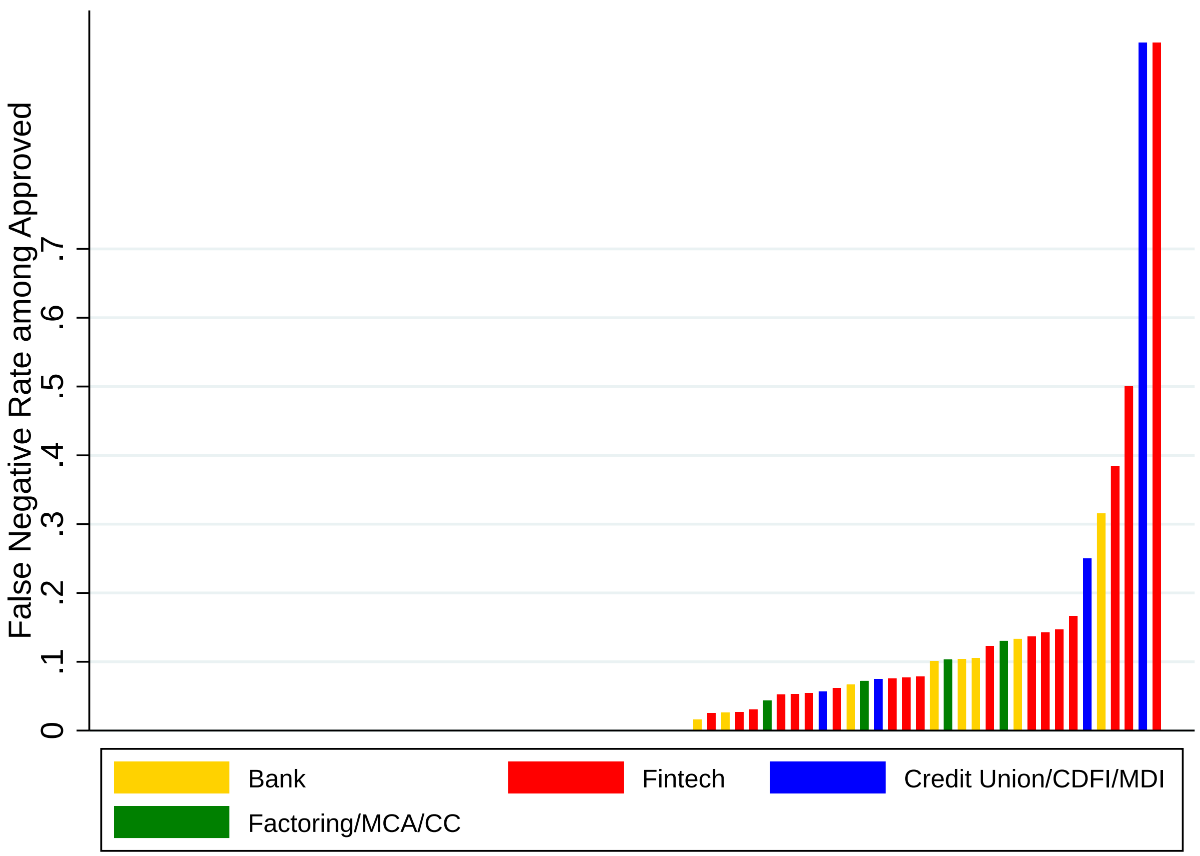
Note: This figure shows the mean of each race variable within the subsample of observations where BISG can predict race—instances where the borrower's last name matches one in the Census database of racial group frequency by last name. N = 13,172. Image-based is a binary variable; e.g., around 67% of our sample are White according to image-based race. BISG Pct is a continuous variable, reflecting the BISG algorithm's calculated probability that the borrower belongs to a specific racial group (Black, White, etc.).

Figure A.2: **Lender-level average false positive rate (Lendio)**



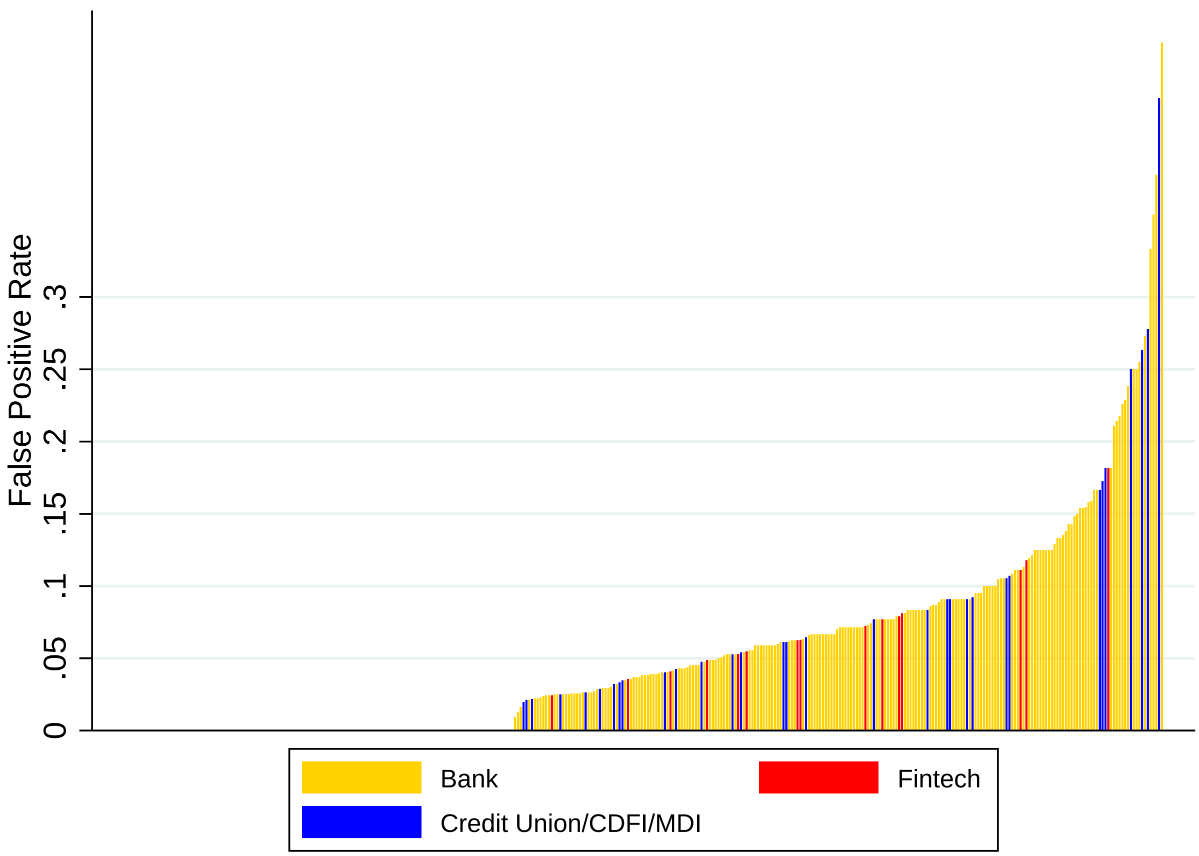
Note: This figure displays the average false positive rate for each lender among the applicants they approved. Each bar represents one unique lender with a non-zero false positive rate. Bars are color-coded according to lender type. Here, false positive refers to a borrower classified as Black by BISG, but identified as non-Black by our image-based algorithm.

Figure A.3: **Lender-level Average False Negative Rate (Lendio)**



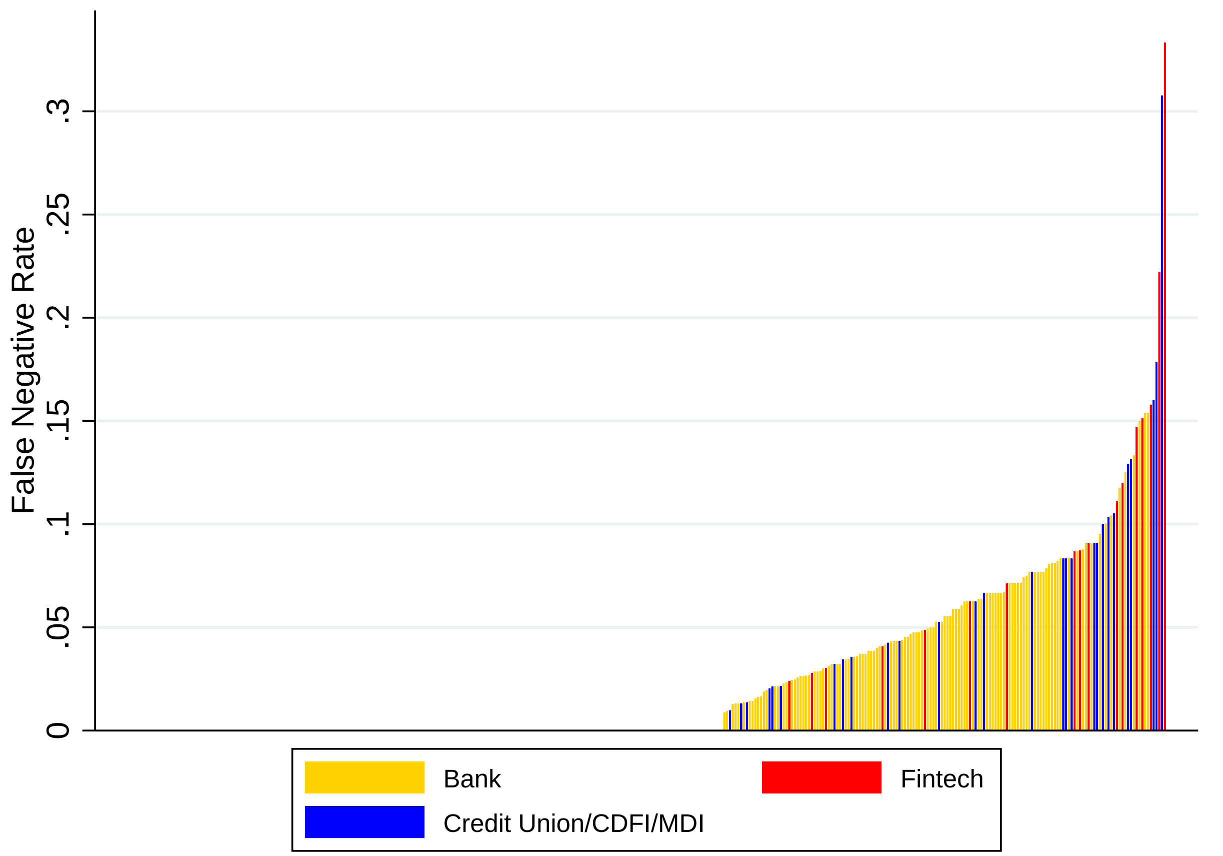
Note:
This figure displays the average false negative rate for each lender among the applicants they approved. Each bar represents one unique lender with a non-zero false negative rate. Bars are color-coded according to lender type. Here, false negative refers to a borrower classified as non-Black by BISG, but identified as Black by our image-based algorithm.

Figure A.4: **Lender-level Average False Positive Rate (PPP)**



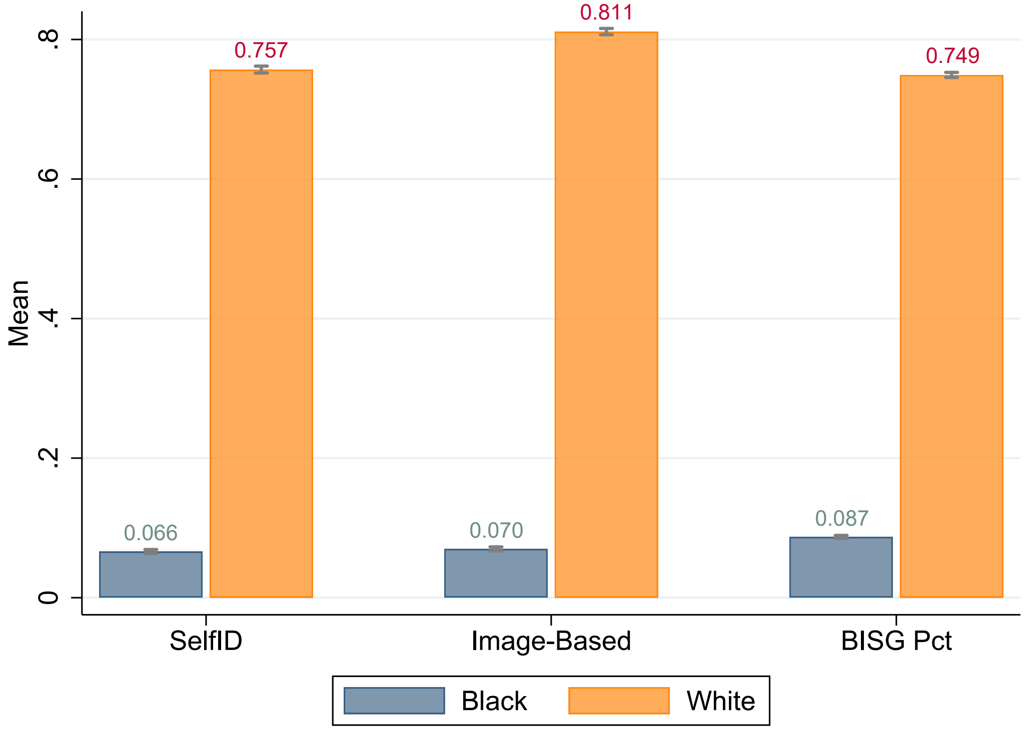
Note: This figure displays the average false positive rate for each lender among the loans they approved. Each bar represents one unique lender with a non-zero false positive rate. Bars are color-coded according to lender type. Here, false positive refers to a borrower classified as Black by BISG, but identified as non-Black by our image-based algorithm.

Figure A.5: **Lender-level Average False Negative Rate (PPP)**



Note: This figure displays the average false negative rate for each lender among the loans they approved. Each bar represents one unique lender with a non-zero false negative rate. Bars are color-coded according to lender type. Here, false negative refers to a borrower classified as non-Black by BISG, but identified as Black by our image-based algorithm.

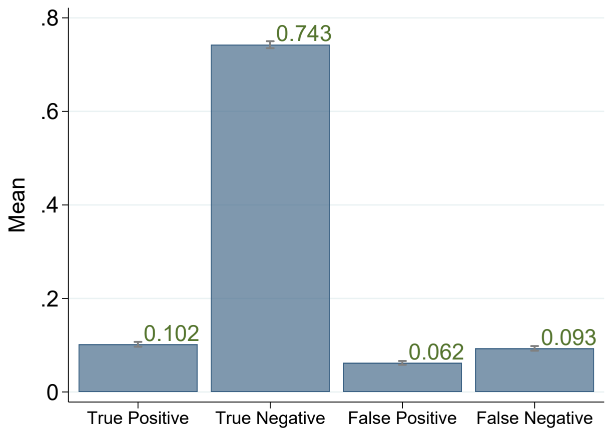
Figure A.6: Borrower Race Variable Statistics (PPP, Unique Borrower–Level)



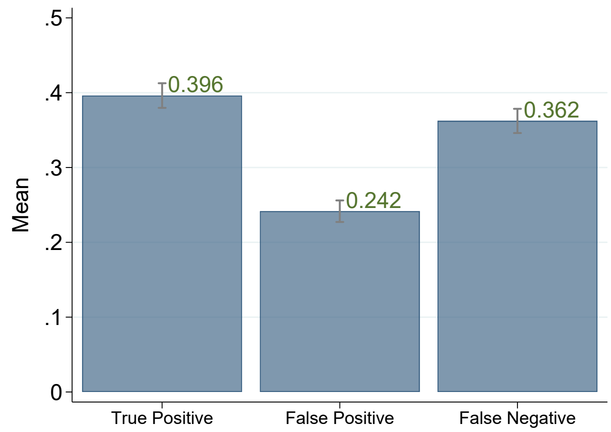
Note: This figure reports the mean of each race variable on the subsample of observations for which BISG is able to predict race, $N = 35,072$. The terms SelfID and image-based are binary variables. For example, 76% of our sample self-identifies as White. The continuous variable BISG Pct is defined as the actual percent chance the borrower is Black / White according to the BISG algorithm.

Figure A.7: BISG Error Rates Lendio Sample

(A) Baseline = Image (Lendio)



(B) Baseline = Image (Lendio, Within Black)



Note: This figure presents mean error types for observations where BISG can predict race. Each error type is defined relative to a benchmark of Image-based race. True Positive (image-based) means both image-based and BISG measures classify the owner as Black. Panels B only comprises the "Within Black" sample, which is the subsample where the borrower is classified as Black by either the algorithm or by the "true race" measure.

Table A.1: 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

Note: This figure shows the mean of each race variable within the subsample of observations where BISG can predict race—instances where the borrower's last name matches one in the Census database of racial group frequency by last name. "Image + BISG Black Population" is the subsample where the borrower is classified as Black by either the BISG algorithm or using our image-based measure.

Table A.2: 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

Note: This figure shows the mean of each race variable within the subsample of observations where BISG can predict race—instances where the borrower's last name matches one in the Census database of racial group frequency by last name. "Image + BISG Black Population" is the subsample where the borrower is classified as Black by either the BISG algorithm or using our image-based measure. "Self Identified + BISG Black Population" is the subsample where the borrower either self-identified as Black or the BISG algorithm classified the borrower as Black.

Table A.3: Correlations Between Race Variables (Lendio)

	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	

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

	Black (Image)
Black (Firstname+Surname)	0.36***
N = 55,362	

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

Note: This table shows correlation coefficients between different measures of race. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table A.4: 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)	0.04***												
	(0.01)												
Animus (Nationscape)		0.02**											
		(0.01)											
Above Median Share Pop Black			0.05**										
			(0.02)										
Log Per Capita Income				-0.04**									
				(0.02)									
Has Postgrad					-0.10***								
					(0.02)								
Has Bachelors						0.01							
						(0.02)							
Number Schools							-0.05***						
							(0.01)						
Has Bachelors of Science								-0.01					
								(0.02)					
Has MBA									-0.08***				
									(0.03)				
Has Masters										-0.08**			
										(0.04)			
Above Median Share Black Pop w/Bachelors											-0.10***		
											(0.02)		
Segregation (Dissimilarity)												-0.07***	
												(0.01)	
Segregation (Isolation)													-0.08***
													(0.01)
Observations	3,629	3,629	3,628	3,629	3,629	3,629	3,629	3,629	3,629	3,629	3,619	3,629	3,629
R-squared	0.004	0.001	0.001	0.001	0.008	0.000	0.009	0.000	0.002	0.001	0.011	0.017	0.020
Y-mean	0.442	0.442	0.442	0.442	0.442	0.442	0.442	0.442	0.442	0.442	0.442	0.442	0.442

Note: This table shows estimates of a regression of an indicator for being classified as a false positive borrower on each row variable using the “Within Black” sample. This sample includes borrowers classified as Black either by the algorithm or by the image-based measure. False Positive Black is 1 if BISG classifies the borrower as Black, but the image-based measure does not. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table A.5: PPP - Regressions with Image-based False Negative Black

Dependent Variable:	False Negative 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			-0.27***										
			(0.02)										
Log Per Capita Income				0.21***									
				(0.02)									
Has Postgrad					0.05***								
					(0.02)								
Has Bachelors						-0.00							
						(0.02)							
Number Schools							0.02***						
							(0.01)						
Has Bachelors of Science								0.04*					
								(0.02)					
Has MBA									0.05*				
									(0.03)				
Has Masters										0.02			
										(0.04)			
Above Median Share Black Pop w/Bachelors											0.12***		
											(0.01)		
Segregation (Dissimilarity)												0.01	
												(0.01)	
Segregation (Isolation)													-0.04***
													(0.01)
Observations	3,629	3,629	3,628	3,629	3,629	3,629	3,629	3,629	3,629	3,629	3,619	3,629	3,629
R-squared	0.002	0.004	0.050	0.039	0.002	0.000	0.003	0.001	0.001	0.000	0.016	0.000	0.005
Y-mean	0.286	0.286	0.286	0.286	0.286	0.286	0.286	0.286	0.286	0.286	0.286	0.286	0.286

Note:
This table shows estimates of a regression of an indicator for being classified as a false negative borrower on each row variable using the "Within Black" sample. This sample includes borrowers classified as Black either by the algorithm or by the image-based measure. False Negative Black is 1 if BISG classifies the borrower as non-Black, but the image-based measure classifies the borrower as Black. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table A.6: 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)	0.05***												
	(0.01)												
Animus (Nationscape)		0.02**											
		(0.01)											
Above Median Share Pop Black			0.05**										
			(0.02)										
Log Per Capita Income				-0.06***									
				(0.02)									
Has Postgrad					-0.12***								
					(0.02)								
Has Bachelors						-0.01							
						(0.02)							
Number Schools							-0.04***						
							(0.01)						
Has Bachelors of Science								-0.02					
								(0.02)					
Has MBA									-0.10***				
									(0.03)				
Has Masters										-0.07*			
										(0.04)			
Above Median Share Black Pop w/Bachelors											-0.12***		
											(0.02)		
Segregation (Dissimilarity)												-0.07***	
												(0.01)	
Segregation (Isolation)													-0.07***
													(0.01)
Observations	3,578	3,578	3,578	3,578	3,578	3,578	3,578	3,578	3,578	3,578	3,576	3,578	3,578
R-squared	0.005	0.002	0.001	0.002	0.011	0.000	0.007	0.000	0.003	0.001	0.013	0.018	0.016
Y-mean	0.464	0.464	0.464	0.464	0.464	0.464	0.464	0.464	0.464	0.464	0.464	0.464	0.464

Note:
This table shows estimates of a regression of an indicator for being classified as a false positive borrower on each row variable using the “Within Black” sample. This sample includes borrowers that either self-identify as Black or were classified as Black by the BISG algorithm. False Positive Black is 1 if BISG classifies the borrower as Black, but the borrower does not self-identify as Black. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table A.7: PPP - Regressions with SelfID-based False Negative Black

Dependent Variable:	False Negative Black												
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Animus (IAT Explicit)	-0.03***												
	(0.01)												
Animus (Nationscape)		-0.04***											
		(0.01)											
Above Median Share Pop Black			-0.26***										
			(0.02)										
Log Per Capita Income				0.21***									
				(0.02)									
Has Postgrad					0.05***								
					(0.02)								
Has Bachelors						0.00							
						(0.02)							
Number Schools							0.02***						
							(0.01)						
Has Bachelors of Science								0.03*					
								(0.02)					
Has MBA									0.06**				
									(0.03)				
Has Masters										0.00			
										(0.04)			
Above Median Share Black Pop w/Bachelors											0.12***		
											(0.01)		
Segregation (Dissimilarity)												0.00	
												(0.01)	
Segregation (Isolation)													-0.04***
													(0.01)
Observations	3,578	3,578	3,578	3,578	3,578	3,578	3,578	3,578	3,578	3,578	3,576	3,578	3,578
R-squared	0.003	0.006	0.046	0.041	0.002	0.000	0.003	0.001	0.002	0.000	0.017	0.000	0.007
Y-mean	0.276	0.276	0.276	0.276	0.276	0.276	0.276	0.276	0.276	0.276	0.276	0.276	0.276

Note: This table shows estimates of a regression of an indicator for being classified as a false negative borrower on each row variable using the “Within Black” sample. This sample includes borrowers that either self-identify as Black or were classified as Black by the BISG algorithm. False Negative Black is 1 if BISG classifies the borrower as non-Black, but the borrower self-identifies as Black. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table A.8: 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)	0.00												
	(0.01)												
Animus (Nationscape)		-0.01											
		(0.01)											
Above Median Share Pop Black			0.03										
			(0.02)										
Log Per Capita Income				-0.00									
				(0.02)									
Has Postgrad					0.00								
					(0.02)								
Has Bachelors						0.02							
						(0.01)							
Number Schools							-0.02*						
							(0.01)						
Has Bachelors of Science								0.05**					
								(0.03)					
Has MBA									-0.01				
									(0.03)				
Has Masters										-0.05			
										(0.04)			
Above Median Share Black Pop w/Bachelors											-0.04**		
											(0.01)		
Segregation (Dissimilarity)												-0.06***	
												(0.01)	
Segregation (Isolation)													-0.07***
													(0.01)
Observations	3,390	3,390	3,389	3,390	3,390	3,390	3,390	3,390	3,390	3,390	3,383	3,390	3,390
R-squared	0.000	0.000	0.001	0.000	0.000	0.001	0.001	0.001	0.000	0.000	0.002	0.017	0.016
Y-mean	0.242	0.242	0.242	0.242	0.242	0.242	0.242	0.242	0.242	0.242	0.242	0.242	0.242

Note: This table shows estimates of a regression of an indicator for being classified as a false positive borrower on each row variable using the "Within Black" sample. This sample includes borrowers classified as Black either by the algorithm or by the image-based measure. False Positive Black is 1 if BISG classifies the borrower as Black, but the image-based measure does not. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table A.9: Lendio - Regressions with Image-based False Negative Black

Dependent Variable:	False Negative Black												
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Animus (IAT Explicit)	-0.01												
	(0.01)												
Animus (Nationscape)		-0.01											
		(0.01)											
Above Median Share Pop Black			-0.37***										
			(0.02)										
Log Per Capita Income				0.29***									
				(0.02)									
Has Postgrad					-0.01								
					(0.02)								
Has Bachelors						-0.01							
						(0.02)							
Number Schools							0.01						
							(0.01)						
Has Bachelors of Science								-0.02					
								(0.03)					
Has MBA									-0.04				
									(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)													-0.09***
													(0.01)
Observations	3,390	3,390	3,389	3,390	3,390	3,390	3,390	3,390	3,390	3,390	3,383	3,390	3,390
R-squared	0.000	0.001	0.087	0.057	0.000	0.000	0.001	0.000	0.000	0.000	0.014	0.000	0.022
Y-mean	0.362	0.362	0.362	0.362	0.362	0.362	0.362	0.362	0.362	0.362	0.362	0.362	0.362

Note: This table shows estimates of a regression of an indicator for being classified as a false negative borrower on each row variable using the "Within Black" sample. This sample includes borrowers classified as Black either by the algorithm or by the image-based measure. False Negative Black is 1 if BISG classifies the borrower as non-Black, but the image-based measure classifies the borrower as Black. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table A.10: Lendio - Correlations Between Application and Approvals by Race

	Share Apps from Black (Image)
Share Apps from Black (Image)	1.00
Approval Rate Black (Image)	0.32***

	Share Apps from White (Image)
Share Apps from White (Image)	1.00
Approval Rate White (Image)	0.17*

	Share Apps from Black (BISG)
Share Apps from Black (BISG)	1.00
Approval Rate Black (BISG)	-0.03

	Share Apps from White (BISG)
Share Apps from White (BISG)	1.00
Approval Rate White (BISG)	0.18*

Note: This table shows correlation coefficients between application and approval rates at the lender level, N = 101. "Share Apps" is the ratio of applications from individuals identified as a particular race to the total applications received by the lender."Approval Rate" is the proportion of applications from applicants of a particular race that are approved. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table A.11: How Image and Firstname+Surname-based Race Measures Predict Loan Approval

Dependent Variable:	Approved			
	(1)	(2)	(3)	(4)
Black (Image)	-0.016***	-0.014**		
	(0.005)	(0.006)		
Black (Firstname+Surname)	-0.008	-0.009		
	(0.006)	(0.006)		
True Positive Black (Firstname+Surname)			-0.029***	-0.029***
			(0.007)	(0.007)
False Positive Black (Firstname+Surname)			-0.001	-0.001
			(0.009)	(0.009)
False Negative Black (Firstname+Surname)			-0.013**	-0.011*
			(0.006)	(0.006)
Observations	42,206	42,206	42,206	42,206
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	0.198	0.265	0.003	0.006
R-squared	0.073	0.074	0.073	0.075
Y-mean	0.08	0.08	0.08	0.08

Note: Columns (1) and (2) in this table report estimates variants of Equation 21, where key independent variables indicate whether the image-based race and Firstname+Surname-based race classify the applicant as Black. We report the P-value from a one-tailed t-test examining if the coefficient on image-based race is larger than the coefficient on Firstname+Surname-based race. Columns 3 and 4 decompose prediction errors by estimating variants of Equation 21, where the key independent variables are four indicators describing how image-based race aligns with Firstname+Surname-based race: True Positive, False Positive, False Negative, and True Negative Black, with True Negative serving as the base group and thus omitted. For example, True Positive means that both image-based and Firstname+Surname-based measures classify the business owner as Black. We report the P-value on a one-tailed t-test for whether the True Positive coefficient is significantly larger in magnitude than the False Positive coefficient. Standard errors are double-clustered by lender and borrower. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table A.12: 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

Note: Columns (1) and (2) of this table present variants of Equation 21, where the key independent variables indicate whether the image-based race and BISFG-based race categorize the applicant as Black. The P-value from a one-tailed t-test is reported, examining whether the coefficient on image-based race is larger than the coefficient on BISFG-based race. Columns (3) and (4) decompose prediction errors, estimating variants of Equation 21. These columns use four indicators that illustrate how image-based race aligns with BISFG-based race: True Positive, False Positive, False Negative, and True Negative Black, with True Negative serving as the base group and thus omitted. For instance, True Positive signifies that both image-based and BISFG-based measures classify the business owner as Black. We report the P-value from a one-tailed t-test for whether the True Positive coefficient is significantly larger in magnitude than the False Positive coefficient. Standard errors, in all cases, are double-clustered by lender and borrower. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table A.13: Predictive Power of All Race Measures on Loan Approval

Dependent Variable:			Approved			
	(1)	(2)	(3)	(4)	(5)	(6)
Black (Image)	-0.018***					
	(0.005)					
Black (BISG)		-0.011**				
		(0.005)				
Black (Geography)			-0.004			
			(0.006)			
Black (Surname)				-0.001		
				(0.005)		
Black (Firstname+Surname)					-0.015***	
					(0.006)	
Black (BIFSG)						-0.016***
						(0.006)
Observations	47,481	47,481	47,481	47,481	42,206	42,197
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	0.073	0.073	0.073	0.073	0.073	0.073
Y-mean	0.08	0.08	0.08	0.08	0.08	0.08

Note: Panel A of this table estimates variants of Equation 21, where the key independent variables are indicators for the various race measures. In both panels, standard errors are double-clustered by lender and borrower. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

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

A Supplemental Details on Data Sources and Construction

B Data Processing

1. Input

   (a) We collect from each LinkedIn profile a suite of variables, including full education and work histories. From a user's education history we are able to parse out school name, start and end date, degree name (such as "Bachelor's of Science"), and field of study (such as "Engineering"). A user's work history consists of a similar set of information: company name, start and end date, description (such as "First Grade Teacher"), and title (such as "CEO"). In our Lendio sample, around 49% of profiles have valid headshots. A further 81% of these images are usable for Deepface, which we use to create our image-based race. Additionally, a subset of companies have their own LinkedIn pages. If a borrower has ever worked at one of these companies, we collect its industry (such as "Construction"), employee count, year founded, and headquarters location.

1. Functions: Two functions process our scraped data.

   (a) The first one does two things: 1) calculate the amount of time the borrower spent in school, and 2) remove borrowers who report locations outside of the United States. Time spent in a school is simply calculated as the end date minus the start date. To determine if a borrower is outside of the United States, the scraped location string is first checked against a list of non-U.S. countries, and removed if it appears in that list. The remaining locations are checked against a list of U.S. states, counties, metropolitan areas, and cities downloaded from the Census Bureau.

   (b) The second one generates a set of binary variables describing the borrower's degree and current job. For each entry in the borrower's reported educational history, we determine if they obtained a bachelor's degree (and if applicable, what type, i.e. arts, business, or science), a master's degree (and if applicable, what type), a law degree, a degree in medicine, or any other post-graduate degree. This is done by checking the scraped degree string against a manually created list of keywords. For example, a degree string that contains any of "JD," "Juris," "Doctor of Law," etc. is coded as a law degree. Time in school is also normalized. That is to say, if a person spent eight years on their bachelor's, the duration is re-coded to four. At this stage, high schools are also dropped from a person's education history. For each person, we keep only their education/work experience data as of the time of the application. We again use a manually created list of keywords to determine if the person is a CEO or a founder.

1. Output

    (a) The output dataset is at the applicant level, where an applicant is a either an applicant for a PPP loan or an applicant in the Lendio data. For each applicant, we have a set of dummies describing their obtained degrees, their undergraduate school, and information on their current job, including company name, job title, and whether they are a CEO or founder, as of the year of the application. In addition, we calculate each borrower's years worked, number of jobs, number of degrees, whether they have a computer science or engineering degree, and their undergraduate school rank. This rank is described further in the School Prestige section.

B.1 Developing Measures of Race

B.2 Processing Images

This section provides more detail on how we classify applicants in the PPP and Lendio samples using images. We began this process by extracting facial embeddings using VGG-Face, a pre-trained image classifier wrapped in the *DeepFace* Python package developed by (Serengil and Ozpinar (2020)).

The VGG-Face model is a convolutional neural network trained on a large dataset of images of faces, along with their associated labels (such as race, gender, age, etc.). When the VGG-Face model is given an image of a face as input, it processes the image through a series of layers to extract features. These layers are organized in a specific way and trained to recognize certain patterns in the input data. For example, the first layer receives the raw pixel values of the image as input and passes them through a series of filters, detecting specific features (such as edges, shapes, and textures). The output of these filters is then passed through additional layers, which use more complex patterns to identify higher-level features like facial features and expressions.

As the data progresses through the layers, the detected features become more abstract and less specific to the raw pixel values of the input image. The final layer produces a fixed-length embedding vector as output, containing a compact representation of the detected features. This embedding vector can then be used as input to machine learning models for race prediction. To summarize, when given a new face image as input, the model processes the image through a series of layers, extracting features and ultimately producing a fixed-length embedding vector as output.

We use this embedding vector to train a random forest model on a dataset of around 170,000 images of venture-backed startup founders with known race information. This first-stage classification achieved an accuracy of 91%. We then applied this model to the combination of facial embeddings for our sample to obtain a preliminary classification.

Although experiments show that these algorithms outperform humans on facial recognition tasks (Phillips et al. (2018)), automated face recognition is not infallible and produces Type I and Type II errors, especially in classifying Black founders. It is worth noting that automated face recognition can result in false positives and false negatives, particularly for Black applicants photographed in very bright lighting or White applicants photographed in settings with very little lighting. At this stage, we also screen out typically Indian names using a name classifier, as dark-skin Indians are often mistakenly classified by the facial recognition software as Black. To address these potential errors, we conducted clerical reviews of all images using the applicant's LinkedIn profile (and other sources of public information) when we were unable to classify them based on the image alone. We classify each applicant as Black or non-Black. In the PPP sample, there were 706 cases where the applicant self-identified as Black, but are unlikely to be perceived as Black based on their profile image, names, and clerical review of their LinkedIn profile and publicly available information.

B.3 Race Prediction from Names

1. Input

   (a) Applicant first name, last name, and zip code of applicant.

2. Output

   (a) We use several demography-based methods to predict race. These methods take in a combination of first name, surname, and geography (zip code) in order to predict the race of an individual. Specifically, we use the Census Bureau's 2010 Last Name Model, the Census Bureau's 2010 Geocode Model, Bayesian Improved Surname Geocoding (BISG), Bayesian Improved First Name Surname Geocoding, and NamePrism.

   Each model produces the probability a borrower is of a certain race/ethnicity. The output dataset is a list of borrowers, the zip code of their business, and the probability they are a certain race given by each model.

B.3.1 BIFSG

We use a Python implementation of BIFSG from the Surgeo library. BIFSG takes as inputs a zip code, first name, and surname. Surname and zip code probabilities (specifically, the percentage of a particular race that lives within the zip code, for example, 2% of all White citizens live in 27106) come from 2010 Census Bureau data, and first name probabilities come from a Harvard University dataset. The probability of a borrower being a certain race is given as

$$q(r|s,f,g) = \frac{P(r|s) \times P(g|r) \times P(f|r)}{\sum_{r=1}^{6} P(r|s) \times P(g|r) \times P(f|r)},$$

where r is a race/ethnicity (Hispanic, White, Black, Asian or Pacific Islander, American Indian / Alaska Native, Multiracial), s is a surname, and g is a zip code.

B.3.2 BISG

Our BISG probabilities also come from Surgeo. BISG takes as inputs a zip code and a surname, using the same files as in BIFSG. The probability of a borrower being a certain race is given as

$$q(i|s,g) = \frac{P(r|s) \times P(g|r)}{\sum_{r=1}^{6} P(r|s) \times P(g|r)}$$

Where r is a race/ethnicity (Hispanic, White, Black, Asian or Pacific Islander, American Indian / Alaska Native, Multiracial), s is a surname, g is a zip code, $P(i|j)$ is the probability of a selected race given surname, and $r(k|i)$ is the probability of a selected zip code given race.

B.3.3 Geocode Model

Our geocode probabilities come from Surgeo, which in turn uses the probability of a race given zip code file from the Census Bureau. That is to say, the file describes the racial makeup of a zip code, for example, 90% of 27106 is White.

B.3.4 Last Name Model

For the last name model, we use an implementation from a Python packaged called ethnicolr. It takes as an input the surname of the borrower. Internally, it uses a 2010 file from the Census Bureau that describes the surname and the probabilities that the surname refers to a Black, White, Asian Pacific Islander, Asian, Hispanic, or multiracial person.

B.3.5 NamePrism

NamePrism is a web service based on Ye et al. (2017) and Ye and Skiena (2019). It takes as inputs a first name and a surname, and returns the probabilities the person is Black, White, Asian Pacific Islander, Asian, Hispanic, or multiracial.

B.4 Variable Definitions

B.4.1 Zip code level characteristics

1. **ACS variables:** We collect per capita income, educational attainment, and total population by race (Asian, Black, Native, Pacific Islander, White, Other) at the zip code level. Data is sourced from the 2020 American Community Survey (ACS) 5-year tables. Since ACS data uses zip code tabulation areas (a Census-specific analogue to zip code), we use a crosswalk provided by Uniform Data System to map ACS data to our zip codes. Educational attainment is broken up into less than high school, high school or equivalency, some college or associates degree, and bachelor's degree or higher.

2. **Zip-code level racial animus:** IAT, Nationscape, Racially charged searches, Segregation

B.4.2 LinkedIn/ACS characteristics

1. Number of Jobs

2. Number of unique jobs

3. Number of years of work experience

4. Has Bachelors

5. Has Masters

6. Has MBA

7. Has Ph.D/JD/MD

8. Number of degrees

9. Has computer science or engineering degree

10. Zip Income: `percap_inc` for total, White, black, asian, etc.

11. Zip fraction bachelors: `ge_ba` for total, White, black, asian, etc.

12. Zip high school or less: `l_hs` for total, White, black, asian, etc.

13. Amount of PPP Funding

C Model Derivations

No Regulation Equilibrium. Taking the derivative of (3) with respect to $\bar{\varepsilon}_j$ yields

$$-s_j(\mu_j - \bar{\varepsilon}_j) f_\varepsilon(\bar{\varepsilon}_j) = 0$$

from which we can immediately obtain (4). For the approval rate, our uniform distribution assumption implies

$$\begin{aligned}
\pi_i^{NR} &= F_\varepsilon(\bar{\varepsilon}_j^{NR}) = F_\varepsilon(\mu_j) = \gamma_0 + \gamma_1 \mu_j \\
&= \gamma_0 + \gamma_1 \mu_A + \gamma_1(\mu_j - \mu_A) \\
&= \gamma_0 + \gamma_1 \mu_A + \gamma_1(\mu_B - \mu_A) B_i \\
&= \text{const} + \gamma_1(\mu_B - \mu_A) B_i.
\end{aligned}$$

which completes the derivation of (5).

Regulatory Constraint on Actual Race. The Lagrangian for the lender's constrained optimization problem is now

$$\mathcal{L} = \max_{\bar{\varepsilon}_A, \bar{\varepsilon}_B} \sum_{j \in \{A,B\}} s_j \int^{\bar{\varepsilon}_j} (\mu_j - \varepsilon)\, dF_\varepsilon(\varepsilon) + \lambda\Big(\kappa - F_\varepsilon(\bar{\varepsilon}_A) + F_\varepsilon(\bar{\varepsilon}_B)\Big).$$

The first order conditions are:

$$\begin{aligned}
(\bar{\varepsilon}_A): &\qquad 0 = s_A(\mu_A - \bar{\varepsilon}_A) f_\varepsilon(\bar{\varepsilon}_A) - \lambda f_\varepsilon(\bar{\varepsilon}_A) \\
(\bar{\varepsilon}_B): &\qquad 0 = s_B(\mu_B - \bar{\varepsilon}_B) f_\varepsilon(\bar{\varepsilon}_B) + \lambda f_\varepsilon(\bar{\varepsilon}_B).
\end{aligned}$$

Solving, we obtain (7) and (8).

For the approval rate, we have

$$\begin{aligned}
\pi_i^{AR} &= F_\varepsilon(\bar{\varepsilon}_j) = F_\varepsilon(\bar{\varepsilon}_A)(1 - B_i) + F_\varepsilon(\bar{\varepsilon}_B) B_i \\
&= \left[\gamma_0 + \gamma_1\left(\mu_A - \frac{\lambda}{s_A}\right)\right](1 - B_i) + \left[\gamma_0 + \gamma_1\left(\mu_B + \frac{\lambda^{AR}}{s_B}\right)\right] B_i \\
&= \gamma_0 + \gamma_1\left(\mu_A - \frac{\lambda^{AR}}{s_A}\right) + \gamma_1\Big((\mu_B - \mu_A) + \lambda^{AR}(s_B^{-1} - s_A^{-1})\Big)
\end{aligned}$$

which completes the derivation of (9).

Regulatory Constraint on Predicted Race. To derive the constraint, substitute for L_j in (12) to obtain

$$s_A^{-1} \sum_{j\in\{A,B\}} s_j \int (1-q) F_\varepsilon(\bar{\varepsilon}_j(q))\, dF_{q,j}(q) - s_B^{-1} \sum_{j\in\{A,B\}} s_j \int q F_\varepsilon(\bar{\varepsilon}_j(q))\, dF_{q,j}(q) \le \kappa.$$

Moving the s_A^{-1} term inside the first integral and the s_B^{-1} term inside the second integral yields

$$\sum_{j\in\{A,B\}} s_j \int \frac{(1-q)}{s_A} F_\varepsilon(\bar{\varepsilon}_j(q))\, dF_{q,j}(q) - \sum_{j\in\{A,B\}} s_j \int \frac{q}{s_B} F_\varepsilon(\bar{\varepsilon}_j(q))\, dF_{q,j}(q) \le \kappa.$$

Combining the two integrals together now delivers (13).

The Lagrangian for the lender's constrained optimization is

$$\begin{aligned}\mathcal{L} = \max_{\bar{\varepsilon}_A(q),\bar{\varepsilon}_B(q)} & \sum_{j\in\{A,B\}} s_j \int \int^{\bar{\varepsilon}_j(q)} (\mu_j - \varepsilon)\; dF_\varepsilon(\varepsilon)\, dF_{q,j} \\ & + \lambda \left\{ \kappa - \sum_{j\in\{A,B\}} s_j \int \left[\frac{1-q}{s_A} - \frac{q}{s_B}\right] F_\varepsilon(\bar{\varepsilon}_j(q))\, dF_{q,j}(q) \right\}.\end{aligned}$$

The optimality condition with respect to $\bar{\varepsilon}_j(q)$ is

$$0 = s_j(\mu_j - \bar{\varepsilon}_j(q)) f_\varepsilon(\bar{\varepsilon}_j(q)) f_{q,j}(q) - \lambda s_j \left[\frac{1-q}{s_A} - \frac{q}{s_B}\right] f_\varepsilon(\bar{\varepsilon}_j(q)) f_{q,j}(q).$$

Canceling terms yields

$$0 = (\mu_j - \bar{\varepsilon}_j(q)) - \lambda \left[\frac{1-q}{s_A} - \frac{q}{s_B}\right].$$

which can be solved to deliver (15).

For the probability of approval, we have

$$\begin{aligned}\pi_i^{PR} &= F_\varepsilon(\bar{\varepsilon}_j(q)) = \gamma_0 + \gamma_1 \bar{\varepsilon}_j(q) \\ &= \gamma_0 + \gamma_1 \left\{ \mu_j + \lambda^{PR} \left[\frac{q_i}{s_B} - \frac{1-q_i}{s_A}\right] \right\} \\ &= \gamma_0 + \gamma_1 \mu_A + \gamma_1(\mu_B - \mu_A) B_i + \gamma_1 \lambda^{PR} \left[\frac{q_i}{s_B} - \frac{1-q_i}{s_A}\right] \\ &= \gamma_0 + \gamma_1 \mu_A - \gamma_1 \lambda^{PR} s_A^{-1} + \gamma_1(\mu_B - \mu_A) B_i + \gamma_1 \lambda^{PR} (s_B^{-1} - s_A^{-1}) q_i\end{aligned}$$

which derives (16).

Counterfactual Experiment Assume that approval policy follows a regression

$$Approved_i = \beta X_i + \varepsilon_i$$

Consider a counterfactual approval policy that tilts this policy with respect to the jth component of X_i while holding the overall approval rate constant:

$$Approved_i^C = Approved_i + h(X_{ij} - \bar{X}_j)$$

for some constant h, and where $\bar{X}_j$ is the sample mean of X_{ij}. Now consider some characteristic Z_i. The share of the approved population with characteristic Z_i (or, alternatively, the average of Z_i in that population) is

$$ApprovedShare_Z^C = \frac{\sum_i Z_i Approved_i^C}{\sum_i Approved_i^C} = \frac{\sum_i Z_i Approved_i^C}{\sum_i Approved_i}$$

where the second step follows from the fact that

$$\sum_i Approved_i^C = \sum_i Approved_i + h\sum_i (X_{ij} - \bar{X}_j) = \sum_i Approved_i.$$

Taking the difference with respect to

$$ApprovedShare_Z = \frac{\sum_i Z_i Approved_i}{\sum_i Approved_i}$$

we obtain

$$\begin{aligned}
ApprovedShare_Z^C - ApprovedShare_Z &= \frac{\sum_i Z_i(Approved_i^C - Approved_i)}{\sum_i Approved_i} \\
&= \frac{h \times \sum_i Z_i(X_{ij} - \bar{X}_j)}{\sum_i Approved_i} \\
&= h \times \frac{\sum_i (Z_i - \bar{Z})(X_{ij} - \bar{X}_j)}{\sum_i Approved_i} \\
&= h \times \frac{\text{Cov}(Z_i, X_{ij})}{\overline{Approved}}
\end{aligned}$$

where $\bar{Z}$ is the sample mean of Z_i, $\overline{Approved}$ is the sample mean of $Approved_i$, $\text{Cov}(Z_i, X_{ij})$ is the sample covariance between Z_i and X_{ij}, and where the third step follows from

$$\sum_i \bar{Z}(X_{ij} - \bar{X}_j) = \bar{Z}\sum_i X_{ij} - \bar{Z}\sum_i \bar{X}_j = 0.$$

Dividing through by h now yields (27).