Journal of Financial Economics 157 (2024) 103857

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/finec

Regulatory arbitrage or random errors? Implications of race prediction algorithms in fair lending analysis[☆]

Daniel L. Greenwald [a], Sabrina T. Howell [a,*], Cangyuan Li [a], Emmanuel Yimfor [b]

[a] Stern School of Business, New York University, United States of America
[b] Columbia Business School, Columbia University, United States of America

ARTICLE INFO

Dataset link: https://data.mendeley.com/datasets/g547vfzybp/2

JEL classification:
G21
G23
G28
J15
C81

Keywords:
Racial disparities
Disparate impact analysis
Race-conscious policies
Race prediction
Small business lending
BISG

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.

There are many high-stakes contexts in which regulators, firms, administrators, and researchers do not directly observe race and instead rely on proxies. 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.[1] 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.[2] Small business lenders do not collect and report applicant race, unlike home mortgage lenders. This setting 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

☆ Dimitris Papanikolaou was the editor for this paper, and we are grateful for his comments. We thank seminar and conference participants at the University of Michigan, MFA 2024, SITE 2023, FOM 2024, MIT Sloan, the U.S. Census Bureau Workshop on Advancing Research on Race, Ethnicity, and Inequality, and the J.H. Carey 2024 conference. We would also like to thank Matt Denes, Heath Witzen, Michaela Pagel, and Scott Nelson, for their feedback. Howell and Yimfor's work on this project is funded by the Alfred P. Sloan Foundation, United States (G-2022-19451).

* Corresponding author.
E-mail addresses: daniel.greenwald@stern.nyu.edu (D.L. Greenwald), sabrina.howell@nyu.edu (S.T. Howell), cl4220@stern.nyu.edu (C. Li), emmanuel.yimfor@columbia.edu (E. Yimfor).

1 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.

2 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".

https://doi.org/10.1016/j.jfineco.2024.103857
Received 30 August 2023; Received in revised form 22 April 2024; Accepted 23 April 2024
Available online 10 May 2024

data.[3] 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.[4]

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.

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.[5] 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, non-mortgage 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.

[3] 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, but several months later it was indefinitely stayed after a lawsuit from a banking association (see here and here).

[4] 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.

[5] 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.

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 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.[6] 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 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 0.54 for BISG and self-reported indicator for being Black. This correlation is somewhat lower than analogous ones reported in other studies, which usually revolve around 70%. We show that this is a direct consequence of BISG being a biased measure of race in the PPP sample because the share of Black individuals in the PPP sample is lower than that in the population upon which BISG is calibrated.

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

[6] 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.

for false negatives. The results are robust to using self-identified 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.[7] 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.

We also show that BISG errors could be influential in determining relative rankings and the identities of lenders most likely to face investigation. Suppose regulators use raw approval rate gaps to rank lenders according to their Black/non-Black approval gaps and then investigate the lenders with the largest gap. How often would the composition of high-gap lenders change based on how race is measured? As one example, if we consider the 10% of lenders with the highest Black/non-Black approval gaps, 40% of them would change their identity if we measured the gap using image-based race rather than BISG-implied race.

In both Lendio and PPP data (where we use loan shares to Black borrowers rather than approval 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., 2024). 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, within the population of Black borrowers, 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.

There are many settings besides lending to which our results are relevant. For example, university admissions officers may seek to evaluate the composition of the incoming class without collecting race data after the 2022 Supreme Court ruling in Students for Fair Admission. Another example is that 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, and can take steps to evaluate whether algorithmic biases might affect their findings.[8]

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, 2024; 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., 2024). To our knowledge, this project is the first effort to focus explicitly on compliance with fair lending laws in small business finance.

[7] 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.

[8] 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. (2024).

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 Appendix B 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 First-name 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 4250 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.

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 (Cook et al., 2022). The model achieved 91% accuracy in a hold-out sample.[13] 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

[9] See the 2010 Decennial Census.

[10] We set a seed for this random assignment to ensure replicability.

[11] See here.

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

[13] Given that about 3 percent of founders in venture-backed startups are Black, a model predicting that "all founders of venture-backed startups are white" would be approximately 97% accurate. However, in the model we implemented, such a prediction would be correct less than 50% of the time. Why? To address the under-representation of Black founders in our dataset, we sampled with replacement. This ensures that our training data has a balanced representation of Black and non-Black founders. For example, suppose we had images of 100 founders with 96 being non-Black and 4 being Black, we would re-sample the images of the Black founders. During this re-sampling, we introduced minor modifications to each image, such as cropping the edges or adjusting the size, so that the resulting images, while similar, were not identical. Consequently, in this hypothetical example we would have a training dataset with 192 images (96 X 2), where half were of Black founders. While the model's initial classification aids in expediting the review process, it is essential to note that we put all the images through a manual verification process. The model serves to facilitate this procedure rather than determining the ultimate outcome.

false negatives, particularly for Black applicants photographed in very bright lighting or White applicants photographed in settings with very little lighting.[14] 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 without 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.

To motivate the linear probability regressions we will run later, we assume the uniform distribution $\varepsilon_i \sim U[0,1]$, where we note that the choice to bound the support between 0 and 1 is without loss of generality, as long as we rescale μ_j accordingly (see Appendix D.2). This parameterization implies the CDF $F_\varepsilon(\varepsilon_i) = \varepsilon_i$, and the PDF $f_\varepsilon(\varepsilon_i) = 1$ on the support $[0,1]$.

No regulation equilibrium. It is clear from Eq. (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

$$V = \sum_{j\in\{A,B\}} s_j \int^{\bar{\varepsilon}_j} \left(\mu_j - \varepsilon\right) dF_\varepsilon(\varepsilon), \tag{2}$$

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

$$\bar{\varepsilon}_j^{NR} = \mu_j, \tag{3}$$

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, and apply our parametric assumption $\varepsilon_i \sim U[0,1]$, we obtain

$$\pi_i^{NR} = \mu_A + \underbrace{(\mu_B - \mu_A)}_{<0} B_i, \tag{4}$$

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

[14] 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.

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 $\kappa \geq 0$:

$$F_\varepsilon(\bar{\varepsilon}_A) - F_\varepsilon(\bar{\varepsilon}_B) \leq \kappa. \tag{5}$$

The lender's problem now is to maximize (2) subject to (5). The optimal policy is to approve loans for borrowers of type j with $\varepsilon_{i,j} < \bar{\varepsilon}_j^{AR}$, where

$$\bar{\varepsilon}_A^{AR} = \mu_A - \frac{\lambda^{AR}}{s_A}, \tag{6}$$

$$\bar{\varepsilon}_B^{AR} = \mu_B + \frac{\lambda^{AR}}{s_B}. \tag{7}$$

Where λ^{AR} is the multiplier on constraint (5) and the superscript AR stands for "actual race". These expressions nest (2) 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

$$\pi_i^{AR} = \underbrace{\left(\mu_A - \frac{\lambda^{AR}}{s_A}\right)}_{\text{const}} + \left[\underbrace{(\mu_B - \mu_A)}_{<0} + \underbrace{\lambda^{AR}(s_A^{-1} + s_B^{-1})}_{>0}\right] B_i. \tag{8}$$

Relative to (4), 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, the optimal policy turns out to be 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

$$\hat{L}_A = \sum_{j\in\{A,B\}} s_j \int (1-q) F_\varepsilon(\bar{\varepsilon}_j(q))\, dF_{q,j}(q), \tag{9}$$

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

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

$$\frac{\hat{L}_A}{s_A} - \frac{\hat{L}_B}{s_B} \leq \kappa, \tag{11}$$

which we show in Appendix D.1 can be written as

$$\sum_{j\in\{A,B\}} s_j \int \left[\frac{1-q}{s_A} - \frac{q}{s_B}\right] F_\varepsilon(\bar{\varepsilon}_j(q))\, dF_{q,j}(q) \leq \kappa. \tag{12}$$

The lender now maximizes

$$V = \sum_{j\in\{A,B\}} s_j \int \int^{\bar{\varepsilon}_j(q)} \left(\mu_j - \varepsilon\right) dF_\varepsilon(\varepsilon)\, dF_{q,j} \tag{13}$$

subject to (12). 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

$$\bar{\varepsilon}_j^{PR}(q) = \mu_j + \lambda^{PR}\left[\frac{q}{s_B} - \frac{1-q}{s_A}\right] \tag{14}$$

where the superscript PR stands for "predicted race". The intuition for (14) 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 (14) becomes (6) and (7) for members of Groups A and B, respectively, nesting our earlier results.

Computing the probability of approval, we obtain

$$\pi_i^{PR} = \underbrace{\left(\mu_A - \frac{\lambda^{PR}}{s_A}\right)}_{\text{const}} + \underbrace{(\mu_B - \mu_A)}_{<0}B_i + \underbrace{\lambda^{PR}\left(s_B^{-1} + s_A^{-1}\right)}_{>0} q_i. \tag{15}$$

Importantly, the loading on B_i is exactly the same as in the no regulation equilibrium (Eq. (4)), 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 (8) and (15), which is equal to

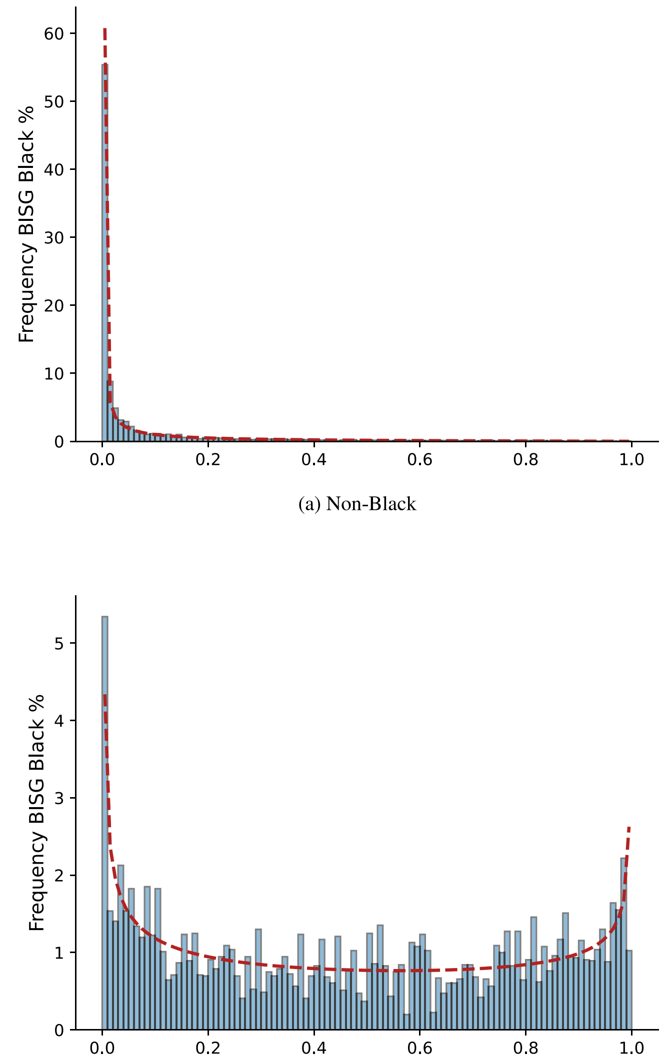
$$\pi_i^{AR} - \pi_i^{PR} = \underbrace{\left(\frac{\lambda^{PR} - \lambda^{AR}}{s_A}\right)}_{\text{const}} + \underbrace{\lambda^{AR}(s_A^{-1} + s_B^{-1})B_i}_{\text{loading on actual race}} - \underbrace{\lambda^{PR}\left(s_B^{-1} + s_A^{-1}\right)q_i}_{\text{loading on predicted race}}. \tag{16}$$

Eq. (16) 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.

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 Fig. 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



(a) Non-Black

(b) Black

Fig. 1. BISG Densities by actual race.
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. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

to equalize the actual or predicted approval rates of the two groups. The approval rates implied by the model can be seen in Fig. 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 (an assumption we relax in Appendix E, see below). In this case, the approval rate for Group B is 4.58%, substantially lower than the approval rate of 9.02% for Group 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

[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.

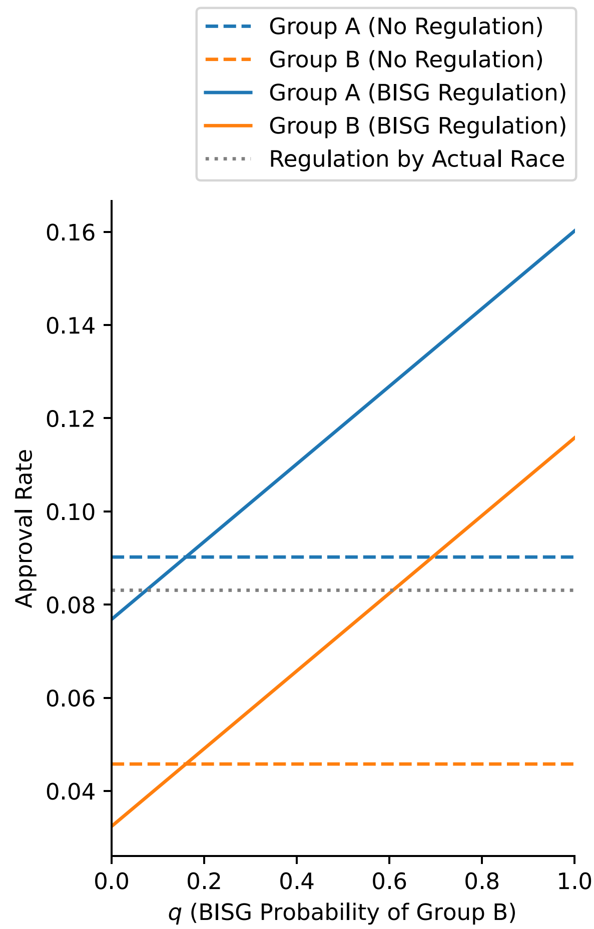


Fig. 2. Numerical example: Lender approval policy.
Note: This figure displays the model's approval rates. The blue dashed line labeled π_A^{NR} displays the approval rate for Group A under the no regulation equilibrium, while the blue solid line labeled π_A^{PR} displays the approval rate for Group A 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. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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 Fig. 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 Fig. 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 Fig. 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.

Extension: Correlation between BISG and fundamentals. In our benchmark model, the BISG-predicted race probabilities embedded in q have no influence on lending except through the regulatory constraint (12). In Appendix E, we extend the model to allow for correlation between BISG and fundamentals.

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

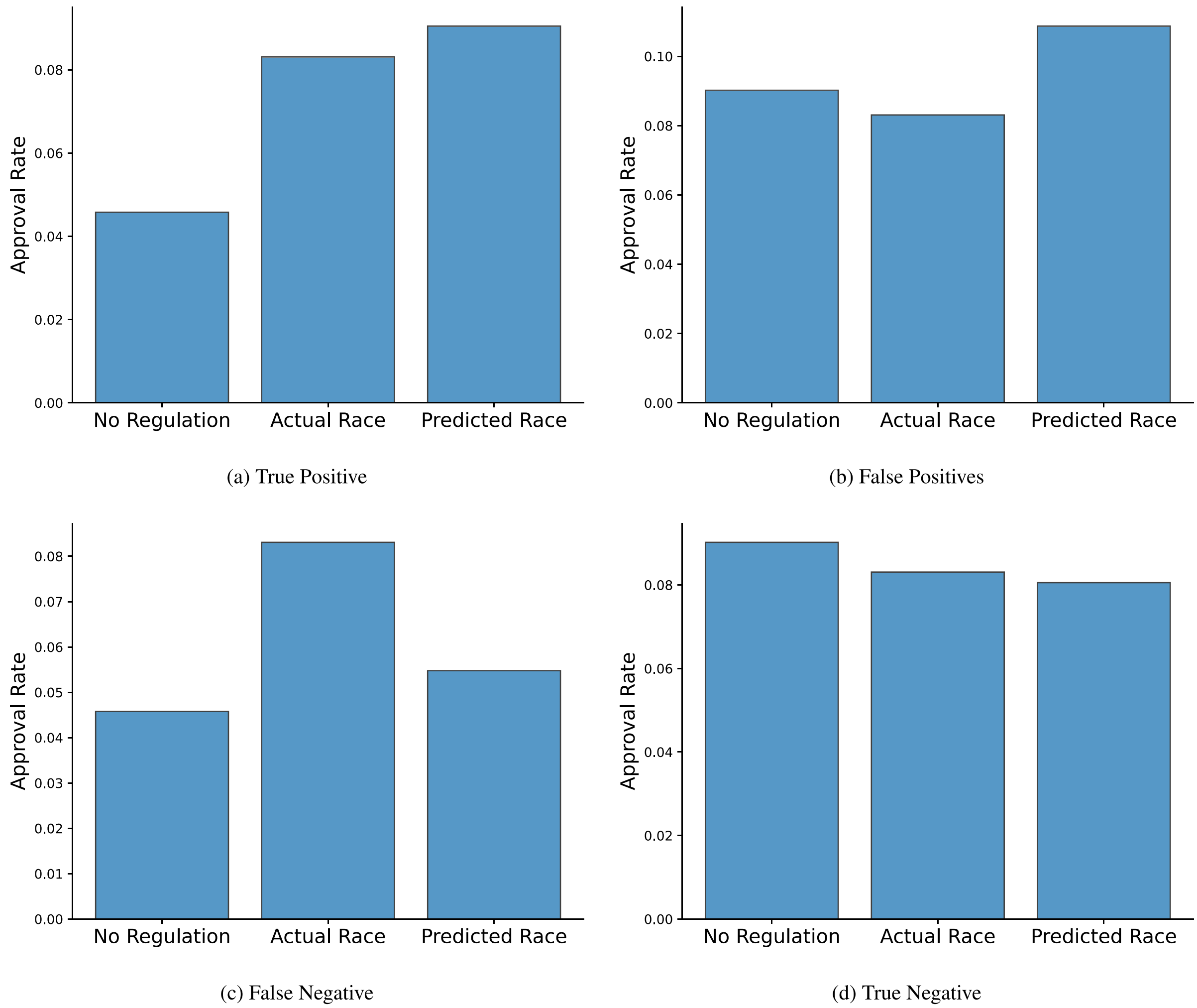


Fig. 3. Model numerical example: Approvals by classification type.
Note: This figure reports approval rates by classification group in the model's numerical example. Each panel displays the approval rates for a given classification of applicants over all three computed equilibria: no regulation, actual 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 Positive" 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.

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]

The raw Lendio dataset has applications from 160,942 unique firms. Of these, 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. Here and below, we consider an "application" to be an instance in which Lendio forwards a loan application to a specific lender. Lendio on average sends an application to about four lenders. Focusing on the application level in Panel A, the average (median) application seeks just over \$100,000 (\$50,000) in funding, but the approved amounts are much lower, at around \$52,000 (\$26,000). The average approval rate is 8.2%.

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,

[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.

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

Table 1
Loan application and lender summary statistics (Lendio).

	N	Mean	Median	SD
Panel A: Application-Level Data				
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.

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 forwards 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.

The firms in the Lendio data are selected on applying to a particular fintech lender. However, consistent with Lendio being a relatively widely used marketplace, they appear roughly representative of U.S. firms along a few important dimensions that we are able to observe. First, the firms have a similar age composition, with 23% in the 6–10 year old range, compared to 16% of all U.S. firms. Second, Lendio applicants have a similar organizational structure; for example, 25% are corporations compared to 18% of U.S. small businesses. Finally, the top 10 states in Lendio are the same as the top 10 states for small businesses overall, with a roughly similar ordering.[18]

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 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.[19] 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.[20] 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. (2024).

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. (2024):

1. Large, medium, and small banks[21];

[18] https://data.census.gov/table/BDSTIMESERIES.BDSFAGE?q=BDSTIMESERIES.BDSFAGE, https://www.nsba.biz/_files/ugd/fec11a_5f7e3afe529c4461970d6f95a6ddb572.pdf,: https://www.iii.org/publications/a-firm-foundation-how-insurance-supports-the-economy/a-50-state-commitment/businesses-by-state

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

[20] 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.

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

Table 2
Loan and lender summary statistics (PPP).

	N	Mean	Median	SD
Panel A: Unique Borrower-Level Data				
Loan Approval:				
Number Loans	22,618	614.636	231.000	761.995
Loan Amt	22,618	138001.568	38461.000	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	44166.648	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.

2. Credit unions, community development financial institutions (CDFIs) and minority depository institutions (MDIs)[22];
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 Tables Appendix A.1 and Appendix 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.[23]

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.

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.[24] 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 Tables Appendix A.1 and Appendix 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.

Selection bias. We collect our race measures for a subsample of the raw PPP and Lendio datasets, raising the question of whether this selection may bias our results. In Appendix F, we explore selection into our analysis sample using a variety of methods. First, we compare all zipcode and firm observables across the key subsamples, such as borrowers who self-report race vs. those that do not in the PPP, and borrowers for whom we observe an image vs. those for whom we do not. These are generally quite similar, especially on key variables such as the zipcode-level Black population share. We also show that the surname and broader geographical distributions are similar in the raw and analysis data. Second, we show that BISG and image-based race

[22] 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.

[23] 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.

[24] We also collected MD, PhD, and JD degrees, but excluded these due to the small sample sizes.

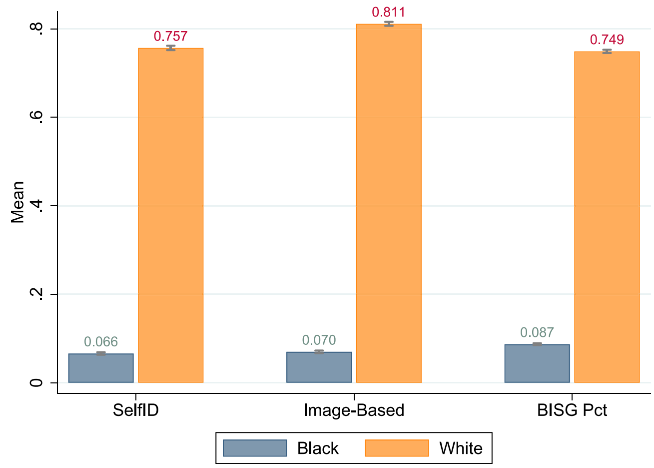


Fig. 4. Borrower race variable statistics (PPP, Unique Borrower–Level).
Note: This figure reports the mean of each race variable on the subsample of observations for which BISG is able to predict race, N = 35,072. The terms SelfID and image-based are binary variables. For example, 76% of our sample self-identifies as White. The continuous variable BISG Pct is defined as the actual percent chance the borrower is Black/White according to the BISG algorithm.

are similarly correlated in Lendio and PPP, and also show that the kernel density of the BISG distribution is very similar across the key subsamples. We conclude that while there are some differences between the analysis and raw samples, they are unlikely to lead to bias that affects our broader insights.

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. Fig. 4 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%.[25]

Correlation rates between the variables are in Table 3. In the first three rows, we present the correlation between indicators for being self-identified Black, image-based Black, and the continuous measure of BISG Black. The correlation between self-identified and image-based race is 0.87, while the correlation between self-identified race and BISG race is just 0.54. In Appendix G, we document that this correlation is somewhat lower than analogous ones reported in other studies, which usually revolve around 70%. We show that this is a direct consequence of BISG being a biased measure of race in the PPP sample because the share of Black individuals in the PPP sample is lower than that in the population upon which BISG is calibrated. 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.[26]

[25] The corresponding figure for Lendio is Appendix Figure Appendix A.1.
[26] The corresponding rates for Lendio are in Appendix Table Appendix A.3.

In the final column, we use a version of BISG that takes into account first name (BISFG). This performs better than BISG, with a correlation of 57% 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, Fig. 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 3630 unique applicants. Fig. 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.6, 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.

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 likely 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

Table 3
Correlations between race variables (PPP).

	Black SelfID	BISG Black Percent	Black Geography	Black Surname	Black Firstname Surname	BIFSG
Black (Image)	0.87***	0.56***	0.38***	0.41***	0.51***	0.60***
Black (SelfID)		0.54***	0.36***	0.38***	0.47***	0.57***
BISG Black Percent			0.72***	0.63***	0.54***	0.86***
Black (Geography)				0.18***	0.20***	0.62***
Black (Surname)					0.69***	0.50***
Black (Firstname+Surname)						0.73***

Note: This table shows correlation coefficients between race variables. BISG Black Percent is the continuous probability of being Black from BISG. For lender-level analysis, we use this continuous probability of being Black as determined by the BISG algorithm. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

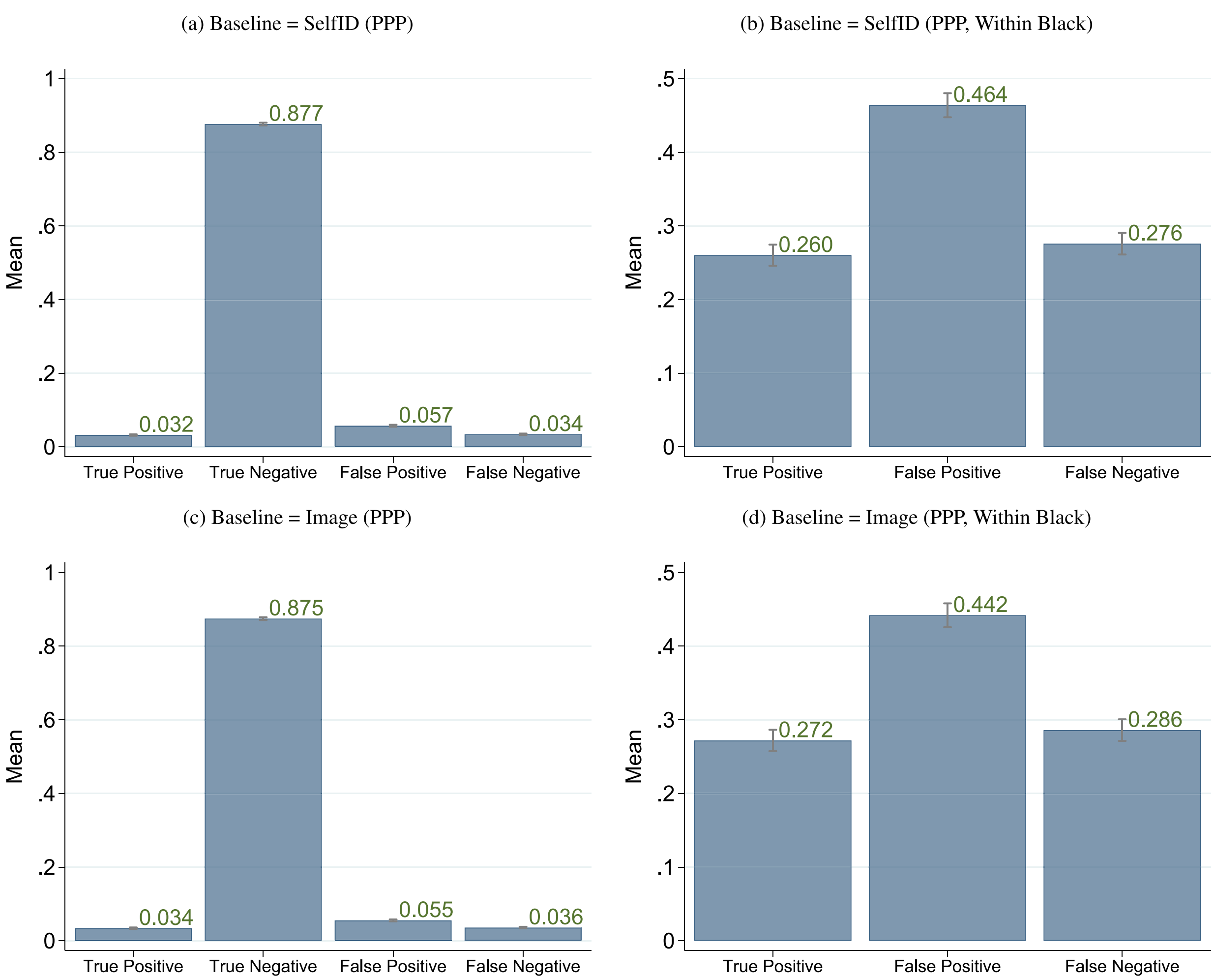


Fig. 5. BISG error rates (Unique Borrower–Level).
Note: This figure presents mean error types for observations where BISG can predict race. Each error type is defined relative to a benchmark "true race". For example, True Positive (SelfID) means that the business owner self-identifies as Black and is 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 4973 observations. Panel D has 4882 observations.

a set of socioeconomic characteristics in our PPP data. Fig. 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 Appendix 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 Fig. 6 represents the regression results using the independent variable of whether an area's share of Black residents is above or below the

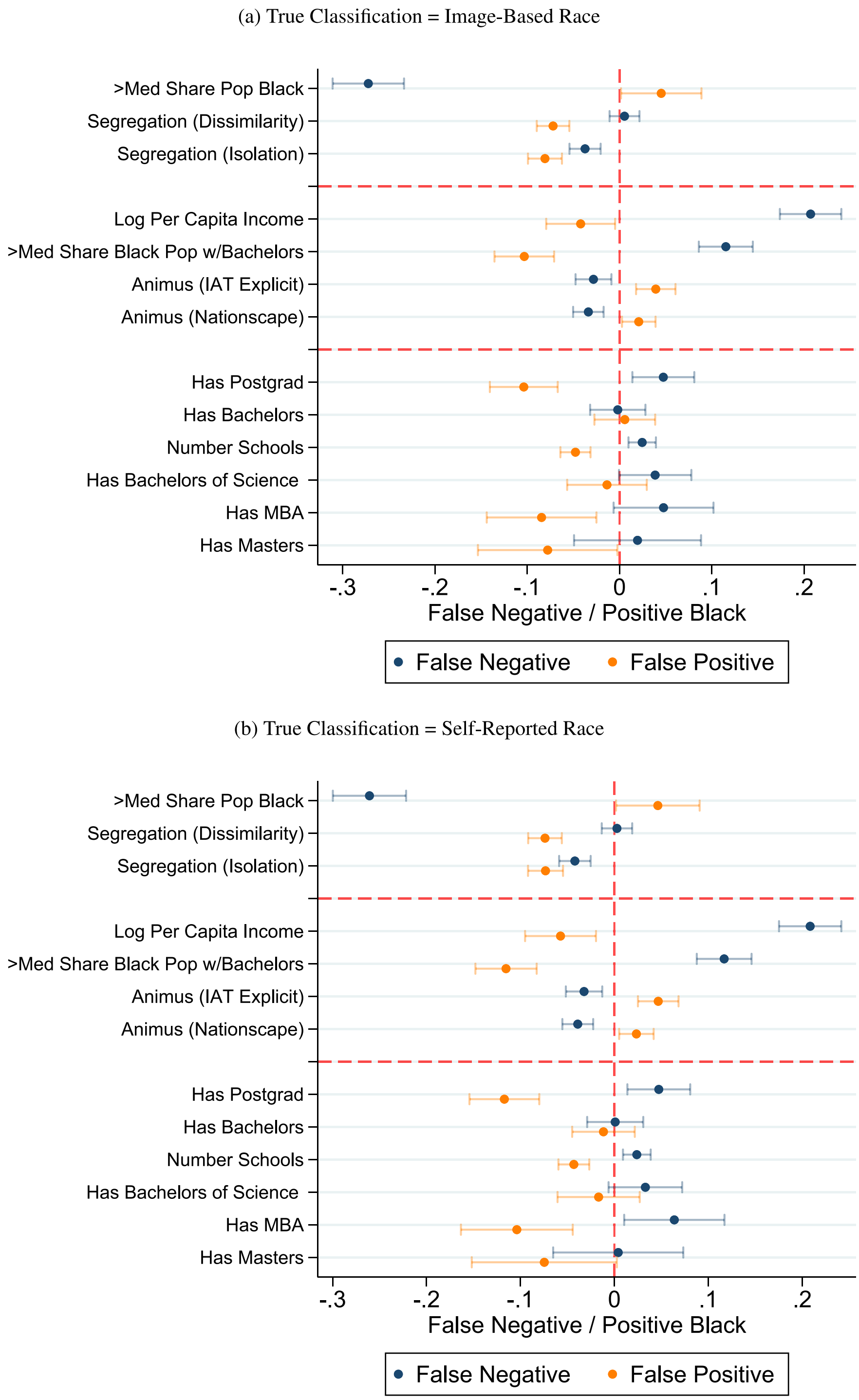


Fig. 6. Effect of socioeconomic covariates on BISG errors (PPP).
Note: This figure shows estimates of a set of regressions, each of either False Negative Black or False Positive Black on one of the socioeconomic characteristics in each row of the *y*-axis, using the "within Black" sample. The indicator 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 ($N = 4882$). Panel B's sample comprises borrowers who either self-identified as Black on their PPP loan application or were algorithmically classified as Black; $N = 4973$.

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 Fig. 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).[27]

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

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.[28]

We consider other socioeconomic characteristics at the geographic (ZIP) level in the following two rows of Fig. 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 Fig. 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 chance of being classified as 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 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 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 Fig. 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

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

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 shows 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.

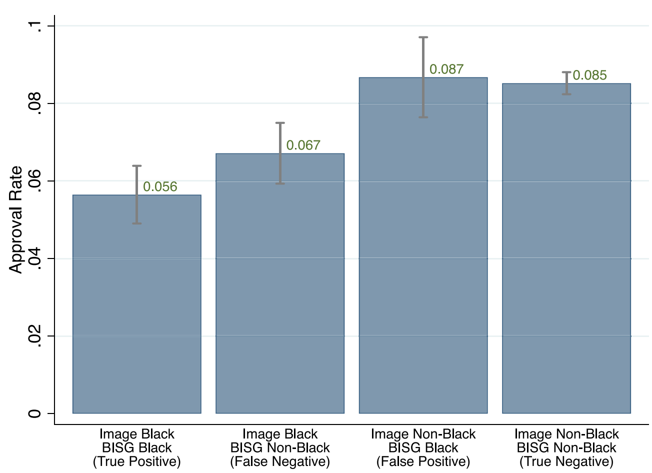


Fig. 7. Lendio approval rate by group.
Note: This figure plots the approval rate by error type, focusing on a subset of 47,481 applicants for whom we could calculate job tenure and confidently determine loan rejection. We are able to calculate job tenure if the firm in the Lendio application matched 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.

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

$$\mathbb{1}(\text{Approved}_{i,l}) = \alpha_l + \alpha_t + \beta \mathbb{1}(\text{Black}_i) + \mathbf{X}_i \delta + \varepsilon_{il}, \tag{17}$$

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 Fig. 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 separately investigate the two components of BISG, surname and geography in Table 6.

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 −1.3pp 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.

Our baseline model does not exactly align with this variation in approval rates in the data. Recall that in the baseline model we assume for parsimony that the BISG probability q has no impact on loan fundamentals, and thus only influences approval rates through the regulatory channel whereby higher-q borrowers can relax a BISG-based regulatory constraint by more. As a result, approval rates should be weakly increasing in q conditional on race, and strictly increasing in q if the regulatory constraint binds. However, in practice (Fig. 7 and Table 5) show that false positive (high-q non-Black) applicants have

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 Eq. (17), 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 Eq. (17), 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 Eq. (17), 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 Eq. (17), 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.

an approval rate that is only barely higher than true negative (low-q non-Black) applicants. Further, true positive (high-q Black) applicants have a *lower* approval rate than false negative (low-q Black) applicants. These are problematic in the baseline model because the first result would imply that the regulatory constraint is close to slack, while the second is strictly impossible in that model. These patterns reflect loan fundamentals pushing in the opposite direction as the regulatory incentive.

In Appendix E, we extend the model by allowing q to be correlated with loan fundamentals, which better explains the observed patterns. This correlation is intuitive because q is correlated with lower local income as well as lower individual educational achievement.[29] Because this negative correlation between q and fundamentals would create

[29] In the absence of underlying frictions, the likelihood of a name being associated with a Black borrower will not be negatively correlated with loan repayment probability. However, exploring these frictions is outside the scope of this paper, so we leave an investigation of the frictions that might cause this negative correlation to future work.

a negative slope of approvals with respect to q in the absence of regulation, our extended model shows that approval rates that are flat or even declining with respect to q can be consistent with a strong influence of regulatory incentives, since approvals would be even more negatively sloped without them.

Approval rate decomposition. In the analysis above, we establish the existence of a bias in the approval rate gap, which appears to be much smaller when race is measured using BISG compared to our image-based measure. We next formally decompose this approval gap into its statistical components to more fully understand the sources of bias.

Define the approval gap to be the difference between the average approval probability of an application from a non-Black borrower and the average approval probability of an application from a Black borrower. Then we can measure the bias in the approval gap as

$$Bias = ApprovalGap^{BISG} - ApprovalGap^{Image} \tag{18}$$

where $ApprovalGap^{Image}$ and $ApprovalGap^{BISG}$ measure this gap using our image-based and BISG-based measures of race, respectively (see Appendix D.1 for a formal definition). In Appendix D.1, we show that the bias term in (18) is approximately equal to

$$Bias \simeq \left((1-\bar{B})^{-1}+\bar{B}^{-1}\right)\left\{\mathrm{Cov}\Big(\mathbf{1}_i, \underbrace{\max(B_i-q_i,0)}_{\text{false neg.}}\Big) - \mathrm{Cov}\Big(\mathbf{1}_i, \underbrace{\max(q_i-B_i,0)}_{\text{false pos.}}\Big)\right\} \tag{19}$$

where $\mathbf{1}_i$ is an indicator for a loan being approved, B_i is an indicator for being Black under our image-based measure, and q_i is the BISG probability of being Black. To understand the false positive and false negative terms, note that $q_i - B_i$ is the BISG prediction error, representing the difference between the BISG-Black probability and the actual (image-based) Black indicator. When $q_i < B_i$, the error goes in the false negative direction (the BISG probability is too low), the false positive term in (19) is zero, and the false negative term is positive. When $q_i > B_i$, the error goes in the false positive direction (the BISG probability is too high), the false positive term in (19) is positive, and the false negative term is zero.

Eq. (19) implies that the measured bias should be higher when more false negative borrowers are approved (higher covariance between approval and false negative errors), and lower when more false positive borrowers are approved (higher covariance between approval and false positive errors). Intuitively, when a false negative borrower is approved, this increases the non-Black approval rate in the BISG-based measure, increasing the BISG-based approval gap, but increases the Black approval rate in the image-based measure, shrinking the image-based approval gap, which together increase the bias term in (19). The reverse logic holds for approval of a false positive borrower, which increases the Black approval rate measured by BISG, but increases the non-Black approval rate using the image-based measure, decreasing the bias in (19).

Eq. (19) allows us to decompose the sources of approval gap bias. In particular, we can separate the bias between what the regulator observes and the true approval rate gap into three components: (i) a component due to the relative approval rates of false negative borrowers, (ii) a component due to the relative approval rates of false positive borrowers, and (iii) a residual due to the fact that (19) holds only approximately.

With this lead-up, we can now quantify our decomposition. Differencing our measured approval rates in Table 4, we find that the overall approval gap bias in (19) is −1.01pp, meaning that regulators using BISG underestimate the true approval gap. This can be additively decomposed into a contribution from false negative errors of −1.09pp, a contribution from false positive errors of −0.13pp, and a contribution from the approximation residual of +0.20pp. Represented as a share of the total, these components explain 107.2%, 12.5%, and −19.7% of the bias, respectively.

These results show that the difference in the approval gap measured by regulators relative to the true (image-based) approval gap is overwhelmingly due to lower approval rates among false negative borrowers. These correspond to borrowers who are Black according to our image-based measure, but are misclassified with high probability as non-Black by BISG due to having either a geographic location or a surname that is shared with relatively few other Black people.

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).[30] 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 Appendix 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.[31] 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

[30] 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".

[31] First names alone are not especially predictive (Appendix Table Appendix A.12). In Appendix Table Appendix 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).

lending regulators 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:

$$\Delta_{\text{Share Black Appr}} = \bar{\pi}_B^{Image} - \bar{\pi}_B^{BISG} = \frac{\#\,\text{Image Black Approved}}{\#\,\text{Image Black Applicants}} - \frac{\#\,\text{BISG Black Approved}}{\#\,\text{BISG Black Applicants}}. \tag{20}$$

The first term on the right hand side of Eq. (20) 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.[32]

In Appendix D.1, we derive an analogue to (19) for this case:

$$\Delta_{\text{Share Black Appr}} \simeq \bar{B}^{-1}\left\{\text{Cov}\left(\mathbf{1}_i, \underbrace{\max(B_i - q_i, 0)}_{\text{false neg.}}\right) - \text{Cov}\left(\mathbf{1}_i, \underbrace{\max(q_i - B_i, 0)}_{\text{false pos.}}\right)\right\}. \tag{21}$$

As with our approval rate bias decomposition, Eq. (21) shows that $\Delta_{\text{Share Black Appr}}$ is increasing in the approval rate among false negative borrowers, and is decreasing in the approval rate among false positive borrowers.[33] 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 $\Delta_{\text{Share Black Appr}}$.[34]

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, Fig. 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 fintechs and large banks are more automated and arms-length than small banks (Howell et al., 2024; Balyuk et al., 2020).

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.

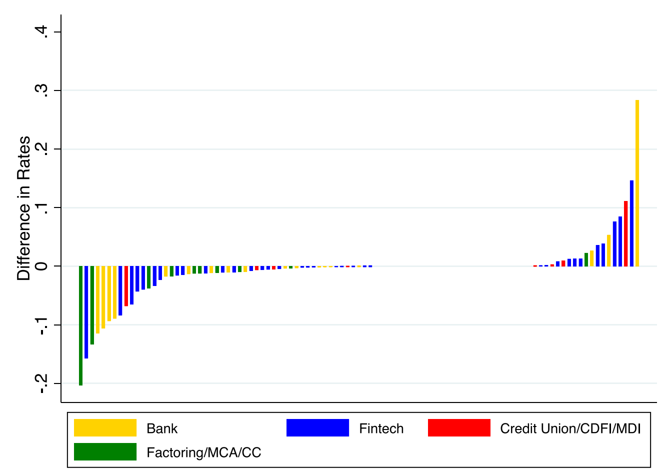


Fig. 8. Lender-level difference in loan approval rates by race measure (Lendio).
Note: This figure plots, for each lender, the $\Delta_{\text{ShareApprovedBlack}}$. This is the difference in the approval rates among Black applicants between image-based and BISG measures, corresponding to $\pi_{i,j}^{AR} - \pi_{i,j}^{PR}$ 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. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

[32] 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.

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

[34] 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.

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 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 analogous 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:

$$\Delta_{\text{Share Loans Black}} = \frac{\#\,\text{Image Black Borrowers}}{\#\,\text{All Borrowers}} - \frac{\#\,\text{BISG Black Borrowers}}{\#\,\text{All Borrowers}}. \tag{22}$$

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 Fig. 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.

This point is directly related to an ongoing policy debate. The Dodd-Frank Act of 2012 required the CFPB to promulgate regulations

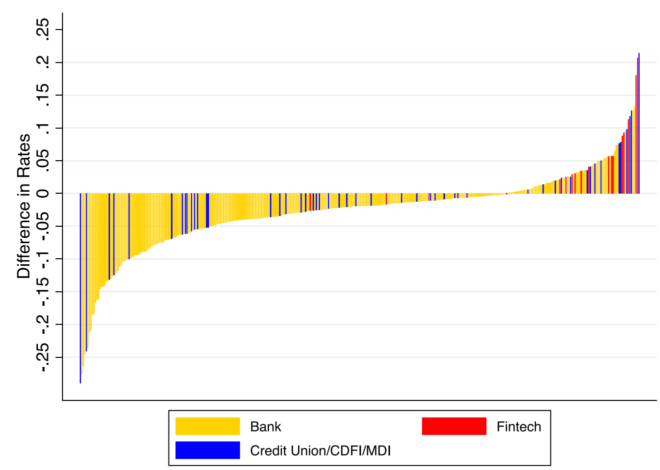


Fig. 9. Lender-Level Difference in lending rates by race measure (PPP).
Note: This figure plots, for each lender, the $\Delta_{\text{ShareLoansBlack}}$. This is the difference in the lending rates to Black firm owners between image-based and BISG measures, corresponding to $\pi^{AR}_{i,j} - \pi^{PR}_{i,j}$ 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. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

on the collection and reporting of self-identified race data in small business lending (referred to as "1071" due to the Section of the Act). There has been stiff opposition from the small banking community to 1071.[35] Banks point out that the new reporting would be costly; indeed, the CFPB estimates the costs of the policy at $46-$100 per loan application.[36] After the CFPB final rulemaking in March 2023 requiring the collection of race among other variables, banking associations sued, and the rule was indefinitely stayed in July 2023.[37]

Our results are also consistent with fintechs serving Black applicants with higher socioeconomic status. This highlights intricate equity

[35] (see here and here).

[36] See p. 760 here: https://files.consumerfinance.gov/f/documents/cfpb_1071-final-rule.pdf.

[37] Texas Bankers Association, et al. v. CFPB, see https://caselaw.findlaw.com/court/us-dis-crt-s-d-tex-mca-div/114780286.html.

implications of shifting from BISG-based race measures to ones more reflective of 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.

Given that raw approval gaps are used by regulators to identify the lenders that warrant more serious investigations, this raises the question of whether the magnitude of the approval rate gap is large enough to shift lenders into or out of the "serious investigation" group. For example, suppose regulators use raw approval rate gaps to rank lenders according to their Black/non-Black approval gaps and then investigate the lenders with the largest gap. How often would the composition of high-gap lenders change based on how race is measured? We calculate that if we consider the 10% of lenders with the highest Black/non-Black approval gaps, 40% of them would change their identity if we measured the gap using image-based race rather than BISG-implied race. Redoing this exercise to look at the top 20%, 30%, 40%, and 50% of lenders implies switching rates of 35%, 26%, 29%, and 26%, respectively. We conclude that BISG errors matter not only for the magnitude of approval gaps among lenders, but are also highly influential in determining relative rankings and the identities of lenders most likely to face investigation.

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, Eq. (17), 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:

$$Approved_i^{BISG\downarrow} = Approved_i - h \times (BISG_i - \overline{BISG}), \tag{23}$$

$$Approved_i^{Image\uparrow} = Approved_i + h \times (Black_i - \overline{Black}), \tag{24}$$

$$Approved_i^{Both} = Approved_i + h \times \left[(Black_i - \overline{Black}) - (BISG_i - \overline{BISG}) \right]. \tag{25}$$

Eq. (23) 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 Eqs. (23) and (17). 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 Eq. (24). 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 Eq. (25). 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

$$Approved\,Share_Z^C = \frac{\sum_i Z_i Approved_i^C}{\sum_i Approved_i^C}. \tag{26}$$

The directional change in that variable under the counterfactual scenario is

$$d\,Approved\,Share_Z^C = \frac{Approved\,Share_Z^C - Approved\,Share_Z}{h} = \frac{\mathrm{Cov}(Z_i, X_{ij})}{\overline{Approved}}, \tag{27}$$

where $Approved\,Share_Z$ is from Eq. (26) using the actual approval rate $Approved_i$ in place of $Approved_i^C$, $\mathrm{Cov}(Z_i, X_{ij})$ 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 $d\,Approved\,Share_Z^C$ 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 D.1, which holds regardless of the value of h.

The results are displayed in Table 9, where columns (1), (2), and (3) correspond to the counterfactuals lowering the weight on BISG-Black (Eq. (23)), increasing the weight on image-Black (Eq. (24)), and applying both changes at once (Eq. (25)), respectively. Because the two variables are positively correlated, these two changes generally have effects of opposite signs. However, because prediction errors can be correlated with socioeconomic characteristics, the magnitudes can differ, leading to a nontrivial net change.

Panel A displays the impact on the shares of approved borrowers by image-based and BISG-based race classification. The first row displays the impact on the share of approved applications to image-Black borrowers. Since BISG scores are positively correlated with image-Black status, reducing the loading on BISG reduces this share, while increasing the loading on image-Black status increases it. Due to prediction errors, however, the effect of increasing the image-Black weight is roughly three times stronger than the effect of decreasing the BISG-Black weight. On net, the paired change would substantially increase the share of loans going to image-Black borrowers, with a 0.99pp increase in the image-Black share for every 1pp change in the approval weights.

The remaining rows of Panel A repeat this exercise by the joint image and BISG classification. Our results in Column (3) show that the vast majority of the net increase in the image-Black share stems from a 0.73pp increase in the False Negative share, compared to only an increase in the True Positive share of only 0.25pp. Examining Columns (1) and (2) reveals that while increasing the image-Black weight has a similar impact on both the True Positive and False Negative shares, reducing the BISG weight has a much stronger countervailing effect on True Positive applicants (whose BISG scores are high) compared to False Negative applicants (whose BISG scores are low), explaining the overall result.

Turning to the bottom two rows of Panel A, representing the effect on non-Black borrowers, we see that the net policy decreases the False Positive share by -0.28pp and the True Negative share by -0.70pp. Because the base rates of these groups are so different, representing 5.8% and 82.0% of approved applications in our data, this is actually a much larger proportional effect on False Positive borrowers, whose high BISG scores are more affected by the decreased loading on BISG compared to True Negative borrowers.

In Panel B of Table 9, we analyze how these counterfactuals affect socioeconomic characteristics at the ZIP code level, using the same set of characteristics used in Fig. 6. Overall the net effects (Column (3)) are modest and not of consistent direction. One the one hand, some effects somewhat favoring more advantaged areas, with the average approved borrower living in an area with \$40 higher local per-capita income,

Table 9
Counterfactual exercise.

Characteristic	(1) BISG Weight ↓	(2) Image Weight ↑	(3) Net Change
Panel A: By Classification, Full Sample			
Image Black Share	-0.66	1.65	0.99
True Positive Share	-0.52	0.77	0.25
False Negative Share	-0.14	0.88	0.73
False Positive Share	-0.17	-0.11	-0.28
True Negative Share	0.84	-1.54	-0.70
Panel B: By Socioeconomic Characteristics, Full Sample			
Local PC Income (Thousands)	0.14	-0.10	0.04
Share Pop Black	-0.36	0.33	-0.03
Share Pop ≥ Bachelors	0.08	-0.06	0.02
Share Black Pop ≥ Bachelors	0.06	-0.02	0.04
Segregation (Dissimilarity)	-0.11	0.31	0.20
Segregation (Isolation)	-0.84	0.97	0.13
Animus (IAT Explicit)	-0.17	0.17	-0.00
Animus (Nationscape)	-0.32	0.33	0.01
Panel C: By Socioeconomic Characteristics, Within Image-Black Subpopulation			
Local PC Income (Thousands)	0.39	0.00	0.39
Share Pop Black	-0.99	0.00	-0.99
Share Pop ≥ Bachelors	0.26	0.00	0.26
Share Black Pop ≥ Bachelors	0.18	0.00	0.18
Segregation (Dissimilarity)	-0.41	0.00	-0.41
Segregation (Isolation)	-1.52	0.00	-1.52
Animus (IAT Explicit)	-0.26	0.00	-0.26
Animus (Nationscape)	-0.21	0.00	-0.21

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 (17) 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.

0.03pp lower Black population share, and 0.02pp and 0.04pp of the overall and Black population having a Bachelor's degree or higher, respectively. On the other hand, Column (3) also shows a net increase in lending to areas with higher segregation, and no differential effect by animus. All told, these results show little effect on socioeconomic characteristics when applied to the full sample.

In sharp contrast, Panel C shows the impact of the change in policy on socioeconomic characteristics *among image-Black borrowers*. Because these borrowers all have the same image-Black classification, changing the image weight has no effect on relative lending propensities across borrowers *within* this subpopulation, leading to a uniform finding of zero effect in Column (2). However, because image-Black borrowers have differing BISG probabilities, the effect of changing this loading in Column (1) will have relative effects across this subsample, which translate into nonzero effects in Column (3). Overall, Panel C reveals changes that are roughly an order of magnitude larger than their full-sample counterparts, and now consistently favor more advantaged areas. In particular, the average approved *image-Black* applicant lives in an area with $390 higher per-capita income, 0.99pp lower Black population share, 0.26pp and 0.18pp higher shares of the overall and Black population with a Bachelor's degree, less segregation, and less racial animus.

To understand why these results differ, note that the change in policy can be decomposed into two effects. First, the policy tilts lending overall toward Black borrowers and away from non-Black borrowers. Second, the policy tilts lending within each racial group toward borrowers with lower BISG scores. Because Black borrowers live in less advantaged areas, the first effect favors disadvantaged areas. But because BISG scores are also correlated with lower socioeconomic advantage, the second effect favors more advantaged areas. Our results in Panel B imply that these two effects roughly cancel out, leading to little change in lending by socioeconomic characteristics overall. Our results in Panel C, however, show that moving from a BISG-based policy to one based on actual or self-reported race that the approval rate gains to the Black population would be disproportionately concentrated among Black borrowers living in more advantaged or affluent 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 is 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.

CRediT authorship contribution statement

Daniel L. Greenwald: Writing – original draft, Conceptualization. **Sabrina T. Howell:** Writing – original draft, Conceptualization. **Cangyuan Li:** Writing – original draft, Conceptualization. **Emmanuel Yimfor:** Writing – original draft, Conceptualization.

Declaration of competing interest

This project received funding from the Alfred P. Sloan Foundation. Sabrina Howell has no conflicts of interest to disclose. Daniel Greenwald has no conflicts of interest to disclose. Emmanuel Yimfor has no conflicts of interest to disclose. Cangyuan Li has no conflicts of interest to disclose.

Data availability

https://data.mendeley.com/datasets/g547vfzybp/2.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jfineco.2024.103857.

References

Ambrose, B.W., Conklin, J.N., Lopez, L.A., 2021. Does borrower and broker race affect the cost of mortgage credit? Rev. Financ. Stud. 34 (2), 790–826.
Anwar, S., Fang, H., 2006. An alternative test of racial prejudice in motor vehicle searches: Theory and evidence. Amer. Econ. Rev. 96 (1), 127–151.
Arnold, D., Dobbie, W., Yang, C.S., 2018. Racial bias in bail decisions. Q. J. Econ. 133 (4), 1885–1932.
Asiedu, E., Freeman, J.A., Nti-Addae, A., 2012. Access to credit by small businesses: How relevant are race, ethnicity, and gender? Amer. Econ. Rev. 102 (3), 532–537.
Athey, S., Karlan, D., Palikot, E., Yuan, Y., 2022. Smiles in Profiles: Improving Fairness and Efficiency Using Estimates of User Preferences in Online Marketplaces. National Bureau of Economic Research.
Balyuk, T., Berger, A.N., Hackney, J., 2020. What is fueling fintech lending? The role of banking market structure. The Role of Banking Market Structure (June 23, 2020).
Bartik, A.W., Nelson, S.T., 2024. Deleting a signal: Evidence from pre-employment credit checks. Rev. Econ. Stat. 1–47.
Bartlett, R., Morse, A., Stanton, R., Wallace, N., 2022. Consumer-lending discrimination in the fintech era. J. Financ. Econ. 143 (1), 30–56.
Bartoš, V., Bauer, M., Chytilová, J., Matějka, F., 2016. Attention discrimination: Theory and field experiments with monitoring information acquisition. Amer. Econ. Rev. 106 (6), 1437–1475.
Bayer, P., Ferreira, F., Ross, S.L., 2018. What drives racial and ethnic differences in high-cost mortgages? The role of high-risk lenders. Rev. Financ. Stud. 31 (1), 175–205.
Begley, T.A., Purnanandam, A., 2021. Color and credit: Race, regulation, and the quality of financial services. J. Financ. Econ. 141 (1), 48–65.
Bellucci, A., Borisov, A., Zazzaro, A., 2013. Do banks price discriminate spatially? Evidence from small business lending in local credit markets. J. Bank. Financ. 37 (11), 4183–4197.
Berg, T., Burg, V., Gombović, A., Puri, M., 2020. On the rise of fintechs: Credit scoring using digital footprints. Rev. Financ. Stud. 33 (7), 2845–2897.
Berger, A.N., Black, L.K., 2011. Bank size, lending technologies, and small business finance. J. Bank. Financ. 35 (3), 724–735.
Bertrand, M., Mullainathan, S., 2004. Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. Amer. Econ. Rev. 94 (4), 991–1013.
Bhutta, N., Hizmo, A., 2021. Do minorities pay more for mortgages? Rev. Financ. Stud. 34 (2), 763–789.
Blanchflower, D.G., Levine, P.B., Zimmerman, D.J., 2003. Discrimination in the small-business credit market. Rev. Econ. Stat. 85 (4), 930–943.
Blattner, L., Nelson, S., 2021. How costly is noise? Data and disparities in consumer credit. arXiv preprint arXiv:2105.07554.
Buchak, G., Matvos, G., Piskorski, T., Seru, A., 2018. Fintech, regulatory arbitrage, and the rise of shadow banks. J. Financ. Econ. 130 (3), 453–483.
Bursztyn, L., Chaney, T., Hassan, T.A., Rao, A., 2021. The Immigrant Next Door: Exposure, Prejudice, and Altruism. Technical Report, National Bureau of Economic Research.
Butler, D.M., Broockman, D.E., 2011. Do politicians racially discriminate against constituents? A field experiment on state legislators. Am. J. Political Sci. 55 (3), 463–477.
Charles, K.K., Guryan, J., 2008. Prejudice and wages: An empirical assessment of Becker's the economics of discrimination. J. Polit. Econ. 116 (5), 773–809.
Chernenko, S., Scharfstein, D.S., 2021. 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.
Cook, L.D., Marx, M., Yimfor, E., 2022. Funding Black High-Growth Startups. Technical Report, National Bureau of Economic Research.
D'Acunto, F., Ghosh, P., Jain, R., Rossi, A.G., 2020. How costly are cultural biases? Available at SSRN 3736117.
Dimmock, S.G., Gerken, W.C., Graham, N.P., 2018. Is fraud contagious? Coworker influence on misconduct by financial advisors. J. Finance 73 (3), 1417–1450.
Dobbie, W., Liberman, A., Paravisini, D., Pathania, V.S., 2021. Measuring bias in consumer lending. Rev. Econ. Stud. 88 (6), 2799–2832.
Egan, M., Matvos, G., Seru, A., 2022. When harry fired sally: The double standard in punishing misconduct. J. Polit. Econ. 130 (5), 1184–1248.
Erel, I., Liebersohn, J., 2020. Does Fintech Substitute for Banks? Evidence from the Paycheck Protection Program. Technical Report, National Bureau of Economic Research.
Fairlie, R.W., Fossen, F., 2021. 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.
Fairlie, R.W., Robb, A.M., 2007. Why are black-owned businesses less successful than White-owned businesses? The role of families, inheritances, and business human capital. J. Labor Econ. 25 (2), 289–323.
Fairlie, R., Robb, A., Robinson, D.T., 2022. Black and white: Access to capital among minority-owned start-ups. Manage. Sci. 68 (4), 2377–2400.
Frame, W.S., Huang, R., Mayer, E.J., Sunderam, A., 2022. The Impact of Minority Representation at Mortgage Lenders. Technical Report, National Bureau of Economic Research.
Fremont, A., Weissman, J.S., Hoch, E., Elliott, M.N., 2016. When race/ethnicity data are lacking: Using advanced indirect estimation methods to measure disparities. Rand Health Q. 6 (1).
Fryer, R.G., Levitt, S.D., 2004. The causes and consequences of distinctively Black names. Q. J. Econ. 119 (3), 767–805.
Fuster, A., Plosser, M., Schnabl, P., Vickery, J., 2019. The role of technology in mortgage lending. Rev. Financ. Stud. 32 (5), 1854–1899.
Gaddis, S.M., 2017. How Black are Lakisha and Jamal? Racial perceptions from names used in correspondence audit studies. Sociol. Sci. 4, 469–489.
Giacoletti, M., Heimer, R., Yu, E.G., 2021. Using high-frequency evaluations to estimate discrimination: Evidence from mortgage loan officers. Available at SSRN 3795547.
Giulietti, C., Tonin, M., Vlassopoulos, M., 2019. Racial discrimination in local public services: A field experiment in the United States. J. Eur. Econom. Assoc. 17 (1), 165–204.
Howell, S.T., Kuchler, T., Snitkof, D., Stroebel, J., Wong, J., 2024. Lender automation and racial disparities in credit access. J. Finance 79 (2), 1457–1512.
Jiang, E.X., Lee, Y., Liu, W.S., 2021. Disparities in consumer credit: The role of loan officers in the FinTech era. Available at SSRN.
Kabir, P., Ruan, T., 2023. Reducing racial disparities in consumer credit: Evidence from anonymous loan applications. Available at SSRN 4364852.
Knowles, J., Persico, N., Todd, P., 2001. Racial bias in motor vehicle searches: Theory and evidence. J. Polit. Econ. 109 (1), 203–229.
Kreisman, D., Smith, J., 2023. Distinctively Black names and educational outcomes. J. Polit. Econ. 131 (4), 877–897.
Massey, D.S., Denton, N.A., 1988. The dimensions of residential segregation. Soc. Forces 67 (2), 281–315.

Milkman, K.L., Akinola, M., Chugh, D., 2012. Temporal distance and discrimination: An audit study in academia. Psychol. Sci. 23 (7), 710–717.
Petersen, M.A., Rajan, R.G., 1994. The benefits of lending relationships: Evidence from small business data. J. Finance 49 (1), 3–37.
Pool, V.K., Stoffman, N., Yonker, S.E., 2015. The people in your neighborhood: Social interactions and mutual fund portfolios. J. Finance 70 (6), 2679–2732.
Price, J., Wolfers, J., 2010. Racial discrimination among NBA referees. Q. J. Econ. 125 (4), 1859–1887.
Robb, A., Robinson, D.T., 2018. Testing for racial bias in business credit scores. Small Bus. Econ. 50 (3), 429–443.
Serengil, S.I., Ozpinar, A., 2020. Lightface: A hybrid deep face recognition framework. In: 2020 Innovations in Intelligent Systems and Applications Conference. ASYU, IEEE, pp. 1–5.
Tang, H., 2019. Peer-to-peer lenders versus banks: Substitutes or complements? Rev. Financ. Stud. 32 (5), 1900–1938.
Tausanovitch, C., Vavreck, L., 2020. Methodology and Representativeness Assessment, Oct 10–17, 2019 (version 20200814). Technical Report, Democracy Fund + UCLA Nationscape, URL https://www.voterstudygroup.org/uploads/reports/Data/Nationscape-User-Guide_2020sep10.pdf.
Tootell, G.M., 1996. Redlining in Boston: Do mortgage lenders discriminate against neighborhoods? Q. J. Econ. 111 (4), 1049–1079.
Voicu, I., 2018. Using first name information to improve race and ethnicity classification. Stat. Public Policy 5 (1), 1–13.
Xu, K., Nosek, B., Greenwald, A., 2014. Psychology data from the race implicit association test on the project implicit demo website. J. Open Psychol. Data 2 (1).