The Limits of Algorithmic Measures of Race in Studies of Outcome Disparities*

Sergey Chernenko† David Scharfstein‡

Purdue University Harvard Business School

April 2023

Abstract

We show that the use of algorithms to predict race has significant limitations in measuring and understanding the sources of racial disparities in finance, economics, and other contexts. First, we derive theoretically the direction and magnitude of measurement bias in estimates of unconditional disparities that use predicted instead of actual race. If their prediction errors were random, existing algorithms such as BIFSG (Voicu, 2018) would underestimate disparities in credit access for Black borrowers by 30–50%. In practice, the algorithms are systematically biased toward identifying minority borrowers who are likely to experience worse outcomes. Second, we show that in many applications the accuracy of predicted race is illusory, as many empirical methodologies call for the inclusion of location fixed effects and comparison of white and minority individuals within a given geography. As a result, estimates of conditional disparities can be dramatically underestimated, in some of our analyses, by up to 60%. While underestimating conditional disparities, predicted race overstates the importance of location in explaining disparities. Finally, because algorithm accuracy can vary across subsamples, predicted race can under- or overestimate interaction effects meant to measure cross-sectional variation in disparities.

Keywords: machine learning, race, racial disparities, Paycheck Protection Program, measurement error

*We are grateful to Yelp for providing data on restaurants and to OpenCorporates for providing data on corporate records. Disclosure: Until April 2022, David Scharfstein was on the board of M&T Bank Corporation, which participated in the Paycheck Protection Program.

†Purdue University, 403 W. State Street, West Lafayette, IN 47907; email: schernen@purdue.edu

‡Harvard Business School, Soldiers Field, Boston, MA 02163; email: dscharfstein@hbs.edu

1 Introduction

A growing number of papers in finance and economics study racial disparities in access to credit, attempting to measure the extent of disparities and understand the factors that give rise to them. Because the racial identity of a potential borrower is often not observable — in many cases because the law prohibits collecting this information — these papers rely on machine learning or Bayesian algorithms to predict a borrower's racial and ethnic identity (Howell et al., 2022; Blattner and Nelson, 2021; Zhang, 2018). These methods — particularly the Bayesian Improved Surname Geocoding (BISG) algorithm — are also used by regulators and lenders to measure credit access for potential borrowers that fall into protected classes according to fair lending laws, notably the Equal Credit Opportunity Act (Consumer Financial Protection Board, 2014).

While it is understood that there is measurement error in race prediction algorithms, what is less well understood is the extent to which such measurement error biases estimates of racial disparities and the factors that drive such disparities. This paper tries to fill this gap by studying loan access in an environment where we can directly observe a borrower's racial and ethnic identity using data from voter records. This allows us to compare estimates of racial disparities based on self-reported race to estimates based on algorithmic predictions of race. These algorithms are based on the names typically associated with a racial group and the population shares of the racial groups in the location where the person lives. We show that there is a very significant underestimation of racial disparities when using predicted race from these algorithms. We also show that using predicted race tends to overestimate the role of borrower characteristics, particularly location, in determining racial disparities.

We start by deriving closed-form expressions for the estimation bias of unconditional outcome disparities. We then examine estimation bias in practice by examining racial disparities in the use of the Paycheck Protection Program (PPP), the $800 billion program introduced during the COVID-19 pandemic to support small businesses. We illustrate what can go wrong using data presented in Chernenko and Scharfstein (2023), which links firm owners to voter records in four states that provide race information on voters. We show that the theoretical estimation bias of unconditional disparities is a function of average outcomes for different groups, algorithm accuracy, and the relative population shares of different groups. Given typical values of population shares and accuracy measures (precision and recall), we estimate that unconditional disparities between Black and white households should be biased towards zero by around 30–50%. The theoretical attenuation bias is particularly large for Black borrowers because race prediction algorithms are generally less accurate for Black borrowers than for Hispanic and Asian borrowers. This estimate is based on the assumption that prediction errors are uncorrelated with individual characteristics. We show that, in practice, this assumption is violated, particularly for Black borrowers: relative to Black borrowers who are incorrectly classified as non-Black, correctly classified Black borrowers tend to have characteristics that reduce PPP access, such as living in lower-income locations or lacking prior bank borrowing relationships. Thus, the attenuation bias in estimated unconditional disparities, while still sizeable, is typically less than predicted for Black borrowers.

Conditional racial disparities, which typically condition on where a borrower lives or works in addition to other characteristics, are also biased toward zero in practice. Researchers often condition on the location of the borrower to understand how much of unconditional racial disparities can be explained by location rather than the race or ethnicity of the borrower. However, race prediction algorithms also use location information to achieve significantly higher precision and recall than algorithms that are purely based on names, particularly for Black Americans. This improvement in prediction accuracy is largely illusory when used to estimate racial disparities conditional on borrower location. Including location fixed effects, say ZIP code, effectively discards much of the predictive power of location in race prediction algorithms. This means that when researchers include location fixed effects, the extent of bias towards zero due to measurement error in predicted race is accentuated. An important implication of this is that researchers and policymakers are likely to attribute an excessive share of unconditional disparities to differences in location, not race. Indeed, in a large sample of potential PPP borrowers, using predicted race in regressions of PPP take-up for Black-owned businesses within a Census block underestimates the true racial disparity by about 30%.

We also show, both theoretically and empirically, that prediction errors in racial identity can inflate estimated disparities. Race-prediction algorithms often misclassify Black and Hispanic borrowers as white. Given that Black and Hispanic borrowers tend to have less credit access, this reduces the measured outcomes for white borrowers; this is part of the reason why disparities for Black and Hispanic borrowers are underestimated. However, this has implications for the estimation of disparities for Asian borrowers. Since the credit outcomes for white borrowers are underestimated, credit outcomes for Asian borrowers — who are less likely to be confused with white borrowers because their names are more distinctively Asian — will look relatively favorable. Thus, measurement error for Black and Hispanic borrowers can make it seem that Asian borrowers have greater access to credit than white borrowers. We present evidence that this is not just a theoretical possibility. Using actual race, Asian-owned firms are unconditionally 3.7 percentage points more likely than white-owned firms to access PPP; this difference increases to 5.2–5.5% when using race-prediction algorithms.

Finally, we examine the use of predicted race to estimate interaction effects that are meant to measure cross-sectional variation in the magnitude of disparities. For example, researchers may be interested in knowing whether disparities in access to credit are smaller for firms with existing borrowing relationships or whether disparities are larger in areas with greater racial bias. We derive closed-form expressions for the estimated coefficients when race is interacted with binary explanatory variables such as whether a firm has a bank borrowing relationship. The estimated interaction terms depend on the true disparities, the relative accuracy of predicted race within the subsamples defined by the binary variable, and the relative shares of different races within the subsamples. Thus, the estimated interaction terms may be biased either downward or upward relative to the true effect, with the direction and magnitude of bias difficult to estimate a priori without knowing both the relative accuracy of the algorithm in the subsamples and the true disparities in the subsamples.

In our empirical analyses of PPP take-up, interactions with measures of existing bank and nonbank borrowing relationships are generally biased downward by 10–40%. But even when bias in measuring interaction terms is small, using predicted race can lead to misleading conclusions. Because predicted race tends to significantly underestimate average disparities, estimates of interaction terms may overstate the relative importance of firm characteristics in explaining cross-sectional variation in disparities. In our analyses, algorithms that deliver the least biased estimate of the interaction of Black with existing bank borrowing relationships suggest that there are no disparities in PPP take-up for firms with such relationships. In reality, when using self-reported race, we find that while bank borrowing relationships significantly mitigate disparities, they do not eliminate them.

This paper is organized as follows. In the next section, we present a simple model that derives estimates of the bias stemming from the use of noisy estimates of race, under the assumption that prediction errors are uncorrelated with individual characteristics. In Section 3, we describe the two algorithms we use to predict race – the Bayesian Improved First Name Surname Geocoding (BIFSG) algorithm and random forest machine learning model, which we train on voter records containing the voter's race or ethnic identity. Section 4 uses these race prediction algorithms to examine disparities in PPP take-up by restaurants in Florida. We compare the estimated disparities based on these algorithms to disparities based on the restaurant owner's actual race as self-reported in voter records, which is the data we used in Chernenko and Scharfstein (2023). Section 5 expands the analysis to a broader sample of potential PPP recipients in three additional states – Georgia, Louisiana, and North Carolina. This broader sample is composed of firms that received EIDL Advance grants through the Small Business Administration's Economic Injury Disaster Loans (EIDL) program. The larger sample allows us to examine the reliability of interaction term estimates. In addition, by analyzing a broader range of geographies, we demonstrate that there can be very different levels of measurement bias even in what appear to be fairly similar sets of firms. Section 6 concludes.

2 Model

In this section, we derive estimates of bias in measured disparities as a function of true disparities, the accuracy of race prediction algorithms, and population shares of different racial groups. Our model assumes that prediction errors are uncorrelated with individual characteristics that are correlated with the outcomes of interest. In our empirical analyses, we will show that estimates of unconditional disparities based on race prediction algorithms tend to be closer to true disparities than would be predicted by the model. The reason for this is that the prediction errors are correlated with individual characteristics that are correlated with outcomes. In particular, the algorithms are more likely to correctly identify Black and Hispanic individuals if they have worse outcomes.

Consider a setting with N types of agents. We will be thinking of racial/ethnic groups, but the results apply to any setting where agent types are not observed and have to be inferred. The share of type i is α_i. The average outcome for an agent of type i is μ_i. We are interested in measuring and understanding differences in outcomes relative to some benchmark type 0.

Although true type is not observable to us as econometricians, suppose we have an imperfect algorithm for predicting the type of each agent. Let

$$\gamma_j^i = Pr(\text{true type} = i | \text{predicted type} = j). \tag{1}$$

Then the average outcome for those predicted to be of type i is

$$\hat{\mu}_i = \frac{\sum_j \alpha_j \gamma_i^j \mu_j}{\sum_j \alpha_j \gamma_i^j} = \sum_j w_i^j \mu_j, \tag{2}$$

a weighted average of outcomes for different types with the weight, w_i^j, being type j's share of all agents predicted by the algorithm to be of type i. This weight depends on the algorithm's accuracy and on the population shares of different types.

The measured difference in average outcomes relative to benchmark type 0 is then

$$\hat{\delta}_i = \hat{\mu}_i - \hat{\mu}_0 = \frac{\sum_j \alpha_j \gamma_i^j \mu_j}{\sum_j \alpha_j \gamma_i^j} - \frac{\sum_j \alpha_j \gamma_0^j \mu_j}{\sum_j \alpha_j \gamma_0^i} = \sum_j (w_i^j - w_0^j) \mu_j. \tag{3}$$

If an algorithm is perfectly accurate, then $w_i^j = 0$ for $i \neq j$ and $w_i^j = 1$ for $i = j$. In this case, the measured difference is equal to the true difference $\mu_i - \mu_0$.

With imperfect algorithms, the measured difference differs from the true difference and is not guaranteed to have the same sign as the true difference. Consider, for example, a setting with three groups: the benchmark group, $G1$, and two minority groups, $G2$, and $G3$. Suppose there is no difference in outcomes for the benchmark group and $G2$. Furthermore, suppose that our algorithm has perfect or close to perfect accuracy in identifying $G2$. Then the measured outcome for $G2$ will be close to its true value. However, the measured outcome for the benchmark group will be biased because some of the agents predicted to belong to the benchmark group will actually belong to $G3$, which has a different average outcome. As a result, $G2$ will appear to have a disparity relative to the benchmark group. This disparity will have the opposite sign of the disparity for $G3$. To make this concrete, suppose the groups are white, Asian, and Black. If Black individuals have worse outcomes than white individuals, and if the algorithm is very good at identifying Asian individuals but struggles to identify Black individuals, then the measured outcome for Asian individuals will be better than for white individuals, even if there is no actual disparity between the Asian and white samples.

2.1 Two Groups

It is helpful to consider the simplified case with just two groups. Motivated by our empirical analysis, we will focus on two groups, one that is Black (b) and the other that is white (w). In this case, we can characterize our algorithm using its precision γ and recall θ. In other words, out of individuals predicted to be Black, share γ are actually Black. Out of individuals who are actually Black, share θ are predicted to be Black. For simplicity and without loss of generality, assume that the average outcome for the white group is zero.

The average outcome for those predicted to be Black can be written as

$$\hat{\mu}_b = \frac{\mu_b \times TP + 0 \times FP}{TP + FP} = \mu_b \times \frac{1}{1 + \frac{FP}{TP}}, \tag{4}$$

where TP is the number of true positives and FP is the number of false positives. Noting that precision $\gamma = \frac{TP}{TP+FP} = \frac{1}{1+\frac{FP}{TP}}$, we can write the estimated average outcome for those predicted to be Black simply as

$$\hat{\mu}_b = \gamma \mu_b. \tag{5}$$

Thus, the estimated average is the true average scaled down by the algorithm's precision.

The average outcome for those predicted to be white is

$$\hat{\mu}_w = \frac{0 \times TN + \mu_b \times FN}{TN + FN} = \frac{1}{1 + \frac{TN}{FN}} \times \mu_b, \tag{6}$$

where TN is the number of true negatives (whites predicted to be white) and FN is the number of false negatives (Blacks predicted to be white). This can be written as

$$\hat{\mu}_w = \frac{1}{1 + \frac{TN}{FN}} = \frac{\alpha\gamma(1-\theta)}{\gamma - \alpha\theta} \times \mu_b. \tag{7}$$

Holding constant algorithm accuracy, the term multiplying μ_b, is increasing in the Black population share α. As the algorithm misclassifies more Black individuals as white, the estimated average outcome for whites is biased more toward the average outcome for Blacks.

The estimated disparity — the difference between the estimated averages for Black and white individuals — is then

$$\hat{\delta} = \hat{\mu}_b - \hat{\mu}_w = \frac{\gamma(\gamma - \alpha)}{\gamma - \alpha\theta} \times \mu_b. \tag{8}$$

Note again that measured disparity is not guaranteed to have the same sign as the true disparity, though in the two-group case, given that α is likely to be small, the measured disparity is likely to have the same sign as the true disparity.

We will report our predicted bias in measured disparities when we report our empirical findings, but for now, it is helpful to consider ballpark estimates. Among registered voters in Florida, Georgia, Louisiana, and North Carolina, BIFSG has precision γ and recall θ for identifying Blacks of around 75%. Black individuals account for about 25% of the share of the population in these states that are either Black or white. This means that estimates of unconditional disparity can be expected to be biased down by about one-third.[1]

2.2 Interactions

We can extend the model to consider interactions with binary variables. For example, in our empirical analyses of PPP we are interested in whether firms with existing bank relationships are more likely to receive PPP loans and whether the effect of bank relationships on PPP take-up is stronger for Black-owned firms than white-owned firms. In other words, do existing bank relationships help reduce the disparity in PPP take-up?

[1] $1 - \frac{0.75 \times (0.75 - 0.25)}{0.75 - 0.25 \times 0.75} = \frac{1}{3}$.

Thus, we are interested in estimating the following model

$$Y_i = \sum_g \beta_g \cdot I(i \in g) + \sum_g \phi_g \cdot I(i \in g) \times I(i \in c) + \varepsilon_i, \tag{9}$$

where g refers to racial and ethnic groups and where $c \in 0, 1$ indicates membership in some other binary group split, for example having a bank loan. We are interested in the difference in the coefficients $\phi_g - \phi_0$.

The key to estimating the interaction term is recognizing that it can be calculated as the difference in the estimated disparities within the two subsamples defined by the binary variable c. Using Eq. 8 for the estimated disparity in the case of two groups, the interaction between Black and bank loan would be estimated as

$$\hat{\phi}_b - \hat{\phi}_w = \frac{\gamma_1(\gamma_1 - \alpha_1)}{\gamma_1 - \alpha_1\theta_1}\delta_1 - \frac{\gamma_0(\gamma_0 - \alpha_0)}{\gamma_0 - \alpha_0\theta_0}\delta_0, \tag{10}$$

where 1 and 0 subscripts indicate the subsamples with and without bank loans. Thus, the interaction term depends on the true disparities in the two subsamples, δ_1 and δ_0, the relative accuracy of the algorithm in the two subsamples, γ_i and θ_i terms, and the relative population shares within the two subsamples, α_i. While it is difficult to make definitive statements regarding how the estimated interaction effect compares with its true magnitude, $\delta_1 - \delta_0$, we can gain some insight by considering a few special cases.

First, if there is no disparity within the first subsample, i.e., if $\delta_1 = 0$, then the estimated interaction is the negative of the estimated disparity in the second subsample. Thus even though we underestimate the magnitude of disparities in the second subsample, we correctly conclude that there are no disparities in the first subsample.

Next, suppose that the algorithm has the same accuracy within the two subsamples and that racial composition is also the same across the two subsamples, $\alpha_1 = \alpha_0$. Then the estimated interaction term is biased towards zero by the same magnitude as in Eq. 8 above:

$$\hat{\phi}_b - \hat{\phi}_w = \frac{\gamma(\gamma - \alpha)}{\gamma - \alpha\theta}\left(\phi_b - \phi_w\right). \tag{11}$$

Finally, suppose that the algorithm is more accurate in the second subsample than in the first one. For example, the sample of firms without bank loans may contain more firms from neighborhoods with a larger minority share. These are the firms that the algorithm may be better at classifying as Black. In the extreme, suppose the algorithm has perfect accuracy within the second subsample. Then the estimated interaction term is

$$\hat{\phi}_b - \hat{\phi}_w = \frac{\gamma_1(\gamma_1 - \alpha_1)}{\gamma_1 - \alpha_1\theta_1}\delta_1 - \delta_0 = \phi_b - \phi_w + \left(1 - \frac{\gamma_1(\gamma_1 - \alpha_1)}{\gamma_1 - \alpha_1\theta_1}\right)\delta_1. \tag{12}$$

Since the term in parentheses is generally positive, we will tend to overestimate the interaction effect. Even if our estimates of the interaction term are close in magnitude to the true effect, we will tend to overestimate the importance of the interaction effect relative to the direct effect or baseline disparity, which will be significantly underestimated.

3 Algorithms

We use two different samples to illustrate the performance of five variations of two different algorithms to predict an individual's racial or ethnic identity.

The first algorithm is the Bayesian Improved First Name Surname Geocoding algorithm (Voicu, 2018). This is a commonly used algorithm for predicting racial and ethnic identity, although the Consumer Financial Protection Bureau and many lenders appear to rely on BISG, which does not use the individual's first name. BIFSG has reasonable accuracy, at least within voter registration and mortgage application data, is easy to implement, and is intuitive and transparent. The algorithm uses an individual's first and last names, along with address, to estimate the probability that the individual belongs to a particular racial or ethnic group. Specifically, the probability that an individual with first name f and surname s, who lives in location l belongs to group r is

$$p(r|f, s, l) = \frac{p(s) \cdot p(r|s) \cdot p(l|r) \cdot p(f|r)}{\sum_i p(s) \cdot p(i|s) \cdot p(l|i) \cdot p(f|i)} \tag{13}$$

The key assumption is conditional independence among first names, surnames, and location, i.e., $p(l|r, s) = p(l|r)$ and $p(f|r, s, l) = p(f|r)$.

The original algorithm uses six groups: Hispanic, non-Hispanic white, non-Hispanic Asian and Pacific Islander, non-Hispanic Black, non-Hispanic American Indian and Alaskan Native, and other/multiracial. Because there are very few non-Hispanic American Indian and Alaskan Native individuals in our data, we include them in the other category along with multiracial individuals.

Estimates of the conditional probabilities of having first name f conditional on belonging to group r, and of belonging to group r conditional on having surname s come from the list of surnames in the 2010 Census and the list of first names assembled by Tzioumis (2018). The list of surnames from the 2010 Census includes all surnames that occur at least 100 times and reports the share of individuals with each surname who self-identify as belonging to different racial/ethnic groups. Tzioumis (2018) uses mortgage application data to compile a list of 4,250 first names with racial shares for each name. Both lists report racial shares for all other surnames and first names that are not broken out individually. We use these shares for the names that cannot be matched to the lists.[2]

We consider two implementations of BIFSG, with one using ZIP-code level population shares of racial groups and the other using Census block group (CBG) level population shares. BIFSG with CBG-level population shares (BIFSG CBG) has greater accuracy but also tends to be more strongly correlated with firm characteristics that are associated with worse outcomes. As a result, BIFSG CBG tends to generate larger estimates of unconditional disparities that are closer to true disparities. But once we control for CBG fixed effects and firm characteristics, BIFSG ZIP and BIFSG CBG tend to deliver similar results.

Posterior probabilities from Eq. 13 are converted into discrete predictions using the maximum a posteriori (MAP) classification scheme. In the other words, we consider an algorithm to classify someone as belonging to group r if that group has the largest posterior probability.

Our second algorithm is a random forest (RF) algorithm that is trained on voter registration data from Florida, Georgia, Louisiana, and North Carolina and that uses as its features essentially the same variables as BIFSG: Census block group level population shares and conditional probabilities of belonging to racial/ethnic groups conditional on having different surnames and first names. The advantage of the random forest algorithm over BIFSG is that it relaxes the assumption of conditional independence among first names, surnames, and locations.

We train our random forest algorithms using `sklearn.ensemble.RandomForestClassifier` Python module, with the default parameters except for using 250 instead of 100 trees in a forest. We make sure to exclude voters who are officers of Florida restaurants, PPP recipients, or EIDL Advance recipients. This ensures that all of our tests are out-of-sample.[3]

We train three different versions of the random forest algorithm. The first two use ZIP-level and CBG-level population shares, just as we did in our BIFSG implementation. The third random forest algorithm uses names only, with no geographic information. While this algorithm is significantly less accurate and is very unlikely to be used in practice, we include it to show that increases in prediction accuracy from using location information are not completely undone by the inclusion of location fixed effects in the empirical analyses.

[2] We find similar results if we drop individuals whose names cannot be matched to the two lists. Dropping these individuals however reduces the restaurant by about 30%.

[3] This is why we cannot use off-the-shelf algorithms as many of them are trained using Florida or North Carolina voter data.

4 PPP Take-Up by Florida Restaurants

In this section, we analyze racial disparities in the take-up of PPP loans by Florida restaurants studied by Chernenko and Scharfstein (2023). Disparities in the PPP have received a lot of attention, making this an interesting setting in which to understand how conclusions about racial disparities and their sources depend on the use of actual versus predicted race. The sample in Chernenko and Scharfstein (2023) consists of Florida restaurants that were licensed as of February 15, 2020, and that were owned by Florida-based for-profit firms whose owner's racial identity can be determined from Florida voter registration data.[4] We match restaurant licenses to Yelp for an extensive list of restaurant characteristics to be used as controls in the regressions of PPP take-up. Hotel and franchise restaurants are excluded as they are frequently owned and operated by affiliated entities, which can make it difficult to determine whether a given restaurant received a PPP loan.

Our sample is about 10% smaller than in Chernenko and Scharfstein (2023) for two reasons. First, we restrict the sample to firms whose officers can be matched to a unique voter in voter registration data, while Chernenko and Scharfstein (2023) include cases with multiple potential matches in voter registration data where all potential matches agree on race. Second, we are unable to get Census block group information for all of the officers in our data.

Table 1 reports the accuracy of the different algorithms in this sample.[5] All algorithms are significantly better at identifying Asian and Hispanic restaurant owners than at identifying Black restaurant owners. For Asian and Hispanic owners, even the random forest algorithm that uses names only has recall of about 85% and precision of 82% precision. For Black owners, the best performing algorithm, RF CBG, has only 62% recall and 60% precision.

Using more granular location information — CBG instead of ZIP — leads to sizable improvements in the precision with which the algorithms identify Black restaurant owners, but not the recall. For example, going from BIFSG ZIP to BIFSG CBG, precision improves from 46.30% to 55.41% while recall only improves from 63.16% to 64.04%. Knowing that someone lives in a Census block group with a majority Black population gives us confidence that this person is likely to be Black. But because many Black individuals do not live in such neighborhoods and do not have distinctly Black names, the algorithm fails to identify them. Having more granular location information has essentially no effect on the ability to identify

[4] While officers and directors listed in corporate records may manage a firm without having any ownership, this seems unlikely for the types of small firms we study. Therefore, for simplicity, we refer to the individuals listed in corporate records as owners.

[5] For brevity, we do not report the other category which is small and has very low accuracy.

Table 1
Accuracy of Different Algorithms in the Florida Restaurants Sample

This table reports the accuracy of different algorithms in predicting the race of Florida restaurant owners. BIFSG is the Bayesian Improved First Name Surname Geocoding algorithm (Voicu, 2018). We use either ZIP or Census block group (CBG) level data on the population shares of different racial/ethnic groups in the 2020 Census. The three random forest (RF) algorithms are trained using data on the distribution of first and last names across racial/ethnic groups and on either ZIP or CBG level population shares. In each cell, the first row reports the algorithm's recall (the share of all observations in the group that are correctly identified by the algorithm); the second row reports the algorithm's precision (the share of predicted observations that are true).

	BIFSG		Random Forest		
	ZIP	CBG	ZIP	CBG	Names
Asian	72.48	73.72	68.76	70.88	66.86
	79.25	78.60	84.03	83.13	83.20
Black	63.16	64.04	59.65	61.84	59.21
	46.30	55.41	49.36	60.13	20.55
Hispanic	89.90	89.71	88.11	88.79	84.90
	78.74	78.74	76.42	77.70	82.27
White	90.58	91.69	91.76	92.85	86.56
	88.96	89.31	88.47	88.87	89.88

Asian and Hispanic restaurant owners. For example, the random forest algorithm with CBG-level population distributions has 70.88% recall and 83.13% precision in identifying Asian restaurant owners, while the random forest with ZIP-level distributions has 68.76% recall and 84.03% precision.

Finally, when identifying Black restaurant owners, the recall of the random forest algorithm tends to be similar to that of BIFSG but the random forest algorithm has higher precision. While BIFSG CBG has 55.41% precision for Black restaurant owners, RF CBG has 60.13% precision.

Table 2 reports the results of linear probability model regressions of PPP take-up on actual versus predicted race. We include dummies for Black, Hispanic, Asian, and Other, but to conserve space report the coefficients on only the first three. The coefficients can be interpreted as disparities relative to white-owned restaurants. Columns 1–4 of Panel A report the results using self-reported race from voter registration. According to the results in column 1, Black, Hispanic-, and Asian-owned restaurants are 26.2pp, 10.5pp, and 2.5pp less likely to receive PPP loans than white-owned restaurants.

Column 2 of Table 2 adds ZIP code fixed effects. Disparities for Black-, Hispanic-,

Table 2
Actual versus Predicted Race in Analysis of PPP Take-up by Florida Restaurants

This table reports the results of linear probability model regressions of the take-up of PPP loans by Florida restaurants. The sample is smaller than in Chernenko and Scharfstein (2023) because a) we require an exact match in voter registration while Chernenko and Scharfstein (2023) include cases where multiple potential matches agree on race, and b) we do not have Census block group information for all voters. BIFSG is the Bayesian Improved First Name Surname Geocoding algorithm (Voicu, 2018). We use either ZIP or Census block group (CBG) level data on the population shares of different racial/ethnic groups in the 2020 Census. The three random forest (RF) algorithms are trained using data on the distribution of first and last names across racial/ethnic groups and on either ZIP or CBG level population shares. Columns 5, 10, 15, 20, and 25 use Eq. 3 to calculate what measured disparity would be if the algorithm's prediction errors were random. Firm characteristics are the controls used by Chernenko and Scharfstein (2023). ZIP code and Census block group (CBG) FEs are based on the firm owner's residential address in voter registration data. Robust standard errors are reported. *, **, and *** indicate statistical significance at 10%, 5%, and 1%. $N = 11,530$.

Panel A

	Voter registration				BIFSG with ZIP					BIFSG with CBG				
					Eq. 3	Empirical estimates				Eq. 3	Empirical estimates			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Black	−0.262***	−0.214***	−0.164***	−0.088*	−0.120	−0.167***	−0.092***	−0.069	−0.036	−0.145	−0.207***	−0.118***	−0.076	−0.045
	(0.024)	(0.027)	(0.050)	(0.049)		(0.021)	(0.024)	(0.047)	(0.046)		(0.022)	(0.026)	(0.054)	(0.051)
Hispanic	−0.105***	−0.085***	−0.080***	−0.067***	−0.076	−0.078***	−0.051***	−0.037	−0.032	−0.077	−0.086***	−0.063***	−0.042*	−0.039
	(0.012)	(0.014)	(0.026)	(0.025)		(0.011)	(0.014)	(0.025)	(0.024)		(0.011)	(0.014)	(0.025)	(0.024)
Asian	−0.025*	−0.013	−0.025	−0.005	−0.020	−0.012	0.005	−0.039	−0.017	−0.020	−0.018	−0.004	−0.043	−0.022
	(0.013)	(0.015)	(0.027)	(0.026)		(0.014)	(0.015)	(0.028)	(0.027)		(0.014)	(0.015)	(0.028)	(0.027)
Adjusted R^2	0.018	0.046	0.207	0.261		0.010	0.039	0.204	0.259		0.012	0.040	0.204	0.259
ZIP FEs		✓					✓					✓		
CBG FEs			✓	✓				✓	✓				✓	✓
Firm controls				✓					✓					✓

Panel B

	RF with ZIP					RF with CBG					RF with name only				
	Eq. 3	Empirical estimates				Eq. 3	Empirical estimates				Eq. 3	Empirical estimates			
	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)	(27)	(28)	(29)
Black	−0.129	−0.182***	−0.109***	−0.120**	−0.089*	−0.156	−0.214***	−0.139***	−0.085	−0.052	−0.058	−0.067***	−0.052***	−0.036	−0.020
		(0.022)	(0.025)	(0.049)	(0.047)		(0.024)	(0.027)	(0.057)	(0.053)		(0.014)	(0.015)	(0.028)	(0.027)
Hispanic	−0.074	−0.088***	−0.067***	−0.051**	−0.046*	−0.076	−0.088***	−0.066***	−0.041	−0.040*	−0.078	−0.091***	−0.067***	−0.054**	−0.048**
		(0.011)	(0.014)	(0.025)	(0.024)		(0.011)	(0.014)	(0.025)	(0.024)		(0.012)	(0.014)	(0.024)	(0.023)
Asian	−0.019	−0.020	−0.003	−0.054*	−0.033	−0.019	−0.019	−0.006	−0.048*	−0.030	−0.018	−0.014	0.001	−0.039	−0.017
		(0.014)	(0.016)	(0.029)	(0.028)		(0.014)	(0.016)	(0.028)	(0.027)		(0.015)	(0.016)	(0.029)	(0.028)
Adjusted R^2		0.011	0.040	0.205	0.260		0.012	0.041	0.204	0.259		0.006	0.040	0.204	0.259
ZIP FEs			✓					✓					✓		
CBG FEs				✓	✓				✓	✓				✓	✓
Firm controls					✓					✓					✓

and Asian-owned restaurants are reduced in magnitude to 21.4pp, 8.5pp, and 1.3pp, with the coefficient on Asian losing its statistical significance. For Black- and Hispanic-owned restaurants, this represents about 20% reduction relative to the estimates in column 1. Column 3 replaces ZIP code with Census block group (CBG) fixed effects. The magnitude of disparities for Black- and Hispanic-owned restaurants is reduced further to 16.4pp and 8.0pp. These results suggest that location can explain up to about one-third of the unconditional disparities.

Column 4 adds a large number of firm characteristics used by Chernenko and Scharfstein (2023). These include the log of firm age, log number of seats, log number of Yelp reviews, average rating, and dummies for whether the firm had secured loans from bank and nonbank lenders. We identify secured loans from state registry of lien filings by lenders under the Uniform Commercial Code (UCC). Most loans to firms that are not backed by real estate have such filings, and we refer to them as UCC loans. Using these controls, Black-owned restaurants face an 8.8pp conditional disparity, or about one-third of the unconditional disparity reported in column 1. Hispanic-owned restaurants face a 6.7pp conditional disparity, about two-thirds of the unconditional disparity in column 1.

Columns 5–9 of Panel A use race predicted by the BIFSG algorithm with ZIP code level population shares. To begin with, column 5 reports the prediction of Eq. 3 for what we should expect measured disparities to be given the accuracy of the algorithm and assuming that prediction errors are uncorrelated with firm characteristics and outcomes. We would expect measured disparities for Black, Hispanic, and Asian-owned restaurants to be 12.0pp, 7.6pp, and 2.0pp. With the exception of Asian-owned restaurants, the actual estimates in column 6 are larger in magnitude: 16.7pp for Black-owned restaurants and 7.8pp for Hispanic-owned restaurants. Using this algorithm, the analysis indicates that there is no unconditional disparity for Asian-owned restaurants.

Columns 7 and 8 of Table 2 add ZIP and CBG fixed effects. We see large reductions in the magnitudes of the coefficients on Black and Hispanic. The magnitude of disparities for Black-owned restaurants drops from 16.7pp in column 6 to 9.2pp with ZIP FEs in column 7 and 6.9pp with CBG FEs in column 8. Once we also include firm characteristics in column 9, none of the estimated coefficients are statistically significant. The disparity for Black-owned restaurants is only 3.6pp or 41% of the conditional disparity of 8.8pp in column 4 which uses actual race from voter records. For Hispanic restaurant owners, the estimated conditional disparity in column 9 is 3.2pp or 48% of the 6.7pp estimate in column 4.

Columns 10–14 of Table 2 use race predicted by the BIFSG algorithm with CBG-level population shares. Column 10 shows that according to Eq. 3, we would expect the measured unconditional disparity for Black-owned restaurants to be around 14.5pp. In fact, the estimate in column 11 indicates a 20.7pp disparity. This is almost 50% larger than what we would expect if the algorithm's prediction errors were random, and it is 20% smaller than the true disparity of 26.2pp in column 1. Once we include ZIP and CBG fixed effects, however, we see dramatic decreases in the magnitude of measured disparities. With CBG fixed effects in column 13, the estimated disparity for Black-owned restaurants is 7.6pp, substantially below the estimate of 16.4pp in column 3. Thus, using predicted race would lead us to dramatically underestimate the magnitude of within-CBG disparities and to significantly overestimate the role of location in explaining unconditional disparities. While CBGs explain about one-third of the true unconditional disparity of 26.2pp for Black-owned restaurants, the use of predicted race in columns 13 and 11 would suggest that CBGs explain about two-thirds of the unconditional disparity.

Panel B of Table 2 reports the results using race predicted by the three random forest algorithms. Columns 15–19 report the results of RF ZIP, the random forest algorithm that uses ZIP code-level distributions. RF ZIP has broadly similar accuracy as BIFSG ZIP, and the prediction of Eq. 3 is of 12.9pp disparity when using RF ZIP versus 12.0pp disparity when using BIFSG ZIP. In fact, the estimates in columns 16–19 are significantly larger in magnitude, and the estimate of conditional disparities in column 19 happens to be essentially the same as in column 4. This is despite the fact that estimates of unconditional disparities and of within ZIP and CBG disparities are much smaller than their true values.

Columns 20–24 of Table 2 report the results using race predicted by RF CBG, the random forest algorithm with CBG-level population distributions. RF CBG has the highest accuracy of the five algorithms. Nevertheless, Eq. 3 predicts a disparity for Black-owned restaurants of 15.6pp or 60% of the estimate based on self-reported race. Its actual estimate of the unconditional disparity is 21.4pp or 82% of its true value. Its estimates of within CBG disparities in column 23 and of conditional disparities in column 24 are significantly smaller, however.

Finally, columns 25–29 of Table 2 report the results using race predicted by the random forest algorithm that uses only name information. This algorithm performs very poorly when used to estimate disparities for Black restaurant owners. It suggests that unconditional disparities are 25% of their true value and that there are no conditional disparities. The algorithm performs relatively well in estimating disparities for Hispanic-owned restaurants, which is not surprising given that according to Table 1, using location does not necessarily improve the ability to identify Hispanic-owned restaurants.

Overall, the results in Table 2 indicate that analyses that use predicted race must be interpreted with great caution. In particular, although algorithms that use granular location information such as BIFSG CBG and RF CBG may deliver similar unconditional disparities, they can lead researchers to significantly overestimate the significance of location in driving disparities and to significantly underestimate the magnitude of disparities within narrow geographies.

4.1 Prediction Errors, Firm Characteristics, and Disparities in Outcomes

The estimated disparities for Black-owned restaurants based on algorithms that use location information tend to be significantly larger than what would be predicted by Eq. 3 if prediction errors were random. This is especially the case when using more granular location information. For example, the disparity estimate based on BIFSG CBG is 20.7pp for Black-owned restaurants, while Eq. 3 predicts a lower disparity of 14.5pp. This suggests that prediction errors are correlated with firm characteristics and with outcomes. Intuitively, the algorithms tend to identify as Black those who live in predominately Black neighborhoods and have more distinctly Black names, which are likely to be correlated with worse socioeconomic outcomes (Fryer and Levitt, 2008).

We examine the correlation between prediction errors, firm characteristics and outcomes in two ways. First, in Table 3 we estimate regressions of PPP take-up on actual race interacted separately with an indicator for whether predicted race matches actual race and an indicator for whether it does not match actual race. For example, we interact Black with *Predicted Black* and with *Predicted not Black*. These analyses show that, especially when algorithms use granular location information, Black-owned businesses that are identified as Black-owned by the algorithm have much worse outcomes than Black-owned businesses that are not identified as Black-owned by the algorithm. Furthermore, once we include controls and thus estimate conditional disparities, there are no differences between Black-owned restaurants that are identified by the algorithm as Black-owned and those that are not identified as Black-owned. Second, in Table 4 we estimate similar regressions but with various firm characteristics as the dependent variables. These regressions show that disparities in firm characteristics tend to be significantly larger for minorities who are identified as such by the algorithms than for minorities whom the algorithms classify as white.

Table 3 reports the results of regressions of PPP take-up on actual race interacted with the algorithm's prediction. These are meant to measure outcomes for minority restaurant owners who are classified as such by the algorithms versus minority restaurant owners whom the algorithms misclassify, which is typically a misclassification as a white restaurant owner Columns 1–5 report the results of unconditional disparities, while columns 6–10 control for CBG fixed effects and firm characteristics. Column 1 reports the results for the BIFSG ZIP

Table 3
PPP Take-up by Florida Restaurants on Actual and Predicted Race

This table reports the results of linear probability model regressions of the take-up of PPP loans by Florida restaurants on both actual and predicted race at the same time. BIFSG is the Bayesian Improved First Name Surname Geocoding algorithm (Voicu, 2018). We use either ZIP or Census block group (CBG) level data on the population shares of different racial/ethnic groups in the 2020 Census. The three random forest (RF) algorithms are trained using data on the distribution of first and last names across racial/ethnic groups and on either ZIP or CBG level population shares. ZIP code and Census block group (CBG) FEs are based on the firm owner's residential address in voter registration data. Firm characteristics are the controls used by Chernenko and Scharfstein (2023). Robust standard errors are reported. *, **, and *** indicate statistical significance at 10%, 5%, and 1%. $N = 11,530$.

	Unconditional					Conditional				
	BIFSG		RF			BIFSG		RF		
	ZIP	CBG	ZIP	CBG	Names	ZIP	CBG	ZIP	CBG	Names
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Black × Predicted Black	−0.292***	−0.315***	−0.302***	−0.307***	−0.262***	−0.066	−0.122*	−0.070	−0.067	−0.076
	(0.030)	(0.029)	(0.031)	(0.030)	(0.031)	(0.066)	(0.067)	(0.067)	(0.070)	(0.061)
Black × Predicted not Black	−0.210***	−0.167***	−0.202***	−0.188***	−0.261***	−0.112*	−0.056	−0.108	−0.111*	−0.110
	(0.039)	(0.039)	(0.037)	(0.038)	(0.037)	(0.068)	(0.065)	(0.067)	(0.063)	(0.071)
Hispanic × Predicted Hispanic	−0.102***	−0.106***	−0.107***	−0.106***	−0.108***	−0.059**	−0.061**	−0.068***	−0.062**	−0.060**
	(0.012)	(0.012)	(0.012)	(0.012)	(0.013)	(0.025)	(0.026)	(0.026)	(0.026)	(0.026)
Hispanic × Predicted not Hispanic	−0.139***	−0.103***	−0.091***	−0.099***	−0.090***	−0.116*	−0.111*	−0.061	−0.094	−0.105*
	(0.034)	(0.033)	(0.031)	(0.032)	(0.028)	(0.062)	(0.062)	(0.058)	(0.058)	(0.055)
Asian × Predicted Asian	−0.026*	−0.027*	−0.025	−0.027*	−0.026	−0.018	−0.016	−0.030	−0.033	−0.020
	(0.015)	(0.015)	(0.016)	(0.015)	(0.016)	(0.030)	(0.030)	(0.030)	(0.030)	(0.030)
Asian × Predicted not Asian	−0.024	−0.020	−0.025	−0.021	−0.024	0.029	0.025	0.051	0.067	0.026
	(0.024)	(0.024)	(0.022)	(0.023)	(0.022)	(0.046)	(0.046)	(0.043)	(0.045)	(0.043)
Adjusted R^2	0.018	0.018	0.018	0.018	0.017	0.261	0.261	0.261	0.262	0.261
CBG FEs						✓	✓	✓	✓	✓
Firm controls						✓	✓	✓	✓	✓

algorithm. The coefficient on *Black* × *Predicted Black* indicates a 29.2pp disparity, while the coefficient on *Black* × *Predicted not Black* indicates a smaller 21.0pp disparity. The difference increases once we consider algorithms that use CBG-level information. With BIFSG CBG, the disparity for Black-owned restaurants that are predicted to be Black-owned (31.5pp) is almost twice as large as the disparity for Black-owned restaurants that the algorithm fails to classify as Black-owned (16.7pp).

We find similar results in columns 3–4 when using the random forest algorithms with ZIP-level and CBG-level population shares. When we use the random forest algorithm with names only, on the other hand, we find that estimated disparities depend only on actual race and not on whether someone was predicted to be Black or not. This suggests that prediction errors of the name-only algorithm are uncorrelated with firm characteristics and outcomes.

Columns 5–10 of Table 3 control for census block group fixed effects and firm characteristics. Although there is some variation, we find similar disparities for those correctly identified by the algorithms and those that are not. This suggests that the effect of predicted race in columns 1–5 is due to its correlation with firm characteristics.

Table 4 reports the results of similar regressions but using ZIP and firm characteristics instead of PPP take-up as the dependent variables. For brevity, we report the results for four characteristics: ZIP-level bank branches per capita, log of median household income at the ZIP code level, log number of seats, and a dummy for whether the firm had a bank UCC loan.

In columns 1–5 of Panel A, the dependent variable is bank branches per capita. In column 1, the coefficient on *Black* × *Predicted Black* is -0.202 versus -0.145 for *Black* × *Predicted not Black*. Thus, Black-owned restaurants that are correctly identified by BIFSG ZIP tend to be in ZIPs with fewer bank branches per capita than Black-owned restaurants that BIFSG ZIP fails to identify as Black. We find similar results for Hispanic- and Asian-owned restaurants. Those identified correctly by the algorithm are located in ZIPs with fewer bank branches per capita than those that the algorithm fails to identify as Hispanic or Asian. In fact, Hispanic- and Asian-owned restaurants that are not identified as such by BIFSG ZIP tend to be located in ZIPs that have the same number of bank branches per capita as white-owned restaurants.

We look at the log of median household income in columns 6–10. The predictions of all four algorithms that use location information are strongly correlated with lower household income for Black- Hispanic- and Asian-owned restaurants, but only when race is correctly predicted. This is not the case for the RF algorithm based only on names. Since this algorithm does not use location information, its predictions are much less correlated with local economic conditions.

Table 4
Disparities in Restaurant Characteristics Based on Actual versus Predicted Race

This table reports the results of regressions of firm characteristics on actual and predicted race. Firm characteristics are size (log seats), whether the restaurant had a bank UCC loan at the start of the COVID pandemic, bank branches per capita, and median household income in the restaurant owner's ZIP. BIFSG is the Bayesian Improved First Name Surname Geocoding algorithm (Voicu, 2018). We use either ZIP or Census block group (CBG) level data on the population shares of different racial/ethnic groups in the 2020 Census. The three random forest (RF) algorithms are trained using data on the distribution of first and last names across racial/ethnic groups and on either ZIP or CBG level population shares. Standard errors are adjusted for clustering by ZIP code. *, **, and *** indicate statistical significance at 10%, 5%, and 1%. $N = 11,530$.

Panel A

	Bank branches per capita					Household income				
	BIFSG		RF			BIFSG		RF		
	ZIP	CBG	ZIP	CBG	Names	ZIP	CBG	ZIP	CBG	Names
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Black × Predicted Black	−0.202***	−0.212***	−0.205***	−0.206***	−0.191***	−0.244***	−0.248***	−0.251***	−0.240***	−0.224***
	(0.020)	(0.020)	(0.021)	(0.021)	(0.022)	(0.019)	(0.019)	(0.020)	(0.020)	(0.020)
Black × Predicted not Black	−0.145***	−0.126***	−0.146***	−0.141***	−0.166***	−0.188***	−0.179***	−0.183***	−0.196***	−0.222***
	(0.028)	(0.032)	(0.026)	(0.028)	(0.025)	(0.026)	(0.027)	(0.025)	(0.026)	(0.025)
Hispanic × Predicted Hispanic	−0.086***	−0.091***	−0.091***	−0.094***	−0.098***	−0.081***	−0.083***	−0.081***	−0.081***	−0.081***
	(0.014)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Hispanic × Predicted not Hispanic	0.004	0.044	0.026	0.058	0.045	−0.027	−0.017	−0.039*	−0.035	−0.044**
	(0.040)	(0.049)	(0.045)	(0.048)	(0.042)	(0.024)	(0.023)	(0.023)	(0.023)	(0.020)
Asian × Predicted Asian	−0.080***	−0.080***	−0.082***	−0.084***	−0.083***	−0.010	−0.009	−0.012	−0.010	−0.011
	(0.014)	(0.014)	(0.015)	(0.014)	(0.015)	(0.012)	(0.012)	(0.012)	(0.011)	(0.012)
Asian × Predicted not Asian	0.030	0.034	0.021	0.034	0.016	0.065***	0.067***	0.061***	0.061***	0.055***
	(0.024)	(0.024)	(0.021)	(0.023)	(0.020)	(0.017)	(0.017)	(0.016)	(0.016)	(0.015)
Adjusted R^2	0.014	0.016	0.015	0.017	0.016	0.028	0.028	0.028	0.027	0.027

(continued)

Table 4—*continued*

	Panel B									
	Log seats					Has bank UCC loan				
	BIFSG		RF			BIFSG		RF		
	ZIP	CBG	ZIP	CBG	Names	ZIP	CBG	ZIP	CBG	Names
	(1)	(2)	(3)	(4)	(5)	(6)				
Black × Predicted Black	−0.832***	−0.823***	−0.829***	−0.820***	−0.777***	−0.170***	−0.167***	−0.173***	−0.173***	−0.162***
	(0.063)	(0.063)	(0.064)	(0.064)	(0.062)	(0.010)	(0.012)	(0.010)	(0.010)	(0.013)
Black × Predicted not Black	−0.482***	−0.489***	−0.517***	−0.514***	−0.596***	−0.057	−0.060	−0.063	−0.055	−0.080*
	(0.076)	(0.084)	(0.071)	(0.080)	(0.076)	(0.050)	(0.052)	(0.046)	(0.050)	(0.045)
Hispanic × Predicted Hispanic	−0.342***	−0.346***	−0.348***	−0.348***	−0.334***	−0.077***	−0.079***	−0.076***	−0.078***	−0.078***
	(0.029)	(0.029)	(0.029)	(0.030)	(0.030)	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)
Hispanic × Predicted not Hispanic	−0.324***	−0.292***	−0.287***	−0.278***	−0.378***	−0.077***	−0.060**	−0.086***	−0.071***	−0.072***
	(0.081)	(0.082)	(0.074)	(0.075)	(0.065)	(0.024)	(0.027)	(0.024)	(0.023)	(0.021)
Asian × Predicted Asian	−0.402***	−0.401***	−0.406***	−0.401***	−0.410***	−0.109***	−0.105***	−0.107***	−0.108***	−0.109***
	(0.042)	(0.041)	(0.043)	(0.042)	(0.044)	(0.014)	(0.014)	(0.015)	(0.014)	(0.015)
Asian × Predicted not Asian	−0.288***	−0.287***	−0.293***	−0.296***	−0.292***	−0.092***	−0.101***	−0.099***	−0.095***	−0.095***
	(0.045)	(0.050)	(0.041)	(0.043)	(0.042)	(0.018)	(0.016)	(0.016)	(0.017)	(0.016)
Adjusted R^2	0.050	0.050	0.050	0.050	0.049	0.016	0.016	0.016	0.016	0.015

In Panel B of Table 4, we look at log seats and bank UCC loans as the dependent variables. We once again find that those who are Black and predicted to be Black have significantly worse outcomes than those who are Black but not identified as Black by the algorithms. The results for bank UCC loans are especially striking. While Black-owned restaurants that are identified by RF CBG as Black are 17.3pp less likely to have a bank UCC loan, Black-owned restaurants that the algorithm fails to identify as Black are only 5.5pp less likely to have a bank UCC loan.

Overall, the results in Table 4 show that predicted race is likely to be strongly correlated with firm characteristics. If these characteristics are correlated with the outcomes of interest, then using predicted race will affect the conclusions about unconditional disparities, the relative importance of location and firm characteristics, and conditional disparities.

5 PPP Take-Up by EIDL Advance Recipients

We next compare the performance of actual versus predicted race in measuring and understanding the sources of disparities in PPP take-up by firms receiving EIDL Advance grants. The Economic Injury Disaster Loan (EIDL) program is an existing Small Business Administration (SBA) program that was significantly expanded to mitigate the adverse economic impact of the pandemic. As part of this program, SBA made non-forgivable low-interest loans directly to small businesses without using financial intermediaries. It also made grants of \$1,000 per employee, up to \$10,000 in total, under the EIDL Advance program.[6] Grant recipients demonstrate awareness of government emergency support programs and demand for such support. Most recipients should have also been eligible for PPP loans, which for most firms would have been significantly larger than the EIDL Advance grant.[7] Thus Chernenko and Scharfstein (2023) find that 80% of Florida restaurants that received EIDL Advance grants also applied for and received PPP loans.

Studying PPP take-up by EIDL Advance recipients allows us to expand our sample to other industries and other states with voter registration data that includes publicly available information on race. These states include Florida, Georgia, Louisiana, and North Carolina. We will show that there are significant differences in algorithm accuracy between the restaurants and EIDL Advance recipients samples, suggesting that the same algorithm such as BIFSG, can perform very differently in different samples and that one cannot rely on estimates of its accuracy in voter registration and mortgage applications data. Second, the

[6] https://data.sba.gov/dataset/covid-19-eidl-advance.

[7] Any EIDL Advance grant would count against the forgivable portion of the PPP loan.

larger sample of EIDL Advance recipients will allow us to study the effect of using predicted race on estimating interaction terms, such as the interactions between Black and indicators for existing bank relationships.

We follow Chernenko and Scharfstein (2023) in constructing the sample of EIDL Advance recipients. We start by matching EIDL Advance recipients to Florida, Georgia, Louisiana, and North Carolina corporate records, which we access through OpenCorporates, an open database of corporations around the world. We limit the resulting sample to for-profit corporations and LLCs that were active as of February 15, 2020. We next match PPP borrowers to these corporate records and use this link to determine which EIDL Advance recipients also received PPP loans.

Because neither EIDL Advance nor OpenCorporates data include industry classification, we match firms to the Dun & Bradstreet database and use the firm's primary SIC code in Dun & Bradstreet. Firms with SIC code 9999, non-classifiable establishments, are excluded from the analysis.

Table 5 reports the accuracy of the algorithms in the sample of EIDL Advance recipients. The algorithms are better at identifying Black-owned firms in this sample than in the sample of Florida restaurants. RF CBG for example has 71.75% recall and 80.00% precision within the EIDL Advance sample but only 61.84% recall and 60.13% precision within the sample of Florida restaurants.

Table 5
Accuracy of Different Algorithms in the EIDL Advance Sample

This table reports the accuracy of different algorithms in predicting the race of EIDL Advance recipients. The sample consists of firms receiving EIDL Advance grants that can be matched to OpenCorporates corporate records, Dun & Bradstreet, and voter registration data. In each cell, the first row reports the algorithm's recall (the share of all observations in the group that are correctly identified by the algorithm); the second row reports the algorithm's precision (the share of predicted observations that are true).

	BIFSG		Random Forest		
	ZIP	CBG	ZIP	CBG	Names
Asian	83.08	83.72	78.93	80.65	76.64
	71.71	71.18	77.53	76.99	77.04
Black	71.28	73.86	69.63	71.75	59.74
	70.62	75.61	75.77	80.00	58.58
Hispanic	88.57	88.78	88.09	88.14	82.91
	81.99	81.86	82.17	82.82	83.06
White	88.95	90.35	91.37	92.60	87.55
	87.08	88.04	86.77	87.38	83.64

Appendix Table A1 reports the accuracy of the different algorithms within each state and shows significant variation across states. The algorithms do better in classifying Black-owned EIDL Advance recipients in Georgia than in the other states. RF CBG for example has 78.23% recall and 84.37% precision in Georgia, but only 62.69% recall and 76.71% precision in North Carolina. On the other hand, they are better at classifying Hispanic individuals in Florida than in the other states. It is worth noting that the lower accuracy in the Florida restaurants sample is driven by differences in industry rather than state. BIFSG ZIP for example has 65.27% recall and 69.05% precision in identifying Black EIDL Advance recipients in Florida but only 63.16% recall and 46.30% precision in identifying Black-owned restaurants in Florida. These differences in accuracy across industries and states highlight the fact that race prediction algorithms can perform very differently in seemingly similar settings and that researchers should not rely on tests of accuracy within voter registration or mortgage borrower samples, which may not be representative of algorithm accuracy in other settings.

Table 6 reports the results using essentially the same layout as Table 2, which studies PPP take-up by Florida restaurants. Columns 1–4 report the results using actual self-reported race from voter registration data. Black- and Hispanic-owned firms are 22.1pp and 14.8pp less likely to receive PPP loans, while Asian-owned firms are 3.7pp more likely to receive PPP loans than white-owned firms. The estimated disparities for Black- and Hispanic-owned firms are broadly similar to the sample of Florida restaurants, where we found a 26.2pp disparity for Black-owned firms and a 10.5pp disparity for Hispanic-owned firms. Among EIDL Advance recipients, however, Asian-owned firms exhibit a positive disparity while they have a negative 2.5pp disparity in the Florida restaurants sample.

Column 2 adds ZIP code and 3-digit SIC fixed effects, while column 3 replaces ZIP code with Census block group (CBG) fixed effects. Both regressions deliver similar results. Focusing on the estimates in column 3, we see a 14.7pp disparity for Black-owned firms and a 9.1pp disparity for Hispanic-owned firms compared to white-owned firms in the same industry and firms in the same Census block group. Thus industry and location can explain about 30–40% of the unconditional disparities for Black- and Hispanic-owned firms. For Asian-owned firms, we find a positive 2.0pp disparity, which is about half of its unconditional value in column 1.

Column 4 adds firm controls: indicators for the number of employees, log firm age, and dummies for having bank and nonbank UCC loans. The magnitudes of the disparities for Black- and Hispanic-owned firms are reduced to 11.3pp and 7.5pp. The positive disparity for Asian-owned firms is reduced to 0.9pp, or about a quarter of its unconditional value. Overall, firm characteristics explain about half of the unconditional disparities for Black-

Table 6
Actual versus Predicted Race in Analysis of PPP Take-Up by EIDL Advance Recipients

This table reports the results of linear probability model regressions of PPP take-up by EIDL Advance recipients in Florida, Georgia, Louisiana, and North Carolina. The sample consists of firms receiving EIDL Advance grants that can be matched to OpenCorporates corporate records, Dun & Bradstreet, and voter registration data. Regressions include but do not report a dummy for other racial groups. Firm controls are dummies for the number of employees (estimated based on the size of the EIDL Advance grant), log firm age, and dummies for having bank and nonbank UCC loans as of February 15, 2020. Census block group (CBG) FEs use the firm owner's residential address in voter registration data. Robust standard errors are reported. *, **, and *** indicate statistical significance at 10%, 5%, and 1%. $N = 251,851$.

Panel A

	Voter registration				BIFSG with ZIP					BIFSG with CBG				
					Eq. 3	Empirical estimates				Eq. 3	Empirical estimates			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Black	−0.221***	−0.162***	−0.147***	−0.110***	−0.141	−0.170***	−0.097***	−0.084***	−0.066***	−0.154	−0.192***	−0.121***	−0.100***	−0.079***
	(0.003)	(0.003)	(0.004)	(0.003)		(0.003)	(0.003)	(0.004)	(0.003)		(0.003)	(0.003)	(0.004)	(0.004)
Hispanic	−0.148***	−0.104***	−0.091***	−0.075***	−0.107	−0.131***	−0.081***	−0.069***	−0.061***	−0.109	−0.136***	−0.090***	−0.075***	−0.065***
	(0.003)	(0.003)	(0.004)	(0.003)		(0.003)	(0.003)	(0.004)	(0.003)		(0.003)	(0.003)	(0.004)	(0.003)
Asian	0.037***	0.021***	0.020***	0.009**	0.029	0.046***	0.037***	0.033***	0.015***	0.027	0.042***	0.030***	0.027***	0.012***
	(0.005)	(0.005)	(0.005)	(0.005)		(0.004)	(0.004)	(0.005)	(0.004)		(0.004)	(0.004)	(0.005)	(0.004)
Adjusted R^2	0.032	0.100	0.108	0.257		0.022	0.094	0.103	0.255		0.026	0.096	0.104	0.255
ZIP FEs		✓					✓					✓		
SIC3 FEs		✓	✓	✓			✓	✓	✓			✓	✓	✓
CBG FEs			✓	✓				✓	✓				✓	✓
Firm controls				✓					✓					✓

Panel B

	RF with ZIP					RF with CBG					RF with name only				
	Eq. 3	Empirical estimates				Eq. 3	Empirical estimates				Eq. 3	Empirical estimates			
	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)	(27)	(28)	(29)
Black	−0.149	−0.177***	−0.098***	−0.084***	−0.065***	−0.160	−0.197***	−0.122***	−0.100***	−0.078***	−0.107	−0.127***	−0.073***	−0.060***	−0.043***
		(0.003)	(0.003)	(0.004)	(0.004)		(0.003)	(0.003)	(0.004)	(0.004)		(0.003)	(0.003)	(0.003)	(0.003)
Hispanic	−0.105	−0.128***	−0.080***	−0.069***	−0.061***	−0.107	−0.134***	−0.090***	−0.073***	−0.065***	−0.099	−0.123***	−0.076***	−0.064***	−0.056***
		(0.003)	(0.003)	(0.004)	(0.003)		(0.003)	(0.003)	(0.004)	(0.003)		(0.003)	(0.003)	(0.003)	(0.003)
Asian	0.038	0.055***	0.042***	0.038***	0.018***	0.036	0.052***	0.037***	0.034***	0.014***	0.044	0.063***	0.046***	0.041***	0.019***
		(0.005)	(0.005)	(0.005)	(0.005)		(0.004)	(0.005)	(0.005)	(0.005)		(0.005)	(0.005)	(0.005)	(0.005)
Adjusted R^2		0.022	0.094	0.103	0.254		0.026	0.096	0.104	0.255		0.016	0.094	0.103	0.254
ZIP FEs			✓					✓					✓		
SIC3 FEs			✓	✓	✓			✓	✓	✓			✓	✓	✓
CBG FEs				✓	✓				✓	✓				✓	✓
Firm controls					✓					✓					✓

and Hispanic-owned firms.

Columns 6–9 of Table 6 report the results using BIFSG ZIP, with column 5 reporting the predictions of Eq. 3. If prediction errors were random, we would expect BIFSG ZIP to estimate 14.1pp, 10.7pp, and positive 2.9pp disparities for Black-, Hispanic-, and Asian-owned businesses. The actual estimates in column 6 are once again larger in magnitude: 17.0pp for Black- and 13.1pp for Hispanic-owned firms. For Asian-owned firms, BIFSG ZIP estimates a positive 4.6pp disparity which is larger than the positive 3.7pp disparity in column 1.

When we add SIC2 and ZIP or Census block group (CBG) fixed effects in columns 7 and 8, we once again see large decreases in the magnitude of the estimated disparities. For Black-owned firms, the estimated disparity is cut in half from 17.0pp in column 6 to 8.4pp in column 8. The estimate in column 8 represents just a bit more than half of the true within CBG disparity of 14.7pp.

Addition of firm controls in column 9 results in a 6.6pp disparity for Black-owned firms and a 6.1pp disparity for Hispanic-owned firms. For Black-owned firms this is 60% of the 11.0pp disparity in column 4. For Hispanic-owned firms the estimate in column 9 is about 80% of the estimate in column 4. We continue to estimate positive disparities for Asian-owned firms that are larger than the estimates in columns 1–4.

The other algorithms — BIFSG CBG in columns 10–14, RF ZIP in columns 15–19, and RF CBG in columns 20–24 — deliver similar results. They all result in underestimates of unconditional disparities for Black- and Hispanic-owned firms. This downward bias is smaller than would be the case if prediction errors were random. Downward bias gets stronger once we control for location and firm characteristics. For Black-owned firms, conditional disparities are underestimated by 30–40%. For Hispanic-owned firms, the downward bias is 15–20%. At the same time, the algorithms tend to overestimate both unconditional and conditional disparities for Asian-owned firms. RF ZIP, for example, estimates a positive conditional disparity of 1.8pp, which is twice as large as the 0.9pp disparity in column 4.

Overall, Table 6 shows that using predicted race can either under- or overestimate the magnitude of unconditional disparities. It tends to underestimate the magnitude of conditional disparities for Black-owned firms and to attribute too much of unconditional disparities to location. It may also result in overestimates of both unconditional and conditional disparities.

5.1 Prediction Errors, Firm Characteristics, and Disparities in Outcomes

To gain some insight into the relative performance of the different algorithms in the sample of EIDL Advance recipients, we regress firm characteristics that may be correlated with outcomes (PPP take-up) on actual race interacted with predicted race. The results are reported in Table 7, with Panel A looking at bank branches per capita and log of median household income in the firm's ZIP code, and Panel B looking at the number of employees and at whether the firm had a bank UCC loan at the start of the COVID pandemic.

In columns 1–5, we see large differences in ZIP-level bank branches per capita between Black-owned firms that are predicted to have Black owners and Black-owned firms that are not predicted to have Black owners. Black-owned firms that are predicted to have Black owners tend to be located in ZIP codes with about 0.140 fewer bank branches per capita. Black-owned firms that are not predicted to have Black owners tend to be located in ZIP codes with about 0.080 fewer bank branches per capita. There is a similar though smaller gap for Hispanic-owned firms.

In columns 6–10 of Table 7, we find even starker differences for the log of median household income in the ZIP code of the firm's owner. Black owners of firms that are predicted to have Black owners by the algorithms that use location information live in locations with 0.172–0.186 lower log median household income than white-owned firms. Black owners that are not predicted to be Black are in ZIP codes with similar and, in some cases, higher log median household income. Hispanic-owned firms that are predicted to be Hispanic are in ZIP codes with 0.047–0.065 lower log median household income than white-owned firms, while Hispanic-owned firms that are not predicted to be Hispanic are in ZIP codes with 0.035–0.042 higher log median household income than white-owned firms. All Asian-owned firms tend to be located in ZIP codes with higher household income, especially if they are predicted to be Asian-owned.

Panel B looks at the number of employees in columns 1–5 and bank UCC loans in columns 6–10. Black- and Hispanic-owned firms tend to have fewer employees than white-owned firms, with the disparity somewhat larger for Black firms that are identified as such by the algorithms than for Black-owned firms that are not predicted to be Black. Black- and Hispanic-owned firms that are identified as such by the algorithms are significantly less likely to have bank UCC loans than Black- and Hispanic-owned firms that the algorithms fail to identify as minority-owned.

Overall the results in Table 7 suggest that predicted race can be correlated in complex and sometimes unexpected ways with firm characteristics that may be correlated with the

Table 7
Disparities in Firm Characteristics Based on Actual versus Predicted Race

This table reports the results of regressions of firm characteristics on actual and predicted race. Firm characteristics are the number of employees, whether a firm had a bank UCC loan at the start of the COVID pandemic, bank branches per capita in the firm's ZIP, and the log of median household income in the firm's ZIP. BIFSG is the Bayesian Improved First Name Surname Geocoding algorithm (Voicu, 2018). We use either ZIP or Census block group (CBG) level data on the population shares of different racial/ethnic groups in the 2020 Census. The three random forest (RF) algorithms are trained using data on the distribution of first and last names across racial/ethnic groups and on either ZIP or CBG level population shares. Standard errors are adjusted for clustering by ZIP code. *, **, and *** indicate statistical significance at 10%, 5%, and 1%. $N = 251,851$.

Panel A										
	Bank branches per capita					Household income				
	BIFSG		RF			BIFSG		RF		
	ZIP	CBG	ZIP	CBG	Names	ZIP	CBG	ZIP	CBG	Names
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Black × Predicted Black	−0.138***	−0.137***	−0.140***	−0.139***	−0.125***	−0.186***	−0.172***	−0.182***	−0.175***	−0.124***
	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.015)	(0.015)	(0.016)	(0.015)	(0.013)
Black × Predicted not Black	−0.080***	−0.078***	−0.081***	−0.081***	−0.117***	0.027***	0.009	−0.001	−0.006	−0.126***
	(0.022)	(0.022)	(0.022)	(0.022)	(0.023)	(0.008)	(0.008)	(0.008)	(0.008)	(0.012)
Hispanic × Predicted Hispanic	−0.062**	−0.063**	−0.064**	−0.066**	−0.065**	−0.047**	−0.047**	−0.048**	−0.048**	−0.048**
	(0.029)	(0.029)	(0.029)	(0.029)	(0.028)	(0.023)	(0.023)	(0.023)	(0.023)	(0.022)
Hispanic × Predicted not Hispanic	−0.038	−0.026	−0.023	−0.011	−0.032	0.042***	0.038***	0.038***	0.035***	0.012
	(0.024)	(0.024)	(0.026)	(0.029)	(0.030)	(0.013)	(0.012)	(0.013)	(0.012)	(0.017)
Asian × Predicted Asian	−0.051**	−0.042**	−0.039**	−0.041**	−0.041**	0.112***	0.115***	0.109***	0.108***	0.099***
	(0.024)	(0.020)	(0.020)	(0.020)	(0.019)	(0.025)	(0.025)	(0.026)	(0.025)	(0.024)
Asian × Predicted not Asian	−0.013	−0.057***	−0.063***	−0.057**	−0.055**	0.032***	0.015*	0.061***	0.058***	0.095***
	(0.017)	(0.022)	(0.022)	(0.022)	(0.023)	(0.010)	(0.009)	(0.011)	(0.011)	(0.015)
Adjusted R^2	0.002	0.003	0.003	0.003	0.002	0.041	0.037	0.038	0.036	0.027

(continued)

Table 7—*continued*

	Panel B									
	Number of employees					Has bank UCC loan				
	BIFSG		RF			BIFSG		RF		
	ZIP	CBG	ZIP	CBG	Names	ZIP	CBG	ZIP	CBG	Names
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Black × Predicted Black	−0.457***	−0.477***	−0.462***	−0.472***	−0.457***	−0.093***	−0.094***	−0.094***	−0.094***	−0.094***
	(0.027)	(0.027)	(0.028)	(0.027)	(0.029)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Black × Predicted not Black	−0.399***	−0.336***	−0.393***	−0.366***	−0.416***	−0.075***	−0.071***	−0.076***	−0.072***	−0.079***
	(0.032)	(0.033)	(0.031)	(0.033)	(0.028)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)
Hispanic × Predicted Hispanic	−0.356***	−0.360***	−0.357***	−0.357***	−0.348***	−0.057***	−0.057***	−0.056***	−0.057***	−0.057***
	(0.020)	(0.020)	(0.021)	(0.021)	(0.021)	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)
Hispanic × Predicted not Hispanic	−0.424***	−0.393***	−0.412***	−0.409***	−0.439***	−0.045***	−0.043***	−0.047***	−0.046***	−0.049***
	(0.049)	(0.049)	(0.045)	(0.044)	(0.041)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
Asian × Predicted Asian	0.547***	0.540***	0.546***	0.558***	0.568***	−0.020***	−0.019***	−0.021***	−0.020***	−0.020***
	(0.046)	(0.045)	(0.046)	(0.047)	(0.049)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Asian × Predicted not Asian	0.124*	0.136*	0.214***	0.139**	0.179***	−0.020***	−0.023***	−0.018***	−0.020***	−0.022***
	(0.073)	(0.073)	(0.066)	(0.065)	(0.063)	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)
Adjusted R^2	0.005	0.005	0.005	0.005	0.005	0.013	0.013	0.013	0.013	0.013

outcomes of interest and that may help explain some of the unconditional disparities.

5.2 Cross-Sectional Variation in Disparities

We next look at how using predicted race affects estimates of interaction terms. As Eq. 10 shows, estimated interaction terms depend on i) true disparities, ii) algorithm accuracy, and iii) relative population shares. Estimated interaction terms may under- or overestimate true effects.

Table 8 reports the results of linear probability model regressions of PPP take-up by EIDL Advance recipients. In these regressions, we interact the indicator for Black-owned firms with measures of existing bank borrowing relationships, number of employees, and firm age.[8] While we are most interested in the interaction with bank borrowing relationships, we control for the interactions with the number of employees and firm age to make sure that the results are not driven by differences in firm size and age.

In column 1, the coefficient on Black is -0.132, while the coefficient on Black interacted with bank UCC borrowing relationships is 0.101. Thus, bank UCC borrowing relationships significantly mitigate but do not eliminate disparities in PPP take-up for Black-owned firms. We find a smaller coefficient of 0.056 on the interaction of Black with nonbank UCC borrowing relationships.

In columns 2–5, which report on results using race predicted by location-based algorithms, the estimated interaction between Black and bank UCC ranges from 0.063 for BIFSG ZIP to 0.094 for RF CBG. Even though RF CBG gets close to the 0.101 estimate in column 1, it can still lead us to misleading conclusions. Because the algorithm underestimates the average disparity (9.5pp versus 13.2pp for actual race shown in column 1), it makes firm characteristics such as bank UCC borrowing relationships appear more important in explaining variation in disparities relative to their mean. More importantly, RF CBG would lead us to conclude that there are no disparities for Black-owned firms with existing bank UCC borrowing relationships, even though according to the estimates in column 1, these firms still face a 3.1pp disparity.

Overall, the results in Table 8 illustrate that both the direct effects and the interaction terms when estimated using predicted race must be interpreted with caution. Algorithms that have seemingly similar overall accuracy can have different accuracy within different subsamples, resulting in biased estimates of interaction terms that may be correlated with

[8] Log number of employees and log of firm age are standardized to have zero mean and unit standard deviation to ensure that the coefficient on Black can be interpreted as the disparity for an average Black-owned firm without previous bank or nonbank borrowing relationships.

Table 8
Interactions with Firm Characteristics

This table reports the results interacting actual and predicted *Black* with firm characteristics in linear probability model regressions of PPP take-up by EIDL Advance Recipients in Florida, Georgia, Louisiana, and North Carolina. The sample consists of firms receiving EIDL Advance grants that can be matched to OpenCorporates corporate records, Dun & Bradstreet, and voter registration data. We include but do not report dummies for all racial groups. Census block group (CBG) FEs use the firm owner's residential address in voter registration data. Robust standard errors are reported. *, **, and *** indicate statistical significance at 10%, 5%, and 1%. $N = 251,851$.

		BIFSG		RF		
	Voter	ZIP	CBG	ZIP	CBG	Names
	(1)	(2)	(3)	(4)	(5)	(6)
Black	−0.132***	−0.078***	−0.094***	−0.080***	−0.095***	−0.054***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)
Bank UCC	0.074***	0.079***	0.079***	0.078***	0.078***	0.081***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Black × Bank UCC	0.101***	0.063***	0.071***	0.084***	0.094***	0.046***
	(0.013)	(0.010)	(0.011)	(0.012)	(0.013)	(0.010)
Nonbank UCC	0.048***	0.050***	0.049***	0.050***	0.049***	0.050***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Black × Nonbank UCC	0.056***	0.038***	0.049***	0.040***	0.049***	0.043***
	(0.008)	(0.008)	(0.008)	(0.009)	(0.009)	(0.008)
Ln(Employees)	0.186***	0.182***	0.183***	0.182***	0.183***	0.179***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Black × Ln(Employees)	−0.089***	−0.065***	−0.074***	−0.074***	−0.081***	−0.048***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Ln(Firm age)	0.062***	0.064***	0.064***	0.064***	0.064***	0.064***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Black × Ln(Firm age)	−0.009***	−0.002	−0.002	−0.004	−0.003	0.000
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Adjusted R^2	0.260	0.256	0.257	0.256	0.257	0.255
SIC3 FEs	✓	✓	✓	✓	✓	✓
CBG FEs	✓	✓	✓	✓	✓	✓

prediction errors.

6 Conclusion

Our results raise significant concerns about the use of predicted race to measure racial disparities and to understand the drivers of these disparities. While we have focused on measuring disparities in credit access, our findings are relevant for other domains where researchers use predicted race to measure disparities, such as in healthcare. Our findings are relevant not just for academic research, but also have implications for how regulators and lenders measure disparities. The Equal Credit Opportunity Act prohibits discrimination in lending on the basis of race among other "protected classes." Under the existing interpretation of the ECOA, lenders can be found to be in violation of the ECOA if lending policies have a "disparate impact" on the ability of protected classes to access credit, especially if there are less discriminatory alternative lending policies that would enhance access and would have little impact on the business. Because lenders are prohibited from collecting race information, the Consumer Financial Protection Bureau relies on BISG to measure disparate impact. Our findings on the limits of BISG and related measures of race raise significant concerns about the use of such measures in assessing disparate impact. They suggest that, in practice, BISG will likely lead to underestimation of disparate impact of lending practices, although this may not always be the case. The extent and direction of the estimation errors can vary by racial group, by location, and by type of borrower, among other factors. Given the limitations of the algorithmic approach to measuring race that we have identified in this paper, we either need better data on the racial identities of borrowers or better methods to address the biases introduced by the algorithmic approach.

References

Blattner, L., and S. Nelson. 2021. How costly is noise? data and disparities in consumer credit. *Working paper* doi:10.48550/ARXIV.2105.07554.

Chernenko, S., and D. Scharfstein. 2023. Racial disparities in the Paycheck Protection Program. *Working paper* https://www.nber.org/papers/w29748.

Consumer Financial Protection Board. 2014. Using publicly available information to proxy for unidentified race and ethnicity. https://files.consumerfinance.gov/f/201409_cfpb_report_proxy-methodology.pdf.

Fryer, Jr., R. G., and S. D. Levitt. 2008. The causes and consequences of distinctly Black names. *The Quarterly Journal of Economics* 119:767–805. doi:10.1162/0033553041502180.

Howell, S. T., T. Kuchler, D. Snitkof, J. Stroebel, and J. Wong. 2022. Automation in Small Business Lending Can Reduce Racial Disparities: Evidence from the Paycheck Protection Program. *Working paper* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3939384.

Tzioumis, K. 2018. Demographic aspects of first names. *Scientific Data* 5. doi:10.1038/sdata.2018.25.

Voicu, I. 2018. Using first name information to improve race and ethnicity classification. *Statistics and Public Policy* 5:1–13. doi:10.1080/2330443X.2018.1427012.

Zhang, Y. 2018. Assessing Fair Lending Risks Using Race/Ethnicity Proxies. *Management Science* 64:178–97. ISSN 0025-1909, 1526-5501. doi:10.1287/mnsc.2016.2579.

Appendix

Table A1
Accuracy of Different Algorithms in the EIDL Advance Sample Across States

This table reports the accuracy of different algorithms in predicting the race of EIDL Advance recipients in each of the four states in our data: Florida, Georgia, Louisiana, and North Carolina. In each cell, the first row reports the algorithm's recall (the share of all observations in the group that are correctly identified by the algorithm); the second row reports the algorithm's precision (the share of predicted observations that are true).

	BIFSG ZIP				BIFSG CBG			
	FL	GA	LA	NC	FL	GA	LA	NC
Asian	77.42	90.35	85.29	86.14	78.33	90.99	82.44	87.64
	68.08	80.15	75.07	63.56	67.74	79.74	75.00	62.18
Black	65.27	77.16	77.56	65.97	68.55	80.52	76.78	67.11
	69.05	76.86	64.26	63.38	74.67	80.84	67.90	70.01
Hispanic	89.62	78.56	67.48	74.10	89.90	77.60	68.53	73.54
	83.25	72.19	50.66	67.42	83.15	71.97	48.40	68.35
White	89.43	85.79	87.36	91.39	90.26	88.43	89.09	93.38
	87.24	82.98	89.07	89.43	88.07	85.38	89.01	90.02
	RF ZIP				RF CBG			
	FL	GA	LA	NC	FL	GA	LA	NC
Asian	72.42	87.51	85.13	79.72	74.67	87.78	85.88	83.73
	74.79	85.92	80.21	65.72	74.18	85.49	80.22	65.71
Black	65.18	75.87	72.68	61.61	67.89	78.23	73.19	62.69
	72.90	80.89	73.53	70.29	78.23	84.37	74.27	76.71
Hispanic	89.23	77.83	61.01	73.22	89.38	75.67	65.21	71.54
	83.20	73.31	59.45	67.25	83.97	72.02	56.34	68.95
White	90.71	90.12	92.58	94.16	91.86	92.10	92.71	95.69
	87.20	82.48	88.02	88.85	87.76	83.77	88.38	89.13
	RF names							
	FL	GA	LA	NC				
Asian	71.20	82.27	83.61	80.69				
	73.46	87.12	78.28	66.65				
Black	61.80	59.77	57.46	55.76				
	49.51	72.76	62.67	53.84				
Hispanic	83.41	78.89	69.76	76.26				
	85.34	66.97	44.88	61.19				
White	86.69	87.47	89.06	89.78				
	85.07	75.01	84.45	87.83				

Table A2
Accuracy of Different Algorithms in Subsamples

This table reports the accuracy of BIFSG and RF algorithms within subsamples of EIDL Advance recipients with and without bank UCC loans. In each cell, the first row reports the algorithm's recall (the share of all observations in the group that are correctly identified by the algorithm); the second row reports the algorithm's precision (the share of predicted observations that are true).

	Bank UCC				No bank UCC			
	BIFSG		RF		BIFSG		RF	
	ZIP	CBG	ZIP	CBG	ZIP	CBG	ZIP	CBG
Asian	83.93	84.10	80.24	82.19	82.89	83.64	78.64	80.30
	72.07	71.18	77.08	76.88	71.63	71.18	77.63	77.02
Black	69.04	70.73	69.91	71.30	71.60	74.31	69.59	71.81
	59.39	65.75	68.62	74.20	72.53	77.20	76.93	80.90
Hispanic	87.43	87.46	87.48	87.53	88.74	88.99	88.18	88.23
	79.40	79.39	80.63	81.83	82.39	82.25	82.41	82.97
White	89.46	91.00	92.36	93.76	88.82	90.18	91.11	92.31
	90.07	90.51	90.42	90.83	86.35	87.43	85.87	86.53

Table A3
Accuracy of BIFSG CBG Algorithm in Different Settings

This table reports the accuracy of the BIFSG CBG algorithm in five different settings, listed in columns. The first setting is Florida registered voters. The second one is the sample of Florida restaurants studied by Chernenko and Scharfstein (2023). The third one is the sample of Florida PPP recipients matched to OpenCorporates corporate records and to Florida voter registration data. The fourth one is the sample of Florida-based mortgage brokers matched to Florida voter registration data. The fifth one is the sample of Florida-based CPAs matched to Florida voter registration data. In each cell, the first row reports the share of observations that belong to the racial/ethnic group; the second row reports the algorithm's recall (the share of observations that belong to a given group and that are identified as much by the algorithm), the third row reports the algorithm's precision (the fraction of predicted observations that are true positives).

	Voters	Restaurant owners	PPP recipients	Mortgage brokers	CPAs
Asian	2.83	11.88	3.09	1.55	2.50
	65.71	73.72	84.17	54.48	65.97
	71.67	78.60	72.20	56.59	72.01
Black	10.91	3.95	4.80	7.43	3.64
	64.54	64.04	59.62	52.96	53.58
	70.95	55.41	65.81	67.38	50.40
Hispanic	18.51	17.87	16.08	22.69	10.97
	87.84	89.71	90.05	87.78	82.30
	81.22	78.74	80.36	82.08	73.92
White	64.85	62.13	73.61	66.01	80.87
	93.36	91.69	93.96	93.39	95.02
	91.05	89.31	93.60	91.35	94.68