

Imputing Race

Alex Albright*
Juliana Gamboa-Arbelaez[†]

June 2026

Abstract

Missing data on racial identity is a fundamental challenge to studying race in the economy. When race is unobserved, researchers often impute it from variables such as names and locations. Although many imputation methods exist, systematic comparisons of their performance are rare. This paper offers one of the first such comparisons, evaluating leading methods across multiple dimensions. No single method dominates – the best choice depends on the object of interest. These findings highlight the limitations of a common empirical practice: choosing methods based on a single performance dimension.

*Albright: Federal Reserve Bank of Minneapolis, Opportunity and Inclusive Growth Institute
(Email: alex@albrightalex.com)

[†]Gamboa-Arbelaez: University of Minnesota, Economics Department (Email: gambo038@umn.edu)

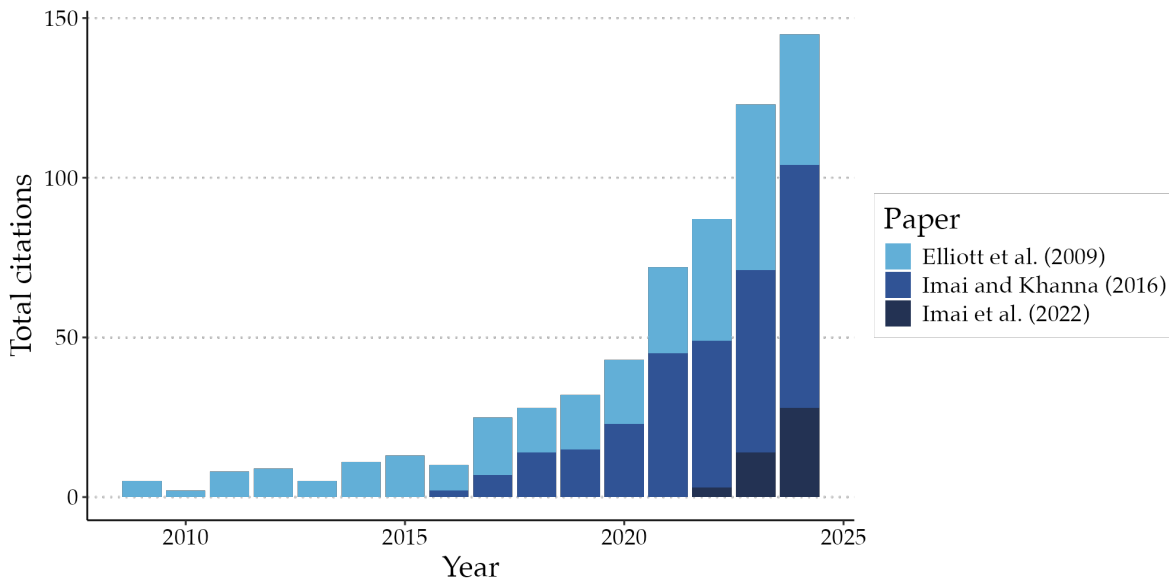
We thank Jeff Larrimore, Tomás Monarrez, Abbie Wozniak, Chi Hyun Kim, Jacob Goldin, and participants at the Federal Reserve System Economic Heterogeneity Conference and the Federal Reserve System Community Development Research Conference for helpful feedback. Weiran Xiao provided excellent research assistance. We thank Kasey Matthews and Sugat Chaturvedi for answering technical questions related to their software packages. We thank James Holt for editorial support. This project would not be possible without the many people who have created and maintained the open-source tools used in this paper. Replication materials are available at https://github.com/apalbright/imputing_race. The views expressed in this paper are our own and do not necessarily represent those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

1 Introduction

Studying economic outcomes across racial groups is crucial for understanding economic inequality in the United States. However, many datasets lack information on race, which creates an obvious challenge for empirical work. How can researchers study racial heterogeneity when race itself is unobserved? In practice, researchers increasingly *impute* race, meaning they predict racial identity using other observables, such as name and geographic location.

Imputing race has enabled recent research in economics on many topics, including taxes (Elzayn et al., 2025), housing (Diamond and Diamond, 2024), lending and credit (Frame et al., 2021; Howell et al., 2024), and the justice system (Feigenberg and Miller, 2022).¹ Research using race imputation is rapidly accelerating: citations to Bayesian Improved Surname Geocoding (BISG), a widely used method, more than tripled between 2020 and 2024 (Figure 1).

Figure 1: Growing Citations to a Common Race Imputation Method



Notes: This figure shows the total Google Scholar citations over time to Bayesian Improved Surname Geocoding (BISG), a commonly used race imputation method. BISG was developed and improved upon in three papers: Elliott et al. (2009), Imai and Khanna (2016), and Imai, Olivella and Rosenman (2022). The figure demonstrates citations across all three papers.

The growing prevalence of race imputation reflects both the expanding availability of imputation methods and shifts in the types of data used in empirical research. In economics,

¹Race imputation has also been used to study the opioid crisis (Cohle and Ortega, 2023), eviction (Hepburn, Louis and Desmond, 2020), policing (Edwards, Lee and Esposito, 2019), and COVID (Labgold et al., 2021).

researchers increasingly rely on administrative and private datasets (Goldsmith-Pinkham, 2024; Einav and Levin, 2014).² These types of datasets often lack direct information on race but contain individual-level variables suitable for imputation. Administrative data examples include tax records (Fisher, 2023), healthcare records (Fremont et al., 2016), and court records (Hepburn, Louis and Desmond, 2020), while private data examples include credit bureau data (Albanesi and Vamossy, 2024), proprietary housing data (Box-Couillard and Christensen, 2024), and cellphone location data (Gong et al., 2025).

Although imputing race underlies a growing body of research, systematic comparisons of imputation methods remain limited. How do researchers impute race, how well do these methods perform, and on which dimensions? The lack of broad comparisons is understandable: methods are proliferating, implementations are fragmented across software environments, and research incentives often favor developing new methods over evaluating existing ones (Christensen and Miguel, 2018; King, 1995). Yet addressing these questions is crucial, because the choice of imputation method can have first-order consequences for empirical conclusions.

Our paper addresses this gap in the literature by conducting a comprehensive comparative analysis on race imputation methods. To enable comparison, we focus on methods that can be applied using publicly available US data that also include self-reported race – specifically, loan data from the Paycheck Protection Program (PPP). We compare imputation methods that rely on names, geographic information, or both, since these are variables commonly used for imputation that are also present in the PPP data.³ We compare standard methods (surname-based imputation, geography-based imputation, BISG, and BIFSG) and recently developed methods (Zest Race Predictor, NamePrism, and BIRDIE) that fit this criterion. For racial gap estimation, we focus on gaps in the rate of obtaining PPP loans from Fintech lenders, first studied by Howell et al. (2024) — an outcome well-suited to illustrating how method choice can affect disparity estimates.

When evaluating performance, we distinguish between three distinct performance dimensions: individual-level classification, racial disparity estimation, and racial demographic estimation. Prior work has typically discussed these dimensions separately. Papers that introduce new methods often emphasize improvements in individual-level classification

²Administrative data now appear in 25% of NBER working papers, up from just 5% in 2005 (Goldsmith-Pinkham, 2024). Similarly, 20% of *American Economic Review* articles in 2014 received exemptions from the journal’s data availability policy due to private data use, compared to 4% in 2006 (Einav and Levin, 2014).

³Race can be imputed using other kinds of data, such as images (Greenwald et al., 2024; Cook, Marx and Yimfor, 2025; Hurtado and Sakong, 2023) or online behavior (Matthew, Miller and Tucker, 2024), but name and geographic data are the most common imputation inputs. More recently, large language models have been used for race imputation without task-specific training (Dasanaike, 2026).

performance (Ye et al., 2017; Chaturvedi and Chaturvedi, 2024; Chalavadi, Pastor and Leitch, 2025; Li, 2023; Dasanaike, 2026), while a growing body of work highlights the importance of imputation methods for estimating racial disparities (Greenwald et al., 2024; McCartan et al., 2025). A few papers have discussed two of these three dimensions simultaneously, specifically when comparing standard BISG to alternative machine learning-based methods (Chaturvedi and Chaturvedi, 2024; Decter-Frain, 2022; Curiel and DeLuca, 2024). Our paper is distinctive in evaluating a range of both common and newly developed methods across all three dimensions.

Consistent with prior work (Greenwald et al., 2024; McCartan et al., 2025; Chernenko and Scharfstein, 2023), we find that commonly used race imputation methods can perform poorly on dimensions of interest. All methods systematically overcount White business owners and undercount Black business owners. More strikingly, imputation methods can both substantially underestimate and overestimate the magnitude of racial gaps, and some even reverse their direction entirely — a finding that contrasts with Rieke et al. (2022), who find that imputed race reliably recovers directional disparities in ride-hailing data. The choice of imputation method can therefore have first-order consequences for the qualitative conclusions of a study.

Our results also shed new light on recently developed methods. BIRDIE (McCartan et al., 2025) stands out for racial gap estimation, closely tracking observed gaps across all pairwise combinations of White, Black, and Hispanic borrowers. However, BIRDIE performs worse when Asian borrowers are involved — a finding we attribute to country-of-origin heterogeneity within the Asian racial category, which violates BIRDIE’s key identification assumption. This contrasts with McCartan et al. (2025), who find BIRDIE substantially outperforms standard methods in the North Carolina voter file, where violations are minimal. BIRDIE is thus a valuable tool for racial gap estimation, but its performance depends on the plausibility of its identification assumptions in a given empirical context.

Our findings also caution against assuming that newer or more complex imputation methods are necessarily better than standard methods. ZRP and NamePrism each rank first on a few metrics, but both are usually outperformed by simpler methods such as BIFSG or surname-based imputation. Greater model complexity does not always guarantee better performance. Indeed, NamePrism (Ye et al., 2017) — cited by over 170 papers as of April 2026 — ranks last on roughly half of all performance metrics despite its use of proprietary data and sophisticated ML techniques. These shortcomings likely reflect its origins as an international nationality classifier, suggesting that methods developed for cross-national contexts may not translate effectively to US racial classification.

Our primary methodological contribution is demonstrating that no single method dominates across all performance dimensions — the best choice depends on the object of interest. Methods can outperform on some dimensions while underperforming on others. Crucially, superior individual-level classification accuracy does not guarantee better estimation of racial demographics or racial disparities. These findings highlight the limitations of a common empirical practice: choosing methods based on a single performance dimension, such as F1 scores.

Our paper contributes to several literatures. First, we directly contribute to research on race imputation methods by providing a comparative analysis of both standard and newly developed approaches using publicly available data. This work builds on early studies of geography-based and surname-based methods ([Fiscella and Fremont, 2006](#)) and the development of BISG ([Elliott et al., 2009](#)), while extending recent evaluations of machine-learning-based imputation ([Curiel and DeLuca, 2024](#); [Ye et al., 2017](#)) and racial disparity adjustment methods ([McCartan et al., 2025](#)).

Second, we contribute to the broader study of race measurement and its implications for empirical research. Since race is a socially constructed category, there is no single correct way to code it ([Rose, 2023](#)). How race is coded affects public understanding of US demographic change ([Starr and Pao, 2024](#); [Ventura and Flores, 2025](#)) and can lead to differences between administrative records and self-reports, with important consequences for estimating racial disparities ([Finlay, Luh and Mueller-Smith, 2024](#); [Baron et al., 2026](#); [Jarrín et al., 2020](#)). Moreover, whether individuals report their race is itself correlated with economic outcomes ([García and Darity Jr, 2022](#)).

Finally, our paper informs research on racial inequality in the US economy from a methodological perspective. Race imputation has been used to study a wide range of economic topics, including lending and credit ([Consumer Financial Protection Bureau, 2014](#); [Hurtado and Sakong, 2023](#)), housing ([Diamond and Diamond, 2024](#)), tax policy ([Cronin, DeFilippes and Fisher, 2023](#); [Costello et al., 2024](#)), and the justice system ([Luh, 2022](#); [Goncalves and Mello, 2021](#)). Our results highlight the underlying importance of race imputation methods in economics research. Studying these methods is increasingly necessary in an era where the proliferation of individual-level data is expanding the possibilities for imputation.

Our paper proceeds as follows. Section 2 provides background on the measurement of race in the US. Section 3 describes the Paycheck Protection Program (PPP) loan data we use. Section 4 outlines the different race imputation methods we compare, while Section 5 describes the three different dimensions of performance we study. Section 6 presents the

performance results across all methods and dimensions. Section 7 concludes.

2 Background

2.1 Measuring Race in the US

Data on race in the US have been collected since the very first decennial Census in 1790 (Jensen and Jones, 2026). That Census was administered by enumerators and included separate columns for "free whites," "all other free persons," and "slaves" (Jensen and Jones, 2026; Medina and Lai, 2023). In contrast, the most recent decennial Census (in 2020) asked respondents to self-identify their race (White, Black or African American, American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander, and Some Other Race) and their ethnicity (Hispanic or Latino, Not Hispanic or Latino) (Jensen and Jones, 2026; Medina and Lai, 2023).

Over the course of 230 years, there have been dozens of changes to the definitions of race and ethnicity in the Census (Jensen and Jones, 2026). These changes were often driven by shifting political climates, demographic patterns, and significant historical events. For example, in 1870, following increased immigration from China, "Chinese" was added as a category (Jensen and Jones, 2026). In 1930, Census instructions incorporated a "one drop rule" for racial classification (Cohn, 2010; Jensen and Jones, 2026). The Hispanic ethnicity question was introduced in 1970 (Jensen and Jones, 2026), and in 2000, respondents gained the ability to select multiple racial categories (Jensen and Jones, 2026).

Not only have the definitions and categories of race and ethnicity shifted dramatically over time, but the methods used to measure them have also evolved throughout the Census's history. From 1790 until 1950, enumerators administered the Census, recording race based on their own observations (Jensen and Jones, 2026). In 1960, the Census transitioned to a mail-based system, enabling respondents to self-report their race for the first time. Today, the principle of self-identification has become central to most official collections of race and ethnicity statistics.

2.2 Self-Reports vs. Administrative Records

Federal guidelines established by the Office of Management and Budget (OMB) designate self-reporting as the preferred method for collecting racial data (U.S. Office of Management and Budget, 2024a). While self-reporting has become the standard in survey research, administrative records are often collected through different mechanisms. Recent research has

documented significant inconsistencies between self-reported and administrative measures of race, which can result in underestimating racial disparities as well as underrepresented populations (Baron et al., 2026; Finlay, Luh and Mueller-Smith, 2024; Jarrín et al., 2020).

What accounts for racial misclassification in administrative records? For one, administrative records often reflect a third party’s perception of an individual’s race rather than the individual’s own identification – similar to the historical practice of Census enumerators recording race based on observation.⁴ This perception can, in turn, depend on the observer’s own identity. For example, Dobre and Witzen (2023) provide evidence that mortgage lenders are more likely to classify an applicant’s race differently from their self-identification when the lender belongs to a different racial group. While third-party reports of race can vary across observers, self-reported race avoids this complication: only one individual’s response matters.

2.3 Linking to Sidestep Imputation

Researchers may avoid imputation altogether if they can link datasets missing race with datasets that contain self-reported race (Ashurst and Weller, 2023). For example, to access self-reported race, researchers have linked bank account records to voter file data (Ganong et al., 2023), credit scores to mortgage data (Albanesi and Vamossy, 2024), and tax data to ACS survey data (Choukhmane et al., 2023).

However, linkage is not always feasible because of administrative costs, data use restrictions, and privacy concerns. Laws protecting confidentiality can limit data sharing, leading some organizations to invest in custom internal imputation methods instead (King et al., 2023).⁵

2.4 Our Definitions

In this paper, we evaluate the performance of imputation methods against self-reported race. We focus on self-reported race because it is defined within a given individual and time, and it is observable in PPP data.

When we discuss race in this paper, we are combining race and ethnicity. We make this

⁴There are other reasons as well: Finlay, Luh and Mueller-Smith (2024) document how operational and bureaucratic complications impede correction of racial misreports in the justice system, while Luh (2022) finds evidence that criminal justice actors (highway patrol officers) may intentionally misreport race to conceal racial bias.

⁵King et al. (2023) describe in detail the privacy-bias trade-off in the context of US federal agencies. See Appendix A.2.1 for examples of US government agencies using race imputation.

choice for two main reasons. First, OMB guidelines now require that race and ethnicity be collected through a single combined question ([U.S. Office of Management and Budget, 2024b](#)), which the Census Bureau expects will improve data quality ([U.S. Census Bureau, 2024a](#)).⁶ Second, for practical purposes, we are constrained by the design of existing imputation methods, all of which already combine race and ethnicity. We therefore follow both updated guidelines and prevailing methodological practice in treating these concepts jointly.

3 Paycheck Protection Program (PPP) Data

We test imputation performance using Paycheck Protection Program loan data. PPP was a COVID-era small business lending program. The data are well-suited for evaluating imputation methods because they include self-reported race alongside the variables commonly used for imputation: first name, last name, and zip code. The only other public US datasets with this combination are six state voter files.⁷ We do not use voter files as our test dataset because several imputation methods were trained on them, making them inappropriate for out-of-sample evaluation.⁸

The raw data are publicly available from the Small Business Administration (SBA) ([U.S. Small Business Administration, 2024](#)). We follow [Greenwald et al. \(2024\)](#) in restricting the sample to “first draw” loans made before February 24, 2021. We further restrict to observations with self-reported race, identifiable person names, and a valid zip code. These steps yield a final sample of 267,212 loans. Appendix [A.3](#) describes sample construction in detail.

Table [A.1](#) shows the number of loans by borrower race and lender type split by Fintech and non-Fintech, following [Howell et al. \(2022\)](#). We split by Fintech status since this will be the outcome variable we consider in comparing methods’ racial gap performance.

4 Race Imputation Methods

We consider race imputation methods implementable with publicly available PPP data, as described in Section 3. Our dataset contains first name, last name, zip code, and

⁶See Appendix [A.2.2](#) for more background on this Census change.

⁷Six US states collect self-reported race from voters: Alabama, Florida, Georgia, Louisiana, North Carolina, and South Carolina, following the 1965 Voting Rights Act ([McCartan et al., 2025](#); [Cruz and Hayes, 2009](#)).

⁸BISG and BIFSG use name dictionaries derived from voter files ([Imai, Olivella and Rosenman, 2022](#)). ZRP is trained on North Carolina, Georgia, and Florida voter files ([Matthews et al., 2024](#)).

self-reported race, and all methods we consider use some combination of these inputs.⁹

We evaluate standard methods – surname-only, zip-only, BISG, and BIFSG – alongside recently developed methods that are growing in usage – Zest Race Predictor, NamePrism, and BIRDIE. We describe each below.

Surname-only imputation: We implement surname-based imputation using the 2000 and 2010 Census surname files (U.S. Census Bureau, 2022, 2021), which include surnames appearing at least 100 times (Kaplan, 2025). Let $p(r|s)$ denote the probability of reporting as race r given surname s , estimated from Census surname files as

$$p(r|s) = \frac{\text{count of people with surname } s \text{ and race } r}{\text{count of people with surname } s}.$$

Zip-only imputation: We impute race using the racial composition of zip codes, drawing on the 2019 5-year ACS.¹⁰ Let $p(r|z)$ denote the probability of being race r given zip code z , estimated from ACS data as

$$p(r|z) = \frac{\text{count of people of race } r \text{ in zip code } z}{\text{count of people in zip code } z}.$$

Bayesian Improved Surname Geocoding (BISG): BISG imputes race using both surname and geographic data (Elliott et al., 2009), combining $p(r|s)$ and $p(z|r)$ under the naive Bayes assumption that surname and geography are mutually independent within racial groups; that is, $p(z|r, s) = p(z|r)$. The posterior probability of race r given surname s and zip code z is

$$p(r|s, z) = \frac{p(r|s) p(z|r)}{\sum_{r'=1}^5 p(r'|s) p(z|r')},$$

where the summation is over five racial categories: White, Black, Hispanic, Asian, and Other. The prior $p(r|s)$ was initially based on Census surname probabilities; Imai, Olivella and Rosenman (2022) improved this by supplementing it with surname dictionaries from

⁹This necessarily excludes methods requiring additional inputs and proprietary methods we cannot replicate. See Appendix A.4 for more details on proprietary imputation methods.

¹⁰The PPP data zip codes are business addresses rather than residential addresses. Ideally, we would impute using residential rather than business addresses, but that would require linking to non-public data (e.g., Infutor, as in Greenwald et al. (2024)). Greenwald et al. (2024) report similar results using business addresses, so we view this as a reasonable alternative that keeps the exercise in the realm of public data.

¹¹We use zip codes, rather than finer geographic units (county, tract, block), to avoid geocoding complications. Zip code is directly available in the PPP data, whereas other units require address geocoding, which introduces additional methodological ambiguity. See Appendix A.5 for details. Zip codes have population distributions similar to census tracts (Clark, Curiel and Steelman, 2022).

six Southern voter files, which substantially reduced unmatched records. We implement BISG using surnames and zip codes (Curiel, 2024).

Bayesian Improved First Name and Surname Geocoding (BIFSG): BIFSG extends BISG by incorporating first name as an additional signal (Voicu, 2018). Let $p(f|r)$ denote the probability of first name f given race r . Under the naive Bayes assumption that surname, first name, and geography are mutually independent within racial groups, the posterior probability of race r given surname s , first name f , and zip code z is

$$p(r|s, f, z) = \frac{p(r|s) p(f|r) p(z|r)}{\sum_{r'=1}^5 p(r'|s) p(f|r') p(z|r')},$$

where $p(f|r)$ values come from first name dictionaries constructed using Southern voter files (Imai, Olivella and Rosenman, 2022). Since no zip code implementation of BIFSG exists, we crosswalk zip to census tract and implement BIFSG at the tract level (Khanna et al., 2024).

Zest Race Predictor (ZRP): ZRP uses XGBoost, a gradient boosting algorithm, trained on 2021 voter files from Florida, Georgia, and North Carolina, with features derived from names and ACS demographic covariates at the zip code, tract, or block group level (Matthews et al., 2024a).¹² We implement ZRP using first name, last name, and zip code (Matthews et al., 2024).¹³

NamePrism: NamePrism uses machine learning to classify nationality and ethnicity from names (Ye et al., 2017). Unlike the other methods, it was developed as an international rather than US-specific classifier.¹⁴ It combines unsupervised and supervised learning to predict one of 39 nationality classes, which are then mapped to US race/ethnicity categories.¹⁵ We implement NamePrism using first and last names (Skiena, 2025).

BIRDIE: BIRDIE (Bayesian Instrumental Regression for Disparity Estimation) is an ad-

¹²XGBoost was selected over alternatives, including logistic regression, naive Bayes, and random forest based on predictive performance on training data (Matthews et al., 2024a). Prior evaluations suggest ZRP can improve on BISG in some contexts: Matthews et al. (2024b) reports higher classification accuracy in Alabama voter records and for small business owners. Curiel and DeLuca (2024) find that ZRP performs especially well in racially diverse areas.

¹³ZRP can use full street addresses and middle names, defaulting to the finest available geographic unit from geocoding, but we limit inputs to first name, last name, and zip code to maintain comparability with other methods evaluated here.

¹⁴The developers show NamePrism outperforms competing international classifiers (Ye et al., 2017), though to our knowledge, it has not been systematically compared to US-specific methods like BISG. As of April 2026, NamePrism has been cited by over 170 papers.

¹⁵The model was trained on 74 million labeled name-nationality pairs from email and Twitter data, plus 57 million contact lists (Ye et al., 2017).

justment method rather than a direct imputation method (McCartan et al., 2025). BISG prediction errors are often correlated with outcomes of interest, which biases disparity estimates. BIRDIE addresses this by using surnames as a high-dimensional instrumental variable for race: McCartan et al. (2025) assume that conditional on race, location, and other observables, surnames have no predictive power over outcomes—an exclusion restriction.¹⁶ We implement BIRDIE using BISG as the baseline with last name and zip code (McCartan, 2023). Since BIRDIE targets disparity measurement rather than individual prediction, we evaluate it only for racial gap estimation, a dimension discussed in Section 5.

Recap: Figure A.1 summarizes which inputs each method uses. All methods except BIRDIE generate predicted probabilities across racial groups, though the specific categories vary by method. We use five categories throughout—White, Black, Hispanic, Asian, and Other—because these are the finest classifications supported by all methods.¹⁷ After probabilities are generated by racial group, we assign each individual to the group with the highest predicted probability.¹⁸

5 Dimensions of Race Imputation Performance

Race imputation methods can be evaluated along multiple dimensions. The canonical one is individual-level classification: Which methods avoid false negatives and false positives? Applied researchers typically impute race to answer one of two questions: How large are racial disparities in outcomes? And what is the racial composition of a population? These three goals—individual-level classification, disparity estimation, and demographic estimation—represent distinct dimensions of performance.

¹⁶Two methods address similar BISG limitations as BIRDIE but require supplementary inputs unavailable in our setting: Greengard and Gelman (2024) modify BISG predictions via raking, which requires knowing population margins, and Argyle and Barber (2024) use an ensemble random forest that incorporates 16 additional political and socioeconomic covariates.

¹⁷Packages for BISG and BIFSG do not distinguish American Indian and Alaskan Native (AIAN) from Multiracial within the Other category, precluding finer comparisons (Khanna et al., 2024). When methods predict six categories, we collapse American Indian and Alaskan Native (AIAN) and Multiracial into Other for consistency.

¹⁸We assign each individual to the group with the highest predicted probability (argmax), following common practice and enabling individual-level classification. Some researchers instead apply probability thresholds before assigning race — for example, assigning an individual to a group only if the predicted probability exceeds 0.80, and classifying those who do not meet this standard as unknown (Siskin, 2014). Researchers may also use the probabilities directly in models. In that case, however, coefficients cannot be interpreted the same way as those estimated with categorical race data, and individual-level classification with meaningful false positive and false negative rates is no longer possible. For discussion of how argmax discretization can itself introduce bias in downstream tasks, see Dong et al. (2025).

Individual-level classification: Individual-level classification is the primary focus in papers that introduce or compare imputation methods (Elliott et al., 2009; Decter-Frain, 2022; Curiel and DeLuca, 2024; Chaturvedi and Chaturvedi, 2024; Chalavadi, Pastor and Leitch, 2025; Li, 2023; Dasanaik, 2026). We evaluate performance using F1 scores, which are well-suited to multi-class classification.

For each racial group g : *recall* is the fraction of group g members correctly imputed as g , *precision* is the fraction imputed as g who truly belong to g , and *F1* is the harmonic mean of recall and precision ($F1_g = \frac{2 \times Precision_g \times Recall_g}{Precision_g + Recall_g}$).

We report F1 scores for each racial group as well as three aggregations: *micro* (equal weight per individual), *macro* (equal weight per group), and *weighted* (weight per group’s population share).

Racial gap estimation: The primary motivation for race imputation in economics is estimating racial disparities across topics including taxes (Elzayn et al., 2025), credit access (Frame et al., 2021), and retirement savings (Coyne, Fadlon and Porzio, 2022). With our PPP data, we estimate racial gaps in Fintech lending, first studied by Howell et al. (2024). We focus on pairwise gaps between the four largest racial groups in the PPP data: Black-White, Hispanic-White, Asian-White, Black-Hispanic, Black-Asian, and Hispanic-Asian.

Racial demographic estimation: Researchers also use race imputation to describe population composition—for example, among evicted people (Hepburn, Louis and Desmond, 2020), people killed by police (Edwards, Lee and Esposito, 2019), and COVID-19 cases (Labgold et al., 2021). We evaluate demographic estimation by comparing imputed and observed racial demographics of people receiving PPP loans. We summarize the overall distance between imputed and observed racial demographics using five-dimensional Euclidean distance.

6 Comparative Performance Results

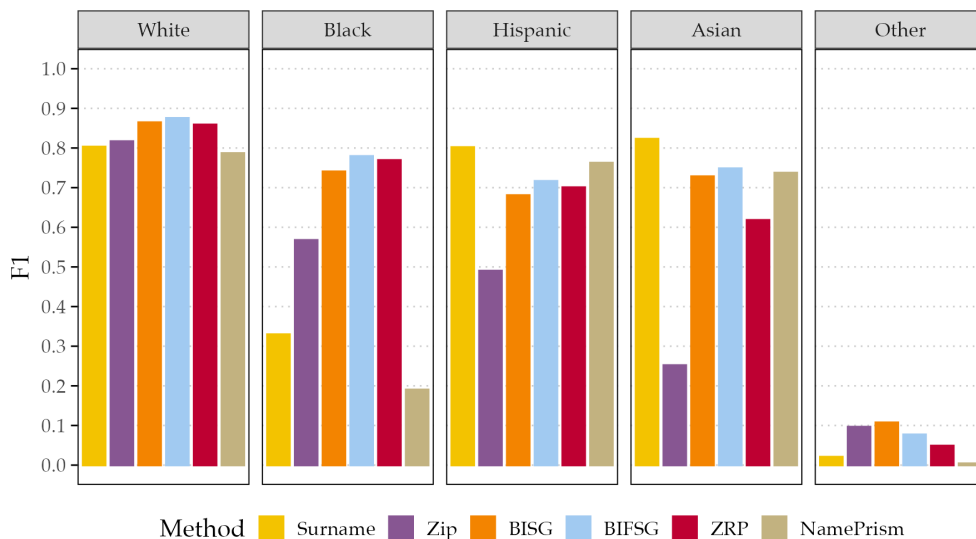
We evaluate race imputation methods across three distinct performance dimensions: individual-level classification, racial gap estimation, and demographic estimation. We discuss each dimension in Sections 6.1–6.3 before synthesizing findings in Section 6.4.

6.1 Individual-Level Classification Performance

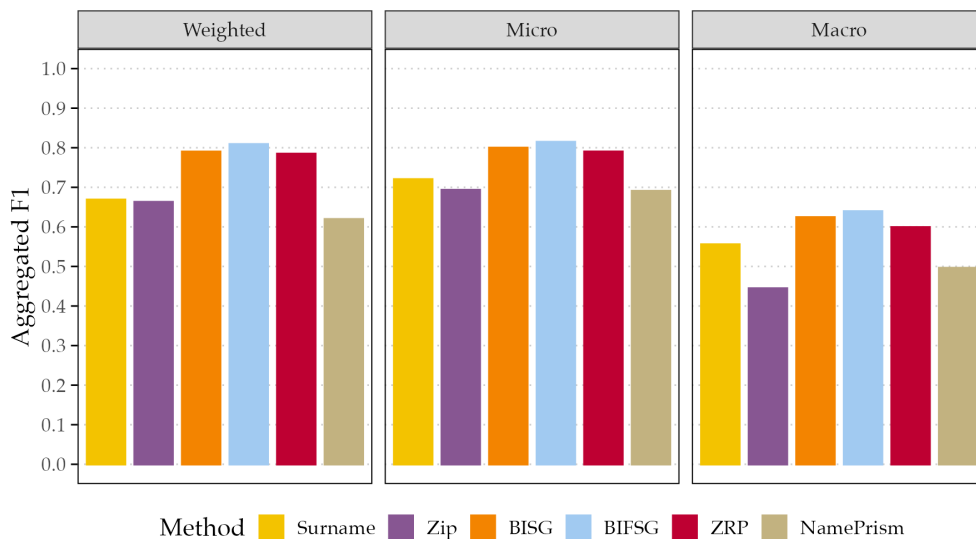
We begin by evaluating individual-level classification performance. For this dimension, we use F1 scores, which combine recall and precision and are well-suited to multi-class classification problems like race imputation. Figure 2a reports F1 scores by racial group, while Figure 2b reports global F1 scores aggregated using three different weighting schemes.

Figure 2: F1 Score Performance Analysis by Race Imputation Method

(a) F1 Scores by Racial Group and Race Imputation Method



(b) Aggregated F1 Scores by Race Imputation Method



Notes: This figure shows F1 scores across racial groups and race imputation methods (Panel A) and aggregated F1 scores by method (Panel B) using PPP data.

Figures 2a and 2b reveal several key patterns in the performance of standard imputation

methods. First, performance varies substantially by the racial group of interest. For example, Figure 2a shows that zip-only imputation outperforms surname-only imputation for the Black population, but the opposite is true for the Hispanic and Asian populations. This is consistent with the established finding that geography-based methods are more effective for imputing Black identity, while surname-based methods are more effective for imputing Hispanic and Asian identity.¹⁹

Second, Figure 2b shows that BISG, which uses both surname and geography data, generates higher aggregate F1 statistics than surname-only or zip-only methods regardless of weighting scheme. However, the aggregate picture masks important differences across racial groups, shown in Figure 2a. For Black and White people, BISG performs better than the two simpler methods, but for Hispanic and Asian people, it underperforms relative to surname-only imputation. BIFSG, which also incorporates first name, improves on BISG for the four largest racial groups — yet surname-only still outperforms BIFSG for Asian and Hispanic populations. This highlights a key finding: more input data do not necessarily improve classification accuracy — the gains depend on the racial group of interest.

Finally, among the newer ML-based methods, ZRP performs similarly to BISG across most racial groups but noticeably worse for Asian people. NamePrism outperforms BISG and BIFSG for Hispanic people and performs on par for Asian people, but generates the lowest F1 scores for White and Black people — most strikingly, an F1 score of roughly 0.2 for Black people (Figure 2a). This suggests that a classifier designed for international nationality prediction may have shortcomings when it comes to US racial classification, particularly for the Black population.

6.2 Performance on Estimating Racial Gaps

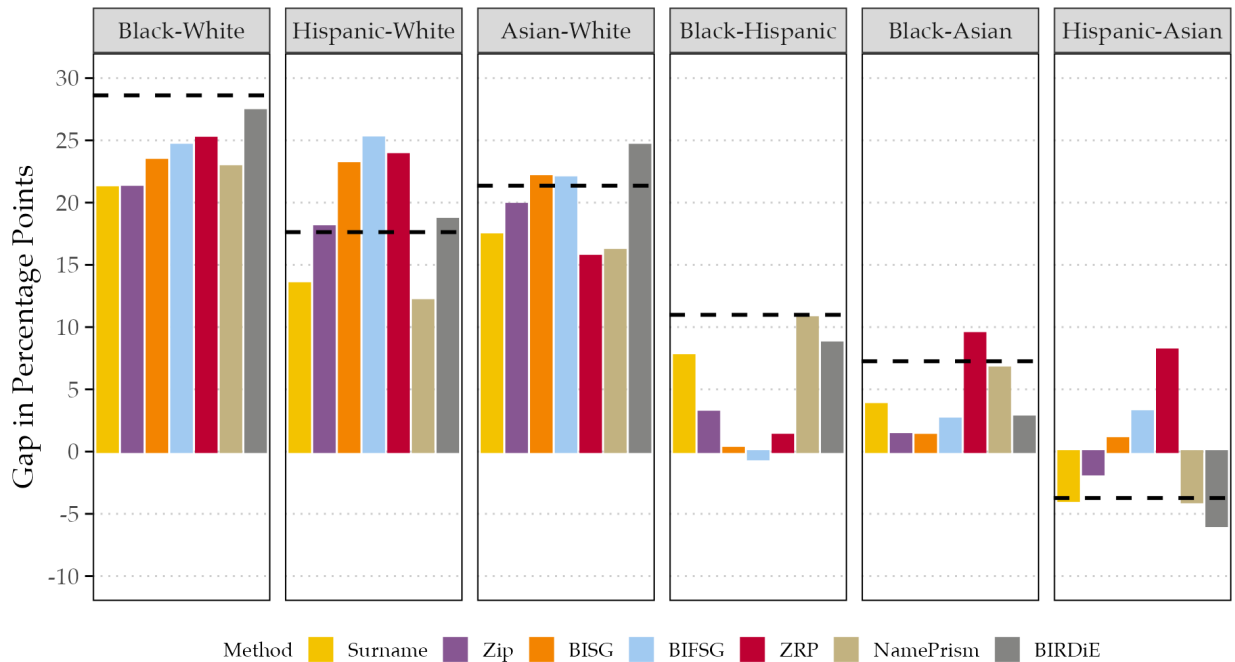
As mentioned above, to evaluate performance on racial gap estimation, we focus on racial gaps in the rate of obtaining PPP loans from Fintech lenders. We examine six pairwise gaps: Black-White, Hispanic-White, Asian-White, Black-Hispanic, Black-Asian, and Hispanic-Asian. Figure 3 illustrates how estimates of these gaps vary across imputation methods. Observed gaps are shown as black dashed horizontal lines.

Several patterns emerge across all six gaps. First, imputation methods can both substan-

¹⁹Elliott et al. (2009) and Fiscella and Fremont (2006) previously documented this finding. Higher Black-White segregation levels relative to other groups lead geography-based methods to work better for inferring Black identity (Iceland et al., 2010; Logan, 2013), while distinctive naming patterns make surname imputation more effective for Asian identity (Fiscella and Fremont, 2006; Comenetz, 2016).

tially underestimate and overestimate the magnitude of racial gaps. For example, BIFSG overestimates the Hispanic-White gap by over 40%, while NamePrism underestimates the same gap by about 30%. Second, and more strikingly, some methods estimate gaps in the wrong direction entirely. BIFSG reverses the direction of the Black-Hispanic gap, and BISG, BIFSG, and ZRP all reverse the direction of the Hispanic-Asian gap. These reversals illustrate that the consequences of method choice are not merely a matter of magnitude — they can fundamentally alter the qualitative conclusions of a study.²⁰ More broadly, as with F1 scores, the relative performance of methods is not consistent across racial groups; for instance, BIFSG has the largest error for the Black-Hispanic gap but performs best for the Asian-White gap.

Figure 3: Racial Gaps in Fintech Lending by Race Imputation Method



Notes: This figure shows the estimated racial gaps in Fintech lending by different race imputation methods. The horizontal black dashed lines demonstrate the observed gaps in the data.

The rightmost gray bar in each panel shows the BIRDIE adjustment (McCartan et al., 2025).²¹ BIRDIE performs well for gaps involving White, Black, and Hispanic business owners: the BIRDIE estimated gaps are within a few percentage points of the observed

²⁰This finding contrasts with those of Rieke et al. (2022), who find that race imputation methods reliably recover directional disparities in ride-hailing data.

²¹Unlike the direct individual-level imputation methods discussed above, BIRDIE is not based on imputing race at the individual-level but instead applies an adjustment specifically for estimating group-level disparities (McCartan et al., 2025). For this reason, we evaluate it only in this performance section.

gaps, outperforming most other methods.²² This is consistent with the results of [McCartan et al. \(2025\)](#), who find BIRDIE substantially outperforms standard methods in the North Carolina voter file.

However, BIRDIE performs worse when Asian business owners are involved. For the Asian-White, Black-Asian, and Hispanic-Asian gaps, BIRDIE estimates deviate more from the observed values and are more consistently outperformed by standard imputation methods.²³ BIRDIE relies on the assumption that conditional on race, geography, and other observables, surname has no predictive power over the outcome ([McCartan et al., 2025](#)). As [McCartan et al. \(2025\)](#) discuss, a key potential violation is the existence of an unobserved confounder that affects both outcome and surname, such as country of origin. Within the Asian racial category, country of origin is strongly encoded in surnames and associated with very different economic outcomes ([Banerjee, 2022](#)). These outcome differences may translate into Fintech lending gaps that surname captures even after conditioning on race and zip code.

Overall, [Figure 3](#) underscores a few key points. First, the choice of imputation method can have first-order consequences for estimates of racial gaps, producing not only different magnitudes but in some cases reversed directions. Second, BIRDIE is a valuable tool for improving racial gap estimation, but it has limitations since performance depends on the plausibility of identification assumptions in a given empirical context, and those assumptions may be better met in some settings than others ([McCartan et al., 2025](#)).

6.3 Performance on Estimating Racial Demographics

Researchers also use race imputation to describe a population's racial composition. To evaluate demographic estimation performance, we compare imputed racial shares against observed shares in the PPP data, summarizing overall accuracy using the Euclidean distance between the imputed and observed distributions in 5-dimensional space.

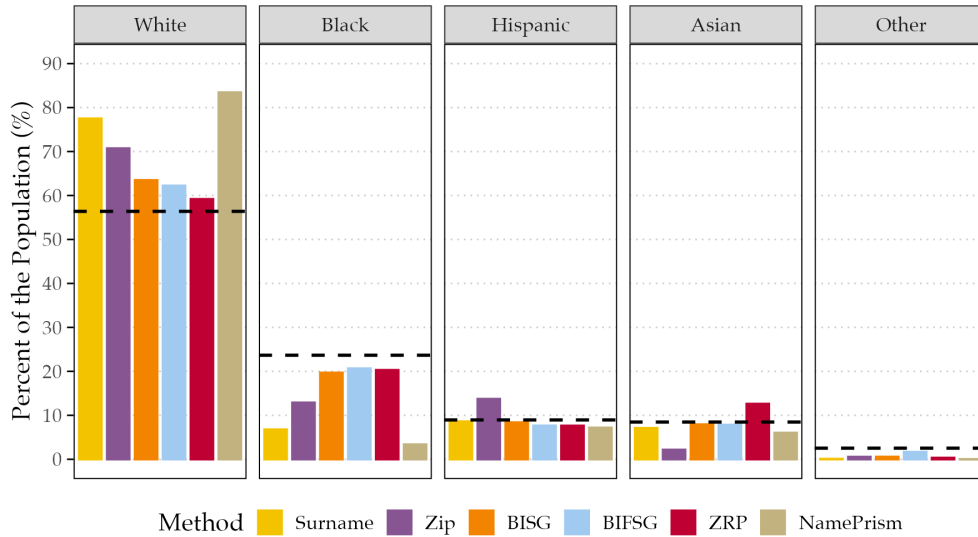
[Figure 4a](#) reveals a consistent pattern: all methods overestimate the share of White business owners and underestimate the share of Black business owners. The most extreme case is NamePrism, which estimates 83% of PPP borrowers are White and only 3% are Black, against observed rates of 56% and 24%. Surname-only and zip-only methods show the same directional error but with smaller magnitude.

²²There are two exceptions: zip-only imputation slightly outperforms BIRDIE for the Hispanic-White gap, and NamePrism outperforms BIRDIE for the Black-Hispanic gap.

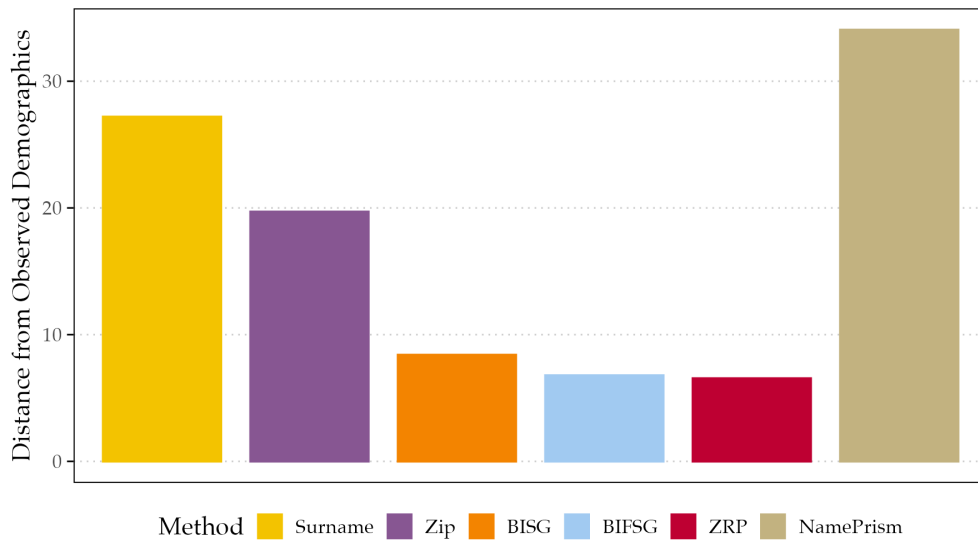
²³For example, BISG and BIFSG outperform BIRDIE for the Asian-White gap. Meanwhile, surname-only and NamePrism outperform BIRDIE for the Hispanic-Asian gap.

Figure 4: Racial Demographics Performance by Race Imputation Method

(a) Racial Demographics by Race Imputation Method



(b) Distance from Observed Demographics by Race Imputation Method



Notes: Panel (a) shows imputed racial demographics of PPP borrowers by imputation method, with horizontal black dashed lines demonstrating observed rates in the data. Panel (b) shows the Euclidean distance between imputed and observed demographic distributions (lower is better).

Figure 4b shows that BIFSG and ZRP perform best in terms of overall distance from the observed population. While both methods underestimate the Black share by a few percentage points, they differ in their remaining errors: ZRP overestimates both the Asian and White shares, while BIFSG overestimates the White share by more but gets the Asian share roughly right (Figure 4a). NamePrism performs worst by a wide margin. Notably, the

two simplest methods — surname-only and zip-only imputation — produce population estimates much closer to the observed rates than NamePrism does.

6.4 Ranking Methods Across Dimensions

Table 1 summarizes how methods rank against one another across all three performance dimensions. Two patterns stand out. First, within a given performance dimension, the best method depends on the racial group of interest. For instance, BIFSG achieves the highest F1 score for White people but not for Asian people. Second, within a given racial group, the best method depends on the performance dimension; one example is that surname-only imputation yields the highest F1 score for Asian people but does not produce the most accurate Asian population share. Together, these patterns confirm a key finding: methods can outperform on one dimension while underperforming on others.

Table 1: Rank-Order of Imputation Method Performance

	Surname	Zip	BISG	BIFSG	ZRP	NamePrism	BIRDIE
F1							
White	5	4	2	1	3	6	–
Black	5	4	3	1	2	6	–
Hispanic	1	6	5	3	4	2	–
Asian	1	6	4	2	5	3	–
<i>F1 Micro</i> [†]	4	5	2	1	3	6	–
<i>F1 Macro</i> [†]	4	6	2	1	3	5	–
<i>F1 Weighted</i> [†]	4	5	2	1	3	6	–
Racial Gaps							
Black–White	7	6	4	3	2	5	1
Black–Hispanic	3	4	6	7	5	1	2
Black–Asian	3	6	7	5	2	1	4
Hispanic–White	3	1	5	7	6	4	2
Hispanic–Asian	1	3	5	6	7	2	4
Asian–White	5	3	2	1	7	6	4
Population Accuracy							
White	5	4	3	2	1	6	–
Black	5	4	3	1	2	6	–
Hispanic	1	6	2	3	4	5	–
Asian	3	6	1	2	5	4	–
<i>Distance</i> [†]	5	4	3	2	1	6	–

Notes: Rank 1 is best (highlighted in teal). For F1, best means highest; for racial gaps, closest to observed gaps; for population accuracy, closest to observed shares; for distance, lowest. [†]Aggregate summary measure. BIRDIE is omitted outside the racial gaps panel.

These patterns have first-order implications for empirical practice. Superior individual-level classification accuracy does not guarantee superior racial gap or demographic estimates. Researchers who select methods based on F1 scores alone — a common empirical practice — risk poor performance on the dimension that matters for their analysis. Our results call for broader evaluation of imputation methods across multiple dimensions, rather than reliance on any single validation criterion.

Table 1 also shows that newer ML-based methods do not consistently outperform simpler methods. ZRP ranks first for the White population estimate and overall distance from the observed population, while NamePrism ranks first for the Black-Hispanic and Black-Asian racial gaps. However, outside these cases, both are consistently outperformed by simpler methods, such as BIFSG or surname-based imputation. Greater model complexity does not always guarantee better performance.

7 Conclusion

Race imputation has become common when studying economic inequality, so much so that citations to standard methods more than tripled between 2020 and 2024. Despite this prevalence and the variety of available methods, systematic comparisons remain rare. This paper addresses that gap by evaluating both standard and newly developed methods across three performance dimensions: individual-level classification, racial gap estimation, and demographic estimation. Our key finding is that no single method dominates — methods can excel on one dimension while underperforming on others.

This finding has first-order implications for empirical practice. As our rankings show, superior classification accuracy does not guarantee better performance on the dimensions that matter most for applied researchers, racial disparities and racial demographics.

Race imputation is one example of a broader trend: data-driven predictions increasingly underlie empirical research across the social sciences. As individual-level data proliferates and imputation of everything from race to income becomes ever more prevalent, scrutinizing the assumptions and limitations of these methods becomes increasingly important.

References

- Albanesi, Stefania, and Domonkos Vamossy.** 2024. "Credit Scores: Performance and Equity." National Bureau of Economic Research, Working Paper 32917.
- Argyle, Bronson, Sasha Indarte, Benjamin Iverson, and Christopher Palmer.** 2025. "Racial Disparities and Bias in Consumer Bankruptcy." National Bureau of Economic Research, Working Paper 33575.
- Argyle, Lisa, and Michael Barber.** 2024. "Misclassification and Bias in Predictions of Individual Ethnicity from Administrative Records." *American Political Science Review*, 118(2): 1058–1066.
- Ashurst, Carolyn, and Adrian Weller.** 2023. "Fairness Without Demographic Data: A Survey of Approaches." *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, 1–12.
- Banerjee, Asha.** 2022. "Understanding Economic Disparities Within the AAPI Community." Economic Policy Institute. <https://www.epi.org/blog/understanding-economic-disparities-within-the-aapi-community/>.
- Baron, Jason, Joseph Doyle, Natalia Emanuel, Peter Hull, and Joseph Ryan.** 2026. "Unwarranted Disparity in High-Stakes Decisions: Race Measurement and Policy Responses." In *Race, Ethnicity, and Economic Statistics for the 21st Century.*, ed. Randall Akee, Lawrence F. Katz and Mark Loewenstein, 207–238. University of Chicago Press.
- Bleemer, Zachary.** 2025. "The Vietnam War and Racial Integration." National Bureau of Economic Research, Working Paper 34900.
- Box-Couillard, Sebastien, and Peter Christensen.** 2024. "Racial Housing Price Differentials and Neighborhood Segregation." National Bureau of Economic Research, Working Paper 32815.
- Chalavadi, Sanjana, Andrei Pastor, and Terry Leitch.** 2025. "LSTM+Geo with xgBoost Filtering: A Novel Approach for Race and Ethnicity Imputation with Reduced Bias." *arXiv*. <https://arxiv.org/pdf/2504.21259>.
- Chaturvedi, Rochana, and Sugat Chaturvedi.** 2024. "It's All in the Name: A Character-Based Approach to Infer Religion." *Political Analysis*, 32(1): 34–49.
- Chernenko, Sergey, and David Scharfstein.** 2023. "The Limits of Algorithmic Measures of Race in Studies of Outcome Disparities." Social Science Research Network, Working Paper 4426161. <https://ssrn.com/abstract=4426161>.

- Choukhmane, Taha, Jorge Colmenares, Cormac O’Dea, Jonathan Rothbaum, and Lawrence Schmidt.** 2023. “Who Benefits from Retirement Saving Incentives in the US? Evidence on Racial Gaps in Retirement Wealth Accumulation.” Michigan Retirement and Disability Center, University of Michigan, Working Paper 483.
- Christensen, Garret, and Edward Miguel.** 2018. “Transparency, reproducibility, and the credibility of economics research.” *Journal of Economic Literature*, 56(3): 920–980.
- Chyn, Eric, and Kareem Haggag.** 2023. “Moved to Vote: The Long-Run Effects of Neighborhoods on Political Participation.” *Review of Economics and Statistics*, 105(6): 1596–1605.
- Clark, Jesse, John Curiel, and Tyler Steelman.** 2021. “New Levels of Racial Estimation.” https://johnacuriel.org/wp-content/uploads/2021/07/clark_curiel_steelman2021pa.pdf.
- Clark, Jesse, John Curiel, and Tyler Steelman.** 2022. “Minmaxing Of Bayesian Improved Surname Geocoding And Geography Level Ups In Predicting Race.” *Political Analysis*, 30(3): 456–462.
- Cohle, Zachary, and Alberto Ortega.** 2023. “The Effect of the Opioid Crisis on Patenting.” *Journal of Economic Behavior & Organization*, 214: 493–521.
- Cohn, D’Vera.** 2010. “Race and the Census: The “Negro” Controversy.” Pew Research Center. <https://www.pewresearch.org/social-trends/2010/01/21/race-and-the-census-the-negro-controversy/>.
- Comenetz, Joshua.** 2016. “Frequently Occurring Surnames in the 2010 Census.” United States Census Bureau. <https://www2.census.gov/topics/genealogy/2010surnames/surnames.pdf>.
- Consumer Financial Protection Bureau.** 2014. “Using Publicly Available Information to Proxy for Unidentified Race and Ethnicity: A Methodology and Assessment.” https://files.consumerfinance.gov/f/201409_cfpb_report_proxy-methodology.pdf.
- Cook, Lisa, Matt Marx, and Emmanuel Yimfor.** 2025. “Funding Black High-Growth Startups.” Social Science Research Network. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4279986.
- Costello, Rachel, Portia DeFilippes, Robin Fisher, Ben Klemens, and Emily Lin.** 2024. “Marriage Penalties and Bonuses by Race and Ethnicity: An Application of Race and Ethnicity Imputation.” *AEA Papers and Proceedings*, 114: 644–648.

- Coyne, David, Itzik Fadlon, and Tommaso Porzio.** 2022. "Measuring Valuation of Liquidity with Penalized Withdrawals." Centre for Economic Policy Research. <https://cepr.org/index%2Ephp/system/files/publication-files/DP17499.pdf>.
- Cronin, Julie-Anne, Portia DeFilippes, and Robin Fisher.** 2023. "Tax Expenditures by Race and Hispanic Ethnicity: An Application of the US Treasury Department's Race and Hispanic Ethnicity Imputation." US Department of the Treasury, Office of Tax Analysis. <https://home.treasury.gov/system/files/131/WP-122.pdf>.
- Cruz, José, and Jacqueline Hayes.** 2009. "Adding Race and Ethnicity: Electoral Data Collection Practice and Prospects for New York State." State University of New York at Albany. https://scholarsarchive.library.albany.edu/cgi/viewcontent.cgi?article=1016&context=nylarnet_docs.
- Curiel, John, and Kevin DeLuca.** 2024. "The Power of Characters: Evaluating Machine Learning-Modified Bayesian Improved Surname Geocoding Inference of Race in Redistricting." *State Politics & Policy Quarterly*, 24(3): 300–321.
- Dasanaïke, Noah.** 2026. "Large Language Models Naively Recover Ethnicity from Individual Records." arXiv. <https://arxiv.org/pdf/2601.21132>.
- Decter-Frain, Ari.** 2022. "How Should We Proxy for Race/Ethnicity? Comparing Bayesian Improved Surname Geocoding to Machine Learning Methods." arXiv. <https://arxiv.org/pdf/2206.14583>.
- Diamond, Rebecca, and William Diamond.** 2024. "Racial Differences in the Total Rate of Return on Owner-Occupied Housing." National Bureau of Economic Research, Working Paper 32916.
- Dobre, Alexandra, and B. Heath Witzén.** 2023. "Mortgage Lender Misclassification of Applicant Race and Ethnicity: Evidence from HMDA Data." Consumer Financial Protection Bureau. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4636883.
- Dong, Evan, Aaron Schein, Yixin Wang, and Nikhil Garg.** 2025. "Addressing discretization-induced bias in demographic prediction." *PNAS Nexus*, 4(2): pgaf027.
- Edwards, Frank, Hedwig Lee, and Michael Esposito.** 2019. "Risk of Being Killed by Police Use of Force in the United States by Age, Race-Ethnicity, and Sex." *Proceedings of the National Academy of Sciences*, 116(34): 16793–16798.
- Einav, Liran, and Jonathan Levin.** 2014. "Economics in the Age of Big Data." *Science*, 346(6210): 715–721.

- Elliott, Marc N, Peter A Morrison, Allen Fremont, Daniel F McCaffrey, Philip Pantoja, and Nicole Lurie.** 2009. "Using the Census Bureau's Surname List to Improve Estimates of Race/Ethnicity and Associated Disparities." *Health Services and Outcomes Research Methodology*, 9: 69–83.
- Elzayn, Hadi, Evelyn Smith, Thomas Hertz, Cameron Guage, Arun Ramesh, Robin Fisher, Daniel E Ho, and Jacob Goldin.** 2025. "Measuring and Mitigating Racial Disparities in Tax Audits." *Quarterly Journal of Economics*, 140(1): 113–163.
- Feigenberg, Benjamin, and Conrad Miller.** 2022. "Would Eliminating Racial Disparities in Motor Vehicle Searches Have Efficiency Costs?" *Quarterly Journal of Economics*, 137(1): 49–113.
- Finlay, Keith, Elizabeth Luh, and Michael Mueller-Smith.** 2024. "Race and Ethnicity (Mis)measurement in the U.S. Criminal Justice System." National Bureau of Economic Research, Working Paper 32657.
- Fiscella, Kevin, and Allen M Fremont.** 2006. "Use of Geocoding and Surname Analysis to Estimate Race and Ethnicity." *Health services research*, 41(4p1): 1482–1500.
- Fisher, Robin.** 2023. "Estimation of Race and Ethnicity by Re-Weighting Tax Data." Office of Tax Analysis, Department of the Treasury. <https://home.treasury.gov/system/files/131/TP-11.pdf>.
- Frake, Justin, Reuben Hurst, and Max Kagan.** 2024. "Partisan Segregation in the US Workplace is Large and Rising." Social Science Research Network. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4639165.
- Frame, Scott, Ruidi Huang, Erik Mayer, and Adi Sunderam.** 2021. "The Impact of Minority Representation at Mortgage Lenders." Cox School of Business, Southern Methodist University. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4054761.
- Fremont, Allen, Joel S Weissman, Emily Hoch, and Marc N Elliott.** 2016. "When Race/Ethnicity Data Are Lacking: Using Advanced Indirect Estimation Methods to Measure Disparities." *RAND Health Quarterly*, 6(1).
- Ganong, Peter, Damon Jones, Pascal Noel, Diana Farrell, Fiona Greig, and Chris Wheat.** 2023. "Wealth, Race, and Consumption Smoothing of Typical Income Shocks." National Bureau of Economic Research, Working Paper 27552.

- García, Raffi E, and William A Darity Jr.** 2022. "Self-Reporting Race in Small Business Loans: A Game-Theoretic Analysis of Evidence from PPP Loans in Durham, NC." *AEA Papers and Proceedings*, 112: 299–302.
- Goldsmith-Pinkham, Paul.** 2024. "Tracking the Credibility Revolution across Fields." arXiv. <https://arxiv.org/pdf/2405.20604>.
- Goncalves, Felipe, and Steven Mello.** 2021. "A Few Bad Apples? Racial Bias in Policing." *American Economic Review*, 111(5): 1406–1441.
- Gong, Xiaofeng, Yuling Han, Susan Parker, Matthew B Ross, and Stephen L Ross.** 2025. "Measuring Racial Disparities in Police Stops: Evidence from Telemetric Mobility Data." <https://static1.squarespace.com/static/5574b6c0e4b0793b94a3401d/t/6940e6e712fb671cbcf06408/1765861095679/MBRoss-DisparitiesTelemetric-12152025.pdf>.
- Greengard, Philip, and Andrew Gelman.** 2024. "An Improved BISG for inferring Race from Surname and Geolocation." arXiv. <https://arxiv.org/html/2304.09126v3>.
- Greenwald, Daniel L, Sabrina T Howell, Cangyuan Li, and Emmanuel Yimfor.** 2024. "Regulatory Arbitrage or Random Errors? Implications of Race Prediction Algorithms in Fair Lending Analysis." *Journal of Financial Economics*, 157: 103857.
- Harris, Ada.** 2020. "Using Bayesian Improved Surname Geocoding (BISG) to Classify Race and Ethnicity in Administrative Employment Data by Industry: A Validation Study [Conference Session]." Joint Statistical Meetings. <https://ww2.amstat.org/meetings/jsm/2020/onlineprogram/AbstractDetails.cfm?abstractid=311006>.
- Hepburn, Peter, Renee Louis, and Matthew Desmond.** 2020. "Racial and Gender Disparities Among Evicted Americans." *Sociological Science*, 7: 649–662.
- Howell, Sabrina T, Theresa Kuchler, David Snitkof, Johannes Stroebel, and Jun Wong.** 2022. "Lender Automation and Racial Disparities in Credit Access." National Bureau of Economic Research Working Paper 29364.
- Howell, Sabrina T, Theresa Kuchler, David Snitkof, Johannes Stroebel, and Jun Wong.** 2024. "Lender Automation and Racial Disparities in Credit Access." *Journal of Finance*, 79(2): 1457–1512.
- Hurtado, Agustin, and Jung Sakong.** 2023. "The Effect of Minority Bank Ownership on Minority Credit." George J. Stigler Center for the Study of the Economy and the State, University of Chicago. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4590142.

- Iceland, John, Kimberly A Goyette, Kyle Anne Nelson, and Chaowen Chan.** 2010. "Racial and Ethnic Residential Segregation and Household Structure: A Research Note." *Social Science Research*, 39(1): 39–47.
- Imai, Kosuke, and Kabir Khanna.** 2016. "Improving Ecological Inference By Predicting Individual Ethnicity From Voter Registration Records." *Political Analysis*, 24(2): 263–272.
- Imai, Kosuke, Santiago Olivella, and Evan TR Rosenman.** 2022. "Addressing Census Data Problems In Race Imputation Via Fully Bayesian Improved Surname Geocoding And Name Supplements." *Science Advances*, 8(49): eadc9824.
- Jarrín, Olga F, Abner N Nyandege, Irina B Grafova, XinQi Dong, and Haiqun Lin.** 2020. "Validity of Race and Ethnicity Codes in Medicare Administrative Data Compared with Gold-Standard Self-Reported Race Collected During Routine Home Health Care Visits." *Medical Care*, 58(1): e1–e8.
- Jensen, Eric, and Nicholas Jones.** 2026. "Measuring the Racial and Ethnic Composition and Diversity of the United States Population: Historical Challenges and Contemporary Opportunities." In *Race, Ethnicity, and Economic Statistics for the 21st Century.*, ed. Randall Akee, Lawrence F. Katz and Mark Loewenstein, 19–55. University of Chicago Press.
- King, Gary.** 1995. "Replication, Replication." *PS: Political Science & Politics*, 28(3): 444–452.
- King, Jennifer, Daniel Ho, Arushi Gupta, Victor Wu, and Helen Webley-Brown.** 2023. "The Privacy-Bias Tradeoff: Data Minimization and Racial Disparity Assessments in US Government." *Proceedings of the 6th ACM Conference on Fairness, Accountability, and Transparency*, 492–505.
- Labgold, Katie, Sarah Hamid, Sarita Shah, Neel Gandhi, Allison Chamberlain, Fazle Khan, Shamimul Khan, Sasha Smith, Steve Williams, Timothy Lash, et al.** 2021. "Estimating the Unknown: Greater Racial and Ethnic Disparities in COVID-19 Burden after Accounting for Missing Race and Ethnicity Data." *Epidemiology*, 32(2): 157–161.
- Li, Cangyuan.** 2023. "Can We Trust Race Prediction?" arXiv. <https://arxiv.org/pdf/2307.08496>.
- Loewenstein, Mark, David Piccone, and Anne Polivka.** 2024. "Measuring Potential Effects of the Proposed Race and Ethnicity Questions in the Current Population Survey." National Bureau of Economic Research, Working Paper 32812.
- Logan, John.** 2013. "The Persistence of Segregation in the 21st Century Metropolis." *City & Community*, 12(2): 160–168.

- Luh, Elizabeth.** 2022. "Not So Black and White: Uncovering Racial Bias from Systematically Misreported Trooper Reports." Social Science Research Network. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3357063.
- Matthew, Abigail, Amalia Miller, and Catherine Tucker.** 2024. "Algorithmic Bias and Historical Injustice: Race and Digital Profiling." National Bureau of Economic Research, Working Paper 32485.
- Matthews, Kasey, Piotr Zak, Austin Li, Christien Williams, Sean Kamkar, and Jay Budzik.** 2024a. "ZRP Model Development Documentation." GitHub. https://github.com/zestai/zrp/blob/main/model_report.rst.
- Matthews, Kasey, Piotr Zak, Austin Li, Christien Williams, Sean Kamkar, and Jay Budzik.** 2024b. "ZRP Validation Experiments' Results." GitHub. https://github.com/zestai/zrp/blob/main/supporting_docs/validation_experiment_results.rst.
- McCartan, Cory, Robin Fisher, Jacob Goldin, Daniel E Ho, and Kosuke Imai.** 2025. "Estimating Racial Disparities When Race Is Not Observed." *Journal of the American Statistical Association*, 120(552): 2140–2153.
- McKinney, Kevin, and John Abowd.** 2024. "Estimating the Potential Impact of Combined Race and Ethnicity Reporting on Long-Term Earnings Statistics." National Bureau of Economic Research, Working Paper 32758.
- Medina, Jennifer, and K.K. Rebecca Lai.** 2023. "An American Puzzle: Fitting Race in a Box." *New York Times*. <https://www.nytimes.com/interactive/2023/10/16/us/census-race-ethnicity.html>.
- NamSor.** 2025a. "Explore Our Plans." <https://namsor.app/prices/>. Accessed: February 4, 2025.
- NamSor.** 2025b. "Find Out the Origin of a Name." <https://namsor.app/features/name-origin/>. Accessed: February 4, 2025.
- NamSor.** 2025c. "Namsor, Name Checker for Gender, Origin and Ethnicity Determination." <https://namsor.app>. Accessed: February 4, 2025.
- Rieke, Aaron, Vincent Southerland, Dan Svirsky, and Mingwei Hsu.** 2022. "Imperfect Inferences: A Practical Assessment." *Proceedings of the 5th ACM Conference on Fairness, Accountability, and Transparency*.
- Rose, Evan K.** 2023. "A Constructivist Perspective on Empirical Discrimination Research." *Journal of Economic Literature*, 61(3): 906–923.

- Rosenman, Evan TR, Santiago Olivella, and Kosuke Imai.** 2023. "Race and Ethnicity Data for First, Middle, and Surnames." *Scientific Data*, 10(1): 299.
- Sebo, Paul.** 2023. "How Well Does NamSor Perform in Predicting the Country of Origin and Ethnicity of Individuals Based on Their First and Last Names?" *Plos ONE*, 18(11): e0294562.
- Siskin, Bernard R.** 2014. "Emerging Trends in Quantitative Analytics [Conference Session]." Consumer Banking Association Live, Washington, DC. https://financialservices.house.gov/uploadedfiles/siskin_presentation_-_red_white_and_banking.pdf.
- Sorbero, Melony, Roald Euler, Aaron Kofner, and Marc Elliott.** 2022. "Imputation of Race and Ethnicity in Health Insurance Marketplace Enrollment Data, 2015-2022 Open Enrollment Periods." RAND Corporation. https://www.rand.org/pubs/research_reports/RRA1853-1.html.
- Starr, Paul, and Christina Pao.** 2024. "The Multiracial Complication: The 2020 Census and the Fictitious Multiracial Boom." *Sociological Science*, 11: 1107–1123.
- U.S. Census Bureau.** 2021. "Frequently Occurring Surnames from the Census 2010." https://www.census.gov/topics/population/genealogy/data/2010_surnames.html. Accessed: August 13, 2025.
- U.S. Census Bureau.** 2022. "Frequently Occurring Surnames from the Census 2000." https://www.census.gov/topics/population/genealogy/data/2000_surnames.html. Accessed: August 13, 2025.
- U.S. Census Bureau.** 2024a. "Updates to Race/Ethnicity Standards for Our Nation." <https://www.census.gov/about/our-research/race-ethnicity/standards-updates.html>. Accessed: May 21, 2025.
- U.S. Census Bureau.** 2024b. "What Updates to OMB's Race/Ethnicity Standards Mean for the Census Bureau." <https://www.census.gov/newsroom/blogs/random-samplings/2024/04/updates-race-ethnicity-standards.html>. Accessed: May 21, 2025.
- U.S. Office of Management and Budget.** 2024a. "OMB Statistical Policy Directive No. 15 on Race and Ethnicity Data Standards: Data Collection and Editing Procedures." <https://spd15revision.gov/content/spd15revision/en/2024-spd15/data-collection-editing-procedures.html> Page Last Revised: April 7, 2026.
- U.S. Office of Management and Budget.** 2024b. "OMB Statistical Policy Directive No. 15 on Race and Ethnicity Data Standards: Question Format." <https://spd15revision.gov/>

<content/spd15revision/en/2024-spd15/question-format.html> Page Last Revised: April 7, 2026.

U.S. Small Business Administration. 2024. "Paycheck Protection Program (PPP) FOIA Data." <https://data.sba.gov/dataset/ppp-foia>. Accessed September 30, 2024.

Ventura, Ilana M, and René D Flores. 2025. "The "Rise" of Multiracials? Examining the Growth in Multiracial Identification in the 2020 US Census." *RSF: The Russell Sage Foundation Journal of the Social Sciences*, 11(1): 44–64.

Voicu, Ioan. 2018. "Using First Name Information to Improve Race and Ethnicity Classification." *Statistics and Public Policy*, 5(1): 1–13.

Whitsel, Eric A, P Miguel Quibrera, Richard L Smith, Diane J Catellier, Duanping Liao, Amanda C Henley, and Gerardo Heiss. 2006. "Accuracy of Commercial Geocoding: Assessment and Implications." *Epidemiologic Perspectives & Innovations*, 3(1): 8.

Ye, Junting, Shuchu Han, Yifan Hu, Baris Coskun, Meizhu Liu, Hong Qin, and Steven Skiena. 2017. "Nationality Classification Using Name Embeddings." arXiv. <https://arxiv.org/pdf/1708.07903>.

Software References

- Curiel, John.** 2024. "zipWRUext: Package for the wru ZIP code extension." Github. <https://github.com/jcuriel-unc/zipWRUext>.
- Kaplan, Jacob.** 2025. "predictrace: Predict the Race and Gender of a Given Name Using Census and Social Security Administration Data." Version 2.0.1. Github. <https://jacobkap.github.io/predictrace/>.
- Khanna, Kabir, Brandon Bertelsen, Santiago Olivella, Evan Rosenman, Alexander R. Hayes, and Kosuke Imai.** 2024. "wru: Who are You? Bayesian Prediction of Racial Category Using Surname, First Name, Middle Name, and Geolocation." <https://github.com/kosukeimai/wru>.
- Matthews, Kasey, Piotr Zak, Austin Li, Christien Williams, Sean Kamkar, and Jay Budzik.** 2024. "Zest Race Predictor." GitHub. <https://github.com/zestai/zrp>.
- McCartan, Cory.** 2023. "BIRDIE: Bayesian Instrumental Regression for Disparity Estimation." Github. <https://github.com/CoryMcCartan/birdie>.
- Skiena, Steven.** 2025. "NamePrism API." <https://www.name-prism.com/api>.
- United States Census Bureau.** 2025. "Census Geocoder." <https://geocoding.geo.census.gov/geocoder/>.

Supplemental Appendix

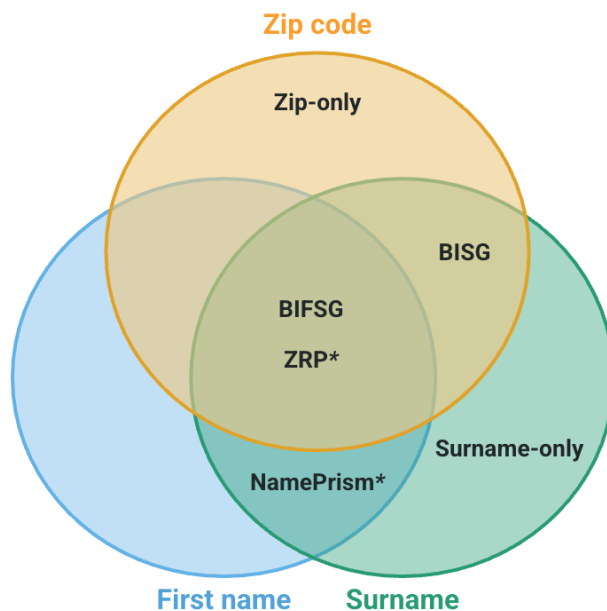
A.1 Supplemental Tables and Figures

Table A.1: Sample Composition by Borrower Race and Lender Type

Borrower Race	Total	Non-Fintech Lender	Fintech Lender
White	150,681	137,354	13,327
Black	63,242	39,552	23,690
Hispanic	23,907	17,578	6,329
Asian	22,627	15,793	6,834
Other	6,755	6,316	439
Total	267,212	216,593	50,619

Notes: This table shows the number of PPP loans by borrower race and lender type. Fintech lenders are defined following [Howell et al. \(2022\)](#). The sample is restricted to first-draw loans made before February 24, 2021 with self-reported race, identifiable person names, and a valid zip code. See Appendix A.3 for more data details.

Figure A.1: Inputs Used for Race Imputation Methods



Notes: This figure shows the input data used by each race imputation method in our context. Recently developed methods are denoted with asterisks. BIRDIE is excluded as it is an adjustment method rather than a direct imputation method. BIFSG requires mapping zip code to census tract before implementation.

A.2 Further Background

A.2.1 Race Imputation in US Government Agencies

Race imputation has been used across multiple US government agencies beyond the US Treasury: the Consumer Financial Protection Bureau (CFPB) has used it to assess fair lending ([Consumer Financial Protection Bureau, 2014](#)), the Equal Employment Opportunity Commission (EEOC) has used it to analyze demographic disparities in labor discrimination cases ([Harris, 2020](#)), and the Department of Health & Human Services (HHS) and the Centers for Medicare and Medicaid Services (CMS) have used it to study healthcare inequities ([Sorbero et al., 2022](#); [Fremont et al., 2016](#)). The US Court of Appeals has also upheld the use of race imputation in federal voting rights cases ([Greengard and Gelman, 2024](#)).

A.2.2 Combined Race and Ethnicity in the US Census

The Census reports that growth in the "Some Other Race" category has been driven largely by Hispanic respondents who do not see their identity reflected in the existing race options ([U.S. Census Bureau, 2024b](#)). A combined race and ethnicity question for the decennial Census is expected to reduce reliance on this residual category by encouraging respondents to select from the available choices. This change is also planned for earlier adoption in the American Community Survey ([U.S. Census Bureau, 2024a](#)). For analyses of how this shift may affect measurement, see [Loewenstein, Piccone and Polivka \(2024\)](#) for impacts on the Current Population Survey (CPS) and [McKinney and Abowd \(2024\)](#) for implications on long-term earnings statistics.

A.3 Data Appendix

Raw data (11.8M observations): We use the Paycheck Protection Program (PPP) FOIA release from the U.S. Small Business Administration (SBA) ([U.S. Small Business Administration, 2024](#)). The dataset is loan-level and covers PPP lending between March 2020 and May 2021. The unit of observation is a PPP loan, and the borrower name field can contain either individual names or business/entity names. The FOIA release is available as 13 CSV files that must be appended to reconstruct the full dataset. The first 12 files cover loans \leq \$150k (10,499,686 observations), and file 13 covers loans $>$ \$150k (968,524 observations). We append all files together, which yields a dataset of 11,468,210 observations.

Simple filtering (yields 1.1M observations): Next, we restrict to first-draw loans before February 24, 2021 with valid zip and self-reported race. The following variables in the raw allow us to make these restrictions:

- BorrowerZip: borrower ZIP code,²⁴ missing is encoded as NaN
- Race: self-reported owner race; missing race is encoded as Unanswered
- Approvedate: used to restrict loans approved before Feb 24, 2021
- ProcessingMethod: identifies first-draw PPP loans, which are coded as PPP

The restricted sample consists of 1,112,727 observations. The primary constraint on the sample is self-reporting race. Most owners do not self-report race.

Name-based filtering (yields 275,007 observations): The final required step is to subset down to observations with an identifiable person name. In practice, many values for the BorrowerName variable are company names. Therefore, we develop an approach to omit those observations, which consists of the following steps:

- **Step 1:** We remove observations whose BorrowerName field contains any digit (0–9) or any of the special characters &, /, @, or #. This step yields 1,005,014 observations.
- **Step 2:** We omit observations where the BorrowerName is likely a business name instead of a person name. We therefore omit (1) names containing any legal entity suffix (see **List 1**); (2) names containing any broad business keywords (see **List 2**); (3)

²⁴Note that the zip code in this data is for the business address. Ideally, we would impute using a residential address, however, that would require a linkage with non-public data (e.g., from Infutor, as done in [Greenwald et al. \(2024\)](#)). [Greenwald et al. \(2024\)](#) report that their results are similar if they use a business address, so we view this as a good alternative, especially given it keeps the exercise in the realm of public data.

names containing the term "DBA," which stands for "Doing Business As." This step yields 277,384 observations.

- **Step 3:** We omit observations where BorrowerName includes the word "and" since we observe these are businesses. This yields a final sample of 275,007 observations with identifiable person names.

List 1 (Legal Suffixes):

- LLC, L.L.C, INC, INCORPORATED, CORP, CORPORATION, LTD, LIMITED, PLLC, P.L.L.C, PC, P.C, PA, P.A, LP, L.P, LLP, L.L.P, PLC

List 2 (Business Keywords):

- *Entity types:* COMPANY, CO, ENTERPRISES, ENTERPRISE, PS
- *Financial:* BANK, BANC, CREDIT, UNION
- *Corporate/Org:* SERVICES, SERVICE, GROUP, HOLDINGS, CONSULTING, CONSULTANTS, SOLUTIONS, PARTNERS, PARTNERSHIP, ASSOCIATES, ASSOCIATION, FOUNDATION, INSTITUTE, CENTER, MANAGEMENT, MGMT
- *Healthcare:* CLINIC, HOSPITAL, MEDICAL, DENTAL, HEALTH, CARE, CHIROPRACTIC, DENTISTRY
- *Food/Hospitality:* RESTAURANT, CAFE, BAR, GRILL, PIZZA, FOOD, KITCHEN, DELI, BAKERY, CATERING, BREWERY
- *Personal Services:* SALON, SPA, BEAUTY, BARBER, NAIL, HAIR, TAILOR, TAILORING
- *Construction/Trades:* CONSTRUCTION, BUILDERS, BUILDING, ROOFING, PLUMBING, ELECTRIC, CONTRACTING, PAINTING, SIDING
- *Auto/Transport:* AUTO, AUTOMOTIVE, MOTORS, TRUCKING, TRANSPORT, LOGISTICS
- *Retail:* SHOP, STORE, MART, MARKET, SUPPLY, SUPPLIES, DEPOT, WAREHOUSE, DISTRIBUTION, MKT
- *Real Estate/Finance:* REALTY, REAL ESTATE, PROPERTIES, PROPERTY, INVESTMENTS, AGENCY, INSURANCE, FINANCIAL, ACCOUNTING, TAX, BOOKKEEPING

- *Tech*: TECHNOLOGIES, TECHNOLOGY, TECH, SYSTEMS
- *Maintenance*: CLEANING, MAINTENANCE, REPAIR, INSTALLATION
- *Education*: SCHOOL, ACADEMY, LEARNING, EDUCATION, TRAINING, UNIVERSITY, COLLEGE
- *Government*: CITY, COUNTY, STATE, TOWNSHIP, DEPARTMENT
- *Religious*: CHURCH, MINISTRIES, MINISTRY, PARISH, ABBEY
- *Landscaping/Agriculture*: LANDSCAPING, LAWN, GARDEN, TREE, FARM, FARMS, RANCH, DAIRY, CATTLE
- *Creative/Media*: PHOTOGRAPHY, STUDIO, DESIGN, CREATIVE, MEDIA, PRODUCTION, IMAGE, IMAGING
- *Entertainment/Arts*: ENTERTAINMENT, ORCHESTRA
- *Rentals*: RENTALS, RENTAL, LEASING
- *Legal*: THE, OFFICE, OFFICES, LAW, LEGAL, ATTORNEY, TRUST, ESTATE
- *Fitness*: FITNESS, GYM, YOGA, WELLNESS
- *Childcare*: DAYCARE
- *Organizations*: CHAMBER, COMMERCE, ASSEMBLY
- *Misc*: INTERNATIONAL, NATIONAL, GLOBAL, WORLDWIDE, PROFESSIONAL, PROFESSIONALS, DETAIL, DETAILING, CARWASH, WASH, TAXIDERMY, GROOMING, PET, VETERINARY, VET, CHARTERS, FISHING, INN, LODGE, LOUNGE, MACHINE, MOTEL, TOURS, TOWING, HVAC
- *Other*: OF, MASSAGE, FABRIC, CERAMIC, TREES, BUSHES, SHRUBS, SCHOOLS, CUPS, CREAM, COUNSELING, JEWELRY

Name cleaning (yields 267,212 observations): The BorrowerName field does not distinguish first, middle, and last name components, but imputation methods require first and last name fields. We strip credentials and suffixes, then drop observations with only one name component or with five or more components. This leaves 267,212 observations with 2-part (72%), 3-part (25%), or 4-part names (3%). We define the first component as the first name and the last component as the last name, leaving any middle components unused.

Racial categories: The original PPP dataset contains the following self-reported race categories: White, Black or African American, Asian, American Indian or Alaska Native,

Native Hawaiian or Other Pacific Islander, Puerto Rican, Multi Group, and Eskimo & Aleut. Since the dataset does not explicitly include a unified Hispanic or Latino race category, we construct it using the borrower's ethnicity field. Specifically, if a borrower reports their ethnicity as Hispanic or Latino, we replace the race value with Hispanic or Latino; otherwise, we retain the borrower's self-reported race.

In addition, for concision, we shorten Black or African American to Black and Hispanic or Latino to Hispanic. We also consolidate several groups to align with standard categories used in race imputation:

- Native Hawaiian or Other Pacific Islander → Asian
- Puerto Rican → Hispanic
- American Indian or Alaska Native → Other
- Eskimo & Aleut → Other
- Multi Group → Other

After this step, the final race groups used in the analysis are White, Black, Asian, Hispanic, and Other.

Fintech indicator variable: We follow [Howell et al. \(2022\)](#) and define a lender as a Fintech if it's one of the lenders in their Table A.1 (**List 3**).

List 3 (Fintech lenders):

- Cross River Bank, Kabbage, Celtic Bank Corporation, Lendio, WebBank, Customers Bank, Readycap Lending, Itria Ventures, Intuit Financing, Newtek Small Business Finance, Fundbox, MBE Capital Partners, FC Marketplace, Harvest Small Business Finance, Fountainhead SBF, CRF Small Business Loan Company, Sunrise Banks National Association, Accion, Fund-Ex Solutions Group, The Bancorp Bank, Centerstone SBA Lending, Grow America Fund, Evolve Bank and Trust, NBKC Bank, immito, Loan Source, BayBank, VelocitySBA

A.4 Black Box Private Sector Methods

A.4.1 Imputation in Purchased Data

Empirical researchers are increasingly using private data for research (Einav and Levin, 2014). Proprietary datasets that include race variables may be derived from their own proprietary race imputation methods.

A prominent example is L2’s national voter file, which has been used in researching bankruptcy (Argyle et al., 2025), the Vietnam war (Bleemer, 2025), political identity (Frake, Hurst and Kagan, 2024), and more. L2 imputes race unless it is self-reported in the underlying state voter file (Frake, Hurst and Kagan, 2024; Bleemer, 2025). The L2 imputation is proprietary (Argyle et al., 2025), but researchers have linked L2 to other data (such as SNAP records or college applications) and documented discrepancies between L2’s imputed race and self-reported race in those administrative sources (Chyn and Haggag, 2023; Bleemer, 2025).

A.4.2 NameSor

NamSor is a private company that sells race/ethnicity imputation (NamSor, 2025c). NamSor advertises that it can provide the origin, ethnicity, the US race classification, country, and gender of names (NamSor, 2025b).

The tool has been used in 364 research papers (as of February 2025), according to NamSor’s website (NamSor, 2025c). Most papers that use NamSor are in the medical field or are papers that analyze academic publications biases. To the best of our knowledge, Sebo (2023) is the only paper to evaluate the performance of NamSor against data with known identification. That said, there has not been a comparative evaluation of NamSor against other methods like BISG or NamePrism.²⁵

We do not compare NamSor against other methods. We make this choice for two reasons. First, NamSor is prohibitively expensive for many research contexts. NamSor’s pricing page (as of February 2025) shows that it costs \$4,167 to process 500,000 names (NamSor, 2025a). Second, the methods used in NamSor are not transparent. In this paper, we compare free methods for race imputation with technical documentation. We do not extend the exercise to include black box tools for purchase.

²⁵NamSor has more in common with NamePrism than other methods in our main text. NamSor and NamePrism are unique in that they have international use cases.

A.5 Geocoding Complications

Geocoding addresses introduces several complications. First, geocoding generates missing data because not all addresses can be successfully geocoded. Second, it introduces its own methods ambiguity: missing rates and accuracy vary across geocoding methods (Clark, Curiel and Steelman, 2021; Whitsel et al., 2006).²⁶ Third, geocoding can be financially expensive (proprietary vendors) or computationally expensive (open source methods) (Clark, Curiel and Steelman, 2021). To maximize data coverage, avoid methodological ambiguity, and minimize cost, we use zip codes instead of geocoding for all imputation methods that use geography.

²⁶Examples of proprietary geocoders are Google, HERE, and ESRI. Meanwhile, free geocoding can be implemented by using the Census API, which bases geocoding on Census Bureau TIGER/Line files (United States Census Bureau, 2025). Separately, some papers studying race imputation with voter file data use L2 proprietary data, which already includes geocoded addresses, sidestepping methodological ambiguity (McCartan et al., 2025; Rosenman, Olivella and Imai, 2023).