Modern Fair Lending Analysis

The Hidden Biases of BISG Proxy-Based Disparity Estimates

By Richard R. Pace

Table of Contents

Executive Summary

Around 2013, certain U.S. federal and state financial regulators (the "financial regulators") began to employ the Bayesian Improved Surname Geocoding ("BISG") race / ethnicity proxy model[1] to expand their supervision and enforcement of federal fair lending laws and regulations to consumer auto loans for which the collection of race / ethnicity data was prohibited.[2] Using these model-based proxies, the financial regulators estimated whether a lender's credit outcomes (e.g., loan prices) exhibited statistically-significant and unfavorable disparities for one or more prohibited basis groups in violation of applicable fair lending laws and regulations.[3] For those lenders where such disparities were present, the financial regulators and enforcement agencies entered into numerous public and non-public agreements alleging illegal price discrimination, requiring on-going fair lending monitoring of such disparities using the BISG-based proxies, and mandating hundreds of millions of dollars in customer remediation payments.

Since that time, the U.S. consumer lending industry has largely adopted the BISG proxy model to expand its fair lending compliance risk management activities – not only for automobile lending, but also for a growing list of other consumer lending products, such as personal loans and credit cards, for which actual race / ethnicity data is not permissibly collected. Additionally, these proxy-based analyses are being applied to a broader range of lending outcomes other than pricing – such as judgmental credit decisions by human underwriters and, more recently, certain model-based estimates used throughout an institution's loan marketing, underwriting, pricing, and servicing processes. In fact, the growing adoption of AI technologies paired with alternative consumer data to disrupt traditional consumer lending has led to increasing concerns over potential "algorithmic" discrimination based on race / ethnicity, sex, age, etc. This emerging fair lending risk has led to further consideration and use of the BISG proxy model to assess the potential algorithmic bias that may be present in such models[4], and for the design and execution of corrective actions to "de-bias" such models should evidence

[1] The BISG proxy model estimates the likelihood that a consumer belongs to one of six mutually-exclusive races / ethnicities – specifically, Non-Hispanic White, Non-Hispanic Black, Hispanic, Non-Hispanic Asian / Pacific Islander, Non-Hispanic American Indian / Alaskan Native, and Non-Hispanic Two or More Races. It was originally developed to explore potential racial / ethnic disparities in health care outcomes and was published in Marc N. Elliott et al., "Using the Census Bureau's Surname List to Improve Estimates of Race/Ethnicity and Associated Disparities," Health Services & Outcomes Research Methodology (2009) 9:69-83 (the "Elliott BISG paper").

[2] Currently, applicant race and ethnicity information is required to be collected and reported, under specific conditions, for certain residential mortgage loan applications under the Equal Credit Opportunity Act ("ECOA") as implemented by Regulation B and the Home Mortgage Disclosure Act ("HMDA") as implemented by Regulation C.

[3] See, for example, "CFPB and DOJ Order Ally to Pay $80 Million to Consumers Harmed by Discriminatory Auto Loan Pricing," CFPB Press Release, December 20, 2013.

[4] For example, the BISG proxy methodology was used to assess potential algorithmic bias in certain cash-flow based credit underwriting models used by a subset of U.S. financial institutions. See "The Use of Cash Flow Data in Automated Credit Underwriting," FinReg Lab, July 23, 2019.

of such bias be discovered.[5]

Given the growing usage of the BISG proxy model for modern fair lending compliance testing across the consumer lending landscape, as well as the significant legal, compliance, and reputational risks that allegations of discrimination based on such proxies can impose on institutions, one would expect that this model, and the corresponding fair lending testing approaches in which it is used, had satisfied rigorous model validation testing before formal adoption and use. Unfortunately, publicly-available research on the appropriateness of the BISG proxies for fair lending analysis has been fairly limited and mainly focused on two specific areas: (1) the aggregate-level and individual-level accuracies of the BISG-based race / ethnicity proxy estimates, and (2) the comparative magnitudes of loan price disparity estimates when measured using different BISG proxy approaches and different sets of control variables.[6]

For the first area, all the referenced studies rely on samples of HMDA-reportable residential mortgage loan applications with self-reported race and ethnicity to assess proxy accuracy. However, given that mortgage applicants tend to be of higher economic quality than the general population of U.S. adults on which the BISG proxy model is based, it is unsurprising that these studies find notable proxy accuracy errors at both the aggregate and individual levels. Additionally, it remains unclear how much of these measured inaccuracies are due to the underlying bias in the demographics of the HMDA data sample, and how much are due to the inherent inaccuracy of the BISG proxy model itself. The lack of answers to these important yet fundamental questions, eight years after the first adoption of the model for fair lending testing and enforcement activities, highlights just one of the important risks of the BISG proxy model for fair lending use cases – the lack of independent validation of the model's accuracy and appropriateness on the types of datasets to which it has (and will be) applied.[7]

For the second area, the Zhang BISG Proxy Paper analyzes disparities in average note rates from a HMDA-reportable residential mortgage loan sample, while the AFSA White Paper analyzes dealer markup disparities from a proprietary indirect auto loan database. In both studies, however, the true

---

[5] While admittedly counterintuitive, for fair lending compliance purposes it is still an open question as to whether an institution's use of actual or proxied consumer demographic data to de-bias an algorithm is technically permissible under current U.S. fair lending laws and regulations.

[6] See: (1) "Using Publicly Available Information to Proxy For Unidentified Race and Ethnicity: A Methodology and Assessment", CFPB, Summer 2014 (the "CFPB BISG Proxy Paper"), (2) "Fair Lending: Implications for the Indirect Auto Finance Market," American Financial Services Association, November 19, 2014 (the "AFSA White Paper"), and (3) Zhang, Yan, "Assessing Fair Lending Risks Using Race/Ethnicity Proxies," Management Science 64 (1), January 2018, pp. 178-197 (the "Zhang BISG Proxy Paper"). We note that our study is not based on a comprehensive assessment of all academic studies on, or related to, this topic. Rather, it only includes references to prominent industry and regulatory studies related to fair lending applications.

[7] As HMDA-reportable residential mortgage loans already have self-reported race / ethnicity data, the BISG proxy model is not practically relevant to this product type. Instead, it is most frequently applied to automobile loans, personal loans, and credit cards – all products for which: (1) independent validation results have not been produced due to the legal prohibition of race / ethnicity data collection for such products, and (2) the underlying sociodemographic characteristics of customers are likely different than for residential mortgage loan customers.

magnitudes of the underlying price disparities are unknown and the results simply provide a comparison of disparity estimates: (1) under different proxy approaches and relative to self-reported actual race / ethnicity indicators, and (2) under different sets of additional control variables – thereby limiting our understanding of potential proxy bias in these estimates.[8]

Given these limitations of existing studies, and the significant high-stakes uses of these proxies for U.S. consumer lenders and financial regulators, our goal in this study is to explore the properties of the BISG proxy model, and the associated fair lending disparity estimates generated therefrom, to identify important risks and limitations of which users should be aware when employing the model for fair lending compliance risk management purposes. However, unlike most previous work in this area, our study is designed: (1) without the demographic and economic biases of HMDA data, (2) with known "ground truth" fair lending price disparities, (3) analyzing both disparate treatment and disparate impact price discrimination scenarios,[9] and (4) using both the "BISG Continuous" and the "BISG Classification" approaches when estimating fair lending price disparities.[10]

This work relies on a large synthetic, but representative, geo-surname sample of U.S. adults with both "known" races / ethnicities and corresponding BISG race / ethnicity probabilities. This synthetic sample is generated via Monte Carlo techniques from the BISG proxy model (as implemented by the CFPB) which is described further in the next section. This approach is consistent with model validation testing practices that seek to explain key properties of models through an analysis of model outcomes under different input scenarios. To this synthetic data sample, we add various disparate treatment and disparate impact price discrimination scenarios with known disparity amounts, employ standard regression-based fair lending testing methodologies to estimate these disparities using the BISG proxy model outputs, and analyze the results to provide insights as to any biases discovered.

---

[8] Because these studies do not know (or have data on) all the underlying causal drivers of potential pricing differences across borrowers, the resulting price disparity estimates are likely tainted somewhat due to omitted variable bias. While this is largely an unavoidable outcome given the data limitations, it does complicate our ability to understand whether differences in disparity estimates based on proxies and actuals are due to omitted variables, the proxy approach itself, or an interaction of the two.

[9] Briefly, "disparate treatment" scenarios are where discrimination is based on an individual's actual race / ethnicity, and "disparate impact" scenarios are where discrimination occurs indirectly due to the uneven impact of a policy or practice on prohibited basis and control groups. Our analysis, detailed later in this study, finds that biases in fair lending price disparity estimates vary significantly based on the specific type of discrimination that is assumed.

[10] The "BISG Continuous" approach uses each customer's set of BISG probabilities, as is, in the fair lending regression analysis while the "BISG Classification" approach converts each individual's set of BISG probabilities into a specific race / ethnicity designation, and uses the resulting set of discrete proxy categories in the fair lending regression model. The former approach is most commonly used by leading federal financial regulators to evidence potential fair lending violations. However, nearly all U.S. consumer lenders with non-HMDA fair lending testing programs – for example, for auto loans, personal loans, credit cards, and small business loans – use the "BISG Classification" approach to identify specific prohibited basis customers that may be eligible for financial remediation should actionable fair lending disparities be identified, and some of these lenders also use the BISG Classification approach to estimate the regression-based fair lending disparities. We also note that the BISG Classification approach is relatively common when testing for algorithmic bias.

Based on this work, we identified the following key findings that raise several concerns about the reliability of BISG-proxy based fair lending disparity estimates and group membership counts within the context of our analysis framework.

Key Findings

Aggregate Race / Ethnicity Distribution Accuracy

The BISG proxy probabilities produce accurate and unbiased aggregate counts of race / ethnicity group membership so long as the underlying socioeconomic characteristics of the analysis sample are aligned with the corresponding characteristics of the U.S. Census data on which the BISG proxy model is based. However, to the extent that the analysis sample departs from this condition – which may occur in consumer lending when the applicant pool is skewed toward particular socioeconomic profiles such as higher income / assets or higher credit quality – then the BISG proxy probabilities can produce materially biased / inaccurate aggregate group membership counts, as well as materially biased / inaccurate fair lending disparity estimates under the BISG Continuous approach (see below for further details on the disparity estimate bias).

Individual-level Race / Ethnicity Classification Accuracy

The BISG proxy probabilities are a direct reflection of the underlying socio-demographic patterns of residential micro-geographies – refined by demographic differences in surnames – summarized by U.S. Census data. Our empirical analysis of how these patterns impact the individual-level accuracy and socioeconomic characteristics of the BISG-based proxies reveals that:

- **The BISG proxy model produces relatively undifferentiated BISG Black probability values for Actual Black sample members.**[11] Rather than a model flaw, this property is simply a reflection of the Census data upon which the BISG proxy probabilities are based – thereby limiting its predictive power at the individual level. In fact, according to the 2010 U.S. Census data underlying the BISG proxy model, Blacks in our broad-based national sample reside in micro-geographies that are, on average, only 39% Black while Whites in our overall sample reside in micro-geographies that are, on average, 76% White. Additionally, 59% of micro-geographies in which our sample Whites reside are at least 80% White, while only 18% of micro-geographies in which our sample Blacks reside are at least 80% Black. This general lack of geographic segregation for the vast majority of sample Blacks – combined with a relatively low degree of surname segregation – causes the poor Black probability differentiation for Black sample members.

- **The relatively undifferentiated BISG Black probability distribution for Black sample members significantly impairs the BISG proxy model's ability to predict accurately**

[11] Ideally, the BISG Black probability distribution for Actual Blacks would be heavily concentrated toward high BISG Black probability values – consistent with the BISG White probability distribution observed for Actual Whites.

the race of individual Black sample members. According to our estimates, the BISG proxy model inaccurately predicts the race of Black sample members 42 - 73% of the time (depending on which individual-level classification rule is employed) – a significantly larger error rate than for Whites (7 - 25%). Hispanics (22 - 52%) and APIs (34 - 58%) also exhibit elevated individual-level error rates – albeit somewhat lower than observed for Blacks. **In terms of aggregate group membership counts based on individual-level classification, Blacks experience the greatest degree of bias with undercounts ranging from -20% to -70%** depending on which individual-level classification rule is employed. This compares to biases of -14% to -54% for APIs, 0% to -48% for Hispanics, and +8% to -20% for Whites.

- **A deeper-dive into the individual-level error rates reveals another hidden bias – the Black False Negatives (i.e., Actual Blacks that are incorrectly predicted to be another race / ethnicity) and the Black False Positives (i.e., Actual Non-Blacks that are incorrectly predicted to be Black) are not random members of the sample**. Instead, we find that individual-level classification eliminates higher income Blacks residing in racially-diverse micro-geographies from the Predicted Black group, and substitutes in lower-income Whites residing in minority-skewed micro-geographies. Accordingly, the socioeconomic characteristics of Predicted Blacks are biased in an adverse manner from the socioeconomic characteristics of true Actual Blacks – displaying much lower average Census Block Group ("CBG") median incomes and being much more concentrated in higher minority micro-geographies. The extent of these biases depends on the specific individual-level classification rule that is employed. As we discuss in the next two sections, these hidden biases have important impacts on downstream fair lending disparity estimation results.

<u>Hidden Biases in Fair Lending Disparity Estimates Under the BISG Continuous Approach</u>

- **The BISG proxy probabilities produce unbiased <u>disparate treatment</u> disparity estimates under the BISG Continuous regression approach, so long as the underlying socioeconomic characteristics of the analysis sample are aligned with the corresponding characteristics of the U.S. Census data on which the BISG proxy model is based**. However, to the extent that the analysis sample departs from this condition – which may occur in consumer lending when the applicant pool is skewed toward particular socioeconomic profiles such as higher income / assets or higher credit quality – then the BISG proxy probabilities can produce materially biased / inaccurate disparate treatment disparity estimates under the BISG Continuous approach. For example, in our testing of disparate treatment scenarios, we found that a +4% skew in the percentage of Whites in our analysis sample (relative to the expected percentage of Whites based on BISG probabilities) can cause a 17% underestimation of the scenario's true Black fair lending disparity under the BISG Continuous approach.

- **Even when the socioeconomic characteristics of the analysis sample align with those of the U.S. Census data underlying the BISG proxy model, the BISG Continuous** approach will produce materially biased / inaccurate disparate impact disparity estimates when the driver of the disparate impact is correlated with the U.S. Census micro-geographies underlying the BISG proxy model**. For example, if discretionary pricing differences across customers are driven by (or correlated with) borrower income levels, then the BISG Continuous approach produces inflated disparity estimates. In our testing where discretionary pricing was correlated with CBG median income levels, we found that the BISG Continuous approach produced disparate impact price disparity estimates that were 80-100% overstated for Hispanic and Black groups (relative to Whites). The range of overstatement was even wider at the individual state-level.

- **We describe two alternative estimation approaches – the Bootstrap Regression Approach and the Proportional Regression Approach – to recover the true disparate impact disparity estimates for analysis samples whose underlying race / ethnicity distributions align with the BISG probabilities produced by the proxy model.** Both approaches are designed to eliminate the source of disparity inflation discussed above and, therefore, produce disparate impact disparity estimates consistent with the "ground truth" amounts.

Hidden Biases in Fair Lending Disparity Estimates Under the BISG Classification Approach

- **Under disparate treatment scenarios – whereby discrimination is based on actual race / ethnicity – all BISG Classification approaches cause underestimation of true minority price disparities by 9 - 34% because of how False Negatives and False Positives interact with the scenario**. For example, if Blacks are subject to disparate treatment price discrimination, then their False Negatives (i.e., Actual Blacks that are predicted to be White) will cause the average price for Predicted Whites to be higher, and the Black False Positives (i.e., Actual Non-Blacks who are predicted to be Black) will cause the average price for Predicted Blacks to be lower. Both of these effects reduce the estimated Black price disparity below the true value.

- **In our disparate treatment scenarios, the BISG Max classification rule[12] generated the largest disparity estimation bias while the BISG 80% Threshold rule[13] generated the smallest bias**. This is due to the significantly smaller influence of False Negatives / False Positives on the BISG 80% Threshold approach in which the exclusion of "Unknowns" precludes the type of average price disparity dilution discussed immediately above. Alternatively, since the

[12] This rule, which we attribute to the Zhang BISG Proxy Paper, classifies an individual to the race / ethnicity category associated with his/her largest BISG probability value. Unlike threshold-based rules, no sample members are excluded from subsequent fair lending testing since every sample member receives an assignment.

[13] The threshold rules assign an individual to a specific race / ethnicity only if that individual's associated BISG probability exceeds a minimum value (the "threshold") such as 80% or 50%. If none of the individual's BISG probabilities exceed the threshold, then the individual is classified as "Unknown" and removed from the fair lending testing sample.

BISG Max approach does not exclude any sample members, the False Negative / False Positive bias is maximized. While this suggests that the BISG 80% Threshold rule provides the least biased disparate treatment disparity estimate under the different BISG Classification approaches considered, it comes at the expense of significantly smaller addressable samples – particularly for minorities – since so many sample members are excluded as "Unknown". Accordingly, while the BISG 80% Threshold rule may produce the least bias in the estimation of the fair lending group disparity, it exhibits the worst Recall Accuracy of all classification methods (i.e., the percentage of true prohibited basis members that are accurately proxied) – thereby limiting its utility in remediating the actual consumers impacted by such disparities.

- **Under our <u>disparate impact</u> scenario, the BISG Classification approaches produced minority price disparity estimates that were biased between -10% and +80% because of how the False Negatives and False Positives interact with the scenario**. For example, if Blacks are subject to disparate impact price discrimination correlated with CBG median income levels, then we find that their False Negatives (i.e., Actual Blacks that are predicted to be White) are concentrated in high income-predominantly White geographies (with corresponding low prices) and their False Positives (i.e., Actual Non-Blacks who are predicted to be Black) are concentrated in lower-income predominantly Black geographies (with corresponding high prices). Therefore, by removing high income-low price Blacks and adding low income-high price Non-Blacks to the Predicted Black group, we obtain an overestimate of the true underlying Black price disparity.

<u>Using BISG Proxies for Fair Lending Testing Below the National-Level</u>

- **Users need to exercise additional caution when analyzing samples using BISG proxies within more localized geographies – such as state-level, MSA-level, or county-level**. Our analysis indicates significant variability in proxy accuracy and disparity biases at these levels – particularly for micro-geographies with small populations of the minority groups on which the testing is focused.

Overall, our work has identified important features of the BISG proxy model that, when combined with common fair lending testing approaches, create heightened risk of disparity estimation bias and material miscounts of group membership. While we have offered both intuitive and analytical explanations for the biases associated with our specific discrimination scenarios, the fair lending community would benefit from additional work to: (1) explore these potential biases under a fuller range of disparate treatment and disparate impact scenarios, (2) determine whether there are other relevant conditions under which the biases we have estimated would be materially different in size (or of a different direction), and (3) investigate additional mitigation steps to address these biases. Notwithstanding this additional work, these results strongly suggest that U.S. consumer lenders and financial regulators should jointly reassess current applications of the BISG proxy model to fair lending supervision, enforcement, and compliance risk management activities to ensure that the hidden biases identified by this study are not causing incorrect conclusions and empirically unsupported decisions in these high-stakes areas.

Overview of the Study

The next section introduces the synthetic dataset we use for our analyses – describing its underlying theory, the method of its construction, certain demographic and geographic distributional properties, and comparing these properties to relevant empirical benchmarks. Next, we focus on the properties of the BISG proxy model that impact its aggregate-level and individual-level accuracies in predicting the race / ethnicity of sample members. In particular, we explore how the BISG proxy probabilities inherit key features of the U.S. Census data on which they are based, and how such features impact the model's inherent predictive power for individual-level race / ethnicity classification. As part of this assessment, we introduce the concepts of Recall Accuracy, Precision Accuracy, False Positives, and False Negatives and compare individual-level accuracy – and certain socioeconomic biases – across three alternative classification rules: BISG Max, BISG 80% Threshold, and BISG 50% Threshold.

With this foundation, the final section links the BISG probabilities to the two corresponding estimation approaches used for fair lending price discrimination testing (i.e., BISG Continuous and BISG Classification). Here we create a series of disparate treatment and disparate impact pricing scenarios in which the pricing disparity amounts are known. We then use both the BISG Continuous regression approach, as well as the BISG Classification regression approach, to estimate these disparities as would typically be done within a regulatory examination, enforcement proceeding, or as part of an institution's fair lending compliance risk management process. We compare the disparity estimates under these estimation approaches to the known "ground truth" disparity amounts to identify potential estimation biases, provide insights to the resultant findings, and – where possible – provide recommendations for bias mitigation.

Caveats and Future Efforts

Let us be the first to acknowledge that the synthetic sample used in these analyses may not be exactly representative of any lender's non-HMDA customer base to which the BISG proxy model would be applied. Nevertheless, the synthetic sample does possess statistical properties – such as aggregate race / ethnicity distributions, correlations between "actual" and proxy races / ethnicities, as well as other accuracy measures – that are generally aligned with those measured in previous studies based on real-world data samples. This provides a degree of confidence in the broad appropriateness of the synthetic sample for the goals set out above. However, the bias estimates generated from this sample should be considered "baseline" estimates associated with a broad geo-surname sample of the U.S. adult population. While the insights derived from these analyses should, at a minimum, generalize in direction to more specific "real world" samples of most U.S. lenders, ultimately we see this work as a first step in: (1) surfacing some specific risks and limitations of the BISG proxy model as used in fair lending testing, and (2) providing an empirical framework in which to assess these risks and limitations. We encourage users of this model to build off of this "baseline" work to assess such risks on their own customer samples and to advance a broader validation assessment of this important tool.

Creating A Synthetic U.S. Adult Sample Using the BISG Proxy Model

One of the key reasons for the BISG proxy model's lack of model validation testing for fair lending analysis purposes is: (1) the lack of broad-based datasets of U.S. consumer loan transactions containing the actual race / ethnicity of each individual, and (2) a set of corresponding lending outcomes in which the specific type and magnitude of discrimination is known (e.g., 25 bps disparate treatment price discrimination of Hispanics). While HMDA-reportable residential mortgage loan applications may appear to satisfy the first need, we would argue that such transactions are actually inappropriate for this model validation testing as they are not the type of transactions to which the BISG proxy model would actually be applied (as they already possess actual race / ethnicity data). Even if we could obtain a non-HMDA dataset for testing, we are also challenged by not knowing the "ground truth" discrimination activity within the transaction sample – which confounds our ability to measure and attribute potential biases in the estimated lending outcome disparities.

For these reasons, we have taken a different path for our analysis. Specifically, we create a large-scale, synthetic dataset from the BISG proxy model by creating pairs of surnames and geographies that are consistent with their respective U.S. Census data distributions. For each synthetic individual, the BISG proxy model provides us with a corresponding set of BISG proxy probabilities. We then apply standard Monte Carlo techniques to simulate the actual individuals from these probabilities – thereby creating our analysis dataset. As a "general" synthetic sample of U.S. adults, this dataset could be considered more broadly representative of the BISG proxy model's input space and, therefore, more representative of the model's key properties of which we have interest. Secondly, with a synthetic dataset, we can design specific discrimination scenarios in which the ground truth discrimination activity is known. This allows us to precisely estimate the potential disparity estimation biases that may be generated using the BISG proxies, as well as analyze the fundamental drivers of such biases. In the rest of this section, we describe in more detail the concept of our synthetic dataset, how it is specifically created, key properties of the resulting sample, and comparisons of these key properties to publicly available benchmarks from real world consumer lending datasets.

Sample Design and Generation Using Monte Carlo Techniques

As brief background to the BISG proxy model, consider the following synthetic adult created by the pairing of an individual with the last name Beech and a Census Block Group ("CBG") near Worcester, MA. According to the BISG proxy model illustrated in **Figure 1** below, U.S. adults with this geo-surname profile are 95.5% likely to be White, 2.5% likely to be Black, 0.5% likely to be API, 0.5% likely to be Hispanic, and 1% likely to be an Other race / ethnicity.[14] What this means, practically, is that if 1,000 individuals were randomly surveyed from this geo-surname segment, we would find that 955 would self-identify themselves as White, 25 would self-identify themselves as Black, 5 would self-identify as API, 5 would self-identify as Hispanic, and 10 would self-identify as Other.

Surname	Block Group ID
Beech	250277309013
Race / Ethnicity	**BISG Probability**
White	95.5%
Black	2.5%
API	0.5%
Hispanic	0.5%
Other	1.0%

Figure 1: BISG Probabilities of Synthetic Individual 1

Now suppose that we have a large sample of individuals and their corresponding BISG probability distributions. However, we do not know each individual's actual race / ethnicity.

How can we simulate an actual race / ethnicity for each individual in a manner that is consistent with each individual's set of BISG probabilities?

That is, how can we create a synthetic dataset where the "actual" race / ethnicity of each sample member – as well as the overall "actual" race / ethnicity distribution of the aggregate sample – is consistent with the BISG probabilities that characterize the sample. One simplistic approach would simply assume that each individual's actual race / ethnicity is equal to the category with the highest BISG probability. For example, since the geo-surname distribution in **Figure 1** is heavily skewed to White (i.e., a 95.5% probability), we simply set this individual's actual race / ethnicity to White. However, this approach can be complicated for individuals whose BISG probability distributions are

[14] Throughout this document, "Black" refers to individuals classified as "Black or African American," "API" refers to individuals classified as "Asian" or "Native Hawaiian or Other Pacific Islander," and "Hispanic" refers to individuals classified as "Hispanic or Latino" as defined by the OMB in Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity (October 30, 1997). "Other" represents a combination of "2 or More Races" and "American Indian / Alaskan Native" – both of which are usually excluded from BISG proxy-based fair lending analyses due to the low predictive power of the proxy model for these groups. See CFPB BISG Proxy White Paper, p. 16. Finally, all Non-Hispanic groups (i.e., White, Black, API, and Other) are referenced without the Non-Hispanic ("NH") prefix.

not so skewed.

For example, **Figure 2** presents another synthetic individual with a last name of Brancheau and an address in a Tampa, FL CBG. According to the BISG proxy model, U.S. adults with this geo-surname profile are 68.9% likely to be White, 0.2% likely to be Black, 0.5% likely to be API, 30.2% likely to be Hispanic, and 0.2% likely to be an Other race / ethnicity. While White is the largest BISG probability for this geo-surname segment, there is also a fairly large probability that an individual from this segment is Hispanic (30.2%).

Surname	Block Group ID
Brancheau	120570118022
Race / Ethnicity	BISG Probability
White	68.9%
Black	0.2%
API	0.5%
Hispanic	30.2%
Other	0.2%

Figure 2: BISG Probabilities of Synthetic Individual 2

From these two individuals, we see that if we simply assign each individual to the race / ethnicity with the highest BISG probability, then we are not being consistent with the uncertainty captured by the BISG probabilities (e.g., in **Figure 2** there is a 30.2% chance that Individual 2 is Hispanic and in **Figure 1** there is a 4.5% chance that Individual 1 is Non-White) and we will obtain a sample whose aggregate actual race / ethnicity distribution will not align with the sample's aggregate expected race / ethnicity distribution based on the underlying sets of BISG probabilities.

Instead, we need to employ an assignment technique that respects the uncertainty of each individual's set of BISG probabilities – which means potentially assigning them to a race / ethnicity that is not the highest probability but is still consistent with their BISG probability profile. This probabilistic-based approach is known as a Monte Carlo assignment process and works as follows:

- First, the 5 BISG probabilities from each individual's geo-surname distribution are sequentially aligned along the unit line (i.e., between 0 and 1). Using the example in **Figure 1** above, the distance between 0 and 0.955 would correspond to White, the distance 0.956 to 0.980 would correspond to Black, the distance 0.981 to 0.985 would correspond to API, the distance 0.986 to 0.99 would correspond to Hispanic, and the distance between 0.991 to 1.0 would correspond to Other.

- A uniform random number between 0 and 1 is then generated and its value is used to assign the actual race / ethnicity of the individual based on what region of the unit line the value falls into. For example, if the random number for Individual 1 is 0.48 (as shown in expanded **Figure 1** below), Individual 1 is assigned to be White. Alternatively, if the random number is 0.96, the individual is assigned to be Black. Similarly, if the random number for Individual 2 is 0.75 (as shown in expanded **Figure 2** below), Individual 2 is assigned to be Hispanic. Using this probabilistic-based assignment technique, we are able to respect the inherent uncertainty in each individual's actual race / ethnicity and, therefore, create an overall sample of "actual" individuals whose demographics are consistent – at both the individual and aggregate levels – with the demographics expected based on the sample's underlying BISG probability distributions.

Surname	Block Group ID
Beech	250277309013
Race / Ethnicity	BISG Probability
White	95.5%
Black	2.5%
API	0.5%
Hispanic	0.5%
Other	1.0%

\+ **0.48 Random Number** = **Simulated Race / Ethnicity** White

Figure 1: Probabilistic Assignment of Synthetic Individual 1's Race / Ethnicity

Surname	Block Group ID
Brancheau	120570118022
Race / Ethnicity	BISG Probability
White	68.9%
Black	0.2%
API	0.5%
Hispanic	30.2%
Other	0.2%

\+ **0.75 Random Number** = **Simulated Race / Ethnicity** Hispanic

Figure 2: Probabilistic Assignment of Synthetic Individual 2's Race / Ethnicity

Using this probabilistic-based method, we still observe a large percentage of the sample being assigned to race / ethnicity categories associated with their largest BISG probability. However, we also observe some individuals being assigned, albeit less frequently, to different race / ethnicity categories than their highest BISG probability – a feature of the synthetic dataset that is both expected and desired.

This is the approach we use to create a 10-million-member synthetic sample of U.S. adults for our subsequent analyses.[15] More technically, we randomly pair 10 million surnames with 10 million census block groups according to population frequencies provided in each Census data file.[16] This ensures

[15] We selected a sample size of 10 million in order to: (1) explore a broad range of the BISG proxy model's input space, (2) generate a sample sufficiently large to represent a broad-based U.S. adult sample, and (3) to minimize the influence of sampling error on our results.

[16] A limitation of this approach is the assumed independence of the surname and census block group combinations. However, we note that: (1) such independence is actually consistent with the underlying assumptions of the Naïve

that more common surnames, and more populated census block groups, are sampled at rates consistent with their relative population shares. For each of these 10 million sample members, the BISG proxy model as described in the CFPB's BISG Proxy White Paper is used to derive the member's set of five BISG race / ethnicity probabilities. Given these probabilities, the Monte Carlo technique described above is used to assign an "actual" race / ethnicity to each sample member. At the end of this process, we have a 10-million-member dataset containing individual BISG probabilities and "actual" races / ethnicities that are collectively representative of the U.S. adult geo-surname population as modeled by the BISG proxy model.

In the next two sub-sections, we explore some basic properties of our synthetic sample in order to provide comfort that it meets the overall analytical objectives we set for it, and to derive some initial high-level insights to the BISG proxy model's demographic estimates.

Bayes modeling approach that underlies the BISG proxy model, (2) descriptive statistics of the synthetic sample presented in subsequent sections generally align closely with similar descriptive statistics reported in other studies based on actual datasets, and (3) any imprecisions that may be introduced into the resulting BISG probabilities due to potential violations of the independence assumption in real data are unlikely to alter the relative ordering of the five individual probabilities – thereby mitigating the impact of such errors.

Demographic and Geographic Properties of the Synthetic Sample

Figure 3 below shows the overall distribution of the actual races / ethnicities for the synthetic sample – with approximately two-thirds of the sample being White, 16% being Hispanic, nearly 11% being Black, about 4% being API, and 3.2% as Other.[17]

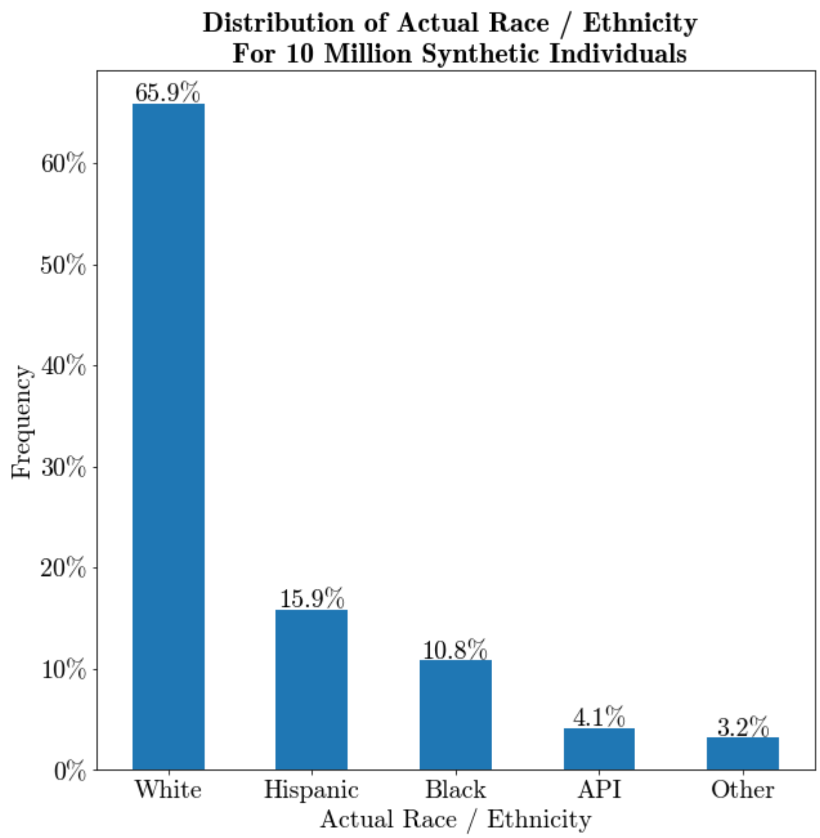
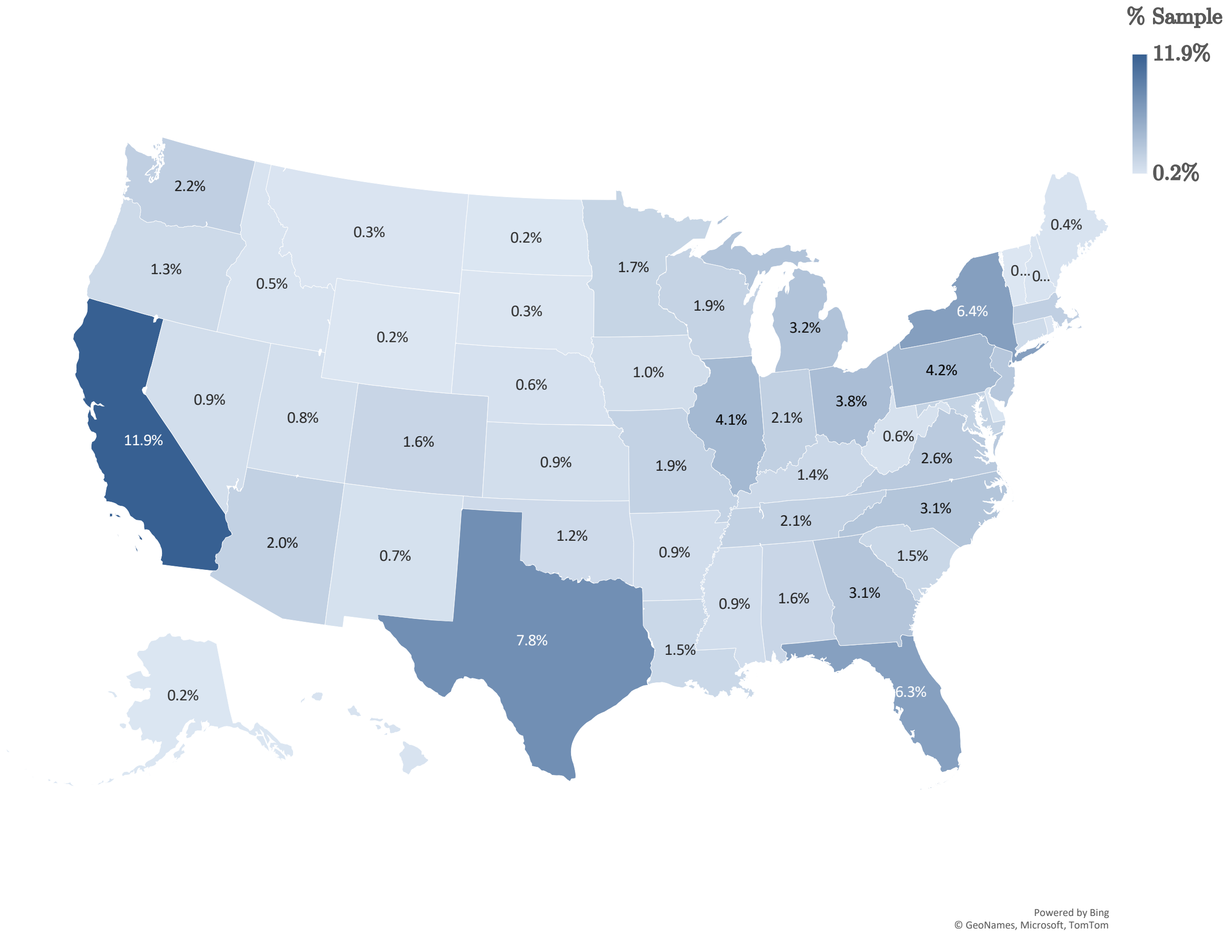


Figure 3

Figure 4 below shows the geographic distribution of the 10 million synthetic U.S. adults – with expected concentrations in high population states such as California, Texas, New York, and Florida, and low populations in the West / Mountain states.

[17] This distribution is consistent with findings from the CFPB BISG Proxy White Paper, p. 20. Specifically, "According to the 2010 Census of Population, 14% of the U.S. adult population was Hispanic; 67% non-Hispanic White; 12% non-Hispanic Black; 5% Asian/Pacific Islander; and 1% American Indian/Alaska Native."

Figure 4: State-Level Distribution: 10M U.S. Adult Synthetic Sample



Figures 5 and 6 show each state's Hispanic and Black distributions (i.e., the percentage of each state's sample that is actually Hispanic and Black) based on the 1.59 million Hispanics, the 1.08 million Blacks, and the 7.33 million other races / ethnicities in the sample. As expected, the West, Southwest, and Florida have the greatest concentrations of Hispanic adults, while the Mid-Atlantic, South, and Southeast have the greatest concentrations of Black adults.

Figure 5: State-Level Hispanic Distributions: 10M U.S. Adult Synthetic Sample

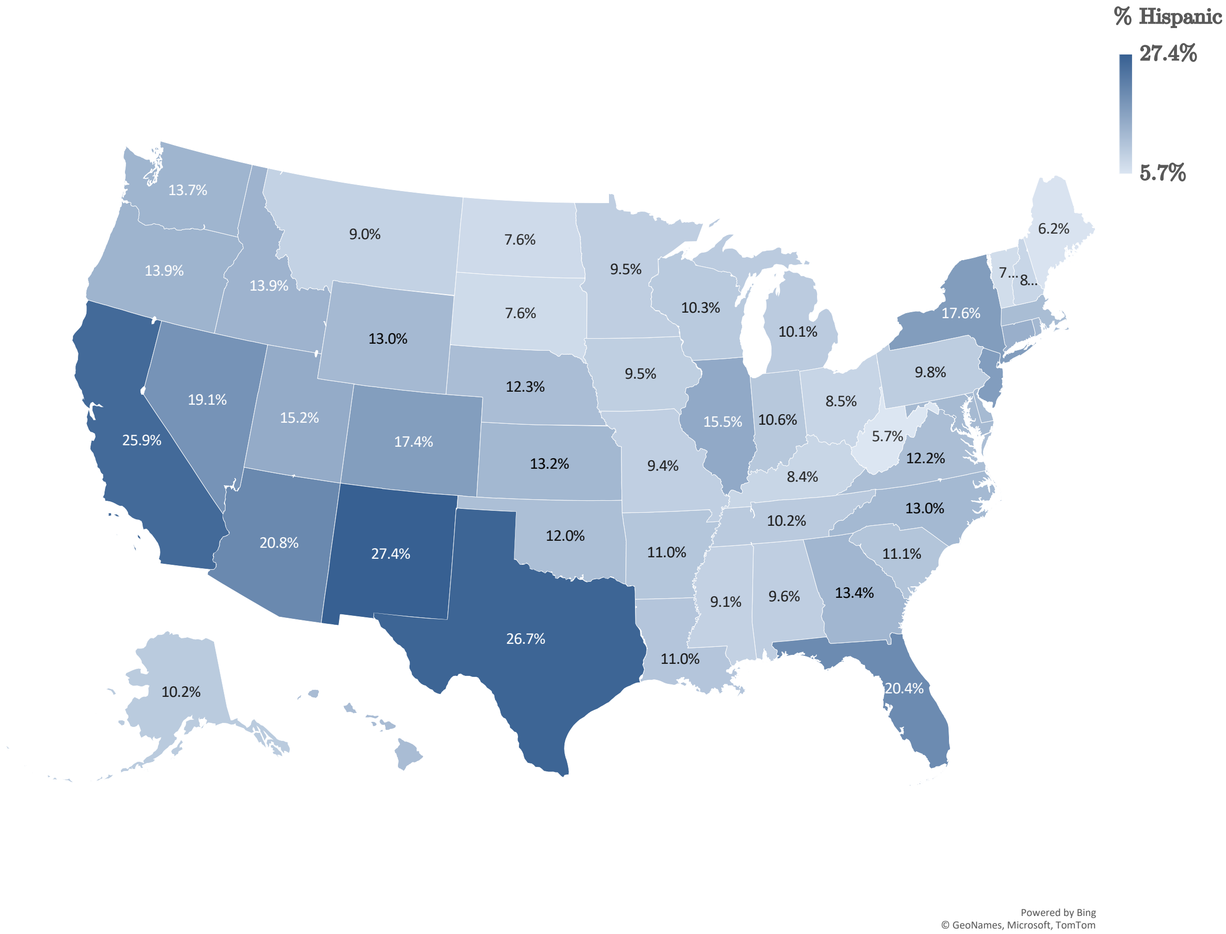
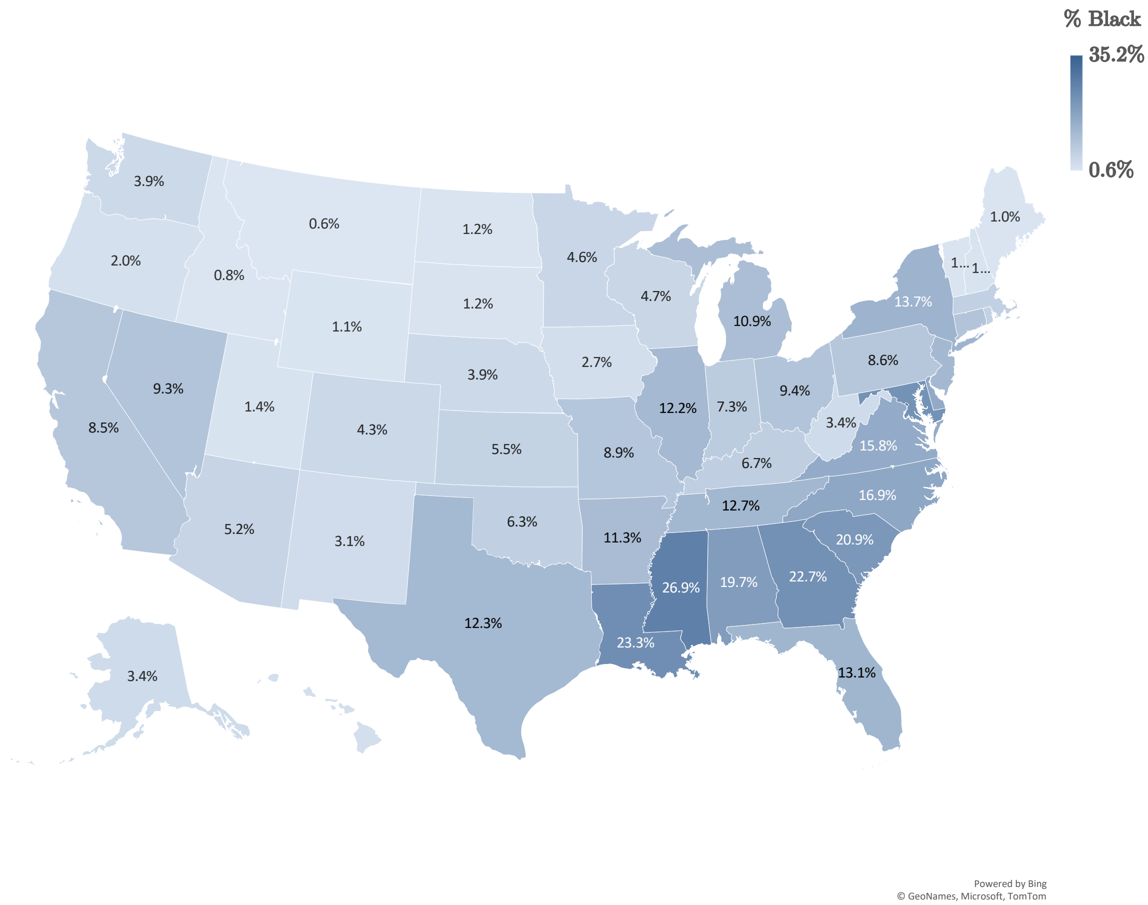


Figure 6: State-Level Black Distributions: 10M U.S. Adult Synthetic Sample



Appendices A and B provide further details on the state-level distributions of the 10-million-member synthetic sample of U.S. adults. Based on these state-level breakdowns (as well as **Figures 5 and 6** above), the following observations are noted:

- There are nine U.S. states where the sample contains less than 1,000 Black individuals: AK, ID, ME, MT, ND, NH, SD, VT, and WY

- There are seven U.S. states where the sample contains less than 1,000 API individuals: AK, DC, MT, ND, SD, VT, and WY

These low sample counts are notable as they are based on an overall national sample size of 10 million – far larger than typical samples used by U.S. consumer lenders. Additionally, these low sample sizes are not a unique property of this study; rather, they reflect both the relative sizes of each state's overall population as well as the specific race / ethnicity distribution within the state as reflected by the Census data. As will be shown in subsequent sections, such low levels of minority "densities" in these states will challenge the accuracy of their BISG proxy classifications and impact relative sizes of certain fair lending disparity estimation biases.

Sample Properties of BISG Proxy Probabilities

Figure 7 below displays the distributions of the BISG proxy probabilities for each "actual" race / ethnicity group in the synthetic sample.

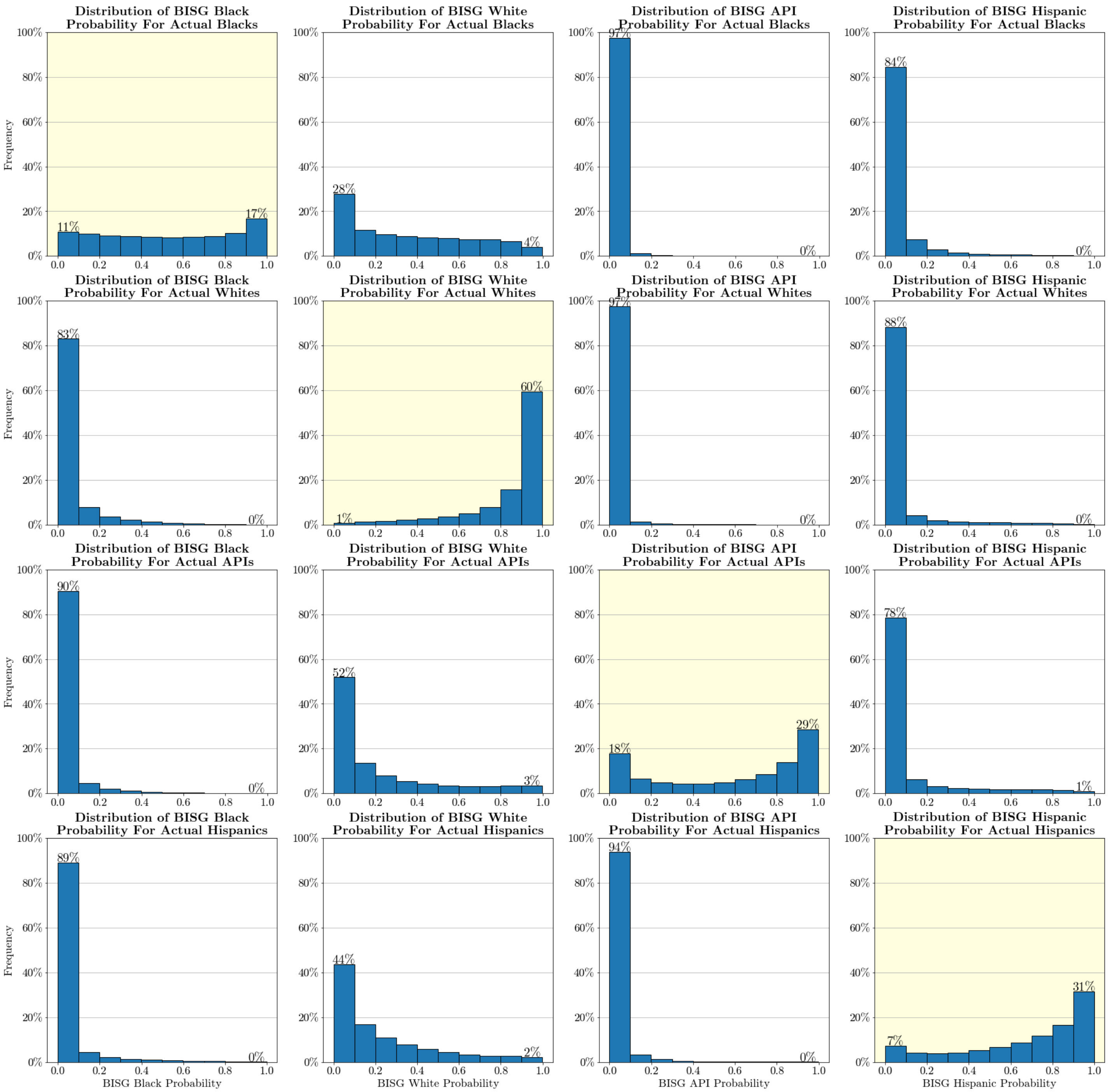


Figure 7: Sample BISG Probability Distributions: Actual Race/Ethnicity (Rows) vs. Estimated Race/Ethnicity (Columns)

Each row corresponds to the BISG proxy probabilities associated with a specific "actual" group (i.e., Actual Blacks (first row), Actual Whites (second row), Actual APIs (third row), and Actual Hispanics (fourth row)) while each column corresponds to a different BISG proxy probability (i.e., BISG Black probabilities (first column), BISG White probabilities (second column), BISG API probabilities (third column), and BISG Hispanic probabilities (fourth column)). The yellow-shaded charts on the diagonal correspond to the matching BISG proxy probabilities for each row's actual group (i.e., Black-Black, White-White, API-API, and Hispanic-Hispanic) while the off-diagonal charts correspond to the non-matching BISG proxy probabilities for each actual group. Ideally, for best predictive power, the matching (diagonal) probabilities should show heavy concentration to the right (i.e., high probability values) while the non-matching (off-diagonal) probabilities should show high concentration to the left (i.e., low probability values).

For example, the chart in the upper left corresponds to the distribution of BISG Black probabilities for Actual Black individuals in the sample (a matching/diagonal chart). Unfortunately, as can be clearly seen here, the distribution of BISG Black probabilities for Actual Blacks is fairly uniform across the unit line with only a small concentration at probability values of 90% (0.9) or greater. This indicates that the BISG proxy model is challenged to differentiate Actual Blacks via the BISG Black probabilities – unlike the matching/diagonal chart for Actual Whites in the second row that shows a heavy concentration (i.e., about 60%) in probability values of 90% or greater.

Moving across the first row, we can see how the model generates non-matching proxy probabilities for Actual Blacks. In particular, we see that the model rarely assigns high API (third column) or Hispanic (fourth column) probabilities to Actual Blacks – but does tend to assign elevated White probabilities (second column) thereby contributing – as discussed later in this document – to Black misclassification errors (i.e., Actual Blacks that are falsely classified as Whites).

Overall, **Figure 7** reveals the following observations:

- Actual Blacks tend to have the most undifferentiated (i.e., flattest) matching BISG probability distribution, Actual Whites the most differentiated matching BISG probability distribution – with Actual APIs and Hispanics somewhere in between. In general, the flatter a group's matching BISG probability distribution, the more difficult it is to predict accurately membership in that race / ethnicity group at the individual level – which we will see in more detail in a later section.

- All three non-White groups tend to have elevated BISG White probability values with Actual Blacks and Hispanics displaying the largest "skew" of BISG White probabilities above 10%. This contributes to the presence of False Positive Whites (and False Negative Blacks and Hispanics) when analyzing the BISG model's individual-level predictive accuracy – a risk that we will also further detail in a later section.

In the next section, we evaluate the BISG proxy model's ability to predict accurately the aggregate race / ethnicity distribution of a sample – as well as the specific race / ethnicity of individual sample members.

BISG Proxy Model Accuracy

BISG Proxy Probabilities: Aggregate-Level Accuracy

Due to the law of large numbers, we would expect the aggregate actual race / ethnicity distribution of our synthetic U.S. adult sample to align closely with the aggregate expected race / ethnicity distribution from the BISG proxy model.[18] **Figure 8** below presents this comparison where the expected race / ethnicity distribution is obtained by summing up the individual BISG race / ethnicity probabilities across the 10 million synthetic U.S. adults and expressing these sums as a percentage of the total sample. For example, summing up the BISG Black probabilities for all 10 million individuals yields a total of 1,085,553 estimated Blacks. Dividing this sum by the 10 million individuals in the total sample yields an expected Black BISG distribution percentage of 10.856%.

As the table below summarizes, the BISG proxy model yields highly accurate aggregate race / ethnicity distributions relative to actuals with estimation variances of <= 0.01% in absolute value, which is consistent with our expectations due to the law of large numbers.

Figure 8: Actual vs. Expected Race / Ethnicity Distribution For 10 Million Synthetic Individuals

	White	Black	API	Hispanic	Other
Actuals	65.920%	10.849%	4.147%	15.899%	3.185%
BISG Proxy	65.930%	10.856%	4.148%	15.889%	3.179%
Difference	0.009%	0.007%	0.001%	-0.010%	-0.007%

What this result means is two-fold:

- The BISG proxy model is not inherently biased or error-prone in estimating aggregate race / ethnicity distributions for samples that are consistent with the model's underlying census data properties. That is, previously reported measures of aggregate proxy error in the CFPB BISG Proxy Paper, in the AFSA White Paper, and in the Zhang BISG Proxy Paper are really due to the authors' application of the BISG proxy model to data samples that are known to differ from the Census data properties underlying the model. More specifically, since these studies rely on samples of residential mortgage loans, the corresponding applicants are likely from wealthier and/or higher income sub-segments of the BISG geo-surname segments to which the applicants are assigned by the BISG proxy model – thereby introducing product-specific socioeconomic bias into

[18] For real world samples, we would expect a material alignment of actual and expected race / ethnicity distributions so long as the samples are not biased within geographies (such as from higher income or wealthier segments of the geographic population) and so long as the Census demographics used in the BISG proxy model are not materially outdated.

the accuracy testing results.[19]

- Aggregate-level BISG proxy biases identified using skewed data samples do not generalize to other more different samples. That is, the fact that the BISG proxy model may underestimate the share of Whites and overestimate the share of Blacks / Hispanics in a given residential mortgage sample does not mean that the model will similarly do so for other consumer lending products with different socioeconomic properties. For example, if the socioeconomic properties of typical credit card applicants are much more similar to the overall U.S. adult population, then the BISG proxy model may produce aggregate race / ethnicity distributions for such samples that are more accurate than those observed for residential mortgage.

This finding is consistent with the well-known model risk management principle that a model should be applied to data samples that are materially aligned with the key properties of the original dataset used to estimate or train the model. To the extent that there are material misalignments in the "production" and "development" data, then elevated model risk exists that may cause the model to produce biased / inaccurate results to model users – thereby requiring appropriate model risk controls and mitigants. In the case of the BISG proxy model's application to consumer lending data, we are unaware of any publicly available studies that demonstrate that the model is sufficiently predictive of aggregate race / ethnicity distributions on the types of consumer lending products to which the model would be applied.[20]

We note that the aggregate inherent accuracy results in **Figure 8** are based on a sample size of 10 million – which is much larger than the typical sample sizes analyzed by most U.S. lenders. Accordingly, **Figure 9** presents the likely observable range of inherent aggregate proxy error distributional differences for each race / ethnicity group based on simulations of alternative sample sizes.[21]

[19] In all three studies, the aggregate representation of Whites in the samples are found to be under-estimated by the BISG proxy model – a result that is generally consistent with a sample skewed towards wealthier / higher income applicants.

[20] "Sufficiently predictive" means that the expected bias / inaccuracy levels are not of a magnitude that would alter the essential conclusions of an analysis.

[21] For each sample size group, we randomly-generate 1,000 samples of that size from the 10 million synthetic U.S. adults. For each of these 1,000 samples, we calculate the aggregate distributional difference as shown in **Figure 8**, and then calculate the mean and standard deviations of these differences across the 1,000 samples. For example, for the 25K sample size group, we randomly generate 1,000 samples of size 25K, calculate the actual and estimated aggregate race / ethnicity distributions for each random sample, calculate the proxy error distributional difference for each random sample, and then calculate the mean and 95% confidence interval of these 1,000 results.

Figure 9: 95% Confidence Intervals of Aggregate Proxy Error Distributional Differences for Alternative Sample Sizes

Sample Size	White	Black	API	Hispanic	Other
100K	+/- 0.2%	+/- 0.14%	+/- 0.08%	+/- 0.14%	+/- 0.1%
50K	+/- 0.27%	+/- 0.2%	+/- 0.11%	+/- 0.19%	+/- 0.14%
25K	+/- 0.41%	+/- 0.28%	+/- 0.16%	+/- 0.29%	+/- 0.2%
10K	+/- 0.67%	+/- 0.44%	+/- 0.27%	+/- 0.45%	+/- 0.33%

This analysis indicates that even at relatively small sample sizes of 10,000 individuals, one is unlikely to observe aggregate proxy error distributional differences of more than +/- 0.67% for any individual race / ethnic group. Thus, the BISG proxy model should produce aggregate race / ethnicity distributions that are materially accurate to true race / ethnicity distributions for the loan application volumes of medium-to-large consumer lenders – again, conditional on the absence of bias in the data sample relative to the applicable U.S. adult census population.[22]

Finally, previous studies also assessed the BISG proxy model's aggregate accuracy through an evaluation of correlation coefficients between the known race / ethnicity of sample members and their corresponding BISG probability values. A correlation coefficient of exactly 1.0 indicates that the BISG proxy probabilities are perfectly aligned with the corresponding actual race / ethnicity of sample members, while a correlation coefficient of 0.0 indicates no alignment. In general, the closer a calculated correlation coefficient is to 1.0, the stronger the predictive power of the associated BISG probability in predicting the actual race / ethnicity of the sample members. **Figure 10** below summarizes the correlation coefficients from these studies and compares them to the correlation coefficients computed from this study's 10-million-member synthetic sample.

Figure 10: BISG Proxy Correlation Coefficient Comparisons

Study	Data Sample	API	Black	Hispanic	White
Present Study	Synthetic Sample	0.75	0.69	0.79	0.74
CFPB (2014) BISG Proxy Paper	Mortgage Sample	0.83	0.70	0.81	0.77
Elliott (2008) BISG Proxy Paper	Health Care Sample	0.77	0.70	0.82	0.76
Zhang (2018) BISG Proxy Paper	Mortgage Sample	0.73	0.74	0.83	0.76

Overall, the correlations range from 0.69 to 0.83 – indicating relatively good alignment of the BISG probabilities with the actual race / ethnicity of sample members. Hispanics tend to have the highest correlations across these studies, followed by Whites and APIs, with Blacks having the lowest

[22] Currently, the BISG proxy model is based on the 2010 Census surname and geo-demographic data. Accordingly, there may also be small biases in the BISG probabilities when applied to current data samples – even if such data samples were otherwise unbiased – due to potential geographic population "drifts" over the past decade. Release of revised geo-demographic data based on the 2020 Census should mitigate such biases in the future. In the meantime, it is possible that the use of a more aggregated geographic unit of analysis – such as census tracts rather than census block groups – may mitigate somewhat such biases (due to relatively less "drift" within the larger area).

correlations across the groups. Importantly, the correlations produced by this study's synthetic dataset – like the demographic and geographic benchmarks discussed previously – are very much in-line with those of the other studies – providing a level of comfort with the representativeness of the synthetic geo-surname sample and its utility for the questions explored further herein.

BISG Proxy Probabilities: Individual-Level Accuracy

A critical, but frequently misunderstood, property of the BISG proxy model is that its precision in the aggregate does not translate to its precision for an individual. That is, even though we showed in **Figures 8, 9, and 10** how closely aligned the aggregate actual and expected race / ethnicity distributions are under the BISG proxy model, this accuracy breaks down at the individual level due to the need to convert each individual's set of BISG probabilities into a single race / ethnicity prediction. For example, in **Figure 1** Individual 1 is estimated to be 95.5% White, 2.5% African American, and 2% some other race / ethnicity. While this multi-probability profile creates no difficulties in deriving the aggregate race / ethnicity distribution of the entire 10 million U.S. adult sample, it does create a difficulty in deriving a specific race / ethnicity for this individual.

A logical and simple way to address this difficulty is to simply choose the race / ethnicity associated with the highest probability value in the set (the so-called "BISG Max" classification rule).[23] In the case of **Figure 1**, this would result in a prediction of White for Individual 1's race which happens to be correct. However, in instances such as **Figure 2** where more than 1 race / ethnicity has a material probability value, our predicted race / ethnicity may be in error. Here, for example, although White has the highest probability at 68.9%, Individual 2 is actually Hispanic (which is not an unlikely observation since its probability is 30.2%).

Within a given sample of individuals, we assess the individual-level accuracy of such a classification rule using the following framework that compares the actual individual race / ethnicities within the sample to those predicted by the classification rule applied to the BISG probabilities of the sample members. To simplify our description of this framework, we assume that there are only two races / ethnicities – Hispanic and White – and we are assessing the individual predictive accuracy for Hispanics.

[23] This is by far the most common classification decision rule used in the machine learning field and is adopted here as a "baseline" classification rule for this reason – as well as the fact that it produces a predicted race / ethnicity for every sample member. It appears to have been first explored in the fair lending context by OCC economist Yan Zhang: See Zhang, Yan, "Assessing Fair Lending Risks Using Race/Ethnicity Proxies," Management Science 64 (1), January 2018, pp. 178-197. However, as we will see in later sections of this study, alternative threshold-based race / ethnicity classification rules are typically used for fair lending compliance risk management purposes and, accordingly, we will explore and compare the accuracy of those rules there.

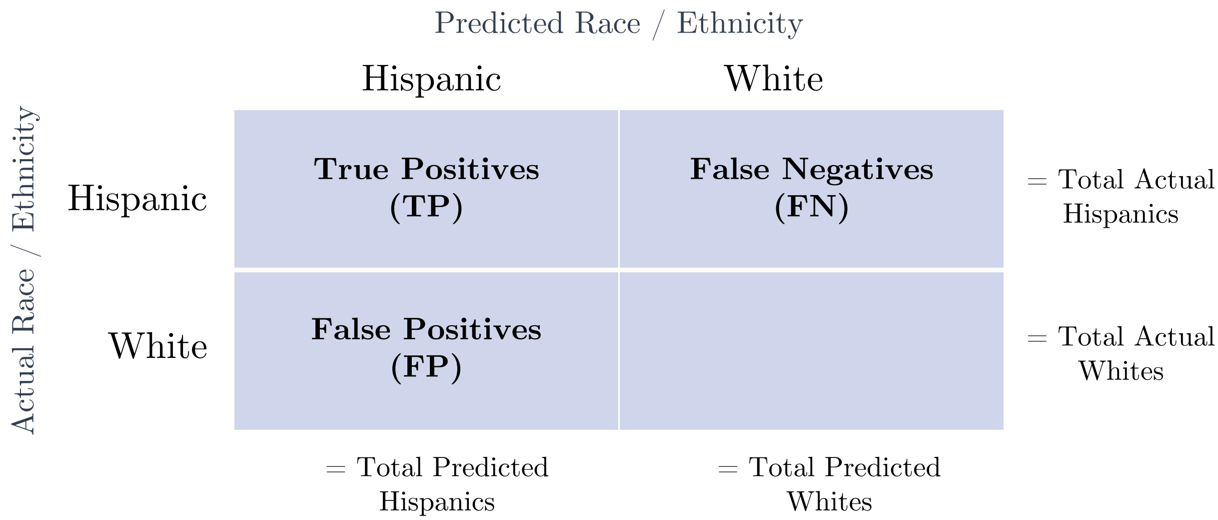


Figure 11: Evaluation of Actual vs. Predicted Individual Race / Ethnicity

Figure 11 is oriented as follows. The actual number of Hispanics in our sample is contained in the top row and the actual number of Whites in our sample is contained in the bottom row. The predicted number of Hispanics in our sample (according to the BISG Max classification rule) is contained in the left column and the predicted number of Whites in our sample (using the same rule) is contained in the right column.

Looking at the top row, we can see that the Total Actual Hispanics can be divided between True Positives (TP) – that is, those Actual Hispanics that the classification rule accurately predicted, and False Negatives (FN) – that is, those Actual Hispanics that were incorrectly predicted to be Whites by the classification rule (since their BISG White probabilities were the largest of the set). However, this is just one measure of predictive accuracy (i.e., how many of our Actual Hispanics were correctly predicted). Looking alternatively at the left column, we see that Total Predicted Hispanics can also be divided into those that are accurate (i.e., correspond to Actual Hispanics – True Positives (TP)) and those that are inaccurate (i.e., they correspond to Actual Whites – False Positives (FP)). This represents a different, but equally important, measure of predictive accuracy (i.e., how many of our Predicted Hispanics were correct).

Therefore, when we evaluate the overall predictive accuracy of an individual-level classification rule, we need to perform the evaluation along two dimensions – one that assesses accuracy relative to total actuals (i.e., TP / (TP + FN)), and one that assesses the accuracy of total predictions (i.e., TP / (TP + FP)). We refer to these two accuracy measures as **Recall Accuracy** and **Precision Accuracy**, respectively. **Recall Accuracy** calculates the percentage of individuals with a given actual race / ethnicity that are correctly predicted by the classification rule. For example, if there are 100,000 Hispanic individuals in the sample and the classification rule correctly predicts a Hispanic ethnicity for 75,000 of these individuals based on their underlying BISG probabilities, then the Recall Accuracy rate would be 75%. This means that the classification rule correctly captures 75% of the sample's

Actual Hispanics. The remaining 25% of Actual Hispanics that are not correctly predicted by the classification rule represents "False Negative" predictions – that is, they are falsely predicted (mis-classified) to be some other race / ethnicity.

Precision Accuracy calculates the percentage of total predictions that are correct. For example, if the classification rule predicted that 100,000 individuals are Hispanic, and 80,000 of these predictions are accurate (TPs), then the Precision Accuracy rate would be 80%. This accuracy measure provides insight into the classification rule's "False Positive" predictions – that is, those individuals predicted to be a specific race / ethnicity but who are actually not. In our example, 20% of Hispanic predictions are False Positives as they do not correspond to Actual Hispanics.

From a fair lending perspective, both accuracy measures – Recall and Precision – are important. For example, suppose we had 95% Precision Accuracy for Hispanics. On the surface, this seems to be a great outcome as 95% of our Hispanic predictions are, in fact, accurate. However, if our Recall Accuracy was only 50% – meaning that half of Actual Hispanics were being misclassified to some other race / ethnicity, then such a result is much less impressive. Alternatively, we may have high Recall Accuracy (say, 90%) but low Precision Accuracy (say 45%) because our classification rule predicts many Non-Hispanic individuals to be Hispanic.

As both Recall and Precision Accuracy measures are considered equally important in the fair lending context, we can combine both accuracy measures into what is called **"F1 Accuracy"** which is a single blended "average".[24] In our examples above, a 95% Precision Accuracy and a 50% Recall Accuracy would yield a 66% F1 Accuracy, and a 90% Recall Accuracy and a 45% Precision Accuracy would yield a 60% F1 Accuracy. In general, the higher a classification rule's F1 Accuracy rate, the more preferred that classification rule would be (again, assuming that Recall and Precision Accuracies are equally important to the problem at hand).

Figure 12 below calculates these individual predictive error rates for our 10 million U.S. adult sample – as well as presents a deeper dive into the errors within each race / ethnicity category.

Figure 12: Actual vs. Predicted Race / Ethnicity at Individual-Level Using BISG Max Classification Rule

	Total	White	Black	API	Hispanic	Other
Overall Predictive Error Rate	16.8%					
Recall Accuracy	61.4%	93.1%	57.6%	66.2%	78.2%	11.9%
Precision Accuracy	75.4%	86.1%	71.6%	77.4%	78.2%	63.5%
F1 Accuracy	64.6%	89.5%	63.8%	71.3%	78.2%	20.1%
Over- / Under-Count of Actuals		8.0%	-19.5%	-14.4%	0.0%	-81.2%

[24] Technically, F1 Accuracy is the harmonic mean of Recall and Precision accuracies – calculated as (2*Recall*Precision) / (Recall+Precision)

Overall, 16.8% of individuals (1.68 million) are of a different race / ethnicity than the BISG proxy model would predict using the BISG Max classification rule – even though, at the aggregate level, actual and predicted races / ethnicities are essentially identical. This result occurs because of the inherent uncertainty of an individual's race / ethnicity when more than one of the five BISG probabilities has a non-zero value[25] – which is the case for nearly all sample members[26]. Importantly, this individual-level uncertainty or measurement error cancels out when aggregated across all sample members (see **Figure 8**) yielding accurate aggregate predicted group membership counts (i.e., 1,085,553 Predicted Blacks) even though the BISG proxy model is unable to specify accurately which individual sample members actually comprise these aggregate predicted groups (i.e., which of the 10 million sample members comprise the 1,085,553 Predicted Blacks).

This is an important limitation of the BISG proxy model that has yet to be sufficiently addressed within the context of fair lending regulatory activities and industry fair lending compliance risk management processes where specific identification of an individual's race / ethnicity is required. The practical impact of this limitation is quite high as evidenced by our sample where **inherent individual-level uncertainty results in incorrect race / ethnicity predictions 16.8% of the time**. However, even this result is somewhat misleading within a fair lending context as it is heavily influenced by the proportionally large (65.9%) White segment of the sample for which individual predictive accuracy errors are relatively less common (89.5% F1 Accuracy rate in **Figure 12**). If we instead evaluate and compare predictive accuracies at the individual race / ethnicity level, we see a much different picture of the BISG proxy model's limitations.[27]

- Between 72% and 86% of predicted races / ethnicities are accurate within specific race / ethnicity categories (see the "Precision Accuracy" row of **Figure 12**) – **with Predicted Non-White groups exhibiting the lowest individual-level accuracy rates (72 - 78%)**. For example, approximately 28% of individuals predicted to be Black (i.e., having "Black" as the highest BISG probability) are actually some other race / ethnicity (i.e., are False Positive Blacks)

- Between 58% and 93% of actual races / ethnicities are accurately predicted within specific race / ethnicity categories (the "Recall Accuracy" row of **Figure 12**) – **with Actual Non-White groups exhibiting the lowest individual-level accuracy rates (58 – 78%)**. For example, only 58% of Actual Blacks are predicted accurately by the BISG proxy model (indicating that 42% of Actual Blacks have BISG Black probability values that are not the highest of the set – the False

[25] Technically, the BISG proxy model generates six individual probabilities; however, as has been done in this study, it is common to consolidate the American Indian / Alaskan Native and 2+ Races categories into a single "Other" category due to the much lower reliability of the BISG proxy model for these groups.

[26] Only 0.13% of the sample exhibits a maximum BISG probability equal to 1 (i.e., perfect certainty of race / ethnicity).

[27] As a reminder, for the purposes of these results, predicted races / ethnicities are determined using the BISG Max classification approach (i.e., predicting race / ethnicity using the highest probability within the set). As we will see in later sections, alternative individual-level classification rules typically employed in fair lending analyses will generate different classification error rates. Additionally, the "Other" category is excluded from these observations.

Negative Blacks).

- **Estimated aggregate counts of each <u>predicted</u> racial / ethnic group are off by between -19.5% ("Blacks") and +8.0% ("White").** For example, if we summed up the number of Predicted Blacks and compared this sum to the number of Actual Blacks in the sample, we would find that the BISG Max classification rule underpredicted the total number of Actual Blacks by nearly 20%.[28]

Given the wide variability in accuracy metrics across individual race / ethnic groups – as well as the disproportionate representation of certain groups in the overall sample (e.g., Whites), the 16.8% overall predictive error rate can be misleading. Accordingly, a more meaningful set of "overall" accuracy metrics – in which an unweighted average of the accuracy metrics across the individual race / ethnicity groups is calculated – is contained in the "Total" column in **Figure 12**. Here we see that the BISG proxy model, using the BISG Max classification rule, achieves an average 61.4% Recall Accuracy, an average 75.4% Precision Accuracy, and an average 64.6% F1 Accuracy across the five individual race / ethnicity groups. **From a practical perspective, these results mean that for a randomly-selected individual:[29] (1) the BISG Max classification rule will accurately predict the individual's actual race / ethnicity only 61% of the time, on average and (2) the predicted race / ethnicity from the BISG Max classification rule for that individual will be accurate 75% of the time, on average**.

In summary, this analysis clearly illustrates the breakdown of the BISG proxy model's predictive accuracy when applied to the individual level – which is a critical requirement for present-day fair lending compliance risk management in which the demographic identification of individual customers is necessary for certain types of fair lending bias testing, as well as for the remediation of potential fair lending biases under lender corrective action policies. This breakdown occurs because: (1) the BISG proxy model is not designed to predict accurately at the individual-level and (2) the inherent uncertainty of each individual's group membership – while offsetting at the overall aggregate sample level – remains present for each individual and, therefore, contributes to individual-level race / ethnicity classification errors.

While different classification rules will produce somewhat different individual race / ethnicity classification errors (as we will see in later sections), the primary driver of individual-level accuracy across all classification rules is how well the BISG proxy model and its underlying U.S. Census data distinguish the five race / ethnicity groups from each other. In particular, as previously discussed in **Figure 7**, strong predictive power is tied to: (1) high BISG probability values for the matching race / ethnicity category (e.g., Hispanic individuals receiving high BISG Hispanic probability values) and

---

[28] For the BISG Max classification rule, the overestimated and underestimated counts net to zero across all five racial / ethnic groups.

[29] That is, for an individual randomly-selected from one of the five race / ethnicity groups, and using the BISG Max classification rule to predict individual race / ethnicity from the BISG proxy model's set of five probabilities.

(2) low BISG probability values for non-matching BISG probabilities (e.g., Hispanic individuals receiving low BISG White, Black, API, and Other probability values). To the extent that the BISG proxy model fails to generate these "desired" BISG probability distributions for material segments of a sample, overall predictive power at the individual level will be adversely impacted. **Figure 13** below helps to illustrate this point.

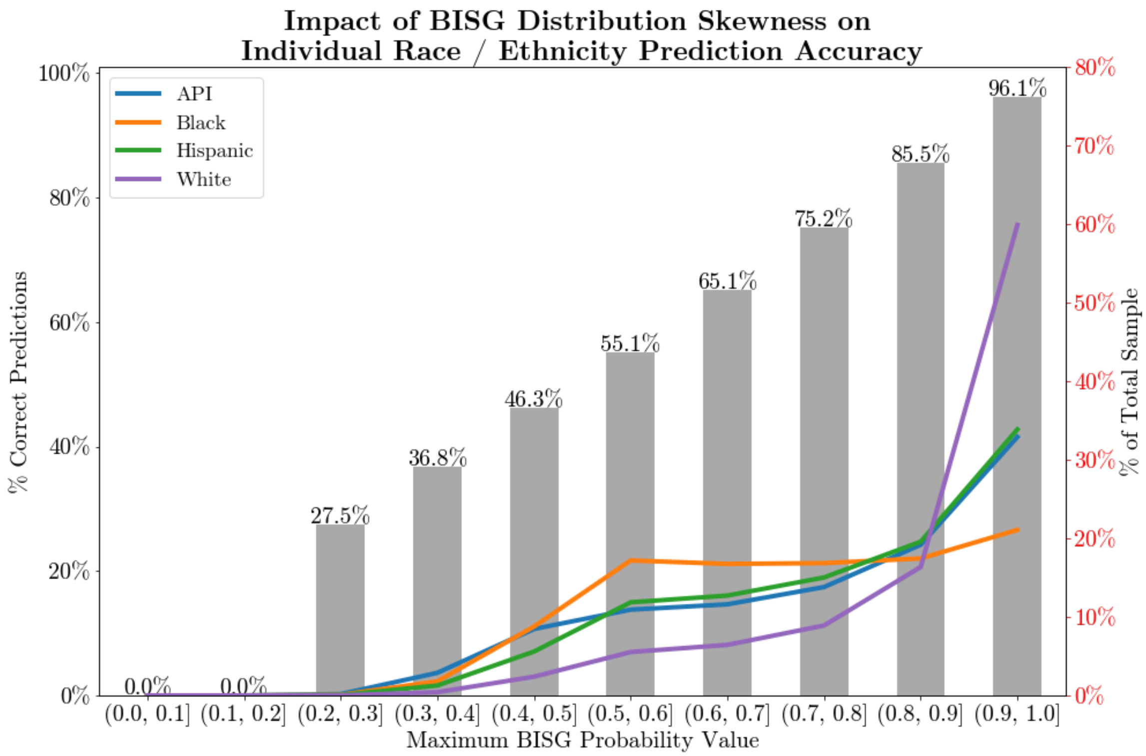


Figure 13

Here we obtain the maximum BISG probability value for all 10 million individuals in our sample and segment these values into 10 ordered groups on the horizontal axis. For those individuals whose maximum BISG probability value falls into a given probability segment, we calculate the percentage of these individuals for whom the predicted race / ethnicity (based on the BISG Max classification rule) matches the individuals' actual race / ethnicity. These accuracy rates are depicted by the grey bars that are aligned to the left-side vertical axis. For example, the last grey bar on the right indicates that 96.1% of individuals with maximum BISG probability values in excess of 90% have correctly predicted races / ethnicities – a very high accuracy rate as we would expect. However, as we move leftward on this chart to individuals for whom the BISG proxy model is less certain (as evidenced by lower maximum BISG probability values), we observe steadily decreasing individual accuracy rates.[30] For example, individuals with maximum BISG probabilities between 50% and 60% have an individual

[30] With five possible BISG races / ethnicities, the lowest possible maximum BISG probability value is 20%. This explains the lack of data for maximum BISG probability values of 0-10% and 10%-20% (i.e., it is technically impossible for there to be maximum BISG probability values in this range).

accuracy rate of just 55.1%.

Given this framework, it is clear that the overall individual accuracy rate for each race / ethnicity group will largely depend on how the maximum BISG probabilities of the individuals in that group are distributed. In particular, race / ethnicity groups whose members have geo-surname segments that are, on average, more racially segregated (i.e., who possess high maximum BISG probability values) will exhibit higher individual accuracy rates, while race / ethnicity groups whose members have geo-surname segments that are, on average, more racially diversified (i.e., who possess lower maximum BISG probability values) will exhibit lower individual accuracy rates. These distributions for each of the four primary race / ethnicity groups are exhibited in **Figure 13** by the colored lines that align to the right-side vertical axis. Here we see that Whites, as a group, have maximum BISG probabilities that are highly-concentrated at the upper end of the distribution (i.e., 60% of Whites have maximum BISG probabilities greater than 90% and 16% have maximum BISG probabilities between 80% and 90%) – thereby contributing to the high individual accuracy rate for the group as summarized in **Figure 12**. On the other hand, Blacks – as a group – have maximum BISG probabilities that are more evenly distributed across the probability segments with about 18% of Blacks having maximum BISG probabilities in each of the top 5 probability segments. This lack of concentration in the upper probability segments contributes to the lower individual accuracy rate for the group as summarized in **Figure 12**.

Some may view these results as counterintuitive as conventional wisdom is that Blacks and Whites both tend to live in more segregated geographies. While this may certainly be true for specific micro-geographies, such segregation is much less extreme when measured at more macro levels – particularly for Blacks. For example, **Appendix C** analyzes the Census Block Group ("CBG") demographics of sample members within 10 racially-diverse MSAs. For some of these MSAs, segregation appears to be quite high; for example, Blacks in Detroit reside in CBGs that are, on average, 66.4% Black, Hispanics in Miami reside in CBGs that are, on average, 60.9% Hispanic, and Whites in Pittsburgh reside in CBGs that are, on average, 91.7% White. Drilling further into specific counties within these 10 MSAs (not shown in Appendix C), we see examples of even greater segregation – with Blacks in Wayne County MI residing in CBGs that are, on average, 78.4% Black, Hispanics in Miami-Dade County FL residing in CBGs that are, on average, 78.2% Hispanic, and Whites in Armstrong County PA residing in CBGs that are, on average, 98.2% White.

Notwithstanding the presence of significant segregation within certain micro-geographies, it is also true that members of different racial / ethnic groups live in many other geographies that are more racially diverse – thereby offsetting some of the high segregation rates when aggregated to the national level. In fact, according to the 2010 Census data, 66% of the Blacks in our overall national sample reside in

CBGs that are less than 50% Black.[31] Furthermore,

- Blacks in our overall sample reside in CBGs that are, on average, only 39% Black while Whites in our overall sample reside in CBGs that are, on average, 76% White.

- 59% of CBGs in which sample Whites reside are at least 80% White while only 18% of CBGs in which sample Blacks reside are at least 80% Black.

In summary, the different distributions of maximum BISG probabilities for the four primary race / ethnicity groups are the primary drivers of the variation in individual-level accuracy rates reported in **Figure 12**, and these different distributions are driven by the socioeconomic demographics of the U.S. adult population. Therefore, from a national perspective, since Whites tend to reside in more racially-segregated geographies and Blacks tend to reside in more racially-diverse geographies, the BISG proxy methodology has higher inherent individual-level accuracy for Whites and lower inherent individual-level accuracy for Blacks. With respect to Hispanics and APIs, as shown in **Figure 13** both groups have maximum BISG probability distributions that lie in between those for Whites and Blacks – which translates to a similar pattern in their relative individual-level accuracy rates shown in **Figure 12**.[32]

One important final point is that individual accuracy rates can vary significantly at more micro-geographic levels. For example, **Figure 14** below provides the F1 Accuracy rates for the 10 racially-diverse MSAs discussed above and presented in **Appendix C**.

Figure 14: F1 Accuracy Rates for Selected MSAs

MSA	API	Black	Hispanic	White
Atlanta	73.3%	72.0%	81.8%	85.1%
Boston	71.8%	55.4%	75.5%	91.9%
Chicago	73.7%	73.7%	79.8%	89.7%
Detroit	72.1%	76.1%	73.0%	90.2%
Los Angeles	63.4%	59.9%	78.0%	81.8%
Miami	77.3%	70.6%	81.3%	81.6%
New York	69.2%	73.4%	79.2%	88.1%
Pittsburgh	72.5%	56.2%	55.3%	93.3%
San Diego	69.0%	44.4%	78.3%	86.7%
San Francisco	60.2%	57.6%	78.6%	84.5%

Here we see that while the F1 Accuracy rate for Blacks at the national level is 63.8% according to

[31] This feature of the sample is consistent with that described for an unnamed U.S. financial institution that was referred by the CFPB to the U.S. Department of Justice for potential fair lending violations in automobile financing. Specifically, according to a February 14, 2014 CFPB Referral Letter to the U.S. Department of Justice, pp. 30-31 "...approximately 72% of ... African-American customers reside in census tracts that are less than 50% African-American." and "Put another way, ... only a small percentage of protected class customers live in areas that are heavily minority."

[32] Hispanics and APIs tend to display more surname segregation than geographic segregation, as we will see later in this study.

Figure 12, it can range as low as 44.4% for Blacks in the San Diego MSA and as high as 76.1% for Blacks in Detroit. This MSA-level variability ties directly to the relative segregation rates we previously discussed in **Appendix C** whereby Blacks in the San Diego MSA reside in CBGs that are, on average, only 10.1% Black, while Blacks in the Detroit MSA reside in CBGs that are, on average, 66.4% Black. **Accordingly, caution needs to be exercised when using the BISG proxy model to predict individual races / ethnicities for micro-geographies – particularly if such geographies have relatively low segregation rates for the particular races / ethnicities of interest.**[33]

Individual-Level Accuracy: Exploring the False Negative Bias

As shown in **Figure 12**, the ability of the BISG proxy model and corresponding BISG Max classification rule to predict accurately the actual members of each race / ethnicity group (i.e., Recall Accuracy) varies significantly across the four primary racial / ethnic groups – with White Recall Accuracy the greatest at 93.1% and Black Recall Accuracy the lowest at 57.6%. As illustrated in the top row of **Figure 15** below, what this means is that 42.4% of Total Actual Blacks in the sample are False Negatives ("FNs") – that is, they are misclassified as some other race / ethnicity and therefore excluded from the Total Predicted Black group (i.e., the sum of the first column) used in downstream fair lending analyses.

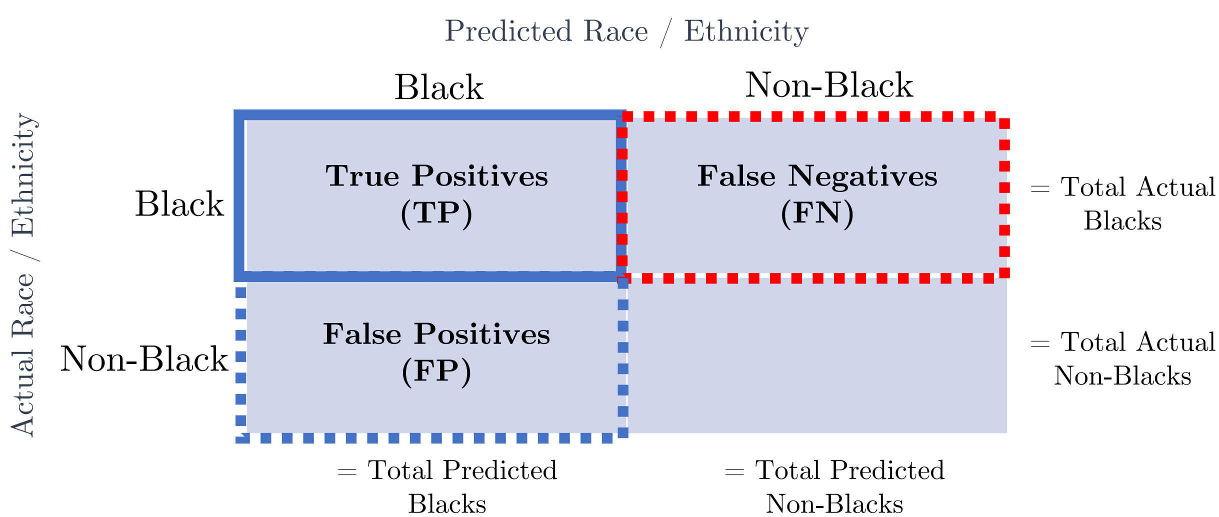


Figure 15

For fair lending analyses relying on individual-level proxies, FNs are important as they represent the segment of actual group members that are <u>excluded</u> from such analyses. In some cases, FNs can be a

[33] **Appendix D** provides state-level accuracy rates for the primary race / ethnicity groups under the BISG Max classification method.

substantial component of actual group members – as we see with the 42.4% of Actual Blacks in our sample. As we will discuss further throughout the next two sections, understanding the properties of both FNs and False Positives ("FPs") is critical to understanding the risks and limitations of the BISG proxy model for fair lending compliance purposes. This is because the accuracy of a predicted group's race / ethnicity is significantly impacted by: (1) the characteristics of the FNs that are excluded from predicted group membership[34] and the characteristics of the FPs that are included in predicted group membership. **To the extent that material differences exist between FNs and FPs relevant to the fair lending analysis, then biases in measured outcome differences can arise, in addition to biases in group membership identification and aggregate membership counts.**

To explore these potential biases, the charts and tables in **Figures 16 and 17** are designed to provide insights to the FNs contained in the individual predicted races / ethnicities within our national geo-surname sample.[35] In the first chart below, we plot the distribution of maximum BISG probabilities for those sample members who are actually White along with the predicted race / ethnicities of those members. For example, nearly 60% of this sample has maximum BISG probabilities greater than 90% (the right-most bar) while only 2.8% has maximum BISG probabilities less than or equal to 50% (the left-most five bars). This is consistent with the corresponding distribution previously discussed in **Figure 13**.

[34] Specifically, the characteristics of the excluded FNs will impact the characteristics of the remaining True Positive ("TP") group members – a group that typically comprises the majority of the predicted group. For example, if the Black FNs are simply a random sample of Actual Blacks, then the impact on the remaining TPs is zero. However, if, instead, the FNs are of materially higher average income or have materially lower average price exceptions, then the characteristics of the Predicted Black group will be biased unless there is an exact offset of the FN bias by the included FPs.

[35] As in prior sections, the BISG Max classification rule is used for this analysis due to its two desirable properties: (1) it generates a predicted race / ethnicity for every individual and (2) its predictions rely on relative BISG probability values (i.e., the largest) and are therefore less sensitive to the use of specific probability threshold values such as 50% or 80%.

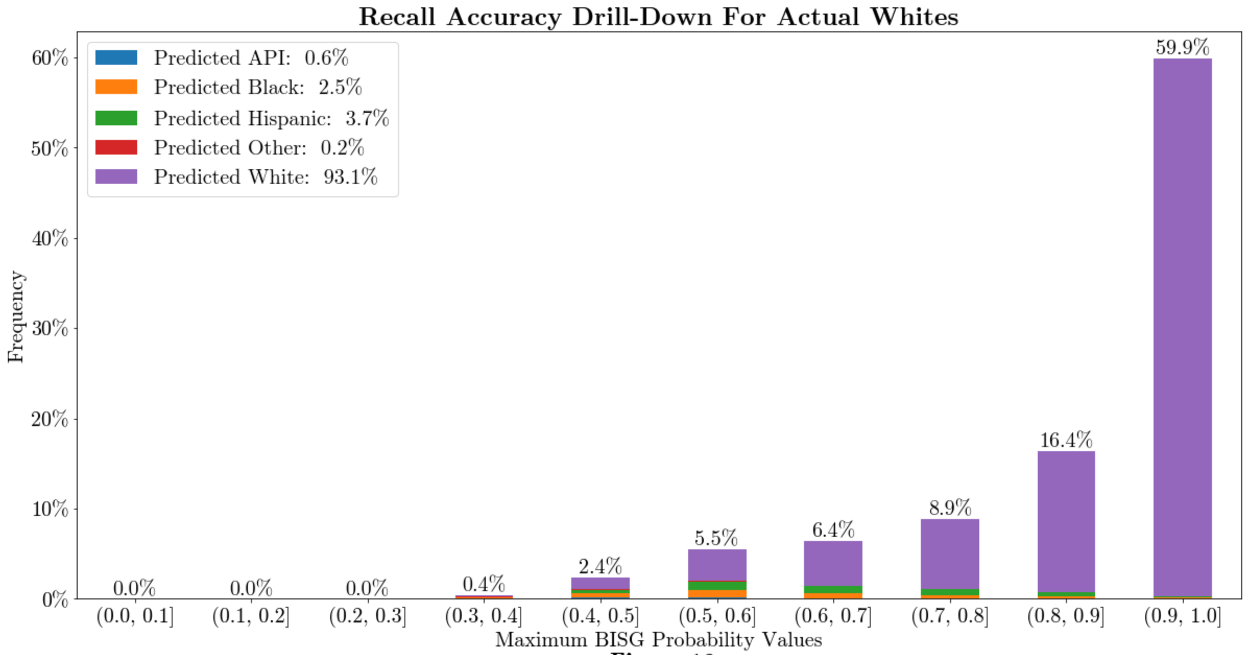


Figure 16a

The White Recall Accuracy rate of 93.1% corresponds to the row-wise True Positive ("TP") Rate in **Figure 15** (i.e., the percentage of Actual Whites correctly predicted to be White). The remaining 6.9% of Actual Whites are the FNs which grow in proportion as we move leftward along the x-axis to lower maximum BISG probability values. As can be seen in the legend box, over half of the FNs (3.7%) are predicted to be Hispanic with another 36% of FNs (2.5%) are predicted to be Black.

But what else do we know about these FNs?

Figure 17a below compares certain characteristics of the White FNs to the White TPs to identify why these White individuals might be misclassified as Non-White, and what relevant biases they may cause.

Figure 17a: Comparative Characteristics of White FNs

Whites	Total Actuals	–	False Negatives	=	True Positives
Average CBG White %	76.3%		55.3%		77.9%
Average CBG Black %	7.3%		18.6%		6.5%
Average CBG Hispanic %	10.3%		19.1%		9.7%
Average CBG API %	4.4%		5.0%		4.3%
Average Surname White %	76.4%		38.5%		79.2%
Average Max Probability	86.9%		63.2%		88.7%
Average Median HH Income	$61,331		$53,681		$61,899
Sample Counts	6,592,038		456,959		6,135,079
% of Actual Whites			-6.9%		93.1%

The first four rows capture the average demographics of the CBGs in which the White FNs and TPs reside, the next row reflects the average surname demographics of group members, the following row reflects the relative degree of certainty associated with the individual race / ethnicity predictions, the next row presents the average median household income of those CBGs,[36] and the final two rows provide comparative sample counts.

Based on this comparison, we can see pretty clearly that the White FNs are not a random selection of Actual Whites; rather, (1) they reside in more racially-diverse CBGs that, on average, have greater Black and Hispanic representation and lower average median household incomes (specifically, about 13% lower than the White TPs), and (2) their surnames are more racially / ethnically diverse – both of which combine to create a greater degree of uncertainty associated with the predicted race / ethnicity of these White individuals as demonstrated by a lower maximum probability average (63.2% for FNs vs. 88.7% for TPs).

For Actual Blacks, on the other hand, the following Recall Accuracy Drill Down chart tells a much different story.

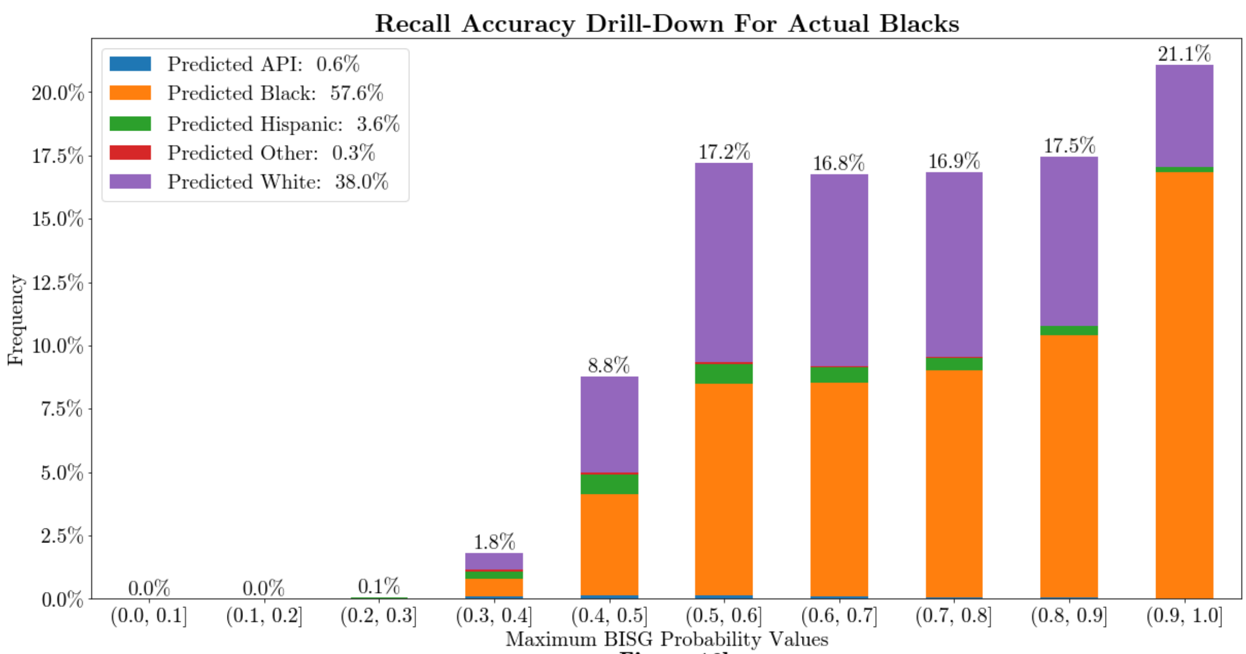


Figure 16b

The Recall Accuracy rate of 57.6% is captured visually by the orange bar segments in the chart (i.e., the TPs) with the FNs depicted by the non-orange bar segments that comprise the remaining 42.4%

[36] The CBG demographics are from the 2010 Census data used to construct the BISG proxy probabilities. The median household income data is from the 2010 American Community Survey Census Block Group dataset and reflects a trailing five-year average (i.e., 2006-2010) expressed in 2010 dollars. For this data, see Steven Manson, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles. IPUMS National Historical Geographic Information System: Version 16.0 [dataset]. Minneapolis, MN: IPUMS. 2021. http://doi.org/10.18128/D050.V16.0

of Actual Blacks. Unlike what we observed with Actual Whites, Actual Blacks have substantial FNs throughout the maximum probability range. Furthermore, 90% of the Black FNs (38%) are misclassified as White with most of the remaining amount misclassified as Hispanic.

Figure 17b below compares certain characteristics of the Black FNs to the Black TPs to identify why these Black individuals might be misclassified as Non-Black, and what relevant biases they may cause.

Figure 17b: Comparative Characteristics of Black FNs

Blacks	Total Actuals	–	False Negatives	=	True Positives
Average CBG White %	38.8%		58.3%		24.5%
Average CBG Black %	38.7%		17.9%		54.0%
Average CBG Hispanic %	16.3%		16.2%		16.3%
Average CBG API %	4.6%		5.8%		3.7%
Average Surname Black %	27.1%		21.9%		30.9%
Average Max Probability	72.6%		68.1%		76.0%
Average Median HH Income	$47,221		$54,906		$41,533
Sample Counts	1,084,853		460,139		624,714
% of Actual Blacks			-42.4%		57.6%

Based on this comparison, we can see pretty clearly that Black FNs are not random either; rather, they reside in CBGs with a much higher White representation (58.3% vs. 24.5% for TPs) which – along with a slightly lower Black surname concentration – combine to create greater uncertainty associated with the predicted race / ethnicity of these Black individuals as demonstrated by a lower maximum probability average (68.1% for FNs vs. 76.0% for TPs). Additionally, the FNs have much higher average median household incomes (specifically, about 32% higher than Black TPs) which is particularly notable as it is the opposite of what we saw for White FNs where average median household incomes were 13% lower than White TPs. **Overall, the combination of a significant Black FN representation (42.4% of Actual Blacks) and significantly higher average median household income for Black FNs indicates a significant risk that the Total Predicted Blacks used for downstream fair lending analyses may be materially biased toward lower-income Black individuals**. However, the extent of such potential bias depends on the characteristics of the Black FPs that will be swapped into the Predicted Black group in place of the Black FNs – which will be explored in the next section on FPs.

The following charts for Actual Hispanics and APIs fall in between those discussed above for Actual Whites and Blacks. Specifically, for Actual Hispanics, the Recall Accuracy rate of 78.2% implies a FN rate of 21.8% – much lower than that for Blacks (42.4%) but three times higher than for Whites (6.9%). As shown in the chart below, like Blacks, the Hispanic FNs occur at all maximum BISG probability levels but are much smaller in magnitude. Over 80% of the Hispanic FNs (17.9%) are misclassified as White with small amounts misclassified as Black (2.7%) and API (1%).

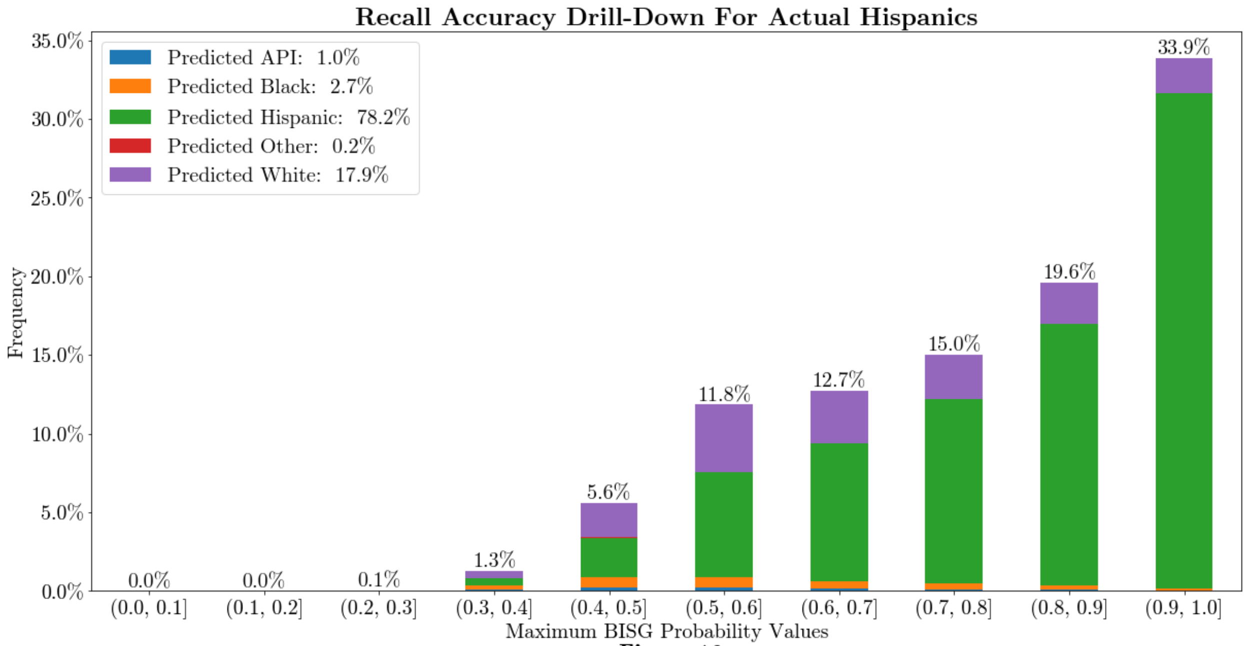


Figure 16c

According to **Figure 17c** below, Hispanic FNs are more likely to be caused by the presence of less-Hispanic surnames than by geographic demographics – which drives the greater prediction uncertainty for Hispanic FNs (66.1% average maximum probability vs. 81.8% for Hispanic TPs).[37] Because the Hispanic FNs are not strongly geographically driven, the observed difference in average median household income between Hispanic FNs and TPs (-4.5%) is not as significant as observed for Whites and Blacks.

Figure 17c: Comparative Characteristics of Hispanic FNs

Hispanics	Total Actuals	–	False Negatives	=	True Positives
Average CBG White %	52.5%		49.8%		53.2%
Average CBG Black %	11.7%		11.0%		11.8%
Average CBG Hispanic %	28.6%		31.1%		27.9%
Average CBG API %	5.6%		6.2%		5.4%
Average Surname Hispanic %	66.9%		22.1%		79.4%
Average Max Probability	78.4%		66.1%		81.8%
Average Median HH Income	$54,823		$52,861		$55,370
Sample Counts	1,589,902		346,636		1,243,266
% of Actual Hispanics			-21.8%		78.2%

For Actual APIs, the Recall Accuracy rate of 66.2% implies a FN rate of 33.8% – lower than that for Blacks but almost five times higher than for Whites. As shown in the chart below, like Blacks and Hispanics, the API FNs occur at all maximum BISG probability levels but are much smaller in

[37] This is consistent with the findings presented in the Elliott BISG paper where geography alone accounted for only 5 – 7% of the BISG probabilities' predictive power for Hispanics and Asians, respectively.

magnitude than we observed for Blacks. The majority of the API FNs (20.9% of the 33.8%) are misclassified as White; however, there are also material misclassifications as Hispanics (9.8%).

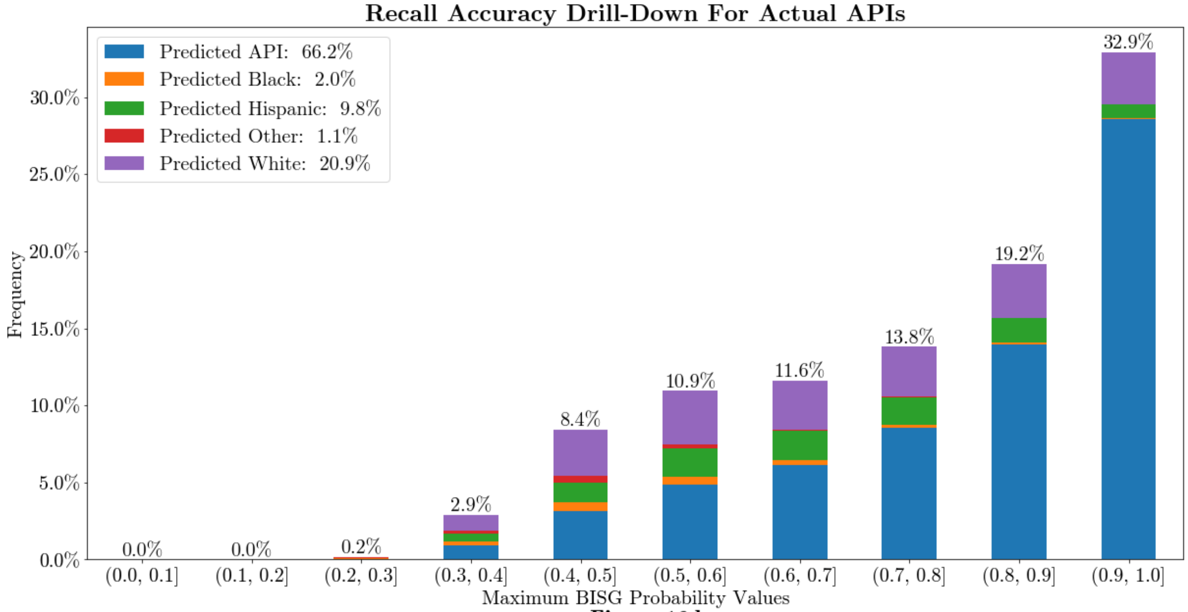


Figure 16d

According to **Figure 17d** below, like Hispanics, API FNs are more likely to be caused by the presence of Non-API surnames than by the geographic drivers noted previously for Whites and Blacks. Additionally, because the API FNs are not strongly geographically driven, the observed difference in average median household income between API FNs and TPs is not as significant as observed for Whites and Blacks (i.e., only +5.6%).

Figure 17d: Comparative Characteristics of API FNs

APIs	Total Actuals	–	False Negatives	=	True Positives
Average CBG White %	59.7%		53.4%		62.9%
Average CBG Black %	9.8%		9.1%		10.1%
Average CBG Hispanic %	15.5%		16.6%		14.9%
Average CBG API %	13.1%		18.7%		10.2%
Average Surname API %	59.4%		11.7%		83.7%
Average Max Probability	76.7%		66.7%		81.8%
Average Median HH Income	$65,667		$68,038		$64,455
Sample Counts	414,671		140,197		274,474
% of Actual APIs			-33.8%		66.2%

In the next section, we explore the properties of the False Positives in a similar manner and then derive some themes regarding overall potential biases in the predicted race / ethnicity groups used in typical downstream fair lending analyses.

Individual-Level Accuracy: Exploring the False Positive Bias

In the last section, we showed how the mapping of individual-level BISG probabilities into specific races / ethnicities created a set of False Negatives – that is, actual members of a given race / ethnicity group who were misclassified into another race / ethnicity group. Through further analysis, we also showed that the False Negatives were not random individuals from the actual race / ethnicity group; rather, they have distinct characteristics that may induce bias in the remaining True Positive actual group members who represent the majority of predicted group members (see the first column of **Figure 15**).

In this section, we complete our exploration of the Predicted Group by analyzing the composition of the False Positive group members – that is, individuals incorrectly classified as members of a given race / ethnicity group. As we will see below, FPs – like FNs – are not random individuals. They have distinct characteristics that may introduce a second source of potential bias into the predicted group members – a bias that has important impacts on measured fair lending outcome disparities.

We start with **Figure 18a** below where we plot the distribution of maximum BISG probabilities for those sample members who are predicted to be White, along with the actual races / ethnicities of those members. Essentially, we are focusing on the group contained in the first column of **Figure 15** (i.e., the predicted members of each race / ethnicity group) and segmenting that group into its TP and FP components.[38] As shown below, about 57% of Predicted Whites have maximum BISG probabilities greater than 90% (the right-most bar) while only 3.2% have maximum BISG probabilities less than or equal to 50% (the left-most five bars).

[38] As should be apparent by now, since TPs are a component of both Actual and Predicted groups, the potential bias between these two groups is driven solely by the comparative properties of the FNs and FPs – which explains our focus in this and the last section.

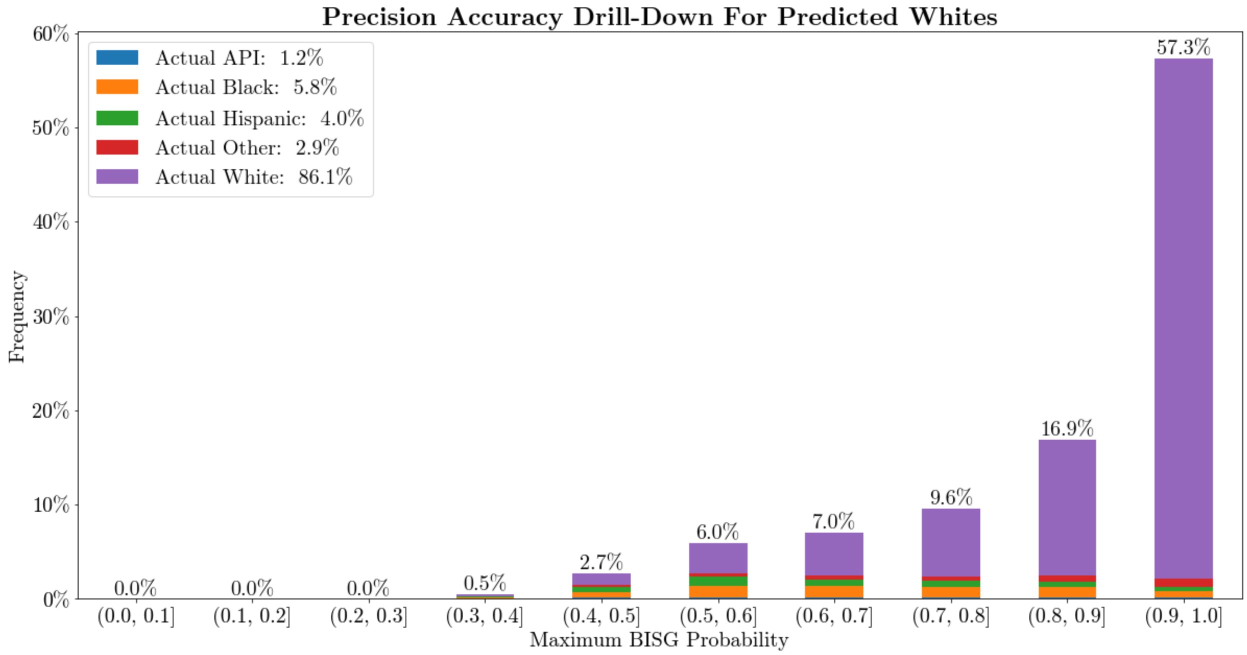


Figure 18a

The Precision Accuracy rate of 86.1% corresponds to the column-wise True Positive ("TP") Rate in **Figure 15** (i.e., the percentage of Predicted Whites who are actually White). The remaining 13.9% of Predicted Whites are the FPs which, as shown above, are a fairly stable component of Predicted Whites at maximum BISG probability values above 50%. Additionally, as can be seen in the legend box, about 40% of the FPs (5.8%) are Actual Blacks, about 30% are Actual Hispanics, and the remainder are Actual APIs and Others.

But what else do we know about these FPs?

Figure 19a expands the comparative characteristics tables from the prior section to walk through the build-up of Predicted Whites – starting from Actual Whites, swapping out the False Negatives, and finally swapping in the False Positives.

Figure 19a: Comparative Characteristics of Actual vs. Predicted Whites

Whites	Total Actuals	–	False Negatives	=	True Positives	+	False Positives	=	Total Predicted
Average CBG White %	76.3%		55.3%		77.9%		60.6%		75.5%
Average CBG Black %	7.3%		18.6%		6.5%		11.4%		7.1%
Average CBG Hispanic %	10.3%		19.1%		9.7%		18.5%		10.9%
Average CBG API %	4.4%		5.0%		4.3%		7.4%		4.8%
Average Surname White %	76.4%		38.5%		79.2%		67.7%		77.6%
Average Max Probability	86.9%		63.2%		88.7%		70.2%		86.1%
Average Median HH Income	$61,331		$53,681		$61,899		$57,221		$61,252
Sample Counts	6,592,038		456,959		6,135,079		987,356		7,122,435
% of Actual Whites			-6.9%		93.1%				
% of Predicted Whites					86.1%		13.9%		

Based on this analysis, we can see the following:

- The White FPs represent, on average, moderately lower income individuals (7.6% lower income vs. White TPs), primarily Black and Hispanic, who reside in more racially diverse but still majority White CBGs (60.6% White vs. 77.9% White for TPs) and who have less ethnically diverse surnames (i.e., a moderately high 70.2% White). They comprise 13.9% of the Predicted White group and have average maximum probability values of just 70.2%.

- White FPs are over twice the size of White FNs – leading to an 8% overestimation of White individuals (7,122,435 vs. 6,592,038). However, a comparison of average income levels between Actual and Predicted Whites reveals a negligible 0.1% difference – indicating that FPs effectively offset FNs from an income perspective (also see **Figures 20a and 21a** below). Average CBG and surname demographics between Actual and Predicted Whites also show minor differences.

Figures 20a and 21a provide a visual comparison of the comparative income distributions.

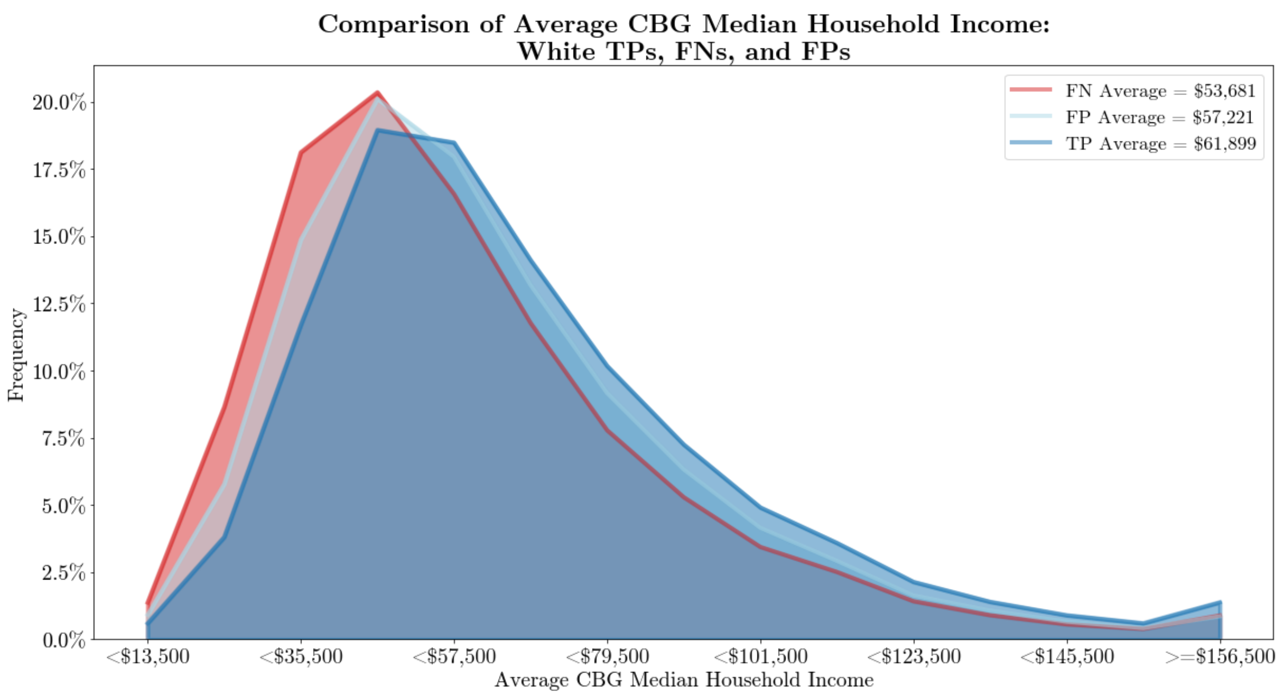


Figure 20a

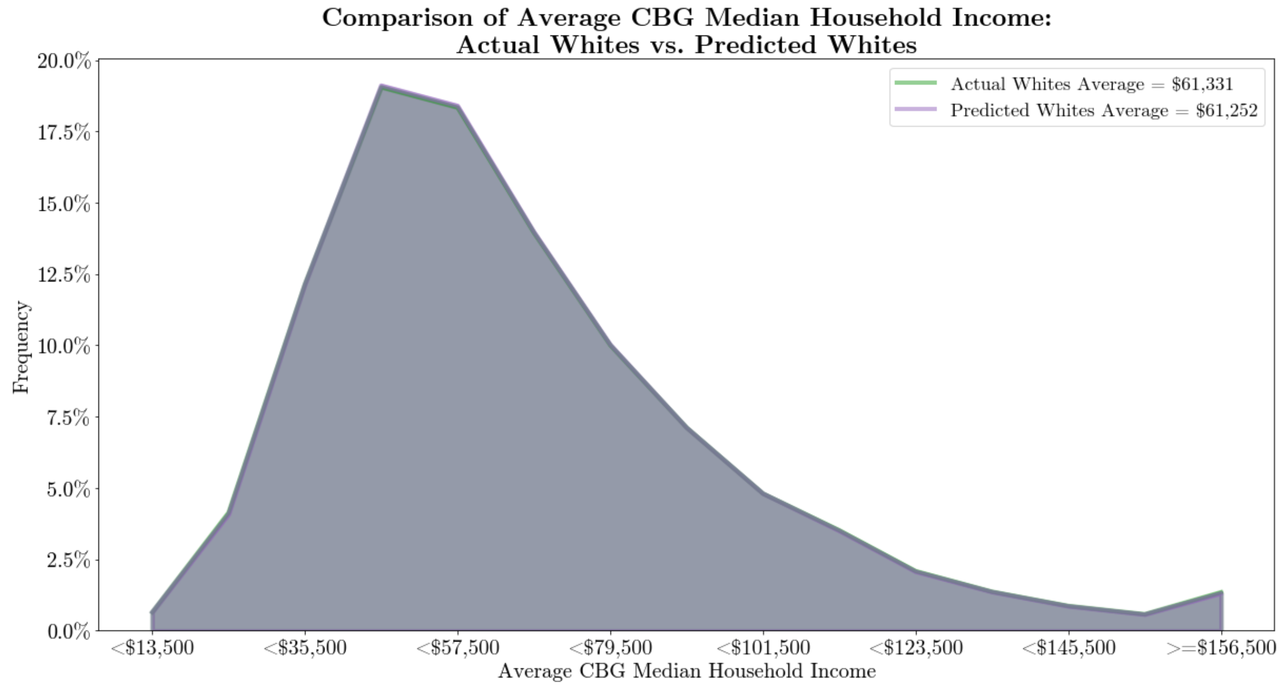


Figure 21a

For Predicted Blacks, the story is very different as shown in the charts below. First, their maximum BISG probabilities are much less concentrated at upper probability levels than observed for Predicted Whites – with only 21.9% of Predicted Blacks having maximum BISG probabilities of at least 90% vs. nearly 60% for Predicted Whites. Additionally, 13.0% of Predicted Blacks come from maximum BISG probability values less than or equal to 50% vs. only 3.2% of Predicted Whites.

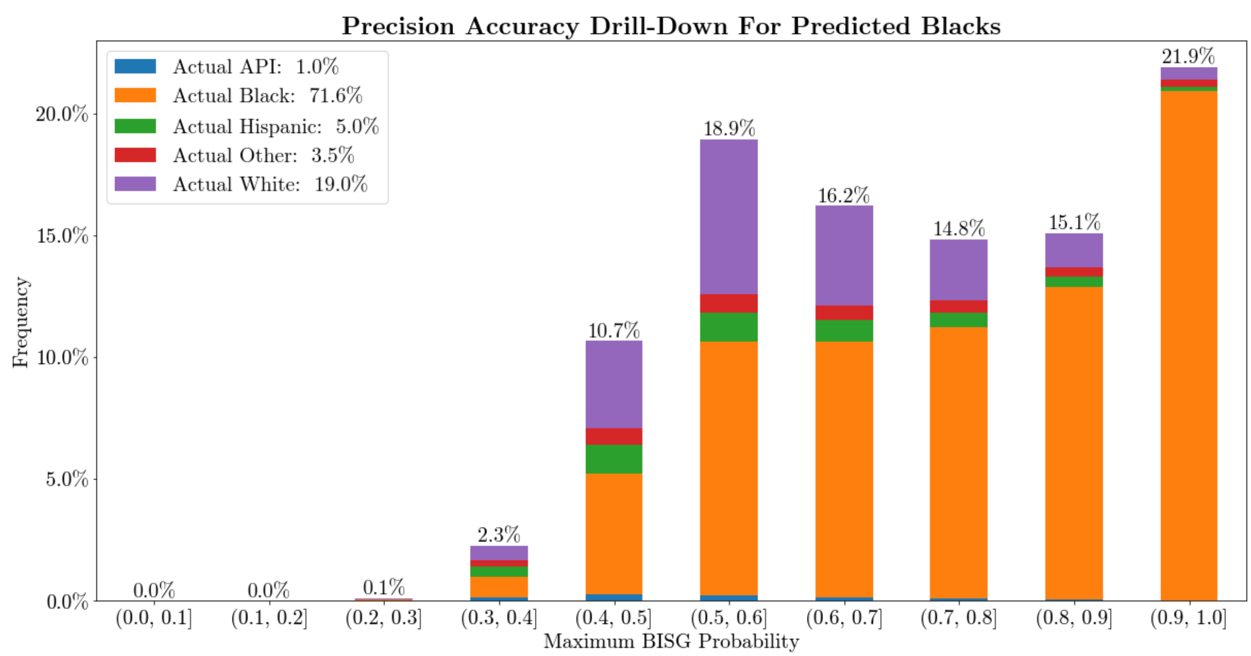


Figure 18b

With a Precision Accuracy rate of 71.6%, the remaining 28.4% of Predicted Blacks are the FPs which, as shown above, are quite prevalent at maximum BISG probability values less than 90%. Additionally, as can be seen in the legend box, about two-thirds of the FPs (19.0%) are Actual Whites, and about 18% are Actual Hispanics.

Figure 19b expands the comparative characteristics tables from the prior section to walk through the build-up of Predicted Blacks – starting from Actual Blacks, swapping out the False Negatives, and finally swapping in the False Positives.

Figure 19b: Comparative Characteristics of Actual vs. Predicted Blacks

Blacks	Total Actuals	–	False Negatives	=	True Positives	+	False Positives	=	Total Predicted
Average CBG White %	38.8%		58.3%		24.5%		32.4%		26.7%
Average CBG Black %	38.7%		17.9%		54.0%		38.1%		49.4%
Average CBG Hispanic %	16.3%		16.2%		16.3%		21.9%		17.9%
Average CBG API %	4.6%		5.8%		3.7%		5.8%		4.3%
Average Surname Black %	27.1%		21.9%		30.9%		28.4%		30.2%
Average Max Probability	72.6%		68.1%		76.0%		60.3%		71.6%
Average Median HH Income	$47,221		$54,906		$41,533		$44,825		$42,469
Sample Counts	1,084,853		460,139		624,714		248,266		872,980
% of Actual Blacks			-42.4%		57.6%				
% of Predicted Blacks					71.6%		28.4%		

Based on this analysis, we can see the following:

- Opposite to what we observed for Whites, the Black TPs represent, on average, <u>lower</u> income Black individuals who live in highly racially segregated CBGs (i.e., 54% Black). In fact, average income for Black TPs is 12% lower than average income for all Actual Blacks. Black TPs represent 71.6% of Predicted Blacks and have an average maximum probability value of 76.0%.

- The Black FPs represent, on average, higher income individuals (7.9% higher income vs. Black TPs), primarily White, who reside in more racially diverse, but still majority minority CBGs (60% Black/Hispanic vs. 70.3% for TPs). While these FPs have higher average income than Black TPs (7.9%), we also note that – relative to White FNs (where they primarily come from – see **Figure 19a**) – their average incomes are about 16% lower. Their average maximum probability of 60.3% is consistent with this misclassification.

- Black FNs are almost twice the size of Black FPs leading to a 20% underestimation of Black individuals (872,980 vs. 1,084,853). Furthermore, a comparison of average income levels between Actual and Predicted Blacks reveals a fairly large 10% reduction in average income for Predicted Blacks – indicating that the loss of higher income FNs dominates the gain of slightly higher income FPs (also see **Figure 20b and 21b** below). Finally, average CBG demographics between Actual and Predicted Blacks show that Predicted Blacks are more concentrated in racially segregated CBGs than Actual Blacks (49.4% Black for Predicted vs. 38.7% Black for Actuals).

Overall, when used for individual-level classification, the BISG proxy model appears to generate biased samples of individual Blacks that: (1) exclude higher income Blacks living in more racially diverse geographies (the FNs), and (2) include lower-income Whites living in high-minority areas (the FPs). Combined together, the Predicted Blacks underrepresent the number of Actual Blacks in the sample by 20% and bias the sample's average income down by 10%.

Figures 20b and 21b provide a visual comparison of the comparative income distributions.

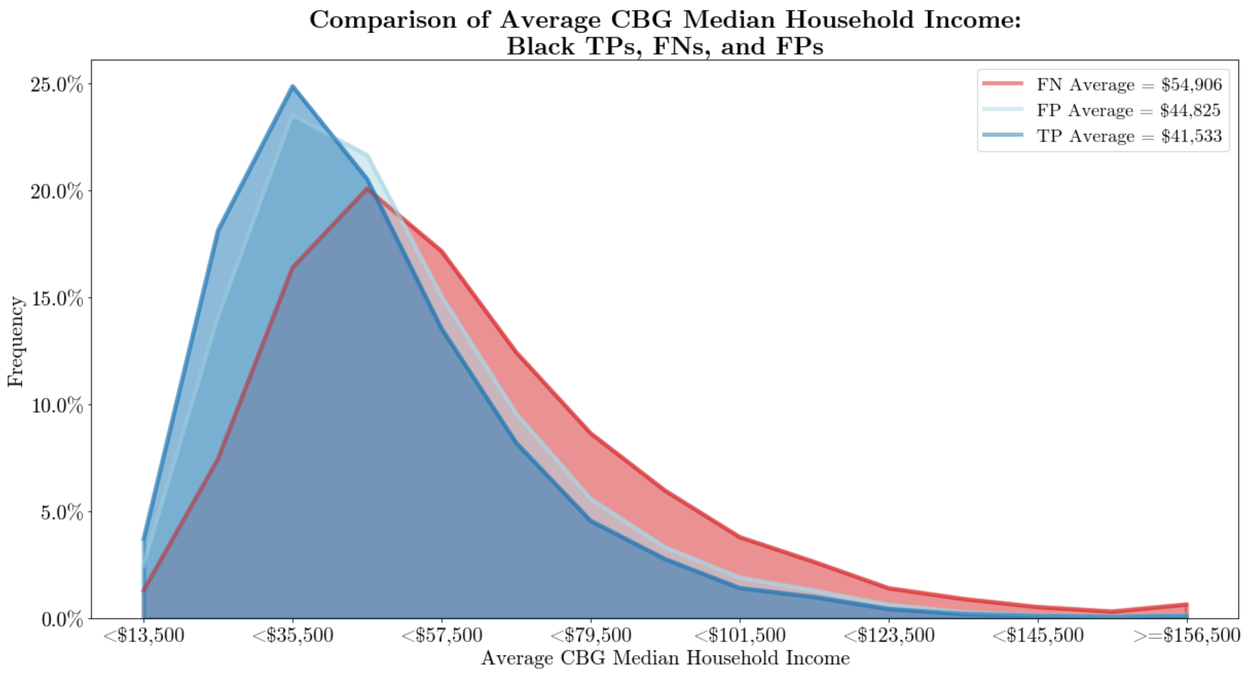


Figure 20b

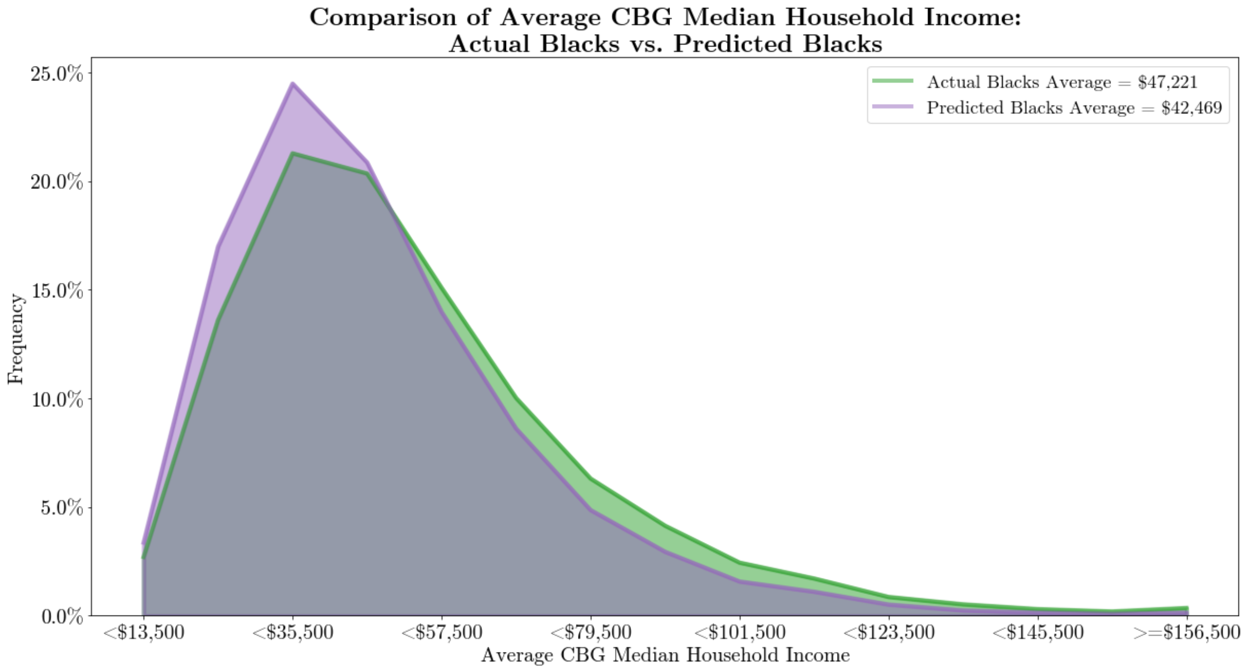


Figure 21b

The following charts for Predicted Hispanics and Predicted APIs are qualitatively similar to that discussed above for Predicted Whites – although less concentrated at the very upper end of the probability range. For Predicted Hispanics in **Figure 18c**, Precision Accuracy is 78.2% and Whites comprise about 70% of the Hispanics FPs – which are, as for Blacks above – rather prevalent at maximum BISG probability values less than 90%.

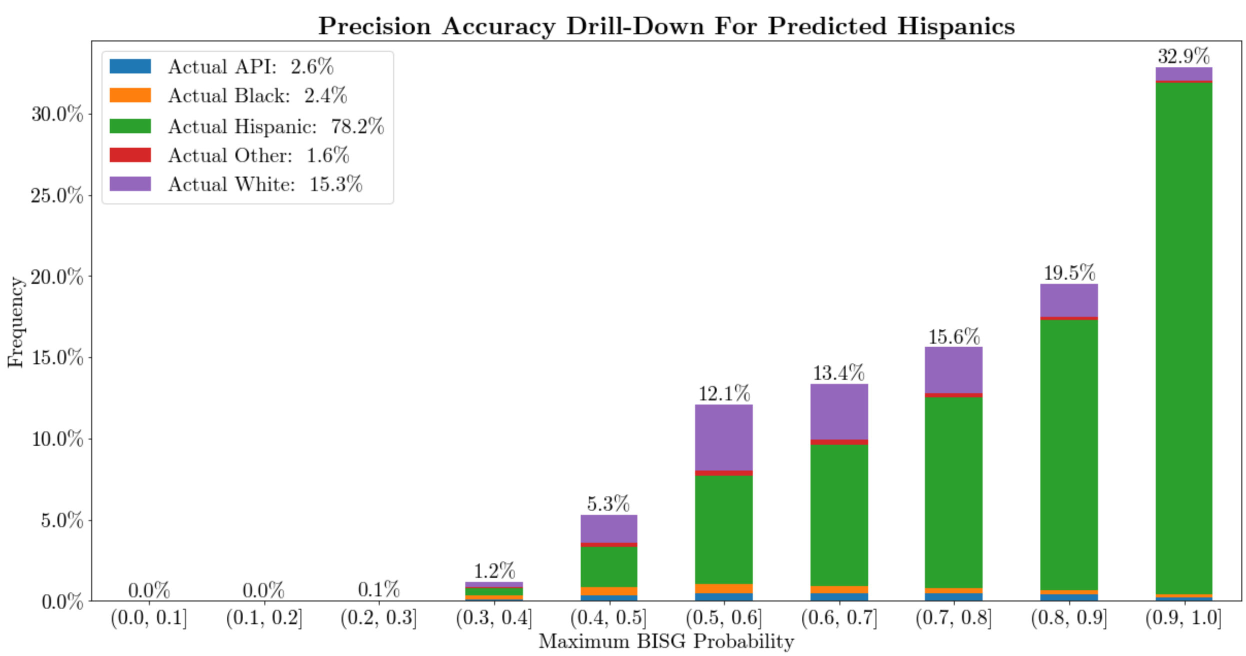


Figure 18c

Figure 19c presents the expanded comparative characteristics table that walks through the build-up of Predicted Hispanics.

Figure 19c: Comparative Characteristics of Actual vs. Predicted Hispanics

Hispanics	Total Actuals	–	False Negatives	=	True Positives	+	False Positives	=	Total Predicted
Average CBG White %	52.5%		49.8%		53.2%		57.8%		54.2%
Average CBG Black %	11.7%		11.0%		11.8%		11.1%		11.7%
Average CBG Hispanic %	28.6%		31.1%		27.9%		24.5%		27.1%
Average CBG API %	5.6%		6.2%		5.4%		5.0%		5.3%
Average Surname Hispanic %	66.9%		22.1%		79.4%		66.8%		76.6%
Average Max Probability	78.4%		66.1%		81.8%		65.1%		78.2%
Average Median HH Income	$54,823		$52,861		$55,370		$56,561		$55,630
Sample Counts	1,589,902		346,636		1,243,266		346,750		1,590,016
% of Actual Hispanics			-21.8%		78.2%				
% of Predicted Hispanics					78.2%		21.8%		

Based on this analysis, we can see the following:

- The geographic demographics of Hispanic TPs are closely aligned to the geographic demographics for Actual Hispanics. However, Hispanic TPs are much more likely to have surnames with high Hispanic concentrations than the broader Hispanic Actual sample (79.4% Hispanic surname demographic for TPs vs. 66.9% for Actual Hispanics). Hispanic TPs represent 78.2% of Predicted Hispanics and have high average maximum probability values (81.8%).

- The Hispanic FPs represent, on average, slightly higher income individuals (+2.2% vs. TPs), primarily White, who reside in slightly more racially diverse CBGs (57.8% White vs. 53.2% White for Hispanic TPs). Although they are Non-Hispanic, they possess surnames whose Hispanic demographics are consistent, on average, with Actual Hispanics. However, as can be seen by their average maximum probability, they have higher inherent uncertainty in their predicted race / ethnicity vs. Actual Hispanics (65.1% average max vs. 78.4% for Actual Hispanics).

- Hispanic FPs are virtually equal to Hispanic FNs in size leading to a negligible difference in the estimated number of Hispanic individuals relative to Actual Hispanics (1,590,016 vs. 1,589,902). Furthermore, a comparison of average income levels between Actual and Predicted Hispanics reveals a minor 1.5% difference – indicating that the FPs largely offset the FNs from an income perspective (also see **Figures 20c and 21c** below). Average CBG demographics between Actual and Predicted Hispanics also show minor differences; however, Predicted Hispanics do have surnames that are more heavily Hispanic, on average, than Actual Hispanics (76.6% vs. 66.9% for Actual Hispanics).

Figures 20c and 21c provide a more visual comparison of the comparative income distributions.

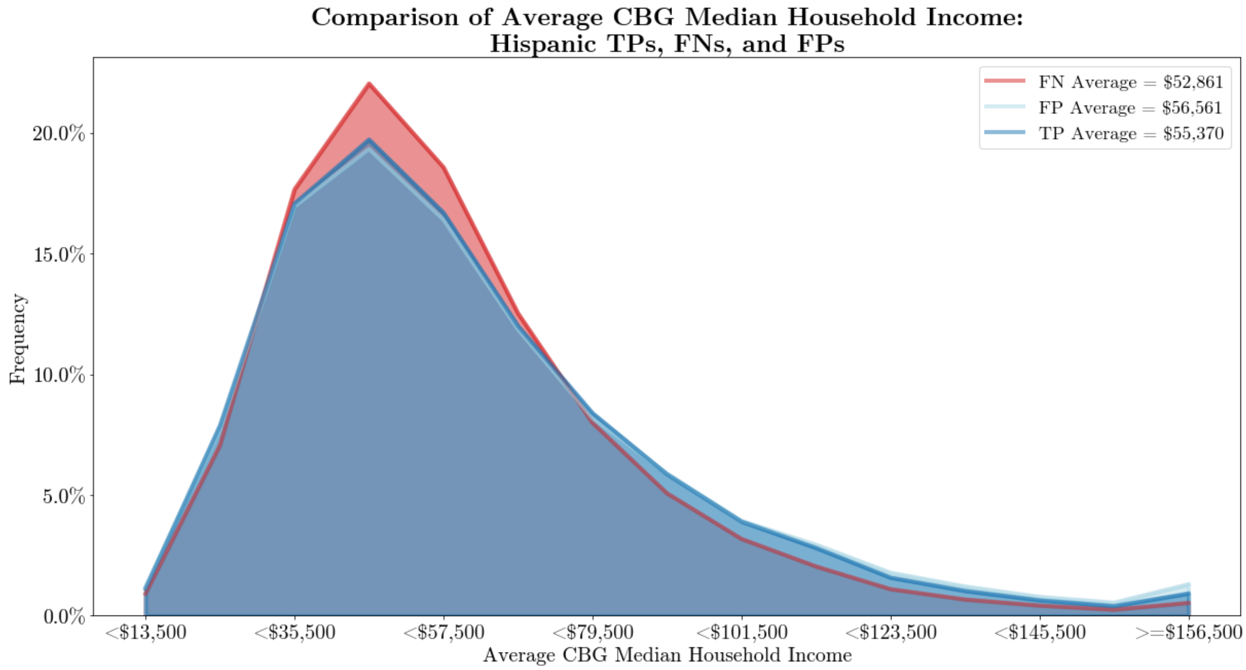


Figure 20c

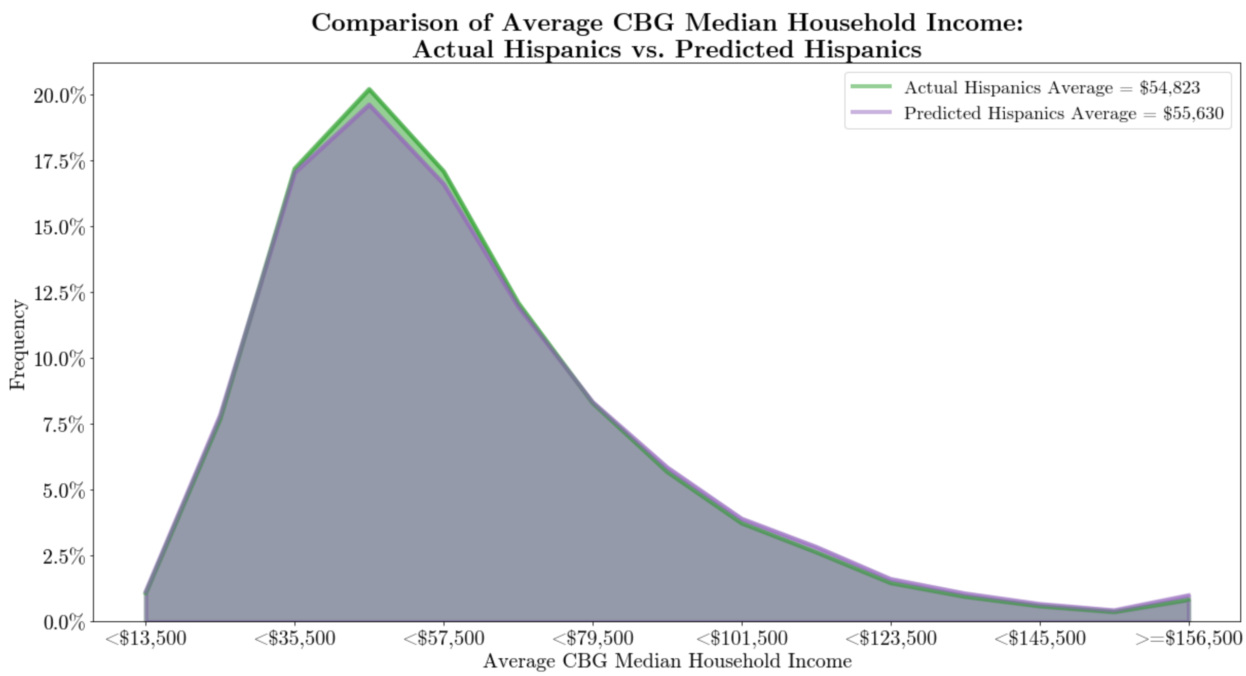


Figure 21c

For Predicted APIs in **Figure 18d**, Precision Accuracy is 77.4% and Whites comprise almost half of the API FPs – which are, as for Blacks above – rather prevalent at maximum BISG probability values less than 90%.

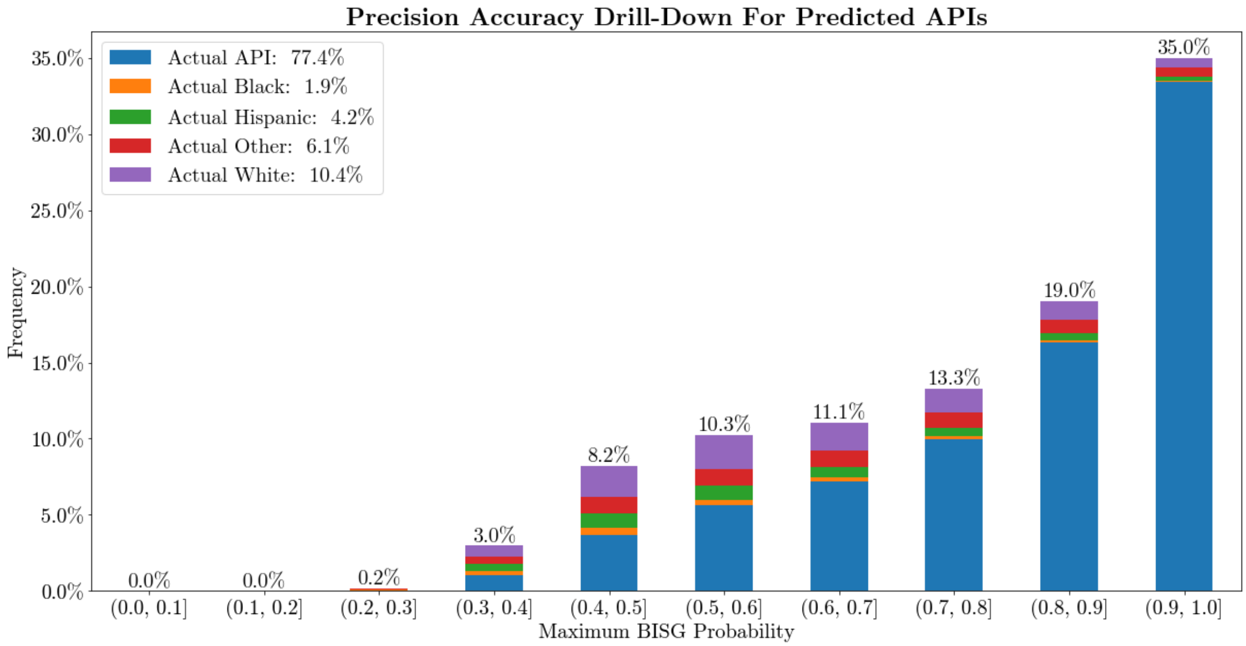


Figure 18d

Figure 19d presents the expanded comparative characteristics table that walks through the build-up of Predicted APIs.

Figure 19d: Comparative Characteristics of Actual vs. Predicted APIs

APIs	Total Actuals	–	False Negatives	=	True Positives	+	False Positives	=	Total Predicted
Average CBG White %	59.7%		53.4%		62.9%		62.1%		62.7%
Average CBG Black %	9.8%		9.1%		10.1%		10.8%		10.3%
Average CBG Hispanic %	15.5%		16.6%		14.9%		14.2%		14.8%
Average CBG API %	13.1%		18.7%		10.2%		11.1%		10.4%
Average Surname API %	59.4%		11.7%		83.7%		68.6%		80.3%
Average Max Probability	76.7%		66.7%		81.8%		62.3%		77.4%
Average Median HH Income	$65,667		$68,038		$64,455		$58,900		$63,197
Sample Counts	414,671		140,197		274,474		80,365		354,839
% of Actual APIs			-33.8%		66.2%				
% of Predicted APIs					77.4%		22.6%		

Based on this analysis, we can see the following:

- The geographic demographics of API TPs are generally aligned to the geographic demographics for Actual APIs – albeit slightly more White. However, API TPs are much more likely to have surnames with high API concentrations than the broader API Actual sample (83.7% API surname demographic for TPs vs. 59.4% for Actual APIs). API TPs comprise 77.4% of the Predicted API group and have high average maximum probability values (81.8%).

- The API FPs represent, on average, somewhat lower income individuals (-8.6% vs. TPs), primarily White, who reside in CBGs with similar demographics to TPs, but possess surnames whose API demographics are higher, on average, than Actual APIs (68.6% vs. 59.4% for Actual APIs).

- API FPs are about 40% smaller in size than API FNs leading to a 14% underestimation of the number of API individuals relative to Actual APIs (354,839 vs. 414,671). Furthermore, a comparison of average income levels between Actual and Predicted APIs reveals a minor 3.8% difference – reflecting the loss of the higher income FNs and the gain of the lower income FPs (also see **Figures 20d and 21d** below). Average CBG demographics between Actual and Predicted APIs also show minor differences; however, Predicted APIs do have surnames that are more heavily API, on average, than Actual APIs (80.3% vs. 59.4% for Actual APIs).

Figures 20d and 21d provide a more visual comparison of the comparative income distributions.

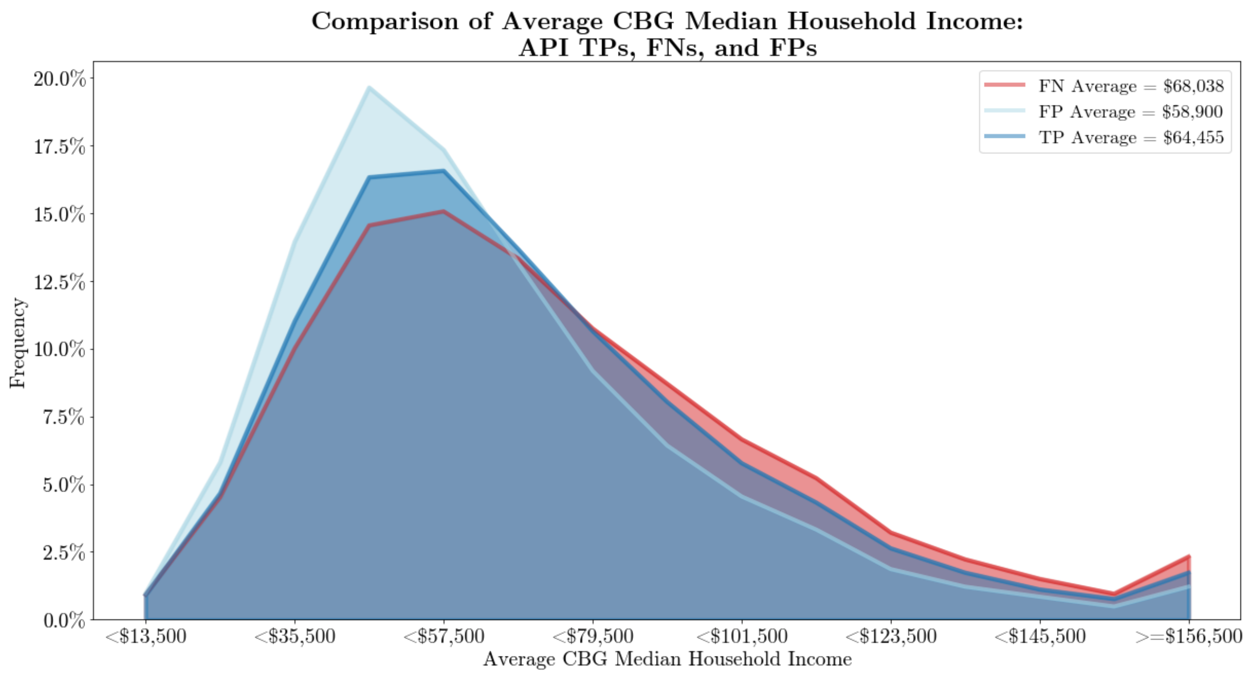


Figure 20d

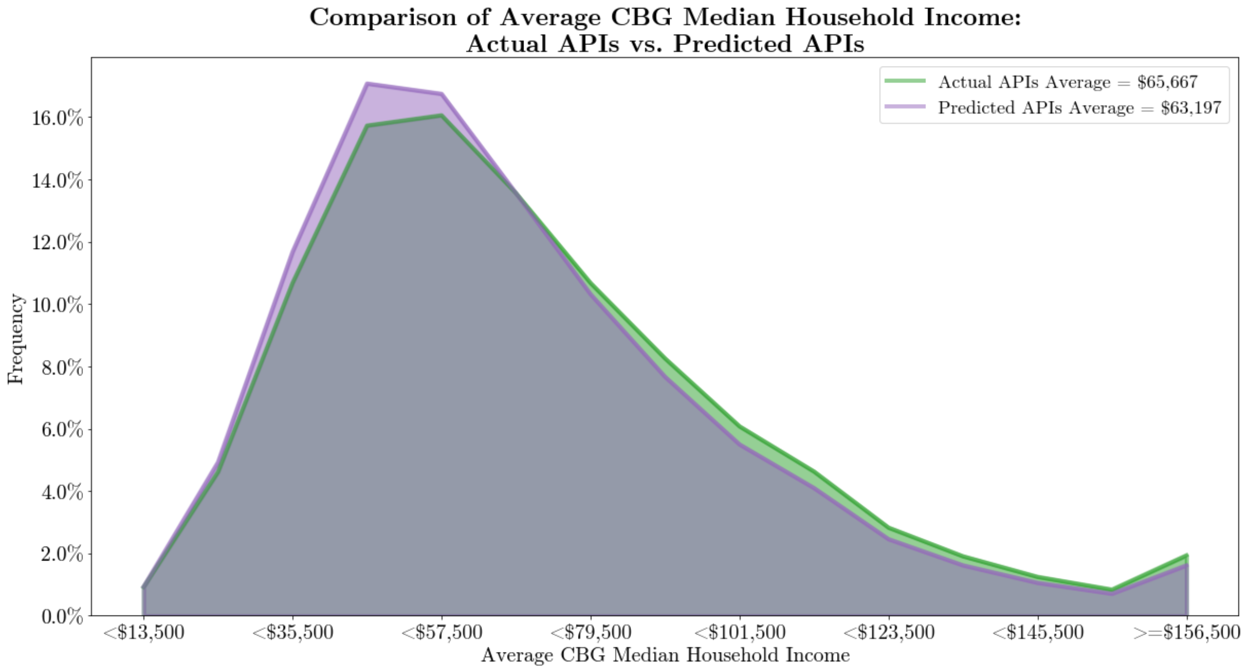
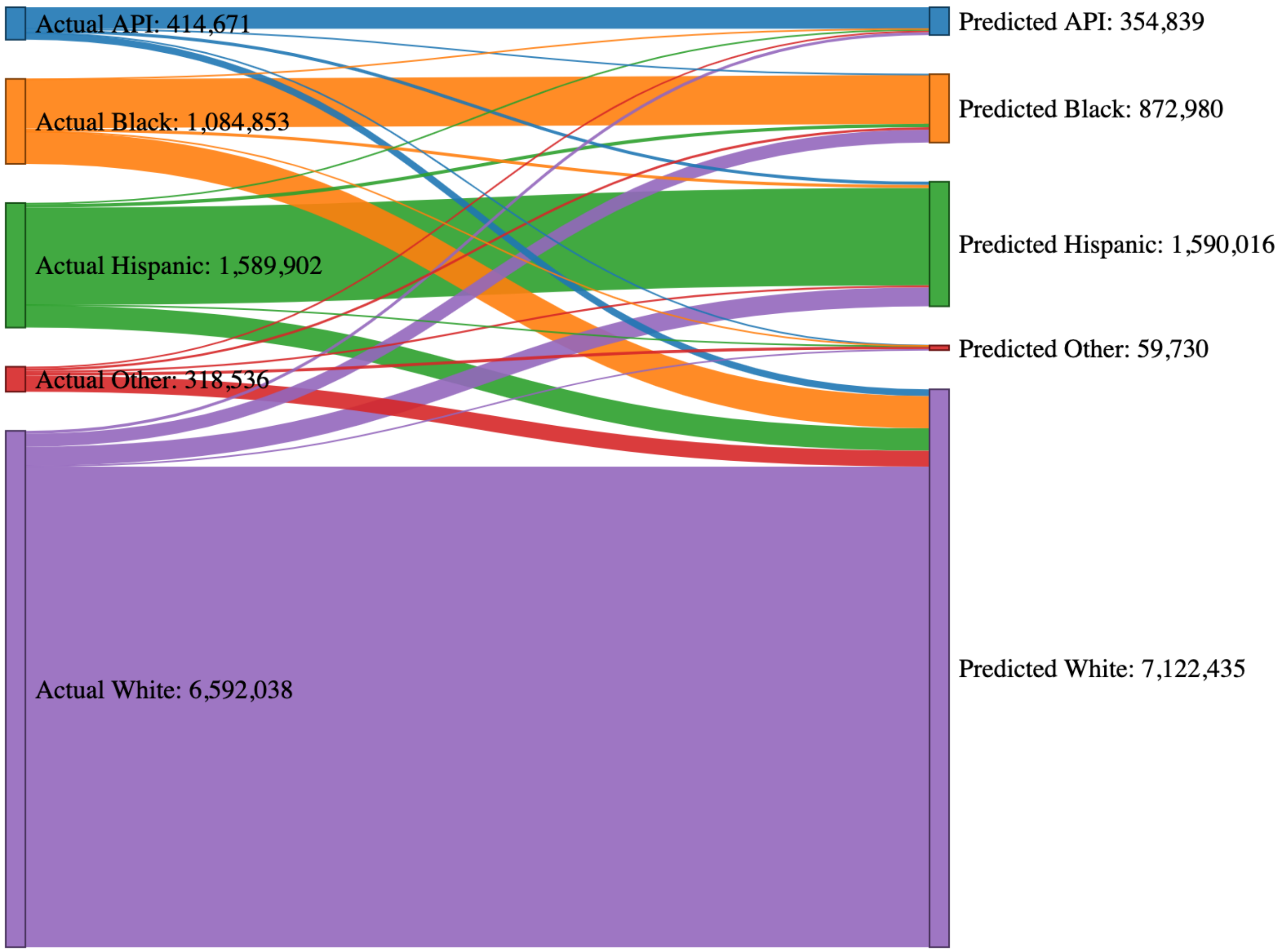


Figure 21d

Figure 22 below uses a Sankey chart to bring the False Negative and False Positive discussions together by illustrating how individual classifications using the BISG Max classification rule changes the composition of predicted race / ethnicity groups (the right-hand side) from their actual levels (the left-hand side). These changes reflect the influences of TPs (flows to and from the same color), FNs (flows from one color to another), and FPs (flows to one color from another).

Figure 22: Sankey Chart of TP, FP, and FN Flows Across Groups Under BISG Max Classification Rule



Based on our analysis of the FNs and FPs across these demographic groups, we note the following overall risks associated with BISG proxy model usage for individual-level race / ethnicity prediction:

- **About one-third of the overall geo-surname sample has maximum BISG probability values less than 80% which introduces elevated predictive uncertainty to a relatively large segment of individual classifications**. This degree of uncertainty is particularly elevated for Blacks (61% with maximum probability values < 80%), Hispanics (47%), and APIs (48%), and generally increases with the degree of racial / ethnic diversity in the CBGs comprising the sample.[39] To the extent that the upcoming 2020 Census data indicate greater systemic racial / ethnic diversity geographically and/or by surname, this risk will only increase.

[39] Essentially, racial / ethnic proxy models work best in geographies with high segregation. In the presence of cultural diversity, these methodologies struggle in predicting an individual's race / ethnicity unless their surname is highly racially/ ethnically segregated.

- 42% of Actual Blacks are misclassified as FNs versus only 7% for Whites – and further analysis of these Black FNs indicate they have significantly higher average CBG median incomes (+32%) than the remaining Black TPs. As the Sankey chart above illustrates, the majority of these Black FNs flow into Predicted Whites. Additionally, the majority of Black FPs come from Actual Whites, where they represent some of the lower-income group members (i.e., 16% lower average CBG median income than the White FNs as a whole). The Black FP's average income is also 18% lower than the Blacks FNs they are replacing. **Overall, the Predicted Black group is about 20% smaller than Actual Blacks and has average CBG median income that is 10% lower than Actual Blacks. To the extent that disparate impact policies or practices are correlated with income levels, and the testing for such disparities fails to control this bias in group composition, estimated fair lending disparities will also be potentially biased**. This risk will be explored more quantitatively in a subsequent section.

- In addition to underestimating the aggregate number of Predicted Blacks and APIs, **individual-level prediction causes heterogeneity to arise in all predicted groups with 14 – 28% of predicted group members being False Positives (this can be seen visually in the Sankey chart). These FPs can confound the estimation of lending outcome disparities driven primarily by group membership (i.e., overt discrimination or disparate treatment) as the measured magnitude of such disparities becomes diluted by the diffusion of the FPs (and their corresponding lending outcomes) across race / ethnic groups**. This risk will also be explored more quantitatively in a subsequent section.

Individual-Level Accuracy: Alternative Classification Rules

In prior sections, we assessed the individual-level accuracy of the BISG proxy model when paired with a traditional machine-learning classification rule in which each individual is assigned the race / ethnicity associated with its highest BISG probability value (the "BISG Max" rule). We note, however, that for the purposes of fair lending compliance risk management, many U.S. lenders also employ alternative classification rules in which each individual is assigned to the race / ethnicity group corresponding to the BISG probability that meets or exceeds a specific fixed threshold such as 50% or 80%. If no probability meets this criterion, then the individual's race / ethnicity is designated as "Unknown" and such individuals are removed from further fair lending analysis. While such exclusions are an undesirable side-effect of the "fixed-threshold" classification rules, compliance analysts tend to have greater confidence in the proxy demographics of the remaining "addressable sample" as they are based on individuals who have attained a minimum BISG probability level (e.g., 80%) – a feature that is not present for the BISG Max classification rule.[40]

As an example of how the fixed threshold classification rule operates, consider an individual with BISG

[40] As we saw with the BISG Max classification rule, an individual can be classified into a specific race / ethnicity group even with a maximum probability value less than 50%. Accordingly, these two fixed threshold rules generate individual-level classifications that are essentially subsets of the overall set of BISG Max classifications.

probabilities of 75% White, 15% Black, 2% API, 5% Hispanic, and 3% Other. Using fixed thresholds of 80% and 50%, we would obtain the following classification results:

- **BISG 80% Threshold** – since no probability meets or exceeds the 80% threshold, the individual's race / ethnicity would be designated as "Unknown" and excluded from further fair lending analysis.

- **BISG 50% Threshold** – the individual would be classified as White since the BISG White probability (75%) exceeds the 50% threshold.

From a purely methodological perspective, we note that the BISG 80% Threshold is the more stringent classification rule and results in a larger number of "Unknown" race / ethnicity classifications – thereby creating a smaller addressable sample than the BISG 50% Threshold. However, of those individuals for which the classification rule does produce a specific race /ethnicity designation, we have more confidence in their accuracy since they are based on a minimum probability of 80%. Alternatively, the BISG 50% Threshold reduces the number of "Unknown" race / ethnicity designations – thereby achieving a greater addressable sample. However, it does this by assigning specific race / ethnicity designations for which we are less confident (i.e., with probabilities between 50% and 80%).

Figure 23 summarizes the individual-level accuracy metrics of these two fixed-threshold classification rules and compares these results to those obtained and discussed previously using the BISG Max classification rule.[41]

Figure 23: Individual-Level Accuracy of BISG 80% and 50% Fixed Threshold Classification Rules

Recall Accuracy Rates: Alternative Classification Rules

Classification Rule	API	Black	Hispanic	White	Total
BISG Max	66.2%	57.6%	78.2%	93.1%	73.8%
BISG 50% Threshold (Excl Unknowns)	70.2%	59.2%	80.9%	94.2%	76.1%
BISG 80% Threshold (Excl Unknowns)	81.6%	70.5%	90.0%	98.6%	85.2%
BISG 50% Threshold (Incl Unknowns)	62.1%	52.9%	75.3%	91.5%	70.4%
BISG 80% Threshold (Incl Unknowns)	42.5%	27.2%	48.1%	75.2%	48.3%

Precision Accuracy Rates: Alternative Classification Rules

Classification Rule	API	Black	Hispanic	White	Total
BISG Max	77.4%	71.6%	78.2%	86.1%	78.3%
BISG 50% Threshold (Excl Unknowns)	81.8%	75.6%	80.5%	87.5%	81.4%
BISG 80% Threshold (Excl Unknowns)	92.0%	91.2%	91.9%	93.8%	92.2%
BISG 50% Threshold (Incl Unknowns)	81.8%	75.6%	80.5%	87.5%	81.4%
BISG 80% Threshold (Incl Unknowns)	92.0%	91.2%	91.9%	93.8%	92.2%

[41] These analyses focus only on the primary race / ethnicity groups typically included in fair lending testing – API, Black, Hispanic, and White. State-level tables are provided in **Appendix E**.

F1 Accuracy Rates: Alternative Classification Rules

Classification Rule	API	Black	Hispanic	White	Total
BISG Max	71.3%	63.8%	78.2%	89.5%	75.7%
BISG 50% Threshold (Excl Unknowns)	75.5%	66.4%	80.7%	90.7%	78.3%
BISG 80% Threshold (Excl Unknowns)	86.5%	79.5%	90.9%	96.2%	88.3%
BISG 50% Threshold (Incl Unknowns)	70.6%	62.2%	77.8%	89.5%	75.0%
BISG 80% Threshold (Incl Unknowns)	58.2%	41.8%	63.2%	83.5%	61.7%

Addressable Sample %

Classification Rule	API	Black	Hispanic	White	Total
BISG Max	100.0%	100.0%	100.0%	100.0%	100.0%
BISG 50% Threshold	88.5%	89.4%	93.1%	97.2%	95.3%
BISG 80% Threshold	52.1%	38.5%	53.5%	76.3%	67.3%

Overall, the BISG 80% Threshold rule excludes a third of the sample from fair lending testing analyses – a significant reduction relative to the BISG 50% Threshold rule which only excludes about 5% of the sample, and the BISG Max classification rule that excludes none (see the "Total" column in the "Addressable Sample %" table). Even more concerning is the disproportionate effect these exclusions have on the addressable samples of traditional minority groups. **Specifically, the BISG 80% Threshold rule excludes 61.5% of Actual Blacks, 47.9% of Actual APIs, and 46.5% of Actual Hispanics versus only 23.7% of Whites**. Alternatively, the BISG 50% Threshold rule excludes 11.5% of APIs, 10.6% of Blacks, and 6.9% of Hispanics versus 2.8% of Whites. While still disproportional, the addressable sample sizes are much higher and the relative disparities are much smaller.

By all accuracy measures, the BISG 80% Threshold rule (excluding Unknowns) yields the highest accuracy of all three classification rules across all three accuracy measures (i.e., Recall, Precision, and F1). However, these higher accuracies arise from the rule's focus on a much smaller subset of individuals for which we have greater confidence in the race / ethnicity predictions (i.e., 80% or greater probability) – thereby creating a very important trade-off between minority inclusion and proxy accuracy.[42] In fact, if we include the Unknowns with the other FNs in order to calculate Recall Accuracy rates on a consistent basis across the alternative classification rules, we clearly see that overall Recall Accuracy drops significantly for the BISG 80% Threshold rule and more modestly for the BISG 50% Threshold rule (the last two rows in the first table labeled (Incl Unknowns)) – with the BISG Max classification rule now exhibiting the best performance across the three (75.7% F1 Accuracy vs. 75.0% for BISG 50% Threshold and 61.7% for BISG 80% Threshold) when evaluated on the same samples.[43]

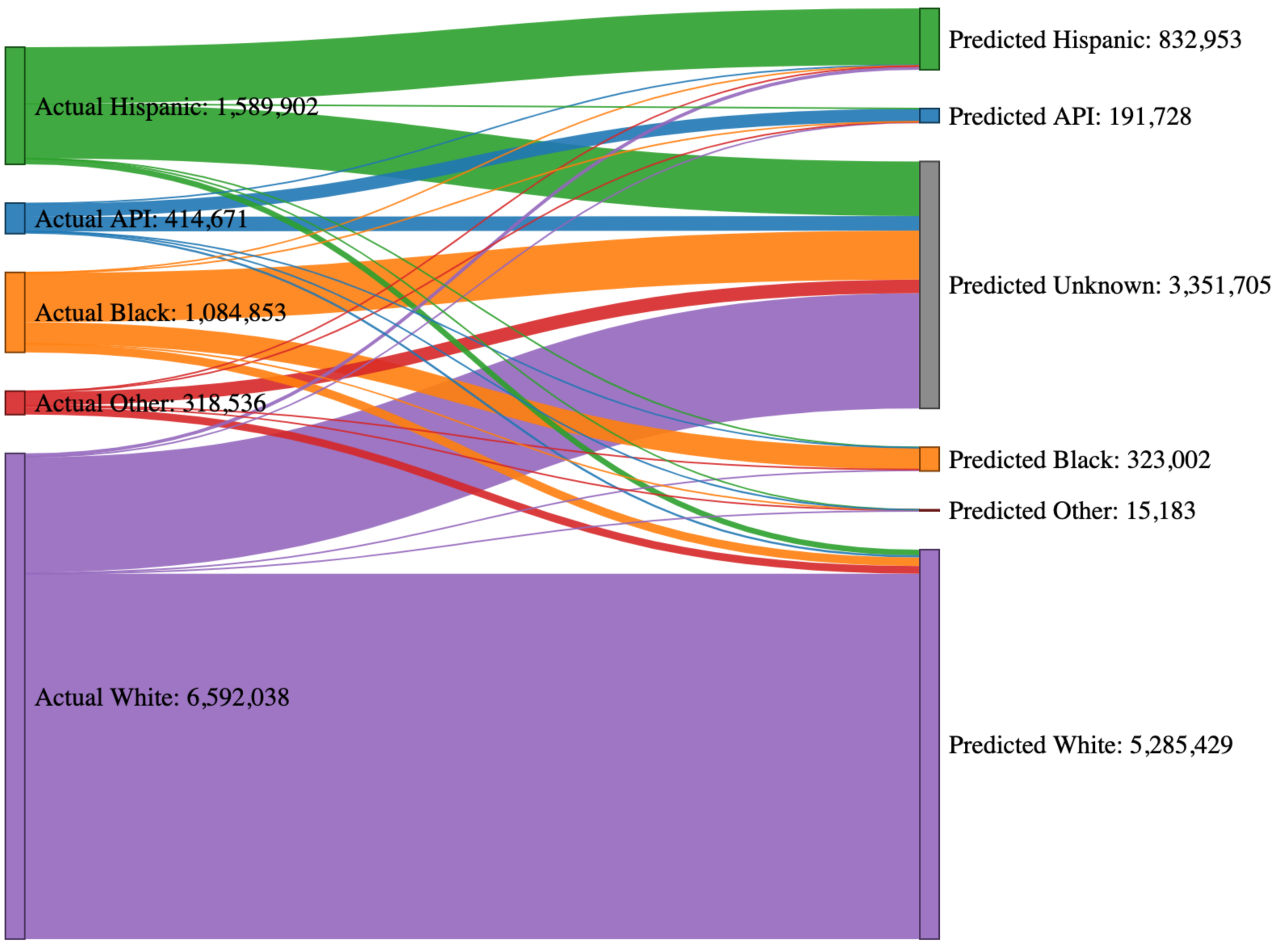
Figure 24 below uses a Sankey chart to illustrate how individual classifications using the BISG 80%

[42] We will also see later in this section that, while more accurate, this small subset of Actual Blacks is significantly biased from a socio-economic perspective – thereby creating additional risks for downstream fair lending testing.

[43] Precision Accuracy rates are unaffected by the Unknown exclusions since they focus only on TPs and FPs.

Threshold classification rule change the composition of predicted race / ethnicity groups (the right-hand side) from their actual levels (the left-hand side). These changes reflect the influences of TPs (flows to and from the same color), FNs (flows from one color to another), and FPs (flows to one color from another).

Figure 24: Sankey Chart of TP, FP, and FN Flows Across Groups Under BISG 80% Threshold Classification Rule



There are three notable features of this chart.

- The significant absorption of Actuals (33.5%) by the Predicted Unknown category due to individuals with maximum BISG probabilities less than 80%.[44] These individuals are typically excluded from downstream fair lending testing.

- The significant reduction in the size of predicted groups relative to original actual group counts.

[44] This percentage is slightly larger than that reflected in the Addressable Sample % table in **Figure 23** due to the inclusion here of the Other race / ethnicity category.

For example, Predicted Blacks are only 30% of the size of Actual Blacks.

- The previous section showed that, under the BISG Max classification rule, all FNs were allocated to other race / ethnicity groups as FPs and, therefore, (1) were still included (albeit with incorrect classifications) in downstream fair lending analyses, and (2) impacted the average characteristics (such as average income) of the predicted groups they entered as FPs. However, under the BISG 80% Threshold rule, almost 90% of FNs are excluded as Unknowns – **meaning that the final predicted groups are largely characterized by the TP segment alone**.

Figure 25 below provides further detail on the transition from actual group members to predicted group members under the BISG 80% Threshold rule to explore these features further.

Figure 25a: Comparative Characteristics of Actual vs. Predicted Whites
BISG 80% Threshold Rule

Whites	Total Actuals	–	False Negatives	=	True Positives	+	False Positives	=	Total Predicted
Average CBG White %	76.3%		56.2%		82.9%		74.9%		82.4%
Average CBG Black %	7.3%		14.4%		5.0%		7.0%		5.1%
Average CBG Hispanic %	10.3%		20.2%		7.1%		10.9%		7.3%
Average CBG API %	4.4%		7.1%		3.5%		5.2%		3.6%
Average Surname White %	76.4%		57.6%		82.6%		77.9%		82.3%
Average Max Probability	86.9%		65.2%		94.1%		89.4%		93.8%
Average Median HH Income	$61,331		$54,898		$63,442		$62,820		$63,404
Sample Counts	6,592,038		1,633,294		4,958,744		326,685		5,285,429
% of Actual Whites			-24.8%		75.2%				
% of Predicted Whites					93.8%		6.2%		

For Whites, the TPs become a more extreme version of the White TPs observed under the BISG Max classification rule – with 20% fewer Actual Whites who reside in more segregated CBGs (82.9% White vs. 77.9% under BISG Max) and possessing slightly higher average CBG median incomes ($63,442 vs. $61,899). FNs now represent almost one-quarter of Actual Whites (vs. 6.9% under BISG Max) while FPs are reduced by two-thirds (326,685 vs. 987,356 under BISG Max). Overall, Predicted Whites are undercounted by 20% (vs. an 8% overcount under BISG Max), primarily come from very high White CBGs, and have average CBG median incomes 3.4% higher than Actual Whites (they were virtually the same under BISG Max).

Figure 25b: Comparative Characteristics of Actual vs. Predicted Blacks
BISG 80% Threshold Rule

Blacks	Total Actuals	–	False Negatives	=	True Positives	+	False Positives	=	Total Predicted
Average CBG White %	38.8%		48.0%		14.1%		18.7%		14.5%
Average CBG Black %	38.7%		26.4%		71.5%		62.5%		70.7%
Average CBG Hispanic %	16.3%		18.2%		11.1%		14.7%		11.4%
Average CBG API %	4.6%		5.6%		1.9%		2.6%		2.0%
Average Surname Black %	27.1%		24.7%		33.5%		31.7%		33.3%
Average Max Probability	72.6%		65.6%		91.6%		87.4%		91.2%
Average Median HH Income	$47,221		$50,575		$38,227		$39,866		$38,371
Sample Counts	1,084,853		790,263		294,590		28,412		323,002
% of Actual Blacks			-72.8%		27.2%				
% of Predicted Blacks					91.2%		8.8%		

For Blacks, the TPs become a more extreme version of the Black TPs observed under the BISG Max classification rule – with 53% fewer Actual Blacks who reside in much more segregated CBGs (71.5% Black vs. 54.0% under BISG Max) and possessing moderately lower (-8%) average CBG median incomes ($38,227 vs. $41,533). FNs now represent almost three-quarters of Actual Blacks (72.8% vs. 42.4% under BISG Max) while FPs are reduced by nearly 90% (28,412 vs. 248,266 under BISG Max). **Overall, Predicted Blacks are undercounted by 70% (vs. a 20% overcount under BISG Max), primarily come from very high Black CBGs, and have average CBG median incomes that are almost 20% lower than Actual Blacks (they were 10% lower under BISG Max).**

Figure 25c: Comparative Characteristics of Actual vs. Predicted Hispanics
BISG 80% Threshold Rule

Hispanics	Total Actuals	–	False Negatives	=	True Positives	+	False Positives	=	Total Predicted
Average CBG White %	52.5%		56.3%		48.3%		58.6%		49.1%
Average CBG Black %	11.7%		10.7%		12.7%		12.4%		12.7%
Average CBG Hispanic %	28.6%		26.0%		31.4%		20.9%		30.5%
Average CBG API %	5.6%		5.3%		6.0%		6.2%		6.0%
Average Surname Hispanic %	66.9%		49.1%		86.0%		82.4%		85.7%
Average Max Probability	78.4%		65.4%		92.3%		87.6%		91.9%
Average Median HH Income	$54,823		$55,702		$53,874		$58,171		$54,223
Sample Counts	1,589,902		824,535		765,367		67,586		832,953
% of Actual Hispanics			-51.9%		48.1%				
% of Predicted Hispanics					91.9%		8.1%		

For Hispanics, the TPs become a more skewed version of the Hispanic TPs observed under the BISG Max classification rule – with 38% fewer Actual Hispanics who reside in more segregated CBGs (31.4% Hispanic vs. 27.9% under BISG Max) and possessing slightly lower (-3%) average CBG median incomes ($53,874 vs. $55,370). FNs now represent an additional 30% of Actual Hispanics (51.9% vs. 21.8% under BISG Max) while FPs are reduced by 80% (67,856 vs. 346,750 under BISG Max). Overall, Predicted Hispanics are undercounted by nearly 50% (vs. virtually no undercount under BISG Max) and have average CBG median incomes that are virtually the same (-1%) as Actual Hispanics (they were 1.5% higher under BISG Max).

Figure 25d: Comparative Characteristics of Actual vs. Predicted APIs
BISG 80% Threshold Rule

APIs	Total Actuals	–	False Negatives	=	True Positives	+	False Positives	=	Total Predicted
Average CBG White %	59.7%		57.6%		62.5%		66.9%		62.9%
Average CBG Black %	9.8%		9.9%		9.7%		10.2%		9.7%
Average CBG Hispanic %	15.5%		15.4%		15.6%		14.8%		15.6%
Average CBG API %	13.1%		15.1%		10.4%		6.4%		10.1%
Average Surname API %	59.4%		35.9%		91.1%		89.1%		91.0%
Average Max Probability	76.7%		65.2%		92.2%		88.1%		91.9%
Average Median HH Income	$65,667		$63,837		$68,139		$64,030		$67,810
Sample Counts	414,671		238,301		176,370		15,358		191,728
% of Actual APIs			-57.5%		42.5%				
% of Predicted APIs					92.0%		8.0%		

For APIs, the TPs become a more skewed version of the API TPs observed under the BISG Max classification rule – with 36% fewer Actual APIs possessing surnames with greater API concentration (91.1% API vs. 83.7% under BISG Max) and possessing moderately higher (+6%) average CBG median incomes ($68,139 vs. $64,455). FNs now represent an additional 24% of Actual APIs (57.5% vs. 33.8% under BISG Max) while FPs are reduced by 80% (15,358 vs. 80,365 under BISG Max). Overall, Predicted APIs are undercounted by nearly 50% (vs. about 15% undercount under BISG Max) and have average CBG median incomes that are slightly higher (+3%) as Actual APIs (they were about 4% lower under BISG Max).

A corresponding set of tables is contained in **Appendix F** for the BISG 50% Threshold classification rule whose results fall in between those presented above for the BISG 80% Threshold rule and the BISG Max rule discussed previously. **Figure 26** below summarizes the impacts that the three alternative classification rules have on four key characteristics of the predicted race / ethnicity groups (relative to the corresponding actual groups).

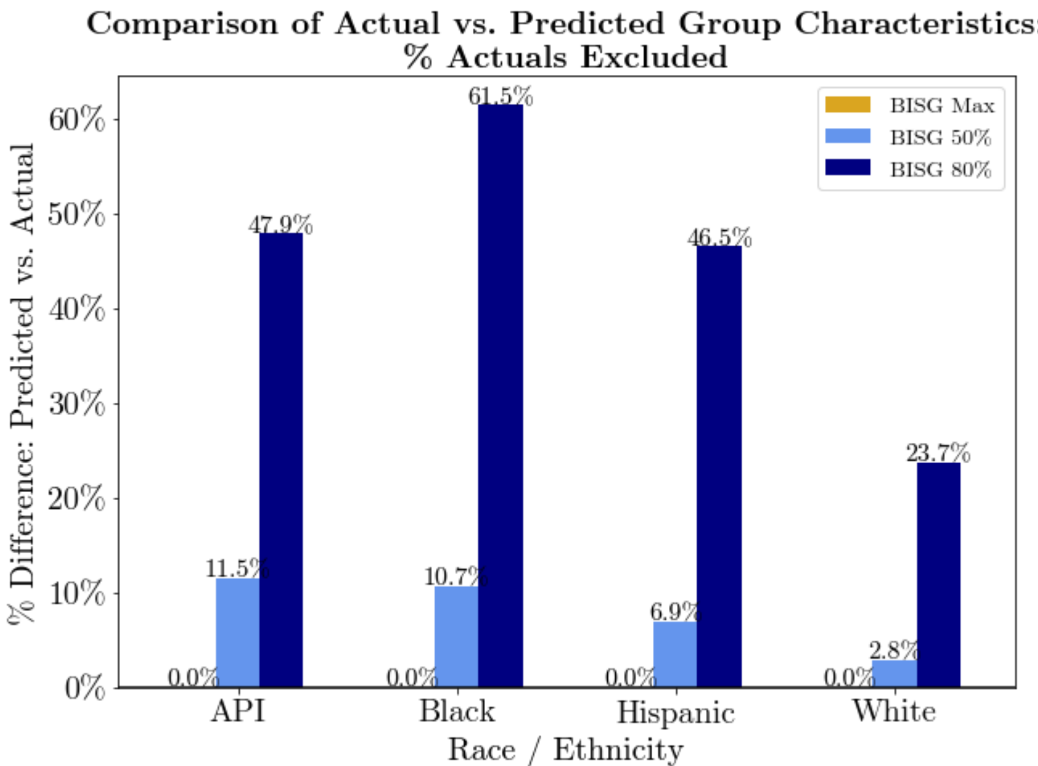


Figure 26a

Figure 26a compares the percentage of actual group members who are excluded as FNs due to their maximum probability falling below classification rule thresholds (the BISG Max results are all zero as the classification rule does not employ any minimum probability threshold). Consistent with the discussion of **Figure 23**, the BISG 80% Threshold causes a significant reduction in the addressable sample for all groups – with disproportionately greater reductions for Non-White groups. Blacks, in particular, suffer the greatest adverse effect of this rule with over 60% excluded due to maximum probabilities less than 80%. The BISG 50% Threshold has a much milder impact on addressable sample sizes although exclusions are still disproportionate.

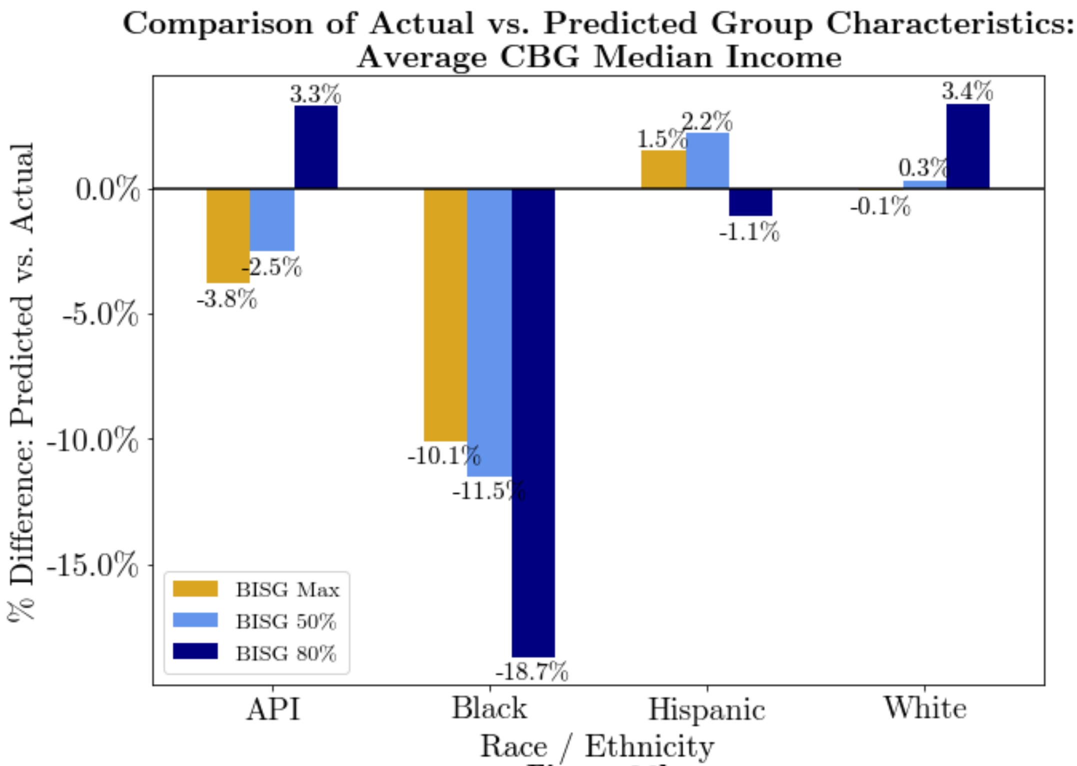


Figure 26b

Figure 26b compares the percentage change in average CBG median incomes between predicted and actual group members under the three alternative classification rules. For example, for Blacks, the BISG Max classification rule generates a Predicted Black group that has a 10.1% lower average CBG median income level than the Actual Black group – which is slightly less than the 11.5% lower average income generated by the BISG 50% Threshold rule. However, the BISG 80% Threshold rule creates a truly significant bias in average CBG median income with an -18.7% difference between Predicted and Actual Blacks.[45]

Overall, the average income bias generated by individual-level classification is greatest for Blacks under all three classification rules and consistent in direction. While the impacts of this bias will be explored more quantitatively in the next section, intuitively such biases can certainly impact measured fair lending outcome disparities if such disparities are related to average borrower income levels – regardless of race / ethnicity – and if such factors are excluded from the disparity measurement process.

[45] Not only is this caused by the higher degree of biased FNs from the BISG 80% Threshold rule, but it is also due to the significant reduction in mitigating FPs due to the outright exclusion of the Unknowns from the analysis sample – that is, unlike with BISG Max, most of the FNs are not recycled as FPs for other race / ethnicity groups.

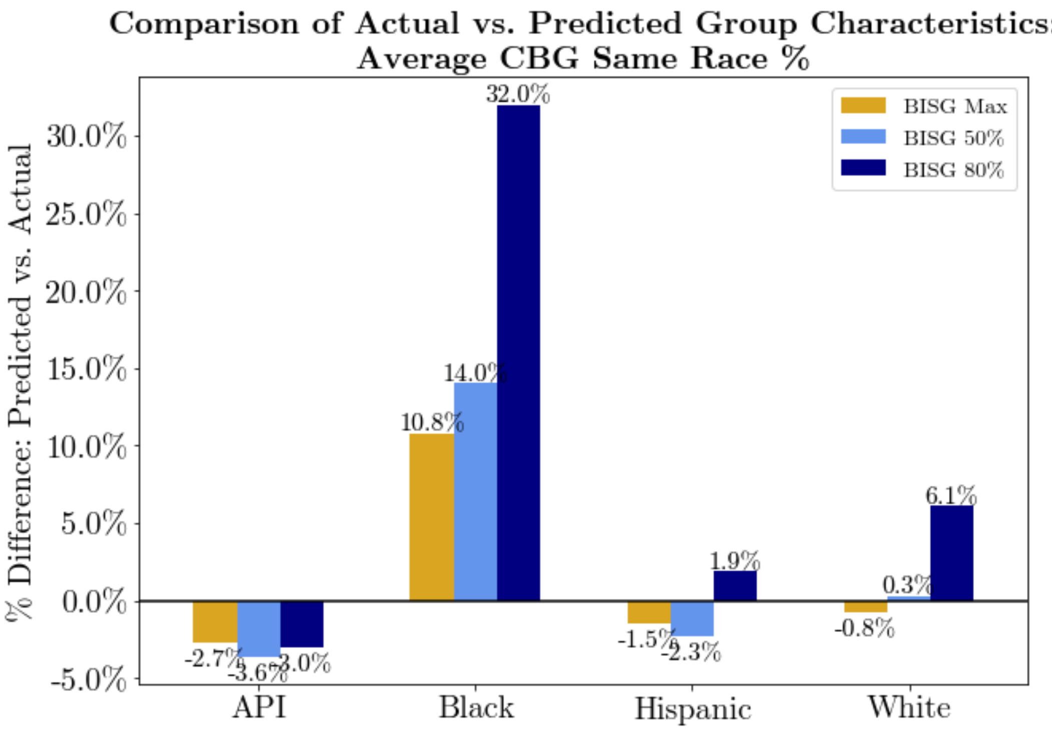


Figure 26c

Figure 26c compares the average "same" racial concentration of CBGs between predicted and actual group members under the three alternative classification rules. For example, for Blacks, the BISG Max classification rule generates a Predicted Black group that is more highly segregated – as it has a 10.8 percentage point higher average Black CBG representation than the Actual Black group. Consistent with the discussion of **Figure 19b**, this occurs due to the bias introduced by the swapping of FPs for FNs. In particular, the Black FNs under the BISG Max classification rule (42.4% of Actual Blacks), were higher-income Blacks residing in more racially-diverse CBGs. By excluding them from the Predicted Black group, the remaining Black TPs were biased towards lower-income Blacks living in more highly segregated CBGs, and this bias was not offset by the much smaller number of Black FPs.

Overall, the average CBG demographic bias generated by individual-level classification is greatest for Blacks under all three classification rules and consistent in direction. Additionally, the bias is truly significant under the BISG 80% Threshold rule where the Predicted Black group comes from very highly segregated CBGs (with significantly lower average CBG incomes as shown above in **Figure 26b**) – with average Black CBG representation that is 32 percentage points higher than the Actual

Black group (70.7% vs. 38.7% for Actual Blacks).[46]

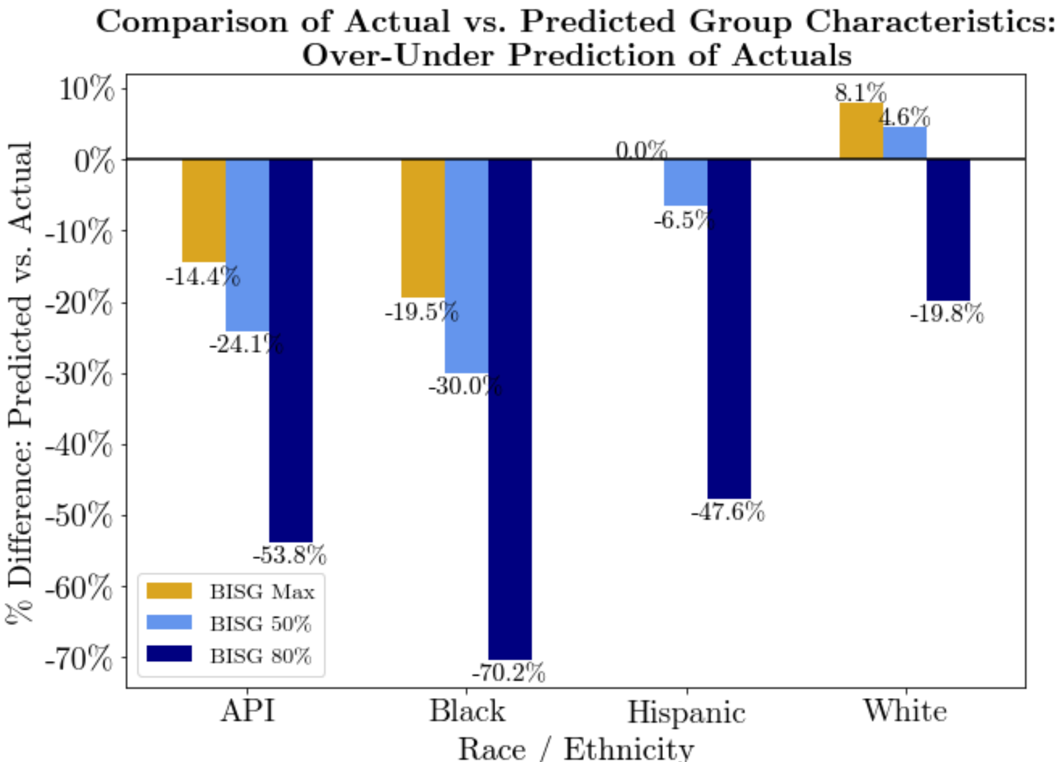


Figure 26d

Finally, **Figure 26d** compares the individual-level membership counts between predicted and actual group members under the three alternative classification rules. For example, for Blacks, the BISG Max classification rule generates a Predicted Black group that is about 20% smaller than the Actual Black group. Consistent with the discussion of **Figure 19b**, this occurs because the number of Black FNs (which are removed from the Predicted Black group) are nearly twice the size of the Black FPs (which are included in the Predicted Black group). The magnitudes of the predicted group undercounts increase with the BISG 50% and 80% Threshold rules due to: (1) the increasingly higher minimum probability thresholds creating increasingly larger FNs, and (2) unlike the BISG Max classification rule where FNs are repurposed into other groups' FPs, the BISG 50% and 80% Threshold rules completely exclude FNs that are below the minimum probability thresholds – thereby creating vastly smaller pools of FPs to mitigate the group count reductions.

In the next section, we explore the impacts of the BISG proxy probabilities and individual-level classifications on the estimation of disparate treatment and disparate impact price disparities.

[46] The bias for Whites under the BISG 80% Threshold rule, while smaller than that observed for Blacks, may potentially increase the bias in measured fair lending disparities for Blacks and Hispanics – particularly if such disparities are related to average income levels. This is because the bias toward higher average income White CBGs may cause an underestimation of the Predicted White group's average outcome measure – thereby widening outcome disparity measurements with Blacks and Hispanics.

Fair Lending Testing Impacts of the BISG Proxy Model

In the previous sections, we analyzed the BISG proxy model's ability to identify accurately the aggregate race / ethnicity distribution of a sample, as well as the specific individual race / ethnicity group membership for individual sample members. Some of the key observations from this analysis are that:

- The model provides highly accurate aggregate distributions of race / ethnicity when applied to samples whose socioeconomic characteristics are aligned with the Census data used as model inputs. Departures from this alignment through population drift, or by using samples with skewed socioeconomic characteristics, will likely generate aggregate proxy errors.

- Several fair lending compliance requirements require the specific identification of an individual's race / ethnicity. When the BISG proxy model is paired with a classification rule to produce such specific identifications, misclassification errors are produced that are non-random in nature. In particular, Blacks tend to exhibit the greatest degree of classification error, the greatest undercount of actual group members, and the most significant amount of socioeconomic bias (i.e., the Predicted Black group is biased towards lower average income Blacks and Non-Blacks).

In this section, we analyze whether these inaccuracies and potential biases have practical impacts on the use of these proxies for two types of fair lending testing scenarios – potential disparate treatment and potential disparate impact in consumer loan pricing.

Hidden Biases in Disparate Treatment Estimates

In its simplest form, disparate treatment involves one group of customers – typically members of a prohibited basis group – receiving relatively unfavorable treatment in certain aspects of a credit transaction relative to similarly-situated control group customers. Disparate treatment can be intentional and overt, or can be unintentional – but still present – based on a comparison of lending outcomes between comparable individuals.

To assess the precise impact of the BISG proxies on the estimation of disparate treatment disparities, we first need a sample of lending outcomes in which the exact form and amount of the disparate treatment is known (i.e., the "ground truth"). To this end, and to avoid overcomplicating this analysis, we simulate a simplified disparate treatment scenario in which a certain minority group (or groups) is charged a discretionary fee amount that is not charged to the corresponding White group – for example, a $100 "processing fee" is charged to Blacks, Hispanics, and / or APIs but not to Whites. Given that we know the ground truth disparate treatment effect ($100), we can then assess whether the corresponding disparate treatment estimates produced by traditional fair lending testing tools on our synthetic dataset are biased in any way by our use of BISG-based proxies.

The specific steps in our analysis are as follows:

- Simulate this disparate treatment scenario on the synthetic dataset using the actual members of each group and the scenario's corresponding fee amounts. For example, assign all Actual Blacks a discretionary fee amount of $100 and assign all other actual group members a discretionary fee amount of $0.

- Estimate the average fee amounts for each BISG-proxied group and calculate the average fee disparity amount for each minority group versus the White group (i.e., the mathematical difference between the estimated average fee for the minority group and the estimated average fee for the White group). Perform this estimation using the BISG Continuous approach – as well as for each of the BISG Classification approaches (i.e., BISG Max, BISG 50% Threshold, and BISG 80% Threshold).

- Compare the estimated average fee disparity amounts to the actual "ground truth" disparity amounts and calculate potential estimation biases.

Figure 27 presents the results of our disparate treatment scenario where Actual Blacks are charged a $100 processing fee and where Actual Whites (and all other minority group members) are charged $0.

Figure 27: Disparate Treatment Scenario Test Results

Scenario: Blacks = $100, All Others = $0

Predicted Race / Ethnicity	BISG 80%	BISG 50%	BISG Max	BISG Continuous
Average Fee $ Amount				
API	$0.47	$1.26	$1.87	
Black	$91.20	$75.60	$71.56	
Hispanic	$0.76	$1.82	$2.44	
White	$2.20	$5.27	$5.78	
Average Fee $ Disparity vs. Whites				
API	-$1.73	-$4.01	-$3.91	-$0.02
Black	$89.01	$70.33	$65.78	$99.97
Hispanic	-$1.44	-$3.45	-$3.35	$0.00
Average Fee Disparity Bias (%)				
API				0.0%
Black	-11.0%	-29.7%	-34.2%	0.0%
Hispanic				0.0%

The first column of this table corresponds to the predicted group members being tested, the second through fourth columns correspond to the fair lending testing results under each of the three classification rules used to predict the individual-level race / ethnicity of each group member, and the last column corresponds to the fair lending testing results using the raw BISG probabilities instead of the individual-level classifications (i.e., the "BISG Continuous" approach). The first set of four rows ("Average Fee $ Amount") contains the estimated average fee amounts of each predicted race / ethnicity group under the three individual-level race / ethnicity prediction methods.[47] The second set of three rows ("Average Fee $ Disparity vs. Whites) calculates the corresponding <u>fair lending disparity</u> – that is, the difference in each minority group's estimated average fee amount vs. the estimated average fee amount for Whites. Finally, the last three rows ("Average Fee Disparity %") converts the estimated fair lending disparity amount into a percentage for the minority group(s) subject to the disparate treatment.

Starting with the first set of rows, we can see that **all three individual classification rules underestimate the $100 fee amount for the Predicted Black group and overestimate the $0 fee amount for all other predicted group members – with Predicted Whites exhibiting the greatest overestimation**. Overall, the BISG 80% Threshold classification rule actually produces average fee amount estimates that are closest to actuals with errors of -11% for Blacks, +$2.20 for Whites, and +$0.47-$0.76 for APIs and Hispanics[48] – even though this classification rule has the smallest addressable sample due to its exclusion of all actual group members with maximum BISG probability values less than 80%. Alternatively, somewhat counterintuitively, the most inclusive individual classification rule – BISG Max – produces estimated average fee amounts that are the most biased with error rates of -34% for Blacks, +$5.78 for Whites, and +$1.87-$2.44 for APIs and Hispanics. The results for the BISG 50% Threshold classification rule fall in between these two.

What is the intuition behind these biases?

Let's start with the fair lending testing results under the BISG 80% Threshold classification rule. For the Predicted Black group, we obtain an average fee amount of $91.20 – which is $8.80 less than the $100 true average fee amount. If we revisit **Figure 25b**, we see that this classification rule excludes 72.8% of Actual Blacks (790,263) which, under this disparate treatment scenario, have an average fee amount of $100. The remaining 294,590 Actual Blacks represent the Black TPs which also have an average fee amount of $100. Therefore, **under pure disparate treatment, regardless of the size of the Black FNs, there is no bias introduced to the average fee amount of the Predicted Black group from excluding the FNs.[49] Therefore, the only bias that occurs is through**

[47] The BISG Continuous method produces measures of relative fee disparity amounts (the second group of rows) rather than the absolute average fee amount for each group.

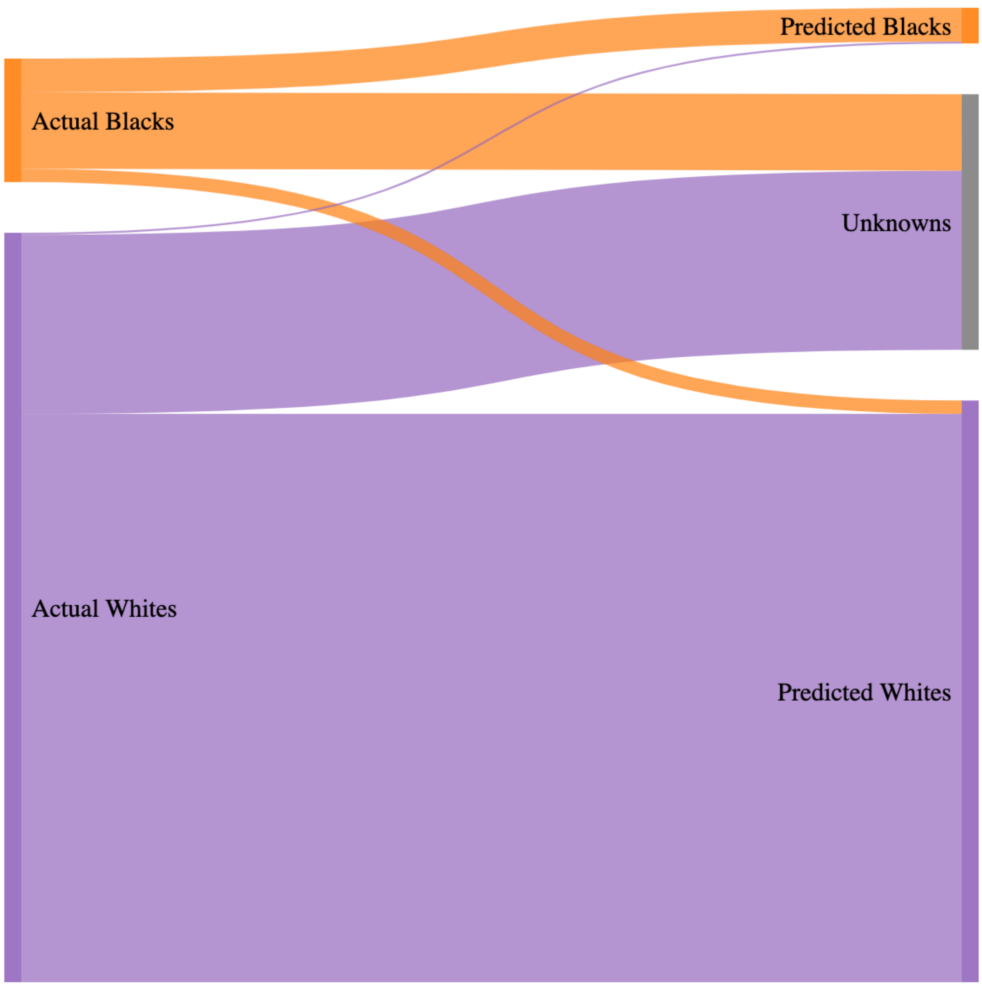
[48] Percentage errors cannot be calculated for Whites, APIs, and Hispanics since the actual fee amount is $0.

[49] This result also holds if the processing fee charged to each Black customer varied randomly around an average. The key is that the borrower's race / ethnicity is the driving factor of the disparate outcome – which is what disparate treatment means. In the next section, we explore potential biases when the disparity in outcomes in driven not by race / ethnicity directly, but by a policy that is correlated with race / ethnicity. As we will see there, since the FNs are not random, they also contribute to fair lending measurement bias.

the Black FPs – that is, the non-Black individuals with $0 fee amounts who are misclassified as Black.

Figure 28 below presents a Sankey Chart illustrating the FP and FN flows for Actual Blacks and Actual Whites under the BISG 80% Threshold rule (we abstract from the other race / ethnicity groups for visual clarity).

Figure 28: FPs and FN Misclassifications Under BISG 80% Threshold



Here we see that a very small number of White FNs (the very thin purple line) flow into the Predicted Black group as FPs – ultimately representing 5.2% of total Predicted Blacks. Alternatively, we see a moderate number of Black FNs (the narrow orange line) flow into the Predicted White group as FPs – representing 2.2% of total Predicted Whites. Overall, as this chart demonstrates for Whites and Blacks, the BISG 80% Threshold rule is characterized by relatively low "cross-contamination" of each Predicted group by FPs as: (1) one group's FPs are another group's FNs, and (2) most FNs are excluded as "Unknowns" under the BISG 80% Threshold rule.[50] This can be clearly seen by the large grey area on the right side of **Figure 28** that absorbs the vast majority of the Black and White FNs.[51]

Additional intuition of the bias measurements in **Figure 27** arises from a decomposition of the estimated average fee amounts for each predicted group.

Revisiting **Figure 25b,** we see that Black FPs represent 8.8% of Predicted Blacks with Black TPs comprising the remaining 91.2% of Predicted Blacks.[52] Simple math proves that the bias from the Black FPs is the cause of the \$8.80 underestimation of the Black average fee amount:

$$\textit{Estimated Average Black Fee Amount} = \underbrace{91.2\% * \$100}_{\text{Black TPs}} + \underbrace{8.8\% * \$0}_{\text{Black FPs}} = \$91.20$$

Similarly, for Whites, we can prove that the bias from the White FPs (more specifically, the Actual Blacks misclassified as Whites) is the cause of the \$2.20 overestimation of the White average fee amount:

$$\textit{Estimated Average White Fee Amount} = \underbrace{93.8\% * \$0}_{\text{White TPs}} + \underbrace{4.0\% * \$0}_{\substack{\text{White FPs} \\ \text{(Non-Black)}}} + \underbrace{2.2\% * \$100}_{\substack{\text{White FPs} \\ \text{(Black)}}} = \$2.20$$

Why does the BISG Max classification rule create more bias than the Threshold classification rules?

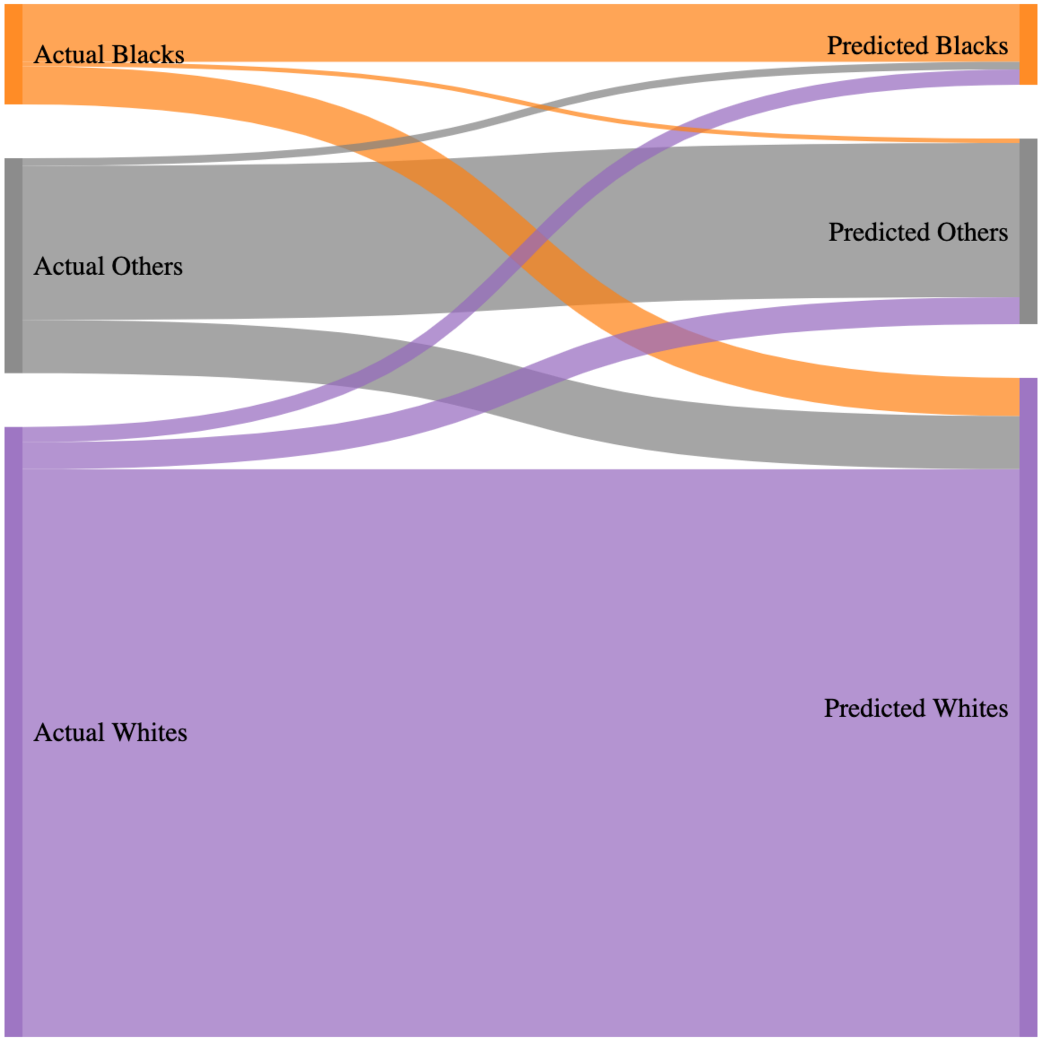
As we saw in **Figure 27**, under the BISG Max classification rule, the \$100 Actual Black fee disparity amount is underestimated by -34% – which is three times the size of the underestimation bias from the BISG 80% Threshold rule. Using the same decomposition framework as above, the intuition for this result should become clear. First, since the BISG Max classification rule produces a predicted race / ethnicity for every sample member, there is no leakage of the "Unknowns" from the predicted groups. Instead, all FNs from one group end up as FPs in the other groups which, unlike the BISG 80% Threshold rule, results in significant "cross-contamination" of each Predicted group by these FPs. This can be seen clearly in **Figure 29**'s Sankey Chart below.

[50] Specifically, 84% of Blacks FNs are Unknown and 96% of White FNs are Unknown.

[51] The very small presence of FPs under the BISG 80% Threshold rule also explains why it has the highest Precision Accuracy of the three alternative classification rules.

[52] Connecting the dots even further, 91.2% is also the Black Precision Accuracy rate under the BISG 80% Threshold (excluding Unknowns) from **Figure 23**.

Figure 29: FPs and FN Misclassifications Under BISG Max Rule



Here we can see: (1) a significantly greater number of Black FNs flowing into Predicted Whites – thereby increasing the Predicted White average fee amount, and (2) a greater number of White FNs and other groups' FNs flowing into Predicted Blacks – thereby decreasing the Predicted Black average fee amount. The linkage of these flows to the estimated biases is shown in the calculations below – which tie to the results in **Figure 27**.

$$\textit{Estimated Average Black Fee Amount} = \underbrace{71.6\% * \$100}_{\substack{\text{Black TPs}\\\text{(Figure 19b)}}} + \underbrace{28.4\% * \$0}_{\substack{\text{Black FPs}\\\text{(Figure 19b)}}} = \$71.56$$

$$\textit{Estimated Average White Fee Amount} = \underbrace{86.1\% * \$0}_{\substack{\text{White TPs}\\\text{(Figure 19a)}}} + \underbrace{8.1\% * \$0}_{\substack{\text{White FPs}\\\text{(Non-Black)}\\\text{(Figure 19a)}}} + \underbrace{5.8\% * \$100}_{\substack{\text{White FPs}\\\text{(Black)}\\\text{(Figure 19a)}}} = \$5.78$$

Figure 30 below summarizes disparate treatment bias measurements over a broader range of scenarios.

Figure 30: Average Fee Disparity Biases Under Alternative Disparate Treatment Scenarios

Predicted Race / Ethnicity	BISG 80%	BISG 50%	BISG Max	BISG Continuous
Individual Disparate Treatment Scenarios				
Black Only	-11.0%	-29.7%	-34.2%	0.0%
Hispanic Only	-9.6%	-23.0%	-25.8%	0.0%
API Only	-8.6%	-19.2%	-23.9%	0.0%
Joint Black & Hispanic Disparate Treatment Scenario				
Black	-10.8%	-29.4%	-33.3%	0.0%
Hispanic	-11.0%	-26.5%	-29.2%	0.0%
Joint Black, Hispanic & API Disparate Treatment Scenario				
API	-10.5%	-23.5%	-27.5%	0.0%
Black	-11.2%	-29.8%	-33.5%	0.0%
Hispanic	-10.3%	-25.2%	-27.8%	0.0%

The first group of rows ("Individual Disparate Treatment Scenarios") imposes the $100 processing fee on each minority group separately (with all other groups receiving a $0 fee). The second group of rows ("Joint Black & Hispanic Disparate Treatment Scenario") imposes the $100 processing fee on both Actual Blacks and Hispanics, and the third group of rows ("Joint Black, Hispanic & API Disparate Treatment Scenario") imposes the $100 processing fee on all three minority groups. Comparing the results across scenarios, we see some minor variations in the individual-level classification results – which are driven by reduced dilution of the minority group average fee amount (from other minority FPs that now have $100 – rather than $0 – fees) and an increase in the White average fee amount (also from more minority FPs with $100 fees).

Appendix G illustrates how these disparate treatment disparity biases can vary at the individual state-level due to different sample sizes and different BISG probability distributions – the latter of which influences the relative frequency of False Negatives and False Positives across states and, therefore, the relative amounts of disparity "cross-contamination". The results for the Black Only disparate treatment scenario show that – even within a given BISG Classification approach – estimation biases can vary significantly across states. For example, the disparate treatment disparity bias for Wyoming Blacks under the BISG 80% classification rule is only -0.8% (vs. national level bias of -11.0%) while that for Idaho Blacks is -34%. Clearly, **model users need to be sensitive to geographic concentrations of their analysis samples that may exacerbate the risks of disparity estimation bias**.

Do the BISG Continuous Results Really Have Zero Bias?

As the results in **Figures 27 and 30** show, the BISG Continuous estimation approach provides an unbiased estimate of the disparate treatment effect – estimating a near-exact $100 fee disparity for

Blacks by regressing the fee amounts on the BISG probability sets.[53] If this result seems surprising given that we are using a set of uncertain probabilities to "identify" the race / ethnicity of each individual, we agree. However, further analysis of these results helps us to understand why.

We first note that these results were generated from a sample of U.S. adults that strictly conformed to the BISG proxy model's Census-based, geo-surname distributions. That is, the aggregate actual race / ethnicity distribution of the sample was statistically the same as the aggregate expected distribution of the sample according to the underlying BISG probabilities assigned to each sample member (see, for example, the discussion of **Figure 8**). This is an important condition in the context of the BISG Continuous approach for the following reason:

- The uncertainty embedded within each BISG probability can be thought of as a form of measurement error that creates "partial" True Positives and "partial" False Negatives for each sample member. That is, rather than a sample member being designated as a "complete" FN or TP as we saw under individual-level classification, each sample member under the BISG Continuous approach is partially both of these designations – with the sum of these partial designations equal to one. For example, consider a Black sample member who has the following set of BISG probabilities:

BISG Black Probability	BISG API Probability	BISG Hispanic Probability	BISG Other Probability	BISG White Probability
0.262	0.003	0.006	0.022	0.707

Partial TP

Partial FNs / FPs

  This sample member can be designated as a 0.262 TP and a 0.738 FN (which both sum to one). Furthermore, since each FN is another group's FP, the partial Black FN (0.738) becomes the other race / ethnicity groups' partial FPs – specifically, a 0.003 API FP, a 0.006 Hispanic FP, a 0.022 Other FP, and a 0.707 White FP.

- With this foundation, this sample member generates two types of BISG measurement error – False Negative and False Positive:

[53] We obtain virtually the same results for disparate treatment scenarios involving Hispanics and APIs. Furthermore, this lack of bias is also present in "imperfect" disparate treatment scenarios – that is, where both the prohibited basis and control groups are assessed fees, but where the frequency of fee assessment is higher for the prohibited basis group.

Measurement Error Type	Calculation	Value
False Negative	BISG Black Probability-1	-0.738
False Positive	BISG API Probability	0.003
False Positive	BISG Hispanic Probability	0.006
False Positive	BISG Other Probability	0.022
False Positive	BISG White Probability	0.707
Total		**0.000**

The FN measurement error is consistently negative (or zero if BISG perfectly predicts the race / ethnicity, which is unlikely) and the FP measurement error is consistently positive (or zero). Furthermore, for each sample member, the errors offset each other when aggregated across all racial / ethnic groups.

The presence of these measurement errors is a critical consideration when evaluating the reliability of potential disparate treatment disparities estimated via the BISG Continuous approach in which the BISG proxy probabilities are used directly as independent variables in an OLS regression.[54] This is because measurement error in the independent variables tends to produce bias in OLS regression coefficient estimates.[55] While there is a vast literature on "classical" measurement error that analyzes the sources and directions of this bias, we note that the BISG measurement errors do not conform to the "classical" definition for the reasons discussed below and, therefore, require further analysis to assess their specific bias properties.

1) **The BISG measurement errors are not random** – in fact, as discussed above, they are consistently negative for the matching BISG probability associated with a sample member's actual race / ethnicity group (reflecting the presence of a partial FN) and they are consistently positive for the sample member's non-matching BISG probabilities (reflecting the presence of partial FPs).

2) **The BISG measurement errors are correlated across each sample member's set of BISG probabilities** – that is, for every sample member, the BISG probability associated with the member's actual race / ethnicity will have negative measurement error and the other BISG probabilities will all have positive offsetting measurement errors.

3) **The BISG measurement errors actually <u>reduce</u> the variances of the regression model's independent variables** (i.e., the BISG probabilities) relative to the variances of

[54] Our analysis of potential bias is limited to the use of the BISG probabilities within ordinary least squares ("OLS") regression models. Analysis of potential bias in other modeling contexts – such as logistic regression – is beyond the scope of this study but is an important area for further research.

[55] See for example Greene, William H., <u>Econometric Analysis</u>, Prentice Hall, 1993, pp. 279-284.

the true underlying independent variable values (i.e., the actual race / ethnicity indicator variables). This is the opposite of what occurs under the classical measurement error scenario where the measurement errors are additive to the regressor variance.

What impact do the BISG measurement errors have on disparate treatment disparity estimates?[56]

If the BISG proxy model is used appropriately – that is, it is applied to a sample whose actual underlying race / ethnicity distribution is statistically aligned to the underlying set of BISG probabilities assigned to it, then the FN and FP measurement errors will effectively be neutralized when aggregated across all sample members – as shown below in **Figure 31** for our synthetic sample.

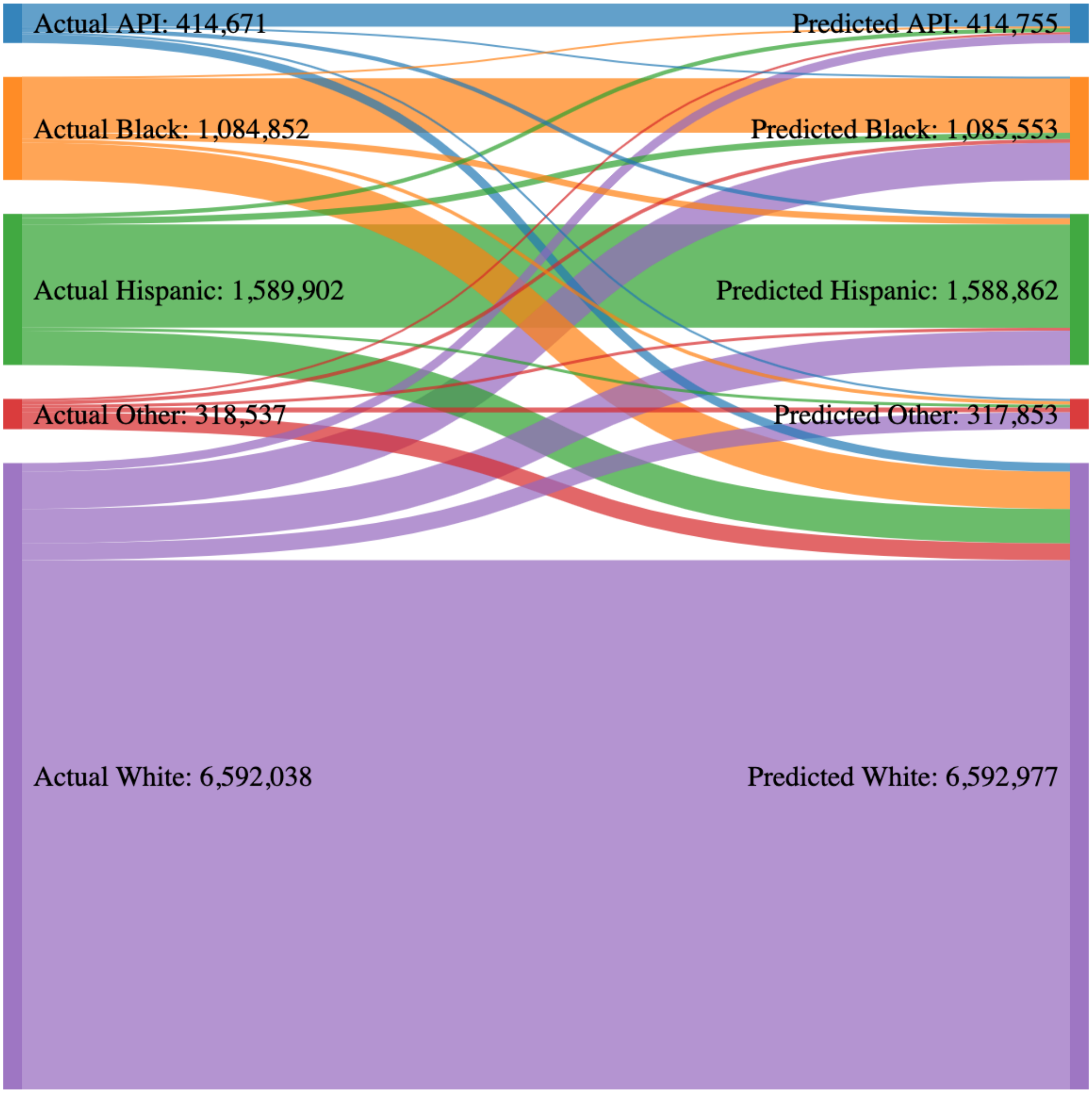
Figure 31: Comparison of Aggregate Actual vs. Predicted Race / Ethnicity Counts

BISG Probability	Aggregate Counts: Actuals	Aggregate Counts: Predicted (Based on Proxies)	Aggregate BISG Measurement Error	Aggregate BISG Measurement Error %
Black	1,084,853	1,085,553	700	0.1%
API	414,671	414,754	83	0.0%
Hispanic	1,589,902	1,588,862	(1,040)	-0.1%
Other	318,536	317,853	(683)	-0.2%
White	6,592,038	6,592,978	940	0.0%
Totals	**10,000,000**	**10,000,000**	**0**	**0.0%**

Here we see that the BISG proxy probabilities replicate the aggregate actual race / ethnicity distribution of our sample – apart from *de minimis* sampling error. This occurs because the alignment of our sample data with the underlying Census data used to construct the BISG probabilities causes the False Negatives in each race / ethnicity group to be completely offset (within sampling error) by False Positives as illustrated in the Sankey chart and table below.

[56] Due to these unique measurement error properties, a deeper theoretical investigation of the sources and directions of potential estimation bias is beyond the scope of this study. However, such research would clearly be useful to the fair lending community to understand the larger set of potential biases that may arise under a broader set of disparate treatment and disparate impact scenarios, as well as the inclusion of additional independent variables into the regression that may be correlated with the BISG measurement errors.

Figure 32: Flows of Partial False Negatives / False Positives Based on BISG Probabilities



BISG Probability	Aggregate Counts: Actuals	Aggregate False Negatives (-)	Aggregate False Positives (+)	Aggregate Counts: Predicted
Black	1,084,853	(511,495)	512,195	1,085,553
API	414,671	(171,802)	171,885	414,754
Hispanic	1,589,902	(502,859)	501,819	1,588,862
Other	318,536	(268,244)	267,561	317,853
White	6,592,038	(1,021,661)	1,022,601	6,592,978
Totals	**10,000,000**	**(2,476,061)**	**2,476,061**	**10,000,000**

In these situations, the BISG probabilities are strong unbiased proxies for the sample's actual race / ethnicity distribution – yielding unbiased estimates of the disparate treatment effects as shown in **Figures 27 and 30**.[57,58]

What if the sample's underlying actual race / ethnicity distribution is not aligned to the BISG-based expected distribution?

If the sample being analyzed differs materially from the BISG model's underlying Census-based distributions – for example, by focusing on higher income or wealthier sub-segments of certain micro-geographies – then this inconsistency introduces bias to the estimates of disparate treatment.[59] To investigate the magnitude and direction of this bias empirically, we modified the actual race / ethnicity distribution of our synthetic sample to deviate from the underlying BISG probability distributions by specific amounts. That is, we "skewed" the composition of the sample to include more or less Actual Whites than expected – with a corresponding offset to Actual Blacks. We varied the magnitude of this skew from -15% to +15% additional Whites across 30 separate skew scenarios.

To implement each skew scenario, we re-performed the Monte Carlo technique to generate a new set of "actual" races / ethnicities from the original set of BISG probabilities that we modified to reflect the

---

[57] Econometrically, the BISG measurement errors have two offsetting effects on the regression coefficient estimates in this scenario. First, they lead to smaller variances of the regression model's race / ethnicity independent variables (relative to the variances that would be obtained using the true race / ethnicity indicator variables). This biases the regression coefficients upward. Second, using BISG proxy probabilities (instead of the true race / ethnicity indicator variables) tends to reduce the covariances between the dependent variable (the fee amount) and the race / ethnicity variables. This is due, for example, to the Black FPs introducing $0 fee amounts and the Black FNs removing $100 fee amounts to the Predicted Black-Fee Amount covariance. This dilution of the true covariance between Actual Blacks and Fee Amounts exerts an offsetting downward bias on the regression coefficients. This is different than the classical measurement error scenario in which the bias inflates the independent variable variances only (since they are, by assumption, uncorrelated with the dependent variable) and, accordingly, biases the regression coefficient estimates toward zero. See p. 88 for a further discussion of how the disparate treatment regression coefficient is impacted by the BISG measurement error.

[58] This is true for more typical sample sizes as well. For example, overall national-level sample sizes as small as 10,000 still produce unbiased disparate treatment disparity estimates, on average, with only +/- 3% sampling variability (i.e., 95% confidence interval). We also obtain unbiased estimates at the state level under distributional alignment. However, states with relatively small minority sample sizes appear to exhibit disparate treatment estimation biases that are not due to sampling error. For example, our ND sample contains approximately 22,000 sample members of which only 267 (1.2%) are Actual Blacks. The Black disparity estimate for this state is $93.20 – indicating a -6.8% disparity bias. Alternatively, our ID sample contains 48,802 sample members of which only 390 (0.8%) are Actual Blacks. The Black disparity estimate for this state is $102.40 – a +2.4% disparity bias. Overall, for states with small populations of certain minority groups, it appears that typical sample sizes (in these cases, 22,000 – 49,000 sample members) may not be sufficient to ensure the type of distributional alignment needed for unbiased disparate treatment disparity estimates.

[59] This is a similar concept to "data appropriateness" in model validation. That is, for a model to be an effective / unbiased prediction tool, the data to which the model will be applied in production must be materially consistent with the data that created the model.

desired skew scenario. During this implementation, we ensured that: (1) only White and Black BISG probabilities could change, and (2) each sample member's Black probability completely offset the change in their White probability to ensure all probabilities still summed to one. Achieving both of these properties required that the desired probability change for certain sample members be attenuated.[60] Accordingly, the overall average skew amount for each scenario may be less than the desired skew amount.

Once the new "skewed" set of actual races / ethnicities were generated, we implemented the $100 Black disparate treatment scenario on these new "actuals" and performed the OLS fair lending regression analysis on these fee amounts using the original (i.e., non-skewed) BISG probabilities as independent variables.[61] The results from this analysis are presented in **Figure 33** below.

[60] For example, a +15% White skew could not be achieved for sample members whose BISG Black probabilities were less than 15% as it would cause the adjusted Black probabilities to be negative. In such cases, the skew amount was attenuated to the maximum amount permitted without causing negative adjusted probabilities (or adjusted probabilities greater than 1 for the other group).

[61] The original unadjusted BISG probabilities are used in the regression as they are the probabilities that would be assigned during a typical fair lending analysis.

Figure 33: Potential Bias of BISG Continuous Disparate Treatment Estimates Under Alternative White Skew Assumptions and $100 Black Fee Disparate Treatment Scenario

Maximum Distribution Skew: Actual Whites	Estimated Black Fee Disparity (Skew Scenario)	Estimated Black Fee Disparity (Ground Truth)	Estimated Bias: Black Fee Disparity %	Change in Actual Number of Blacks	Black Distribution %: Skewed	Black Distribution %: Original	Change in Actual Number of Whites	White Distribution %: Skewed	White Distribution %: Original
-15%	$93.01	$99.89	-6.9%	1,341,718	24.3%	10.8%	(1,341,718)	52.5%	65.9%
-14%	$93.69	$99.90	-6.2%	1,257,637	23.4%	10.8%	(1,257,637)	53.3%	65.9%
-13%	$94.33	$99.89	-5.6%	1,172,602	22.6%	10.8%	(1,172,602)	54.2%	65.9%
-12%	$94.95	$99.87	-4.9%	1,087,115	21.7%	10.8%	(1,087,115)	55.0%	65.9%
-11%	$95.56	$99.89	-4.3%	1,001,124	20.9%	10.8%	(1,001,124)	55.9%	65.9%
-10%	$96.16	$99.91	-3.8%	914,306	20.0%	10.8%	(914,306)	56.8%	65.9%
-9%	$96.75	$99.95	-3.2%	826,634	19.1%	10.8%	(826,634)	57.7%	65.9%
-8%	$97.27	$99.95	-2.7%	739,220	18.2%	10.8%	(739,220)	58.5%	65.9%
-7%	$97.78	$99.96	-2.2%	650,944	17.4%	10.8%	(650,944)	59.4%	65.9%
-6%	$98.24	$99.96	-1.7%	561,506	16.5%	10.8%	(561,506)	60.3%	65.9%
-5%	$98.67	$99.97	-1.3%	470,443	15.6%	10.8%	(470,443)	61.2%	65.9%
-4%	$99.05	$99.97	-0.9%	379,104	14.6%	10.8%	(379,104)	62.1%	65.9%
-3%	$99.39	$99.97	-0.6%	286,837	13.7%	10.8%	(286,837)	63.1%	65.9%
-2%	$99.69	$99.98	-0.3%	192,872	12.8%	10.8%	(192,872)	64.0%	65.9%
-1%	$99.90	$99.98	-0.1%	97,686	11.8%	10.8%	(97,686)	64.9%	65.9%
0%	$99.97	$99.97	0.0%	-	10.8%	10.8%	-	65.9%	65.9%
1%	$99.15	$99.97	-0.8%	(62,076)	10.2%	10.8%	62,076	66.5%	65.9%
2%	$97.99	$99.97	-2.0%	(108,684)	9.8%	10.8%	108,684	67.0%	65.9%
3%	$96.69	$99.97	-3.3%	(149,291)	9.4%	10.8%	149,291	67.4%	65.9%
4%	$95.32	$99.97	-4.7%	(185,634)	9.0%	10.8%	185,634	67.8%	65.9%
5%	$93.89	$99.97	-6.1%	(218,986)	8.7%	10.8%	218,986	68.1%	65.9%
6%	$92.43	$99.99	-7.6%	(249,862)	8.3%	10.8%	249,862	68.4%	65.9%
7%	$90.95	$100.00	-9.1%	(278,647)	8.1%	10.8%	278,647	68.7%	65.9%
8%	$89.42	$99.98	-10.6%	(306,017)	7.8%	10.8%	306,017	69.0%	65.9%
9%	$87.89	$99.98	-12.1%	(331,747)	7.5%	10.8%	331,747	69.2%	65.9%
10%	$86.36	$99.98	-13.6%	(356,127)	7.3%	10.8%	356,127	69.5%	65.9%
11%	$84.81	$99.97	-15.2%	(379,519)	7.1%	10.8%	379,519	69.7%	65.9%
12%	$83.26	$99.95	-16.7%	(401,894)	6.8%	10.8%	401,894	69.9%	65.9%
13%	$81.73	$99.96	-18.2%	(423,381)	6.6%	10.8%	423,381	70.2%	65.9%
14%	$80.17	$99.94	-19.8%	(444,188)	6.4%	10.8%	444,188	70.4%	65.9%
15%	$78.63	$99.93	-21.3%	(464,105)	6.2%	10.8%	464,105	70.6%	65.9%

The first column "Maximum Distribution Skew: Actual Whites" denotes the desired skew amount associated with each scenario. The next set of three columns reports the estimated Black fee disparity and bias amounts under the BISG Continuous estimation approach. In particular, the second column presents the estimated Black fee disparity amount from the fair lending regression analysis while the third column presents the corresponding unbiased Black fee disparity amount estimate.[62] Comparing these two values yields the estimated Black fee disparity bias % – which is displayed in the fourth column. Finally, the second group of columns in the middle displays the impacts of each skew scenario on Actual Blacks while the last group of columns displays the impacts on Actual Whites. The middle row shaded in light blue corresponds to our baseline scenario where no skew is present; accordingly, in this baseline scenario, there are no changes to the Black and White Actuals and the unbiased results

[62] Theoretically, the third column value should always be $100; however, even with aligned BISG proxy probabilities, there will always be some negligible amount of estimation error (in this case, a maximum of 0.13% error).

correspond to those reported in **Figure 27** under the "BISG Continuous" column.

Scanning these scenario results, we note that the bias results are consistently negative regardless of whether the White skew is positive or negative – which was initially surprising. However, a deeper mathematical dive reveals that while the bias could be positive, such an outcome requires fairly specific conditions that would appear to be infrequently encountered. In particular, positive bias could occur if the positive White skew comes predominantly from sample members with below average BISG Black probabilities, or if the negative White skew is concentrated on sample members with above average BISG Black probabilities.[63] Within our synthetic sample, we note that there are 958,765 Blacks with above average BISG Black probabilities and 126,088 Blacks with below average BISG Black probabilities – an uneven, but sensible, distribution.[64] Working through the details of two of our White skew scenarios helps to provide insight as to why positive bias, while theoretically possible, does not arise across our scenarios.

- **+15% White Skew Scenario** – under this scenario, we are seeking to expand significantly the representation of Actual Whites with a corresponding offset to Actual Blacks – which leads to a reduction in the number of Actual Blacks by 464,105 (with attenuation, see Footnote 60) and an increase in Actual Whites by the same amount. This is illustrated in the last row of **Figure 33**. As previously noted, these additional Whites must come from the sample members who were designated as Black under the original non-skewed distribution. What we find is that 126,088 come from Actual Blacks with below average BISG Black probabilities and 338,017 come from Actual Blacks with above average BISG Black probabilities.[65] Intuitively, since there are 7.6x more Blacks with above average BISG Black probabilities than below average BISG Black probabilities (i.e., 958,765 vs. 126,088), it makes sense that the additional Whites will also be skewed toward the former. Therefore, with the additional Whites skewed toward the above average

[63] Remember, since we are maintaining the same overall sample size, we are flipping Actual Blacks to Actual Whites under positive White skew, and flipping Actual Whites to Actual Blacks under negative White skew, and this change in group membership impacts the covariance component of the regression coefficient (the variance component – which is based on the original BISG probabilities – is unaffected). Intuitively, under this disparate treatment scenario, you expect a positive covariance between BISG Black probability values and fee amounts (i.e., since Blacks are charged the fees, there should be a positive association between BISG Black probabilities and fee values). However, if some Blacks have low BISG Black probabilities, then these Blacks will tend to dilute the overall covariance term (since their $100 fee values are associated with relatively low BISG Black probability values). If these Blacks are flipped to Whites, as under our skew scenario, then the covariance term will rise as the low BISG Black probabilities are now associated with $0 fee amounts – not $100. This is the source of the potential positive bias.

[64] These two numbers sum to 1,084,853 which equals the total amount of Actual Blacks in our sample. The distribution is sensible since: (1) the average BISG Black probability level is calculated across the entire sample (not just over the subset of Actual Blacks), and (2) one is more likely to observe Actual Blacks at higher BISG Black probability levels than at lower BISG Black probability levels.

[65] In fact, the 126,088 Actual Blacks with below average BISG probabilities are all of the Actual Blacks with below average BISG Black probabilities – so the rest of the additional Whites must come from the Actual Blacks with above average BISG Black probabilities.

BISG Black probabilities, the net bias becomes negative.

- **-5% White Skew Scenario** – under this scenario, we are seeking to decrease the representation of Actual Whites by 5% with a corresponding offset to Actual Blacks – which leads to an increase in the number of Actual Blacks by 470,443 (with attenuation, see Footnote 60) and a decrease in Actual Whites by the same amount. This is also illustrated in **Figure 33**. Once again, these additional Blacks must come from the sample members who were designated as Actual Whites under the original non-skewed distribution. What we find is that 363,502 come from Actual Whites with below average BISG Black probabilities and 106,941 come from Actual Whites with above average BISG Black probabilities. Intuitively, since there are 1,048,186 Actual Whites with above average BISG Black probabilities and 5,543,852 Actual Whites with below average BISG Black probabilities, it makes sense that the additional Blacks will also be skewed toward the latter. Therefore, with the additional Blacks skewed toward the below average BISG Black probabilities, the net bias becomes negative.

Turning now to the magnitude of the bias, **we see that even under modest White skew amounts, the estimated disparate treatment bias under the BISG Continuous approach can be significant.** For example, under the +12% White skew scenario, Actual Whites increase from 65.9% of the sample to 69.9% – a change of 4 percentage points which is in line with those reported by the other researchers for mortgage-based samples.[66] Moving over to columns two through four, we see that this distribution misalignment causes a -16.7% understatement of the Black fee disparity.

Of course, there are other skew scenarios that may differ from the structure assumed here, and where the bias results may potentially differ. For example, rather than our assumed structure where the overall sample size remains the same, there could be a scenario where Blacks in more diverse geo-surname segments are simply excluded due to the lender's discrimination – e.g., by refusing to accept applications (or consistently denying applications) from Blacks in geo-surname segments with low BISG Black probabilities.[67] Our testing reveals that while such scenarios could produce positive bias in disparate treatment disparity estimates, the extent of the bias appears limited; for example, at the extreme where all 126,088 Blacks with below average BISG Black probabilities were excluded from the sample, the positive bias would only be approximately 1.7%.

Finally, **Appendix H** presents additional results illustrating how skewed samples can affect estimated disparate treatment disparities at more disaggregated geographic levels of testing – in this case, at the state level. This analysis assumes that the maximum White skew is +2% which, at the national level according to **Figure 33**, would generate a 2% understatement of the Black fee disparity amount when aggregated to the national level. The results are sorted from largest understatement to smallest and

[66] Attenuation causes the +12% desired White skew to become +4% actual White skew.

[67] Obviously, it is unlikely that the discriminatory conduct would explicitly occur in this manner (i.e., based on geo-surname segment probability); however, if the lender is biased against Blacks who reside in White-dominant CBGs, the effect may be very similar.

demonstrate that – for a given maximum White skew amount – estimated biases can vary across geographies depending on their relative race / ethnicity distributions. In this analysis, **those states with relatively smaller Black populations experience greater biases in disparate treatment estimates, while those with relatively larger Black populations experience smaller biases in disparate treatment estimates.**

In summary, our exploration of the BISG proxy model's impacts on disparate treatment estimates reveals that:

- **Applying the BISG proxy model to consumer credit transactions whose underlying socioeconomic characteristics differ from those associated with the Census data used to construct the BISG proxies will likely lead to biased disparate treatment disparity estimates.** Such biases are present whether the disparate treatment disparities are estimated using the "BISG Continuous" approach or using alternative individual-level race / ethnicity classifications. In our specific testing scenarios, which were admittedly simplified to maximize insights, the biases were consistently negative – thereby resulting in understated estimates of true disparate treatment disparity amounts. While theoretically possible under very specific conditions, positive biases appear less likely to be present given the current properties of the Census data on which the BISG proxy model is based.

- **The impacts of the socioeconomic misalignments between the analysis sample and the BISG proxy model's Census data are greater for geographies with relatively smaller minority populations, and vice versa.** This indicates that caution should be exercised when using the BISG proxy model to estimate disparate treatment disparities on potentially skewed samples in certain geographies – such as in low minority states, or in low minority MSAs, counties, or census tracts.

- **Using individual-level BISG race / ethnicity classifications in disparate treatment testing – even in the absence of skewed samples – appears to cause downward bias in disparate treatment disparity estimates.** In our specific testing scenarios, which were admittedly simplified to maximize insights, the BISG Max classification rule generated the largest bias while the BISG 80% Threshold rule generated the smallest bias due to the greater influence of the False Negatives / False Positives under the BISG Max approach. However, the reduced bias of the BISG 80% Threshold rule comes at the expense of significantly smaller addressable samples – particularly for minorities.

Given these results, the recommendations are challenging.

- If there is no distributional misalignment between the analysis sample and the Census data underlying the BISG proxies, then the BISG Continuous approach produces unbiased disparate treatment disparity estimates. However, if distributional misalignment is likely present, users will need to determine whether the risk of potential biases precludes responsible use of the BISG proxy-based estimates. If potential biases are not considered material to the decisions derived from the model, then our testing indicates that the BISG Continuous approach produces the least amount of bias (see **Appendix I**). Nevertheless, proper and sufficient disclosure should be provided with the estimated disparities to alert users to the associated risks and limitations of the estimates.

- If individual-level race / ethnicity identification is needed, users should again determine whether the potential misclassifications are of such a magnitude as to preclude responsible use of the BISG proxy-based estimates. While the BISG Max methodology has the highest F1 accuracy and Recall accuracy metrics for the entire sample, potential misclassifications can still be quite large. At the very least, proper and sufficient disclosure should be provided with these results to alert users to the associated risks and limitations of the individual race / ethnicity assignments.

Hidden Biases in Disparate Impact Estimates

In its simplest form, disparate impact involves one group of customers – typically members of a prohibited basis group – receiving relatively unfavorable terms, conditions, or actions in certain aspects of a credit transaction due to the effects of a consistently-applied lending policy or practice. For example, a lender may have a policy not to originate residential mortgage loans below a minimum amount – say $50,000. Even though that policy is applied consistently across all applicants, there may still be a disparate impact if it results in a disproportionate rejection of credit applications from prohibited basis customers. Because disparate impact is not driven by differences in treatment (whether overt or unintentional), allegations typically are based solely on the presence of statistically-measured credit outcome disparities between prohibited basis and control groups.

In recent years, a more extreme disparate impact theory has emerged in which the lender's "policy" of price discretion is targeted. That is, by permitting employee- or third-party loan originators to vary pricing – even within defined limits – on a discretionary basis, a disparate impact is alleged since the empirical evidence indicates that prohibited basis groups have higher average discretionary prices than control groups. This theory of discrimination formed the basis of most federal fair lending enforcement activity over the last 10 years – including the significant enforcement actions and customer restitution payments involving wholesale mortgage and indirect auto lenders. Under these "policy of discretion" allegations, federal bank regulators and U.S. enforcement agencies measure the associated disparate impacts at the highest level of customer aggregation; that is, unlike with disparate treatment, very few – if any – customer or transaction characteristics are considered in the statistical analysis to account for some of the differences in average fees between the prohibited basis and control group customers.[68]

[68] The argument here is that under a "policy of discretion" – such as in the assessment of discretionary fees, every customer has the same likelihood of being assessed such a fee; therefore, no additional statistical controls are justified. However, in reality, those who impose the discretionary fees typically do so based directly or indirectly on certain objective, non-demographic borrower or transaction characteristics – such as loan amount, income, credit score, etc. While, theoretically, these practices should be included in the statistical analysis to ensure an appropriate comparison

Due to the legal prohibition on collecting demographic data on non-mortgage credit applicants, one of the side effects of this disparate impact expansion outside of the mortgage area was to develop race / ethnicity and gender proxies to facilitate empirical fair lending testing, at scale, across the consumer lending industry. This begat the CFPB's application of the BISG proxy model to non-HMDA fair lending testing – such as for auto loans and credit cards – and the corresponding adoption of this proxy methodology across the industry in fair lending compliance risk management programs.

Today, the disparate impact theory and the BISG proxy model remain alive and well in the fair lending domain – with even greater usage due to the rise of AI algorithms in all stages of the consumer lending process. Such algorithms, while mitigating disparate treatment risk due to the automated nature of the decisioning processes in which they are embedded, increase the disparate impact risk due to their potential to include unintentional hidden biases. This has led to a new application of the BISG proxy model to assist consumer lenders in testing their algorithms for such biases, as well as to assist in modifying the algorithms to remove such biases in a manner that has the least impact on overall predictive performance.

In this final section, we explore how the BISG proxy model impacts traditional estimates of disparate impact. Similar to the prior section on disparate treatment, we adopt a simplified approach with known "ground truth" disparities in order to maximize insights. Specifically, we assume that disparate impact occurs by charging a discretionary fee amount that varies with the general income level of the customer. While all customers with the same general income level are treated equally, certain prohibited basis groups – such as Blacks – that have lower average income levels will have higher average fee amounts than the corresponding control group customers (Whites). This is the source of the disparate impact.

To implement this disparate impact scenario, we segment our national-level U.S. adult sample into 10 deciles based on average CBG median income and assume that fees are assessed according to the following discretionary fee schedule:

of lending outcomes between "similarly-situated" customers, the federal bank regulators and U.S. enforcement agencies have taken the position that such practices are not legitimate controls unless they are codified into lender policies or procedures.

Average CBG Median Income Decile	Discretionary Fee Amount
1	$100
2	$90
3	$80
4	$70
5	$60
6	$50
7	$40
8	$30
9	$20
10	$10

In this scenario, loan originators tend to charge lower income borrowers – regardless of their race / ethnicity – higher average fees than higher income borrowers. Importantly, disparate treatment is not present here since all borrowers within a given income range receive the same fee amount. However, because certain minority groups have lower average incomes than Whites, a disparate impact can be alleged.

The next steps in this process are as follows:

- Apply this fee scenario to all members of the synthetic dataset according to their average CBG median income decile.[69]

- Estimate the average fee amounts for each BISG proxied group and calculate the average fee disparity amount for each minority group versus the White group (i.e., the mathematical difference between the estimated average fee for the minority group and the estimated average fee for the White group). Perform this estimation using the BISG Continuous approach – as well as for the three alternative BISG classification rules (BISG Max, BISG 50% Threshold, and BISG 80% Threshold).

- Compare the estimated average fee disparity amounts to the actual "ground truth" average fee disparity amounts and calculate potential estimation bias.

Figure 34 presents the results of this disparate impact estimation scenario.

[69] 0.4% of the 10 million sample members do not have a matching CBG median income value. Accordingly, these sample members are excluded from the downstream fair lending analyses.

Figure 34: Disparate Impact Scenario Results: Discretionary Fee Schedule Based on Income

Race / Ethnicity	BISG 80%	BISG 50%	BISG Max	BISG Continuous	Actuals
Average Fee $ Amount					
API	$46.39	$49.91	$50.78		$48.65
Black	$77.42	$73.39	$72.55		$67.25
Hispanic	$59.39	$58.11	$58.56		$59.10
White	$50.11	$51.89	$52.18		$52.17
Disparate Impact Estimates					
API	($3.72)	($1.98)	($1.40)	($8.25)	($3.51)
Black	$27.30	$21.49	$20.37	$31.15	$15.09
Hispanic	$9.28	$6.22	$6.38	$12.63	$6.94
Disparate Impact Estimate $ Bias					
API	($0.21)	$1.53	$2.11	($4.73)	
Black	$12.22	$6.40	$5.28	$16.06	
Hispanic	$2.34	($0.72)	($0.56)	$5.70	
Disparate Impact Estimate % Bias					
API	-6.0%	43.6%	60.2%	-134.9%	
Black	81.0%	42.4%	35.0%	106.4%	
Hispanic	33.7%	-10.4%	-8.1%	82.1%	

The first column of this table corresponds to the predicted group members being tested, the second through fourth columns correspond to the fair lending testing results under each of the three classification rules used to predict the individual-level race / ethnicity of each group member, the fifth column corresponds to the fair lending testing results using the BISG Continuous approach, and the last column displays the actual "ground truth" results.

The first set of four rows ("Average Fee $ Amount") contains the estimated average fee amounts of each predicted race / ethnicity group under the three BISG Classification prediction methods as well as the ground truth average fee amounts.[70] The second set of three rows ("Disparate Impact Estimates") calculates the corresponding fair lending disparity – that is, the difference in each minority group's estimated average fee amount vs. the estimated average fee amount for Whites – with the last column displaying the ground truth disparate impact amount. Finally, the next three rows ("Disparate Impact Estimate $ Bias") compare the estimated disparate impact amounts to the ground truth amounts and calculates the dollar amount of the bias, while the final set of three rows converts these dollar-based bias amounts into bias percentages.

[70] The BISG Continuous method produces measures of relative fee disparity amounts (the second group of rows) rather than the absolute average fee amount for each group.

Overall, **these results indicate that all BISG proxy approaches – including the BISG Continuous approach – yield biased disparate impact estimates even in the absence of any misalignment between actual and expected race / ethnicity distributions. In fact, these results indicate that the BISG Continuous approach yields the largest biases with magnitudes ranging from -135% for APIs to +106% for Blacks**.

Unlike the simplicity of insights from the previous section's disparate treatment scenario, the insights here are a little more complex due to the fact that all demographic groups are assessed the full range of fees in our scenario – just in different proportions. Accordingly, we will walk through in more detail the results for Blacks under the BISG Classification approaches to highlight the drivers of these estimation biases.

What is the intuition behind these biases?

Figure 35b below reproduces certain data contained in **Figure 25b** from our analysis of FPs and FNs associated with the BISG 80% Threshold rule. To this data, we add new calculations associated with average fee estimates associated with our disparate impact scenario.

Figure 35b: Comparative Characteristics of Actual vs. Predicted Blacks
BISG 80% Threshold Rule

Blacks	Total Actuals	–	False Negatives	=	True Positives	+	False Positives	=	Total Predicted
Average Median HH Income	$47,221		$50,575		$38,227		$39,866		$38,371
Average $ Fee Amount	$67.25		$63.40		$77.57		$75.79		$77.42
Sample Counts	1,084,853		790,263		294,590		28,412		323,002
% of Actual Blacks			-72.8%		27.2%				
% of Predicted Blacks					91.2%		8.8%		

As we discussed previously when evaluating potential biases in the alternative classification rules, the BISG 80% Threshold rule imposes stringent requirements for a sample member to be assigned to a specific race / ethnicity class – namely, a minimum BISG probability of 80%. Because of this high bar for membership, FNs tend to be prevalent – particularly for Blacks where, as shown here, 72.8% of Actual Blacks are falsely excluded from Predicted Black membership. Also discussed in this previous section was the fact that these Black FNs are not a random sample of Actual Blacks but are, in fact, concentrated in more racially-diverse, higher income CBGs (i.e., $50,575 average CBG median income vs. $38,227 for Black TPs). Accordingly, this 32% higher average income skew means that the average fees charged to the Black FNs under our disparate impact scenario are lower than the remaining Black TPs – specifically 18% lower (i.e., $63.40 vs. $77.57). Alternatively, the relatively smaller group of Black FPs (3.6% of Black FNs) tend to come from CBGs with average CBG median income levels that are only slightly higher than Black TPs (i.e., $39,866 vs. $38,227 – a 4% difference). Given this socioeconomic profile, the average fee amount for Black FPs is only slightly lower than for Black TPs at $75.79 (vs. $77.57).

Overall, by excluding a significant portion of higher-income / lower-fee Actual Blacks as FNs and substituting in a small portion of similar-income / similar-fee Non-Blacks as FPs, the BISG 80% Threshold rule results in a group of Predicted Blacks whose average fee amount is biased upward by 15% (i.e., $77.42 vs. $67.25). This is consistent with the results contained in the first set of rows in **Figure 34** for Blacks. However, we now need to compare this Black average fee bias to any bias present in White average fees under the BISG 80% Threshold rule in order to arrive at the overall bias present in the Black disparate impact estimate under our scenario.

Figure 35a below presents identical information for White sample members.

Figure 35a: Comparative Characteristics of Actual vs. Predicted Whites BISG 80% Threshold Rule

Whites	Total Actuals	–	False Negatives	=	True Positives	+	False Positives	=	Total Predicted
Average Median HH Income	$61,331		$54,898		$63,442		$62,820		$63,404
Average $ Fee Amount	$52.17		$58.57		$50.06		$50.88		$50.11
Sample Counts	6,592,038		1,633,294		4,958,744		326,685		5,285,429
% of Actual Whites			-24.8%		75.2%				
% of Predicted Whites					93.8%		6.2%		

Following the same analysis as above for Blacks, here we see that 24.8% of Actual Whites are excluded from the Predicted White group due to insufficient probability levels (i.e., less than 80%). These White FNs come from more racially-diverse CBGs with 13% lower average CBG median income levels – and, accordingly 17% higher average fees – than the White TPs. 20% of these White FNs are offset by Non-White FPs who come from CBGs with average CBG median income levels that are close to those of the White TPs (i.e., $62,820 vs. $63,442 – a -1% difference) and bearing average fees that are only slightly higher than the White TPs (i.e., $50.88 vs. $50.06 – a 2% difference).

Overall, by excluding a significant portion of lower-income / higher-fee Actual Whites as FNs and substituting in a small portion of similar-income / similar-fee Non-Whites as FPs, the BISG 80% Threshold rule results in a group of Predicted Whites whose average fee amount is biased downward by 4% (i.e., $50.11 vs. $52.17). This is consistent with the results contained in the first set of rows in **Figure 34** for Whites. Combining the Black and White results together, we obtain the following estimate of the disparate impact bias for Blacks under the BISG 80% Threshold rule:

Race / Ethnicity Group	Actual Average Fee Amount	Predicted Average Fee Amount	Estimated $ Bias	Estimated % Bias
Black	$67.25	$77.42	$10.16	15.1%
White	$52.17	$50.11	($2.05)	-3.9%
Estimated Disparate Impact Fee Disparity	**$15.09**	**$27.30**	**$12.22**	**81.0%**

The actual difference in average fees between the two groups is $15.09 and the BISG 80% Threshold rule results in an estimate of $27.30 – **an overestimate of $12.22 or 81.0%.** Over 80% of this overestimate is driven by the overstatement of average Black fees (for the reasons stated above) and about 20% is due to the understatement of average White fees.

Under the more inclusive BISG Max classification rule, **Figures 36a and 36b** below show the differences between Actual Blacks / Whites and Predicted Blacks / Whites (leveraging original data from **Figures 19a and 19b**):

Figure 36b: Comparative Characteristics of Actual vs. Predicted Blacks
BISG Max Rule

Blacks	Total Actuals	–	False Negatives	=	True Positives	+	False Positives	=	Total Predicted
Average Median HH Income	$47,221		$54,906		$41,533		$44,825		$42,469
Average $ Fee Amount	$67.25		$58.61		$73.65		$69.77		$72.55
Sample Counts	1,084,853		460,139		624,714		248,266		872,980
% of Actual Blacks			-42.4%		57.6%				
% of Predicted Blacks					71.6%		28.4%		

Similar to the data under the BISG 80% Threshold rule, we see that: (1) the number of Blacks FNs dominate the number of Black FPs (although by a smaller margin than under the BISG 80% Threshold rule), (2) the Black FNs also come from more racially-diverse CBGs with higher average CBG median incomes (and, therefore, lower average fees) than the Black TPs, and (3) the Black FPs come from CBGs with slightly higher average CBG median incomes (and slightly lower average fees) than the Black TPs. Combined together, we see that Predicted Blacks under the BISG Max classification rule have average fees that are biased <u>upward</u> by about 8% over Actual Blacks – about half the amount of bias observed under the BISG 80% Threshold rule.

For Whites, we see the following under the BISG Max rule:

Figure 36a: Comparative Characteristics of Actual vs. Predicted Whites
BISG Max Rule

Whites	Total Actuals	–	False Negatives	=	True Positives	+	False Positives	=	Total Predicted
Average Median HH Income	$61,331		$53,681		$61,899		$57,221		$61,252
Average $ Fee Amount	$52.17		$60.56		$51.54		$56.16		$52.18
Sample Counts	6,592,038		456,959		6,135,079		987,356		7,122,435
% of Actual Whites			-6.9%		93.1%				
% of Predicted Whites					86.1%		13.9%		

Consistent with the pattern observed under the BISG 80% Threshold rule, (1) Whites FNs come from more racially-diverse CBGs with average median incomes that are lower than White TPs, and (2) White FPs come from CBGs with slightly less average CBG median incomes than White TPs. However, a key difference with the BISG 80% Threshold rule is that White TPs dominate the White FNs – thereby having a greater influence on the final Predicted White group. Combined together, we see that Predicted Whites under the BISG Max classification rule have average fees that are virtually unbiased relative to Actual Whites – differing only by $0.01 – which is due to the larger number of White FPs with modestly higher average fees offsetting the exclusion of a smaller number of White

FNs with much higher average fees.

Overall, the upward bias to the estimated average fees for Blacks and the *de minimis* bias for Whites leads to the following impacts on the estimated disparate impact fee disparity under the BISG Max classification rule:

Race / Ethnicity Group	Actual Average Fee Amount	Predicted Average Fee Amount	Estimated $ Bias	Estimated % Bias
Black	$67.25	$72.55	$5.30	7.9%
White	$52.17	$52.18	$0.02	0.0%
Estimated Disparate Impact Fee Disparity	**$15.09**	**$20.37**	**$5.28**	**35.0%**

The actual difference in average fees between the two groups is $15.09 and the BISG Max rule results in an estimate of $20.37 – **an overestimate of $5.28 or 35.0%.** All of this overestimate is driven by the overstatement of average Black fees (for the reasons stated above).

The intuition for the disparate impact estimation biases for Hispanics and APIs – as well as under the BISG 50% Threshold rule – follow similar logic so we will not dive into the underlying details here. However, **Appendices J, K, and L** contain the associated detailed tables for those interested.

One important final point. Although we discuss the results above using the term "disparate impact", we do this to be consistent with our knowledge and experience with U.S. federal bank regulator practice during fair lending examinations and enforcement proceedings. Whether these average fee differences – which, under our scenario, are driven by income levels – are truly actionable under federal and state fair lending laws and regulations is a much more complex legal matter for which expert counsel should be consulted. Furthermore, when we use the term "bias" in the context of these results, we are not assuming that the baseline disparate impact effect should be zero. Rather, the bias that we calculate assumes as its baseline the true average fee difference between these groups – regardless of whether this legally would be considered actionable "disparate impact".

What about the BISG Continuous results?

Interestingly, even in the absence of any skew between the socioeconomic characteristics of the analysis sample and those of the Census data on which the BISG probabilities are based, **our analysis reveals that estimated disparate impact disparities are significantly biased upward for Blacks and Hispanics, and downwards for APIs, when using the BISG Continuous approach**. Specifically, as shown in **Figure 34**, the BISG Continuous approach estimates a Black disparate impact disparity of $31.15 vs. a true disparity of $15.09 (a 106% upward bias) and a Hispanic disparate impact disparity of $12.63 vs. a true disparity of $6.94 (an 82% upward bias). For APIs, the estimate is -$8.25 vs. a true disparity of -$3.51 (a 135% downward bias).

While the precise mechanism for these biases is beyond the scope of this study, we do note the following impacts of the partial FNs and partial FPs on the average fee amounts of the four primary race / ethnicity groups:

Figure 37: Comparative Characteristics of Actual and Predicted Race / Ethnicity Groups: BISG Continuous Approach

Whites	Total Actuals	–	False Negatives	=	True Positives	+	False Positives	=	Total Predicted
Average Median HH Income	$61,331		$56,781		$62,164		$56,768		$61,329
Average $ Fee Amount	$52.17		$56.91		$51.30		$56.93		$52.17
Sample Counts	6,592,038		1,021,661		5,570,377		1,022,601		6,592,978
% of Actual Whites			-15.5%		84.5%				
% of Predicted Whites					84.5%		15.5%		

Blacks	Total Actuals	–	False Negatives	=	True Positives	+	False Positives	=	Total Predicted
Average Median HH Income	$47,221		$52,261		$42,714		$52,310		$47,247
Average $ Fee Amount	$67.25		$61.60		$72.31		$61.56		$67.23
Sample Counts	1,084,853		511,495		573,358		512,195		1,085,553
% of Actual Blacks			-47.1%		52.9%				
% of Predicted Blacks					52.8%		47.2%		

Hispanics	Total Actuals	–	False Negatives	=	True Positives	+	False Positives	=	Total Predicted
Average Median HH Income	$54,823		$55,097		$54,696		$55,060		$54,811
Average $ Fee Amount	$59.10		$58.79		$59.25		$58.82		$59.11
Sample Counts	1,589,902		502,859		1,087,043		501,819		1,588,862
% of Actual Hispanics			-31.6%		68.4%				
% of Predicted Hispanics					68.4%		31.6%		

APIs	Total Actuals	–	False Negatives	=	True Positives	+	False Positives	=	Total Predicted
Average Median HH Income	$65,667		$65,553		$65,748		$65,653		$65,708
Average $ Fee Amount	$48.65		$48.95		$48.45		$48.87		$48.62
Sample Counts	414,671		171,802		242,869		171,885		414,754
% of Actual APIs			-41.4%		58.6%				
% of Predicted APIs					58.6%		41.4%		

What's interesting about these results is that the partial FNs and partial FPs offset each other for each race / ethnicity group – leading to near identical characteristics of the Actual and Predicted groups. While this result may seem odd at first glance, it is, in fact, logical under both the BISG Continuous approach and the disparate impact scenario structure. In particular, unlike the BISG Classification approaches whereby the sample is partitioned into a large number of FN / FP subgroups with distinctly different characteristics (i.e., different average CBG median incomes and different average fee amounts), such partitioning is absent under the BISG Continuous approach since <u>every sample member contributes a partial FP to all other sample members</u>.

When the aggregate actual and predicted race / ethnicity distributions of the sample are aligned, then it must be the case that the aggregate number of actual and predicted sample members must be equal (within *de minimis* sampling error) for each race / ethnicity group. This can occur only if the aggregate partial FNs and partial FPs for each race / ethnicity group offset each other (which we showed previously in **Figure 32**). Finally, it can be proven mathematically, under these conditions, that the average income and fee amount must also be the same between the aggregate partial FP and FN sub-segments (again, within *de minimis* sampling error) and, therefore, be the same between the Actual and Predicted groups – all of which is borne out in **Figure 37** above for our synthetic sample.

So what is the cause of the BISG Continuous disparate impact disparity bias under these conditions?

When we evaluated the disparate treatment disparity bias under the BISG Continuous approach in the last section, we noted that the regression coefficient associated with the estimated disparate treatment disparity was unaffected by the BISG measurement error when the sample's aggregate actual and expected race / ethnicity distributions were aligned. Mathematically, this was because the two components that comprise the BISG Continuous regression coefficient (shown below) – the variance of the regressors X (i.e., the BISG probabilities) and the covariance between the regressors and the fee amounts Y were impacted proportionally.

$$\textit{Estimated Disparity} = \textit{Regression Coefficient} = \frac{Cov(Y,X)}{Var(X)}$$

where *Cov()* refers to covariance and *Var()* refers to variance.

The variance of the regressors is reduced because we use the raw BISG probabilities under the BISG Continuous approach instead of the actual 0-1 race / ethnicity indicator variables. Since the BISG probabilities vary between 0 and 1 (instead of only taking values of 0 or 1), the variance term is necessarily smaller. In terms of the covariance term, we note that, under disparate treatment, the fee amounts vary by the individual's actual race / ethnicity – regardless of their geo-surname segment (i.e., within a geo-surname segment, fee amounts can vary based on each sample member's specific race / ethnicity). Therefore, there was a high covariance between actual race / ethnicity and fee amounts. However, when we use the BISG probabilities instead of the actual race / ethnicity indicators, the covariance between race / ethnicity and fee amounts becomes diluted due to the partial FNs and partial FPs. For example, for Blacks, the partial FNs remove $100 fee amounts from Predicted Blacks and substitute in $0 fee amounts from the partial FPs. For Non-Blacks, the partial FNs remove $0 fee amounts and substitute in some $100 fee amounts from partial FPs (some of which come from the Black FNs). Overall, our analysis showed that the reduction in the covariance term was proportionally equal to the reduction in the variance term – yielding an unbiased estimate of the disparate treatment disparity.

Under the current <u>disparate impact</u> scenario, however, the impacts are decidedly different. While the reduction of the variance term remains the same, for the same reasons discussed above, the covariance term is now unaffected – as is illustrated in the tables in **Figure 37** above. There is no dilution in average fee amounts due to the partial FNs and partial FPs – which occurs because, under disparate impact, the disparity is driven by a socioeconomic attribute that is correlated across geographic segments – not by each sample member's specific race / ethnicity. Since all sample members within a given geographic segment receive the same fee amount – regardless of their actual race / ethnicity – the partial FPs are no longer distinctly different than the partial FNs and the covariance between fee amounts and BISG probabilities is unaffected. Overall, the unaffected covariance and the smaller variance causes the estimated disparate impact disparity to be larger than the true disparity – which is consistent with the empirical results we obtained throughout this section.

Finally, **Appendix M** presents the state-level disparate impact estimates under the BISG Continuous approach – along with the corresponding ground-truth disparate impact values and the calculations of state-level biases for Blacks, Hispanics, and APIs. Based on these results, we observe that the significant upward bias for Blacks is consistently present at the state level – ranging from a low of 61% for New York to a high of nearly 700% for Hawaii, with 37 states displaying bias levels greater than the national-level estimate in **Figure 34** (106.4%). Similarly, all but two states (Hawaii @ -31% and West Virginia @ -159%) display significant positive biases for Hispanics – ranging from a low of 54% for New York to a high of 509% for Virginia, with 36 states displaying bias levels greater than the national-level estimate. For APIs, five states exhibit positive biases ranging from 3% for New York to 142% for Nevada, with the remaining 46 states exhibiting negative biases ranging from -21% (Utah) to -7,245% (New Hampshire) and 37 states displaying bias levels greater than the national-level estimate. **Clearly, the disparate impact biases associated with the BISG Continuous approach can be even more severe when testing focuses on more disaggregated geographic areas.**

It is important to note that our disparate impact disparity estimates are based on a specific disparate impact scenario that is tied directly to the socioeconomic attribute of CBGs – the key micro-geography unit that drives the BISG proxy model. We expect other disparate impact scenarios that operate off characteristics that are less directly correlated with the BISG proxy model's underlying socioeconomic attributes will not have the same exact results. However, to the extent that there is some association between the two, it is logical that the covariance term will suffer some dilution – yielding positive, but perhaps smaller, bias than exhibited in our scenario. This is an area for which future research would be fruitful.

Can this disparate impact estimation bias be mitigated?

Based on our research of this disparate impact scenario structure, the answer to this question is yes, and we describe two alternative estimation approaches below – both of which are designed to shift the BISG measurement error out of the regression model's independent variables. However, we note up-front that while these approaches neutralize the measurement error bias from the BISG Continuous approach, they likely will not neutralize estimation biases caused by other factors – such as distributional misalignment. This would be another fruitful area for further research.

Alternative Estimation Approach #1: Bootstrap Regression

This estimation approach uses the set of BISG probabilities for each sample member to create multiple simulated datasets of "actual" sample members – such that the aggregate "actual" race / ethnicity distribution of the simulated sample is consistent with the expected distribution based on the BISG probabilities.[71] Each simulated dataset is used to estimate the disparate impact disparity estimates via OLS regression; however, instead of using the BISG probabilities as the regressors we instead use the "actual" indicator variables created by the simulation. The regression coefficients from each simulated dataset are accumulated and stored. After completion of the multiple iterations, the final disparity estimate (and its standard error) are computed from the mean and standard deviation of the accumulated regression coefficient dataset.

For example, consider the following sample member:

Fee Amount	BISG Black Probability	BISG API Probability	BISG Hispanic Probability	BISG Other Probability	BISG White Probability
$20	0.262	0.003	0.006	0.022	0.707

Rather than use directly this sample member's BISG probabilities in the OLS regression on fee amounts, we instead simulate (or "bootstrap") an actual sample member from this distribution using these probabilities. If we were to simulate this sample member's actual race / ethnicity 1,000 times, then we would expect the sample member to be designated as: (1) Black in 262 of those simulations, (2) API in 3 of those simulations, (3) Hispanic in 6 of those simulations, (4) Other in 22 of those simulations, and (5) White in 707 of those simulations. This is the process that we use with every sample member in our synthetic dataset. Therefore, each simulation yields a regression dataset of 10 million sample members and a unique set of simulated "actual" races / ethnicities associated with these sample members (i.e., a set of race / ethnicity dummy variables). Each set of race / ethnicity dummy variables are unique as they are randomly-drawn from the underlying set of BISG probabilities. For example, in the first simulation, the above sample member may be designated as White, but in the second simulation they may be designated as Black. However, each simulation's overall simulated race / ethnicity distribution is consistent with the underlying set of BISG probabilities.

Figure 38 below presents the results from the first ten bootstrap regressions on our synthetic dataset under the disparate impact scenario previously described.

[71] This is essentially the same Monte Carlo process we used to create the synthetic dataset for this study; however, it is executed multiple times – say 1,000.

Figure 38: Bootstrap Regression Coefficient Estimates

Simulation Number	Regression Constant Term	Regression Coefficient: API	Regression Coefficient: Black	Regression Coefficient: Hispanic
0	52.17	-3.52	15.06	6.95
1	52.17	-3.53	15.06	6.93
2	52.16	-3.52	15.11	6.97
3	52.17	-3.61	15.08	6.94
4	52.17	-3.59	15.03	6.95
5	52.17	-3.57	15.04	6.92
6	52.17	-3.57	15.09	6.93
7	52.17	-3.59	15.05	6.94
8	52.17	-3.59	15.06	6.94
9	52.17	-3.49	15.05	6.93
10	52.17	-3.55	15.06	6.97
Average	**52.17**	**-3.56**	**15.06**	**6.94**
True DI Disparities		**-3.51**	**15.09**	**6.94**

Here you can see how the bootstrap regression method generates coefficient estimates that are very close to the actual disparate impact disparities contained in **Figure 34** – even with only ten simulations. This is due to the very large sample size (10 million) that we used which produces an OLS estimator with high precision. For more typical sample sizes of, say, 10,000 – 50,000, a larger number of simulations will likely be needed in order to more fully "flesh out" the distributions of the coefficient estimates – as well as to provide enough granularity in the distributions to calculate accurately the standard errors for statistical significance testing. **Figure 39** below illustrates the distribution of the Black disparate impact disparity estimate for a sample size of 10,000 and using 5,000 dataset simulations.[72]

[72] Unfortunately, there is no hard and fast rule as to the number of simulations that should be performed. In the present example, the 5,000 simulations took only 2.5 minutes to execute on a laptop computer and, as seen in **Figure 39**, produced a fairly granular distribution with a well-defined mean and standard deviation.

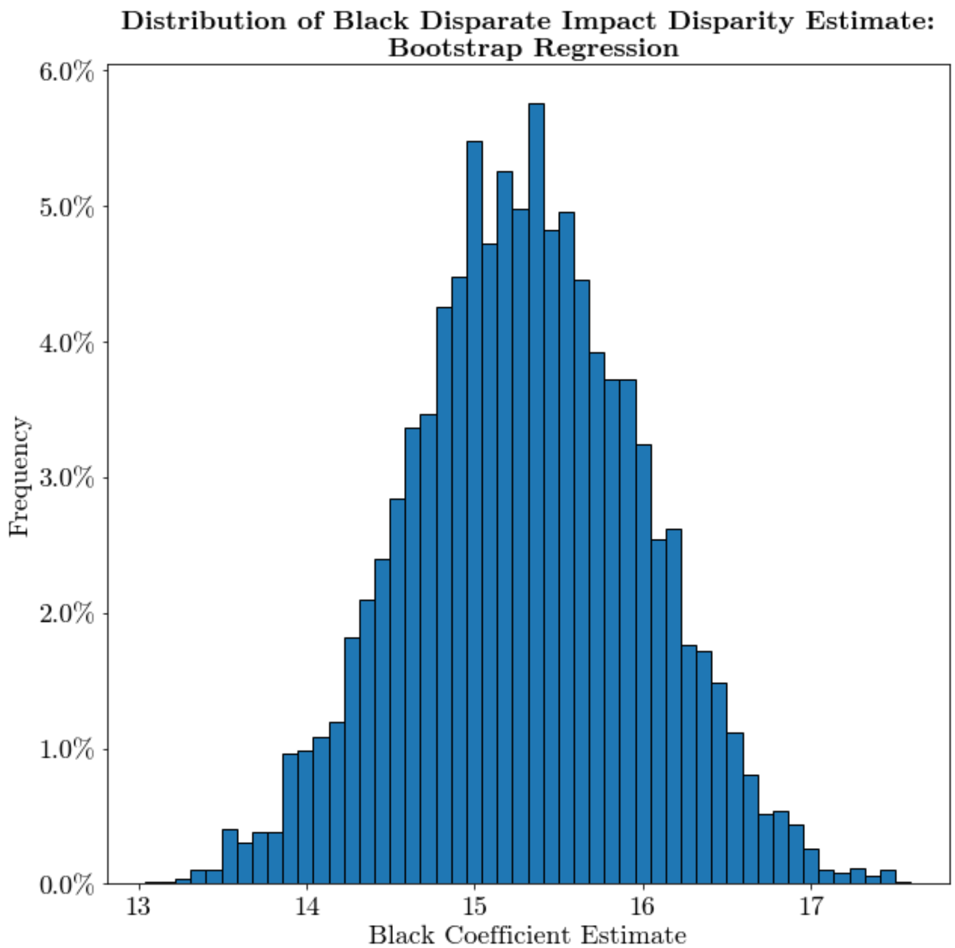


Figure 39

Alternative Estimation Approach #2: Proportional Regression

This estimation approach is conceptually similar to the bootstrap regression approach in that it also replaces the BISG probabilities with "actual" race / ethnicity indicators. However, instead of iterating over hundreds or thousands of simulated datasets of "actual" sample members, this approach requires only one regression – albeit one that uses an "exploded" dataset in which each sample member is represented five times – each corresponding to a different "actual" race / ethnicity which is then weighted by its corresponding BISG probability. As an example, let's return to the sample member above,

Fee Amount	BISG Black Probability	BISG API Probability	BISG Hispanic Probability	BISG Other Probability	BISG White Probability
$20	0.262	0.003	0.006	0.022	0.707

Under the proportional regression approach, we would replicate this sample member's record five times as shown below.

Figure 40: Representation of Sample Member Under Proportional Regression Approach

Assumed Race / Ethnicity	Weight	Fee Amount	Actual API	Actual Black	Actual Hispanic	Actual Other	Actual White
API	0.003	$ 20	1	0	0	0	0
Black	0.262	$ 20	0	1	0	0	0
Hispanic	0.006	$ 20	0	0	1	0	0
Other	0.022	$ 20	0	0	0	1	0
White	0.707	$ 20	0	0	0	0	1
	1.000						

Essentially, each "whole" individual is being disaggregated into five "pieces" with each piece representing the individual's probability of being one of the five BISG races / ethnicities. This process is applied to every sample member – so it results in an "exploded" dataset that is five times the size of the original dataset (i.e., a dataset with 100,000 sample members would increase to 500,000 records). Then, a weighted least squares regression model is executed on the exploded dataset with the BISG probabilities serving as the weights. **Figure 41** below presents the results of this proportional regression approach on our synthetic dataset under the disparate impact scenario.

Figure 41: Proportional Regression Coefficient Estimates

Method	Regression Constant Term	Regression Coefficient: API	Regression Coefficient: Black	Regression Coefficient: Hispanic
Proportional Regression	52.17	-3.55	15.06	6.94
True DI Disparities		**-3.51**	**15.09**	**6.94**

Here we can see that this approach produces disparate impact disparity estimates that are uncontaminated by the BISG measurement error.

Do these alternative estimation approaches produce the same results?

To answer this question, we applied the two alternative estimation approaches to 500 random samples of the synthetic dataset of size 10,000. For each sample, the Black disparate impact disparity estimate was estimated under both approaches and then compared to the sample's "ground truth" Black disparate impact disparity. **Figure 42** below plots the distributions of residual disparate impact estimation biases under the two methods across the 500 samples.

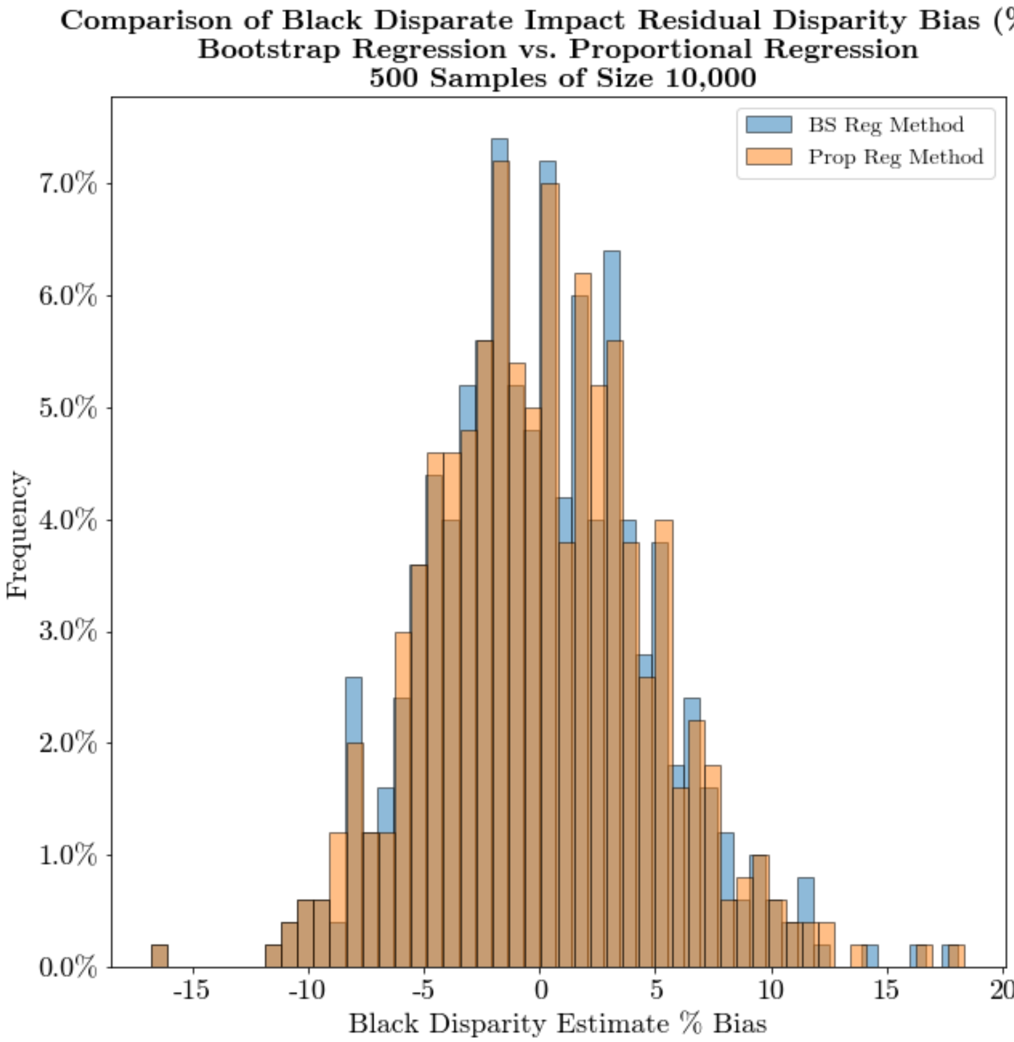


Figure 42

Here we see that, on average across the 500 random samples, both approaches produce unbiased Black disparate impact disparity estimates – that is, the average residual bias amounts are zero for both approaches. Additionally, we see that the distributions of these residual bias amounts are fairly identical across the two approaches – indicating no benefit of one versus the other.

We do note, however, that these distributions – while unbiased – indicate the presence of sampling error that may cause individual results based on relatively small sample sizes (10,000 here) to display either positive or negative residual bias amounts of +/- 9.2% with 95% confidence. Such uncertainty should be explicitly considered and communicated when drawing conclusions from the results of these alternative estimation approaches. Of course, with larger analysis sample sizes, these confidence intervals will shrink accordingly.

Do these alternative estimation approaches work at more disaggregated geographic levels?

We applied the Proportional Regression approach to the state-level disparate impact scenario data to assess whether the biases detailed in **Appendix M** could be eliminated. The short answer to this question is yes; however, as we have seen elsewhere in this study, caution needs to be exercised when analyzing geographies with relatively small sample sizes. In general, as shown in **Appendix N**, we observed significant "near complete" bias reduction when using the Proportional Regression approach at the state-level – with most states exhibiting relatively minor residual bias. Furthermore, even in those states where the bias reduction was not as "complete", we consistently found that the remaining bias levels were significantly less than the original biases.

So which of these two alternative estimation approaches is "better"?

Neither approach is clearly better or worse than the other, but there are certain considerations that may help guide one's selection.

- The Bootstrap regression approach produces an exact distribution of the regression coefficient standard errors – which is useful when evaluating the statistical significance of the disparate impact disparity estimate. Alternatively, while the Proportional Regression approach also generates the regression coefficient standard errors, extreme caution needs to be exercised to ensure that the standard errors are not misstated due to the use of an exploded weighted dataset (i.e., a dataset that is five times the size of the original dataset and where each record represents a fraction of an individual) since both dataset size and the weights should be used in the standard error formula.[73] Users should ensure that the statistical software generating the Proportional Regression results is specified correctly to produce accurate standard errors, or implement their own code to estimate them correctly.

- Depending on typical analysis sample sizes, users may find that one approach is relatively faster to execute than the other – although with modern computer hardware and software tools, differences in execution times may be mere seconds or minutes.

*   *   *

[73] Misstated regression coefficient standard errors will bias the corresponding statistical significance levels – thereby increasing the risk of a biased conclusion regarding the statistical significance of the estimated disparate impact disparity (or an incorrect measurement of the degree to which the result is statistically significant – i.e., the magnitude of the p-value).

Appendix A: State-Level Sample Counts of Actual Race / Ethnicity

State	White	Black	API	Hispanic	Other	Total
AK	14,026	755	929	2,258	4,229	**22,197**
AL	102,521	30,579	3,695	14,879	3,867	**155,541**
AR	67,692	10,643	2,354	10,304	2,944	**93,937**
AZ	131,175	10,593	7,051	42,139	12,044	**203,002**
CA	642,788	101,320	88,906	308,274	50,528	**1,191,816**
CO	116,829	6,964	5,670	28,389	5,035	**162,887**
CT	82,817	10,954	4,951	17,262	2,310	**118,294**
DC	8,743	7,457	798	3,231	951	**21,180**
DE	19,247	4,786	1,138	3,818	660	**29,649**
FL	384,502	82,262	21,588	128,531	13,432	**630,315**
GA	177,678	69,521	11,025	41,077	6,957	**306,258**
HI	16,839	1,074	9,347	5,661	12,476	**45,397**
IA	82,415	2,715	2,840	9,437	1,751	**99,158**
ID	38,575	390	1,365	6,798	1,674	**48,802**
IL	273,731	50,332	16,463	64,117	8,468	**413,111**
IN	161,207	15,197	5,789	22,062	4,171	**208,426**
KS	67,991	5,014	2,829	12,003	3,087	**90,924**
KY	113,755	9,428	3,506	11,852	2,949	**141,490**
LA	87,765	33,767	4,025	15,959	3,546	**145,062**
MA	163,334	13,711	10,240	27,215	4,429	**218,929**
MD	107,876	42,118	8,704	24,220	5,545	**188,463**
ME	39,014	467	1,210	2,765	1,344	**44,800**
MI	233,445	35,127	9,830	32,465	9,994	**320,861**
MN	135,526	7,856	6,810	16,355	4,996	**171,543**
MO	147,587	17,355	5,682	18,305	5,469	**194,398**
MS	56,544	25,317	2,068	8,548	1,745	**94,222**
MT	26,073	179	706	2,912	2,528	**32,398**
NC	198,132	52,191	9,507	40,248	9,089	**309,167**
ND	18,422	267	512	1,697	1,417	**22,315**
NE	45,775	2,301	1,697	7,161	1,348	**58,282**
NH	37,250	597	1,494	3,634	853	**43,828**
NJ	180,346	35,339	15,483	50,359	5,413	**286,940**
NM	37,392	2,040	1,570	18,106	6,941	**66,049**
NV	53,966	8,054	4,692	16,574	3,565	**86,851**
NY	388,684	87,557	33,914	112,644	17,251	**640,050**
OH	289,280	35,457	10,734	31,858	8,852	**376,181**
OK	78,842	7,612	2,825	14,425	16,195	**119,899**
OR	95,901	2,545	4,915	17,584	5,756	**126,701**
PA	322,241	36,519	14,608	41,510	7,939	**422,817**
RI	25,831	2,100	1,388	4,561	970	**34,850**
SC	95,344	31,567	4,163	16,796	3,203	**151,073**
SD	20,761	315	614	1,971	2,332	**25,993**
TN	148,934	26,266	5,557	21,056	4,599	**206,412**
TX	431,681	95,748	27,624	208,005	17,107	**780,165**
UT	61,983	1,102	3,045	12,278	2,343	**80,751**
VA	170,405	41,488	11,189	32,033	7,430	**262,545**
VT	18,221	244	590	1,525	627	**21,207**
WA	157,798	8,598	11,375	30,022	11,331	**219,124**
WI	146,618	8,740	5,974	19,072	4,695	**185,099**
WV	54,008	2,125	1,206	3,567	1,432	**62,338**
WY	14,528	200	476	2,380	719	**18,303**
Total	**6,592,038**	**1,084,853**	**414,671**	**1,589,902**	**318,536**	**10,000,000**

Appendix B: State-Level Actual Race / Ethnicity Distributions

State	White	Black	API	Hispanic	Other
AK	63.2%	3.4%	4.2%	10.2%	19.1%
AL	65.9%	19.7%	2.4%	9.6%	2.5%
AR	72.1%	11.3%	2.5%	11.0%	3.1%
AZ	64.6%	5.2%	3.5%	20.8%	5.9%
CA	53.9%	8.5%	7.5%	25.9%	4.2%
CO	71.7%	4.3%	3.5%	17.4%	3.1%
CT	70.0%	9.3%	4.2%	14.6%	2.0%
DC	41.3%	35.2%	3.8%	15.3%	4.5%
DE	64.9%	16.1%	3.8%	12.9%	2.2%
FL	61.0%	13.1%	3.4%	20.4%	2.1%
GA	58.0%	22.7%	3.6%	13.4%	2.3%
HI	37.1%	2.4%	20.6%	12.5%	27.5%
IA	83.1%	2.7%	2.9%	9.5%	1.8%
ID	79.0%	0.8%	2.8%	13.9%	3.4%
IL	66.3%	12.2%	4.0%	15.5%	2.0%
IN	77.3%	7.3%	2.8%	10.6%	2.0%
KS	74.8%	5.5%	3.1%	13.2%	3.4%
KY	80.4%	6.7%	2.5%	8.4%	2.1%
LA	60.5%	23.3%	2.8%	11.0%	2.4%
MA	74.6%	6.3%	4.7%	12.4%	2.0%
MD	57.2%	22.3%	4.6%	12.9%	2.9%
ME	87.1%	1.0%	2.7%	6.2%	3.0%
MI	72.8%	10.9%	3.1%	10.1%	3.1%
MN	79.0%	4.6%	4.0%	9.5%	2.9%
MO	75.9%	8.9%	2.9%	9.4%	2.8%
MS	60.0%	26.9%	2.2%	9.1%	1.9%
MT	80.5%	0.6%	2.2%	9.0%	7.8%
NC	64.1%	16.9%	3.1%	13.0%	2.9%
ND	82.6%	1.2%	2.3%	7.6%	6.3%
NE	78.5%	3.9%	2.9%	12.3%	2.3%
NH	85.0%	1.4%	3.4%	8.3%	1.9%
NJ	62.9%	12.3%	5.4%	17.6%	1.9%
NM	56.6%	3.1%	2.4%	27.4%	10.5%
NV	62.1%	9.3%	5.4%	19.1%	4.1%
NY	60.7%	13.7%	5.3%	17.6%	2.7%
OH	76.9%	9.4%	2.9%	8.5%	2.4%
OK	65.8%	6.3%	2.4%	12.0%	13.5%
OR	75.7%	2.0%	3.9%	13.9%	4.5%
PA	76.2%	8.6%	3.5%	9.8%	1.9%
RI	74.1%	6.0%	4.0%	13.1%	2.8%
SC	63.1%	20.9%	2.8%	11.1%	2.1%
SD	79.9%	1.2%	2.4%	7.6%	9.0%
TN	72.2%	12.7%	2.7%	10.2%	2.2%
TX	55.3%	12.3%	3.5%	26.7%	2.2%
UT	76.8%	1.4%	3.8%	15.2%	2.9%
VA	64.9%	15.8%	4.3%	12.2%	2.8%
VT	85.9%	1.2%	2.8%	7.2%	3.0%
WA	72.0%	3.9%	5.2%	13.7%	5.2%
WI	79.2%	4.7%	3.2%	10.3%	2.5%
WV	86.6%	3.4%	1.9%	5.7%	2.3%
WY	79.4%	1.1%	2.6%	13.0%	3.9%
Total	**65.92%**	**10.85%**	**4.15%**	**15.90%**	**3.19%**

Appendix C: US Census Demographics For 10 Racially-Diverse MSAs

MSA	Actual Race / Ethnicity	Avg CBG % API	Avg CBG % Black	Avg CBG % Hispanic	Avg CBG % White
Atlanta	API	8.7%	26.4%	10.1%	53.5%
	Black	4.0%	53.6%	8.7%	32.3%
	Hispanic	5.3%	30.8%	13.3%	49.3%
	White	5.0%	20.3%	7.9%	65.6%
Boston	API	10.8%	6.4%	7.3%	74.3%
	Black	8.0%	26.8%	15.1%	48.3%
	Hispanic	6.7%	9.3%	15.0%	67.7%
	White	6.2%	4.2%	5.9%	82.7%
Chicago	API	11.2%	9.7%	15.5%	62.6%
	Black	3.2%	57.2%	14.4%	24.1%
	Hispanic	5.0%	14.1%	30.3%	49.7%
	White	6.2%	7.1%	14.7%	71.0%
Detroit	API	6.4%	14.2%	2.6%	75.4%
	Black	2.1%	66.4%	2.5%	27.3%
	Hispanic	3.0%	20.0%	5.7%	69.8%
	White	3.4%	9.9%	3.0%	82.3%
Los Angeles	API	28.5%	4.8%	34.4%	30.8%
	Black	11.8%	21.3%	46.3%	19.0%
	Hispanic	13.0%	6.7%	56.7%	22.3%
	White	16.7%	4.2%	28.8%	48.3%
Miami	API	3.1%	16.9%	38.8%	40.3%
	Black	2.2%	42.5%	31.3%	22.8%
	Hispanic	1.9%	13.0%	60.9%	23.6%
	White	2.6%	11.9%	31.9%	52.7%
New York	API	21.6%	11.4%	19.6%	46.1%
	Black	6.9%	43.8%	28.5%	19.3%
	Hispanic	9.4%	18.3%	32.9%	38.3%
	White	10.3%	7.1%	15.1%	66.6%
Pittsburgh	API	4.1%	6.6%	1.3%	87.2%
	Black	1.8%	34.5%	1.5%	60.6%
	Hispanic	2.2%	9.7%	1.4%	85.7%
	White	1.7%	4.9%	1.0%	91.7%
San Diego	API	18.6%	5.2%	27.0%	46.8%
	Black	14.8%	10.1%	37.6%	35.1%
	Hispanic	11.1%	5.2%	37.6%	43.8%
	White	10.8%	3.8%	22.9%	59.9%
San Francisco	API	36.5%	6.8%	17.1%	37.2%
	Black	24.1%	19.5%	25.5%	28.2%
	Hispanic	23.0%	9.2%	25.7%	39.6%
	White	22.8%	5.6%	15.9%	53.3%

Appendix D: State-Level Recall (Left) and Precision (Right) Accuracy Rates Under BISG Max Classification Rule

State	API	Black	Hispanic	White	Average
AK	63.8%	13.0%	86.1%	93.3%	64.0%
AL	73.2%	63.3%	78.9%	90.5%	76.4%
AR	74.1%	57.6%	81.4%	92.6%	76.4%
AZ	76.4%	26.2%	75.7%	95.2%	68.4%
CA	55.3%	48.0%	75.0%	92.4%	67.7%
CO	75.8%	34.0%	78.8%	96.0%	71.1%
CT	73.4%	54.8%	81.3%	93.6%	75.8%
DC	71.9%	82.9%	86.0%	86.3%	81.8%
DE	75.0%	51.1%	87.7%	91.0%	76.2%
FL	76.9%	55.7%	81.8%	91.8%	76.6%
GA	70.8%	66.3%	84.6%	88.4%	77.5%
HI	45.3%	8.7%	78.3%	76.3%	52.2%
IA	72.0%	24.2%	74.4%	95.9%	66.6%
ID	78.7%	11.8%	85.4%	96.1%	68.0%
IL	68.4%	65.4%	77.9%	93.7%	76.4%
IN	72.0%	52.3%	77.8%	94.8%	74.2%
KS	72.0%	34.6%	82.0%	94.9%	70.9%
KY	72.7%	39.0%	70.8%	95.4%	69.5%
LA	74.5%	67.2%	84.4%	88.6%	78.7%
MA	67.3%	42.2%	78.6%	94.8%	70.7%
MD	66.8%	70.9%	82.7%	89.7%	77.5%
ME	71.5%	15.0%	41.8%	97.4%	56.4%
MI	69.9%	64.0%	80.0%	93.4%	76.8%
MN	68.6%	33.0%	76.8%	95.4%	68.4%
MO	73.8%	56.9%	76.6%	94.2%	75.4%
MS	74.4%	67.5%	77.9%	87.0%	76.7%
MT	71.7%	8.9%	80.3%	94.5%	63.8%
NC	74.0%	57.1%	85.2%	90.4%	76.7%
ND	69.9%	12.7%	65.6%	95.8%	61.0%
NE	73.9%	38.0%	80.6%	95.4%	72.0%
NH	72.6%	14.9%	65.1%	96.2%	62.2%
NJ	63.6%	60.9%	79.8%	92.8%	74.3%
NM	76.9%	15.6%	66.4%	94.5%	63.4%
NV	66.6%	36.8%	76.8%	95.1%	68.8%
NY	63.0%	71.5%	77.1%	92.8%	76.1%
OH	71.3%	56.4%	71.0%	94.5%	73.3%
OK	67.0%	36.9%	85.8%	93.4%	70.8%
OR	71.5%	19.1%	86.2%	95.9%	68.2%
PA	69.4%	57.9%	73.1%	95.0%	73.9%
RI	74.3%	39.5%	75.2%	94.6%	70.9%
SC	76.3%	59.8%	85.2%	88.5%	77.4%
SD	71.5%	12.7%	69.5%	95.5%	62.3%
TN	72.4%	59.2%	78.7%	93.2%	75.9%
TX	69.5%	55.9%	78.2%	90.4%	73.5%
UT	76.6%	14.5%	84.8%	96.4%	68.1%
VA	66.8%	53.2%	82.5%	91.4%	73.5%
VT	71.9%	15.2%	57.4%	96.6%	60.3%
WA	63.2%	24.9%	85.1%	95.5%	67.2%
WI	73.6%	54.7%	76.6%	95.4%	75.1%
WV	70.3%	24.6%	46.0%	97.4%	59.6%
WY	75.4%	16.5%	86.9%	96.1%	68.7%

State	API	Black	Hispanic	White	Average
AK	76.3%	50.8%	69.3%	80.5%	69.2%
AL	71.3%	75.9%	68.0%	85.6%	75.2%
AR	69.0%	72.9%	72.3%	88.3%	75.6%
AZ	77.9%	56.0%	84.3%	84.2%	75.6%
CA	79.6%	60.5%	81.3%	77.8%	74.8%
CO	77.2%	59.8%	85.4%	88.4%	77.7%
CT	81.7%	68.3%	78.4%	89.1%	79.4%
DC	79.7%	80.5%	83.1%	81.3%	81.2%
DE	81.1%	69.2%	77.7%	84.7%	78.2%
FL	78.3%	70.9%	82.6%	84.8%	79.2%
GA	76.6%	74.3%	77.2%	83.4%	77.9%
HI	68.0%	55.7%	74.0%	62.5%	65.1%
IA	73.6%	65.8%	70.5%	92.7%	75.7%
ID	74.4%	56.1%	80.3%	93.4%	76.1%
IL	79.3%	77.7%	78.3%	87.9%	80.8%
IN	73.4%	72.9%	73.5%	90.9%	77.7%
KS	74.0%	66.6%	76.5%	89.2%	76.6%
KY	72.6%	70.6%	68.2%	90.3%	75.4%
LA	74.0%	76.2%	71.7%	84.6%	76.6%
MA	80.1%	68.7%	74.2%	89.6%	78.2%
MD	80.9%	76.4%	76.1%	83.9%	79.3%
ME	71.7%	57.9%	58.3%	92.0%	70.0%
MI	73.5%	77.8%	68.5%	89.8%	77.4%
MN	78.7%	64.9%	69.6%	90.6%	76.0%
MO	73.7%	75.8%	68.1%	90.0%	76.9%
MS	72.6%	75.6%	67.1%	83.2%	74.6%
MT	65.3%	61.5%	64.7%	92.2%	70.9%
NC	74.8%	70.5%	77.6%	85.1%	77.0%
ND	68.6%	61.8%	63.2%	91.8%	71.3%
NE	75.4%	70.1%	76.2%	91.9%	78.4%
NH	75.9%	60.5%	65.2%	93.0%	73.6%
NJ	81.7%	71.7%	79.8%	86.2%	79.9%
NM	70.7%	54.3%	85.5%	77.7%	72.1%
NV	83.6%	59.6%	87.1%	81.1%	77.9%
NY	78.4%	73.9%	78.8%	86.8%	79.5%
OH	74.2%	74.5%	67.0%	89.9%	76.4%
OK	69.3%	63.5%	72.4%	80.8%	71.5%
OR	76.8%	61.1%	79.0%	90.6%	76.9%
PA	77.2%	74.8%	71.8%	90.5%	78.6%
RI	78.9%	56.1%	76.1%	88.8%	75.0%
SC	73.5%	72.0%	72.3%	84.1%	75.5%
SD	70.6%	54.8%	64.5%	90.8%	70.2%
TN	73.4%	76.2%	70.6%	88.6%	77.2%
TX	77.3%	67.8%	83.6%	81.2%	77.5%
UT	79.5%	61.3%	82.7%	92.5%	79.0%
VA	79.9%	69.2%	75.7%	83.8%	77.1%
VT	72.0%	68.5%	60.9%	92.5%	73.5%
WA	80.5%	59.6%	76.9%	87.6%	76.1%
WI	77.4%	74.3%	72.1%	92.0%	78.9%
WV	67.6%	64.2%	63.4%	91.5%	71.7%
WY	73.6%	66.0%	78.8%	93.6%	78.0%

Appendix D: State-Level F1 Accuracy Rates Under BISG Max Classification Rule

State	API	Black	Hispanic	White	Average
AK	69.5%	20.7%	76.8%	86.4%	63.3%
AL	72.2%	69.0%	73.1%	88.0%	75.6%
AR	71.5%	64.3%	76.5%	90.4%	75.7%
AZ	77.1%	35.7%	79.8%	89.4%	70.5%
CA	65.3%	53.5%	78.0%	84.5%	70.3%
CO	76.5%	43.3%	82.0%	92.1%	73.5%
CT	77.4%	60.9%	79.9%	91.3%	77.3%
DC	75.6%	81.7%	84.6%	83.7%	81.4%
DE	77.9%	58.8%	82.4%	87.7%	76.7%
FL	77.6%	62.4%	82.2%	88.2%	77.6%
GA	73.6%	70.1%	80.7%	85.9%	77.6%
HI	54.4%	15.0%	76.1%	68.8%	53.6%
IA	72.8%	35.4%	72.4%	94.3%	68.7%
ID	76.5%	19.5%	82.8%	94.7%	68.4%
IL	73.4%	71.0%	78.1%	90.7%	78.3%
IN	72.7%	60.9%	75.6%	92.8%	75.5%
KS	73.0%	45.5%	79.1%	92.0%	72.4%
KY	72.6%	50.3%	69.5%	92.8%	71.3%
LA	74.3%	71.4%	77.5%	86.5%	77.5%
MA	73.1%	52.3%	76.4%	92.1%	73.5%
MD	73.2%	73.6%	79.3%	86.7%	78.2%
ME	71.6%	23.8%	48.7%	94.6%	59.7%
MI	71.7%	70.2%	73.8%	91.6%	76.8%
MN	73.3%	43.8%	73.0%	92.9%	70.7%
MO	73.8%	65.0%	72.1%	92.1%	75.7%
MS	73.5%	71.3%	72.1%	85.1%	75.5%
MT	68.3%	15.6%	71.6%	93.3%	62.2%
NC	74.4%	63.1%	81.2%	87.7%	76.6%
ND	69.2%	21.1%	64.4%	93.7%	62.1%
NE	74.6%	49.3%	78.3%	93.6%	74.0%
NH	74.2%	23.9%	65.2%	94.6%	64.5%
NJ	71.5%	65.9%	79.8%	89.4%	76.6%
NM	73.7%	24.3%	74.7%	85.3%	64.5%
NV	74.1%	45.5%	81.6%	87.5%	72.2%
NY	69.9%	72.6%	77.9%	89.7%	77.5%
OH	72.7%	64.2%	69.0%	92.2%	74.5%
OK	68.1%	46.7%	78.5%	86.6%	70.0%
OR	74.1%	29.1%	82.4%	93.2%	69.7%
PA	73.1%	65.3%	72.4%	92.7%	75.9%
RI	76.5%	46.3%	75.7%	91.6%	72.5%
SC	74.9%	65.3%	78.2%	86.2%	76.2%
SD	71.0%	20.6%	66.9%	93.1%	62.9%
TN	72.9%	66.7%	74.4%	90.8%	76.2%
TX	73.2%	61.3%	80.8%	85.5%	75.2%
UT	78.0%	23.5%	83.7%	94.4%	69.9%
VA	72.8%	60.1%	78.9%	87.5%	74.8%
VT	71.9%	24.8%	59.1%	94.5%	62.6%
WA	70.8%	35.1%	80.8%	91.3%	69.5%
WI	75.5%	63.0%	74.3%	93.7%	76.6%
WV	68.9%	35.5%	53.3%	94.4%	63.0%
WY	74.5%	26.4%	82.7%	94.8%	69.6%

Appendix E: State-Level Accuracy Rates Under BISG 80% Classification Rule

State-Level Recall Accuracy Rates
BISG 80% Classification Rule

State	API	Black	Hispanic	White	Average
AK	81.6%	15.3%	86.9%	98.5%	70.6%
AL	88.0%	79.2%	87.1%	98.4%	88.2%
AR	85.5%	68.0%	87.3%	98.8%	84.9%
AZ	84.1%	15.6%	89.0%	98.3%	71.7%
CA	78.2%	62.6%	91.7%	98.0%	82.6%
CO	82.5%	19.7%	86.4%	98.4%	71.8%
CT	85.3%	54.1%	89.7%	98.9%	82.0%
DC	82.7%	94.3%	93.8%	95.4%	91.5%
DE	87.3%	62.2%	91.8%	98.2%	84.9%
FL	86.2%	69.4%	91.8%	97.7%	86.3%
GA	84.7%	81.8%	92.1%	97.2%	89.0%
HI	89.3%	27.8%	96.6%	95.7%	77.4%
IA	84.1%	14.2%	82.7%	99.5%	70.1%
ID	84.2%	1.5%	86.1%	99.0%	67.7%
IL	79.7%	80.1%	88.9%	98.6%	86.8%
IN	83.2%	56.0%	86.6%	99.1%	81.2%
KS	82.0%	34.1%	87.9%	99.0%	75.7%
KY	86.1%	39.6%	80.1%	99.5%	76.3%
LA	87.9%	82.0%	90.0%	97.7%	89.4%
MA	81.2%	43.3%	87.8%	99.2%	77.9%
MD	82.1%	85.9%	91.9%	97.7%	89.4%
ME	84.6%	7.0%	40.2%	99.9%	57.9%
MI	83.7%	77.2%	85.1%	99.2%	86.3%
MN	83.7%	23.2%	84.3%	99.4%	72.7%
MO	86.3%	65.2%	82.0%	99.3%	83.2%
MS	89.4%	83.3%	88.2%	97.5%	89.6%
MT	85.0%	2.2%	69.8%	99.6%	64.1%
NC	86.1%	66.3%	91.3%	98.1%	85.4%
ND	86.1%	5.3%	58.3%	99.7%	62.4%
NE	82.9%	35.3%	85.2%	99.1%	75.6%
NH	83.8%	5.6%	74.5%	99.7%	65.9%
NJ	78.1%	72.6%	90.3%	98.3%	84.8%
NM	83.4%	10.6%	86.8%	98.0%	69.7%
NV	77.4%	36.4%	90.1%	98.4%	75.6%
NY	78.7%	82.4%	90.1%	98.5%	87.4%
OH	84.7%	63.9%	81.8%	99.3%	82.4%
OK	83.0%	52.1%	91.3%	98.9%	81.3%
OR	79.2%	6.1%	86.6%	98.9%	67.7%
PA	83.9%	66.1%	85.2%	99.3%	83.6%
RI	86.8%	20.6%	89.5%	99.2%	74.1%
SC	90.0%	72.4%	90.8%	97.9%	87.8%
SD	86.5%	5.3%	70.6%	99.6%	65.5%
TN	85.1%	71.0%	87.8%	99.0%	85.7%
TX	82.2%	70.3%	91.7%	97.3%	85.4%
UT	81.9%	4.5%	86.5%	98.6%	67.9%
VA	81.0%	65.2%	90.8%	98.4%	83.9%
VT	84.9%	5.5%	43.0%	99.8%	58.3%
WA	76.6%	11.6%	87.9%	99.0%	68.8%
WI	85.5%	61.9%	86.4%	99.4%	83.3%
WV	83.0%	15.3%	49.6%	99.9%	61.9%
WY	80.6%	2.7%	87.2%	98.8%	67.3%

State-Level Addressable Sample Sizes
BISG 80% Classification Rule

State	API	Black	Hispanic	White	Average
AK	46.8%	14.7%	27.3%	64.6%	38.4%
AL	40.4%	43.0%	29.7%	72.2%	46.3%
AR	42.1%	39.6%	36.9%	79.8%	49.6%
AZ	58.9%	23.5%	68.3%	76.2%	56.7%
CA	51.0%	19.0%	60.2%	60.0%	47.6%
CO	58.5%	26.3%	73.9%	84.6%	60.8%
CT	62.4%	33.1%	52.3%	81.8%	57.4%
DC	60.0%	60.8%	64.4%	63.5%	62.2%
DE	55.3%	31.0%	54.3%	70.4%	52.7%
FL	57.7%	36.6%	63.7%	73.0%	57.8%
GA	52.5%	41.6%	51.2%	67.1%	53.1%
HI	25.1%	3.4%	38.3%	18.1%	21.2%
IA	47.3%	46.4%	34.2%	88.5%	54.1%
ID	48.6%	66.7%	62.3%	93.8%	67.8%
IL	57.4%	51.4%	54.2%	79.9%	60.7%
IN	45.2%	43.5%	41.3%	85.0%	53.7%
KS	52.5%	37.4%	49.5%	83.9%	55.8%
KY	41.2%	41.7%	24.8%	84.0%	47.9%
LA	46.7%	45.1%	37.3%	69.6%	49.7%
MA	58.8%	40.1%	44.3%	83.9%	56.8%
MD	56.4%	48.2%	46.9%	69.4%	55.2%
ME	34.4%	58.2%	8.7%	88.8%	47.5%
MI	46.2%	51.5%	27.0%	82.8%	51.9%
MN	55.4%	36.0%	31.3%	85.4%	52.0%
MO	45.2%	48.6%	24.9%	83.9%	50.7%
MS	42.0%	42.0%	24.4%	65.1%	43.4%
MT	28.3%	51.4%	11.1%	87.9%	44.7%
NC	50.4%	32.6%	52.9%	72.0%	52.0%
ND	39.3%	56.6%	8.5%	87.7%	48.0%
NE	51.6%	39.7%	47.3%	87.3%	56.5%
NH	50.1%	63.0%	18.1%	89.8%	55.2%
NJ	60.3%	41.3%	57.9%	76.5%	59.0%
NM	45.9%	16.6%	62.7%	57.6%	45.7%
NV	65.3%	17.1%	72.9%	67.8%	55.8%
NY	53.6%	48.1%	55.9%	78.4%	59.0%
OH	45.2%	44.8%	23.6%	82.9%	49.1%
OK	37.0%	23.6%	39.8%	64.1%	41.1%
OR	57.4%	38.8%	59.1%	89.6%	61.2%
PA	51.1%	48.9%	33.9%	84.5%	54.6%
RI	56.9%	28.0%	44.7%	83.0%	53.1%
SC	45.8%	34.4%	38.8%	68.3%	46.8%
SD	35.0%	47.6%	15.7%	86.0%	46.1%
TN	44.5%	46.9%	33.6%	80.3%	51.3%
TX	54.5%	29.5%	64.7%	65.2%	53.5%
UT	64.8%	50.6%	69.4%	92.2%	69.3%
VA	56.9%	31.5%	47.7%	70.0%	51.5%
VT	39.3%	66.8%	7.5%	89.7%	50.8%
WA	58.9%	27.8%	50.8%	83.0%	55.1%
WI	51.7%	49.1%	34.7%	87.7%	55.8%
WV	33.3%	43.8%	13.7%	87.1%	44.5%
WY	34.7%	56.0%	60.9%	93.4%	61.3%

Appendix E: State-Level Accuracy Rates Under BISG 80% Classification Rule (Continued)

State-Level Precision Accuracy Rates
BISG 80% Classification Rule

State	API	Black	Hispanic	White	Average
AK	92.4%	89.5%	83.8%	89.4%	88.8%
AL	89.7%	92.3%	87.1%	94.1%	90.8%
AR	87.4%	90.2%	88.6%	94.4%	90.1%
AZ	90.4%	85.8%	93.1%	92.1%	90.3%
CA	93.9%	87.6%	93.5%	90.0%	91.3%
CO	90.6%	85.3%	91.2%	93.1%	90.1%
CT	92.7%	91.4%	91.5%	94.8%	92.6%
DC	92.5%	92.9%	90.7%	91.7%	92.0%
DE	92.4%	91.0%	88.8%	93.4%	91.4%
FL	90.4%	91.2%	92.2%	93.1%	91.8%
GA	91.0%	91.5%	89.5%	93.1%	91.3%
HI	94.8%	90.9%	86.4%	87.8%	90.0%
IA	90.1%	89.1%	87.7%	96.4%	90.8%
ID	90.6%	66.7%	90.2%	95.0%	85.6%
IL	92.7%	94.0%	91.8%	94.4%	93.2%
IN	90.5%	89.9%	89.5%	95.6%	91.4%
KS	89.9%	89.2%	89.3%	94.1%	90.6%
KY	89.5%	90.4%	88.0%	95.6%	90.9%
LA	89.8%	92.1%	88.3%	93.4%	90.9%
MA	92.3%	89.4%	90.9%	94.8%	91.9%
MD	93.2%	92.1%	89.5%	93.2%	92.0%
ME	90.3%	95.0%	82.9%	96.7%	91.2%
MI	90.4%	93.2%	88.4%	95.2%	91.8%
MN	91.7%	88.9%	87.6%	95.4%	90.9%
MO	89.5%	92.3%	87.2%	95.1%	91.0%
MS	89.8%	91.5%	86.8%	93.8%	90.5%
MT	88.1%	66.7%	85.3%	95.6%	83.9%
NC	90.3%	89.7%	89.2%	93.5%	90.7%
ND	89.6%	80.0%	86.6%	96.0%	88.1%
NE	90.9%	88.0%	88.6%	95.7%	90.8%
NH	90.0%	100.0%	90.9%	96.8%	94.4%
NJ	93.4%	91.1%	91.6%	93.5%	92.4%
NM	89.2%	85.7%	95.9%	88.9%	89.9%
NV	92.8%	87.1%	92.9%	90.6%	90.9%
NY	93.0%	90.6%	92.3%	94.1%	92.5%
OH	90.2%	91.3%	88.1%	95.5%	91.3%
OK	89.7%	89.5%	88.7%	90.0%	89.5%
OR	90.6%	92.3%	89.4%	93.1%	91.4%
PA	91.5%	91.7%	89.9%	95.8%	92.2%
RI	90.7%	89.6%	91.6%	95.0%	91.7%
SC	90.1%	90.4%	88.4%	93.4%	90.6%
SD	89.0%	88.9%	85.2%	95.5%	89.6%
TN	90.1%	91.9%	89.5%	94.9%	91.6%
TX	91.9%	90.0%	93.8%	91.7%	91.8%
UT	89.1%	92.6%	90.1%	94.6%	91.6%
VA	92.5%	89.8%	89.1%	93.0%	91.1%
VT	87.6%	100.0%	83.1%	96.4%	91.8%
WA	92.1%	88.2%	89.1%	92.3%	90.4%
WI	92.0%	90.9%	89.6%	96.3%	92.2%
WV	86.5%	91.6%	91.0%	96.3%	91.4%
WY	92.4%	100.0%	87.5%	95.4%	93.8%

State-Level F1 Accuracy Rates
BISG 80% Classification Rule

State	API	Black	Hispanic	White	Average
AK	86.7%	26.2%	85.3%	93.7%	73.0%
AL	88.8%	85.2%	87.1%	96.2%	89.3%
AR	86.4%	77.5%	87.9%	96.5%	87.1%
AZ	87.1%	26.4%	91.0%	95.1%	74.9%
CA	85.3%	73.0%	92.6%	93.8%	86.2%
CO	86.4%	32.0%	88.8%	95.7%	75.7%
CT	88.9%	68.0%	90.6%	96.8%	86.1%
DC	87.3%	93.6%	92.2%	93.5%	91.7%
DE	89.8%	73.9%	90.3%	95.7%	87.4%
FL	88.3%	78.8%	92.0%	95.4%	88.6%
GA	87.7%	86.4%	90.8%	95.1%	90.0%
HI	92.0%	42.6%	91.2%	91.6%	79.3%
IA	87.0%	24.5%	85.1%	97.9%	73.6%
ID	87.3%	3.0%	88.1%	96.9%	68.8%
IL	85.7%	86.5%	90.3%	96.5%	89.7%
IN	86.7%	69.0%	88.0%	97.4%	85.3%
KS	85.7%	49.3%	88.6%	96.5%	80.0%
KY	87.8%	55.1%	83.8%	97.5%	81.1%
LA	88.9%	86.7%	89.1%	95.5%	90.1%
MA	86.4%	58.4%	89.3%	97.0%	82.7%
MD	87.3%	88.9%	90.7%	95.4%	90.6%
ME	87.3%	13.0%	54.2%	98.3%	63.2%
MI	86.9%	84.4%	86.7%	97.1%	88.8%
MN	87.5%	36.8%	85.9%	97.4%	76.9%
MO	87.9%	76.4%	84.5%	97.1%	86.5%
MS	89.6%	87.2%	87.5%	95.6%	90.0%
MT	86.5%	4.2%	76.7%	97.6%	66.3%
NC	88.1%	76.2%	90.2%	95.7%	87.6%
ND	87.8%	9.9%	69.7%	97.8%	66.3%
NE	86.7%	50.4%	86.9%	97.4%	80.3%
NH	86.8%	10.6%	81.9%	98.2%	69.4%
NJ	85.1%	80.8%	90.9%	95.9%	88.2%
NM	86.2%	18.9%	91.1%	93.2%	72.4%
NV	84.4%	51.4%	91.5%	94.3%	80.4%
NY	85.3%	86.3%	91.2%	96.2%	89.7%
OH	87.3%	75.2%	84.8%	97.4%	86.2%
OK	86.2%	65.9%	90.0%	94.2%	84.1%
OR	84.5%	11.4%	87.9%	95.9%	69.9%
PA	87.5%	76.8%	87.5%	97.5%	87.3%
RI	88.7%	33.5%	90.5%	97.1%	77.5%
SC	90.1%	80.4%	89.6%	95.6%	88.9%
SD	87.7%	10.1%	77.2%	97.5%	68.1%
TN	87.5%	80.1%	88.6%	96.9%	88.3%
TX	86.8%	79.0%	92.7%	94.4%	88.2%
UT	85.4%	8.5%	88.2%	96.6%	69.7%
VA	86.3%	75.6%	90.0%	95.6%	86.9%
VT	86.2%	10.5%	56.6%	98.1%	62.9%
WA	83.6%	20.5%	88.5%	95.6%	72.0%
WI	88.6%	73.6%	88.0%	97.8%	87.0%
WV	84.7%	26.2%	64.2%	98.1%	68.3%
WY	86.1%	5.2%	87.4%	97.1%	68.9%

Appendix F: Comparative Characteristics of Actual vs. Predicted Race / Ethnicities
BISG 50% Threshold Rule

Whites	Total Actuals	–	False Negatives	=	True Positives	+	False Positives	=	Total Predicted
Average CBG White %	76.3%		52.7%		78.5%		63.6%		76.6%
Average CBG Black %	7.3%		18.0%		6.3%		10.8%		6.9%
Average CBG Hispanic %	10.3%		21.0%		9.3%		16.8%		10.3%
Average CBG API %	4.4%		6.2%		4.2%		6.6%		4.5%
Average Surname White %	76.4%		42.2%		79.6%		69.1%		78.3%
Average Max Probability	86.9%		59.9%		89.4%		74.0%		87.5%
Average Median HH Income	$61,331		$53,504		$62,054		$57,901		$61,536
Sample Counts	6,592,038		559,101		6,032,937		860,833		6,893,770
% of Actual Whites			-8.5%		91.5%				
% of Predicted Whites					87.5%		12.5%		

Blacks	Total Actuals	–	False Negatives	=	True Positives	+	False Positives	=	Total Predicted
Average CBG White %	38.8%		55.6%		23.9%		33.4%		26.2%
Average CBG Black %	38.7%		18.9%		56.3%		41.7%		52.7%
Average CBG Hispanic %	16.3%		17.6%		15.1%		18.6%		15.9%
Average CBG API %	4.6%		6.1%		3.3%		4.6%		3.6%
Average Surname Black %	27.1%		22.4%		31.3%		29.2%		30.8%
Average Max Probability	72.6%		65.8%		78.8%		65.9%		75.6%
Average Median HH Income	$47,221		$54,138		$41,034		$44,044		$41,767
Sample Counts	1,084,853		510,922		573,931		185,238		759,169
% of Actual Blacks			-47.1%		52.9%				
% of Predicted Blacks					75.6%		24.4%		

Hispanics	Total Actuals	–	False Negatives	=	True Positives	+	False Positives	=	Total Predicted
Average CBG White %	52.5%		48.7%		53.7%		61.8%		55.3%
Average CBG Black %	11.7%		11.3%		11.8%		10.6%		11.6%
Average CBG Hispanic %	28.6%		32.0%		27.5%		21.4%		26.3%
Average CBG API %	5.6%		6.3%		5.4%		4.6%		5.2%
Average Surname Hispanic %	66.9%		24.1%		80.9%		72.7%		79.3%
Average Max Probability	78.4%		63.6%		83.2%		69.3%		80.5%
Average Median HH Income	$54,823		$52,585		$55,558		$57,856		$56,006
Sample Counts	1,589,902		392,948		1,196,954		289,589		1,486,543
% of Actual Hispanics			-24.7%		75.3%				
% of Predicted Hispanics					80.5%		19.5%		

APIs	Total Actuals	–	False Negatives	=	True Positives	+	False Positives	=	Total Predicted
Average CBG White %	59.7%		53.5%		63.5%		65.6%		63.9%
Average CBG Black %	9.8%		9.4%		10.0%		10.4%		10.1%
Average CBG Hispanic %	15.5%		16.4%		14.9%		13.8%		14.7%
Average CBG API %	13.1%		18.5%		9.7%		8.4%		9.5%
Average Surname API %	59.4%		15.9%		85.9%		76.5%		84.2%
Average Max Probability	76.7%		64.2%		84.3%		70.3%		81.8%
Average Median HH Income	$65,667		$66,809		$64,969		$59,912		$64,049
Sample Counts	414,671		157,297		257,374		57,220		314,594
% of Actual APIs			-37.9%		62.1%				
% of Predicted APIs					81.8%		18.2%		

Appendix G: State-Level Disparate Treatment Disparity Biases Under BISG Classification
Disparate Treatment Scenario: Blacks = $100 Fee, All Others = $0 Fee

State	Actual Blacks in Sample	Estimated Black Fee Disparity Bias (BISG 80%)	Estimated Black Fee Disparity Bias (BISG 50%)	Estimated Black Fee Disparity Bias (BISG Max)
National	**1,084,853**	**-11.0%**	**-29.7%**	**-34.2%**
AK	755	-11.4%	-42.2%	-52.9%
AL	30,579	-11.1%	-31.5%	-33.4%
AR	10,643	-12.1%	-31.0%	-33.0%
AZ	10,593	-16.1%	-38.4%	-48.9%
CA	101,320	-13.9%	-35.2%	-45.2%
CO	6,964	-16.0%	-38.0%	-43.7%
CT	10,954	-10.8%	-32.6%	-37.1%
DC	7,457	-10.4%	-26.3%	-30.2%
DE	4,786	-12.5%	-38.5%	-41.4%
FL	82,262	-11.7%	-32.3%	-37.0%
GA	69,521	-12.3%	-34.2%	-36.8%
HI	1,074	-9.6%	-29.1%	-46.5%
IA	2,715	-12.4%	-31.3%	-36.6%
ID	390	-34.0%	-31.2%	-44.7%
IL	50,332	-8.1%	-23.7%	-27.5%
IN	15,197	-12.0%	-28.9%	-31.1%
KS	5,014	-12.8%	-32.8%	-37.7%
KY	9,428	-11.9%	-31.3%	-34.0%
LA	33,767	-11.9%	-32.3%	-34.4%
MA	13,711	-12.7%	-30.2%	-35.6%
MD	42,118	-11.2%	-29.5%	-32.9%
ME	467	-5.7%	-35.0%	-43.1%
MI	35,127	-8.8%	-23.8%	-26.8%
MN	7,856	-12.9%	-34.6%	-38.6%
MO	17,355	-9.9%	-26.0%	-28.6%
MS	25,317	-12.9%	-35.0%	-36.8%
MT	179	-33.7%	-31.3%	-39.0%
NC	52,191	-13.9%	-36.7%	-39.3%
ND	267	-20.8%	-38.4%	-39.3%
NE	2,301	-13.4%	-28.5%	-32.8%
NH	597	-1.0%	-32.4%	-40.8%
NJ	35,339	-11.4%	-29.4%	-34.7%
NM	2,040	-15.5%	-38.3%	-49.0%
NV	8,054	-15.0%	-39.8%	-48.1%
NY	87,557	-11.5%	-26.0%	-31.1%
OH	35,457	-10.9%	-27.7%	-30.1%
OK	7,612	-12.0%	-34.5%	-41.4%
OR	2,545	-8.7%	-34.3%	-40.8%
PA	36,519	-10.3%	-26.0%	-29.3%
RI	2,100	-12.4%	-40.9%	-48.4%
SC	31,567	-13.8%	-37.7%	-39.5%
SD	315	-11.9%	-42.2%	-46.4%
TN	26,266	-10.8%	-28.0%	-30.0%
TX	95,748	-12.6%	-34.1%	-40.0%
UT	1,102	-8.3%	-26.8%	-40.1%
VA	41,488	-13.6%	-37.3%	-40.6%
VT	244	-0.9%	-18.5%	-32.5%
WA	8,598	-13.2%	-36.0%	-44.0%
WI	8,740	-10.3%	-24.2%	-28.2%
WV	2,125	-10.0%	-36.1%	-38.5%
WY	200	-0.8%	-32.3%	-35.1%

Appendix H: Potential Bias of State-Level BISG Continuous Disparate Treatment Estimates
Assumptions: Maximum +2% White Skew & $100 Black Fee Disparate Treatment

State	Estimated Black Fee Disparity (Skew Scenario)	Estimated Black Fee Disparity (Ground Truth)	Estimated Bias: Black Fee Disparity	Change in Actual Number of Blacks	Black Distribution %: Skewed	Black Distribution %: Original	Change in Actual Number of Whites	White Distribution %: Skewed	White Distribution %: Original
MT	$93.38	$104.66	-10.8%	(110)	0.2%	0.6%	110	80.8%	80.5%
ID	$94.51	$103.85	-9.0%	(193)	0.4%	0.8%	193	79.4%	79.0%
WY	$94.57	$103.35	-8.5%	(89)	0.6%	1.1%	89	79.9%	79.4%
ND	$85.06	$92.44	-8.0%	(108)	0.7%	1.2%	108	83.0%	82.6%
UT	$95.18	$103.29	-7.9%	(444)	0.8%	1.4%	444	77.3%	76.8%
VT	$96.64	$104.36	-7.4%	(109)	0.6%	1.2%	109	86.4%	85.9%
HI	$95.92	$103.44	-7.3%	(334)	1.6%	2.4%	334	37.8%	37.1%
SD	$84.58	$91.21	-7.3%	(123)	0.7%	1.2%	123	80.3%	79.9%
NH	$90.38	$97.28	-7.1%	(269)	0.7%	1.4%	269	85.6%	85.0%
ME	$87.99	$94.60	-7.0%	(203)	0.6%	1.0%	203	87.5%	87.1%
NM	$92.19	$98.28	-6.2%	(561)	2.2%	3.1%	561	57.5%	56.6%
OR	$97.90	$103.87	-5.8%	(739)	1.4%	2.0%	739	76.3%	75.7%
AK	$92.01	$97.44	-5.6%	(176)	2.6%	3.4%	176	64.0%	63.2%
IA	$94.50	$99.01	-4.6%	(630)	2.1%	2.7%	630	83.8%	83.1%
WV	$93.79	$97.98	-4.3%	(433)	2.7%	3.4%	433	87.3%	86.6%
WA	$97.14	$101.24	-4.0%	(1,862)	3.1%	3.9%	1,862	72.9%	72.0%
AZ	$96.77	$100.75	-3.9%	(1,963)	4.3%	5.2%	1,963	65.6%	64.6%
MN	$97.09	$100.71	-3.6%	(1,442)	3.7%	4.6%	1,442	79.8%	79.0%
NE	$97.80	$101.32	-3.5%	(404)	3.3%	3.9%	404	79.2%	78.5%
CO	$96.58	$99.94	-3.4%	(1,285)	3.5%	4.3%	1,285	72.5%	71.7%
KS	$96.35	$99.33	-3.0%	(840)	4.6%	5.5%	840	75.7%	74.8%
OK	$96.19	$99.11	-3.0%	(1,249)	5.3%	6.3%	1,249	66.8%	65.8%
RI	$94.87	$97.56	-2.8%	(358)	5.0%	6.0%	358	75.1%	74.1%
KY	$96.52	$99.25	-2.8%	(1,378)	5.7%	6.7%	1,378	81.4%	80.4%
WI	$97.10	$99.73	-2.6%	(1,247)	4.0%	4.7%	1,247	79.9%	79.2%
NV	$97.53	$100.12	-2.6%	(1,023)	8.1%	9.3%	1,023	63.3%	62.1%
MA	$97.91	$100.32	-2.4%	(2,156)	5.3%	6.3%	2,156	75.6%	74.6%
IN	$97.08	$99.45	-2.4%	(1,921)	6.4%	7.3%	1,921	78.3%	77.3%
CA	$97.31	$99.57	-2.3%	(12,968)	7.4%	8.5%	12,968	55.0%	53.9%
AR	$97.34	$99.44	-2.1%	(1,013)	10.3%	11.3%	1,013	73.1%	72.1%
MO	$98.28	$100.40	-2.1%	(1,902)	7.9%	8.9%	1,902	76.9%	75.9%
OH	$97.96	$100.02	-2.1%	(3,877)	8.4%	9.4%	3,877	77.9%	76.9%
PA	$97.76	$99.77	-2.0%	(4,120)	7.7%	8.6%	4,120	77.2%	76.2%
CT	$99.78	$101.82	-2.0%	(1,212)	8.2%	9.3%	1,212	71.0%	70.0%
MI	$98.13	$100.03	-1.9%	(3,183)	10.0%	10.9%	3,183	73.7%	72.8%
TN	$98.09	$99.87	-1.8%	(2,495)	11.5%	12.7%	2,495	73.4%	72.2%
TX	$98.25	$100.02	-1.8%	(9,028)	11.1%	12.3%	9,028	56.5%	55.3%
IL	$98.50	$100.22	-1.7%	(4,292)	11.1%	12.2%	4,292	67.3%	66.3%
FL	$98.45	$100.10	-1.7%	(7,756)	11.8%	13.1%	7,756	62.2%	61.0%
NY	$98.12	$99.73	-1.6%	(7,308)	12.5%	13.7%	7,308	61.9%	60.7%
NJ	$98.23	$99.85	-1.6%	(3,338)	11.2%	12.3%	3,338	64.0%	62.9%
VA	$98.09	$99.61	-1.5%	(3,635)	14.4%	15.8%	3,635	66.3%	64.9%
NC	$98.60	$100.12	-1.5%	(4,394)	15.5%	16.9%	4,394	65.5%	64.1%
DE	$98.81	$100.28	-1.5%	(399)	14.8%	16.1%	399	66.3%	64.9%
AL	$99.02	$100.46	-1.4%	(2,198)	18.2%	19.7%	2,198	67.3%	65.9%
SC	$99.26	$100.51	-1.2%	(2,341)	19.3%	20.9%	2,341	64.7%	63.1%
GA	$98.62	$99.79	-1.2%	(4,657)	21.2%	22.7%	4,657	59.5%	58.0%
MD	$98.98	$100.15	-1.2%	(2,757)	20.9%	22.3%	2,757	58.7%	57.2%
LA	$99.14	$100.32	-1.2%	(2,269)	21.7%	23.3%	2,269	62.1%	60.5%
MS	$98.82	$99.86	-1.0%	(1,574)	25.2%	26.9%	1,574	61.7%	60.0%
DC	$98.46	$99.32	-0.9%	(319)	33.7%	35.2%	319	42.8%	41.3%

Appendix I: Comparison of Disparate Treatment Biases Under Alternative Distributional Misalignments and Race / Ethnicity Proxies

Maximum Distribution Skew: Actual Whites	Estimated Black Fee Disparity Bias				Method With Least Bias
	BISG Continuous	BISG Max	BISG 80% Threshold	BISG 50% Threshold	
-15%	-7.0%	-38.7%	-20.9%	-34.7%	BISG Continuous
-14%	-6.3%	-38.3%	-20.0%	-34.2%	BISG Continuous
-13%	-5.7%	-37.8%	-19.0%	-33.7%	BISG Continuous
-12%	-5.1%	-37.4%	-18.2%	-33.3%	BISG Continuous
-11%	-4.4%	-37.0%	-17.3%	-32.8%	BISG Continuous
-10%	-3.8%	-36.6%	-16.5%	-32.4%	BISG Continuous
-9%	-3.3%	-36.3%	-15.6%	-32.0%	BISG Continuous
-8%	-2.7%	-35.9%	-14.9%	-31.6%	BISG Continuous
-7%	-2.2%	-35.6%	-14.2%	-31.2%	BISG Continuous
-6%	-1.8%	-35.3%	-13.5%	-30.9%	BISG Continuous
-5%	-1.3%	-35.0%	-12.9%	-30.6%	BISG Continuous
-4%	-0.9%	-34.8%	-12.3%	-30.3%	BISG Continuous
-3%	-0.6%	-34.6%	-11.8%	-30.1%	BISG Continuous
-2%	-0.3%	-34.4%	-11.4%	-29.9%	BISG Continuous
-1%	-0.1%	-34.3%	-11.1%	-29.7%	BISG Continuous
0%	0.0%	-34.2%	-11.0%	-29.7%	BISG Continuous
1%	-0.9%	-34.6%	-11.4%	-30.0%	BISG Continuous
2%	-2.0%	-35.1%	-12.1%	-30.6%	BISG Continuous
3%	-3.3%	-35.7%	-12.8%	-31.2%	BISG Continuous
4%	-4.7%	-36.4%	-13.6%	-31.8%	BISG Continuous
5%	-6.1%	-37.1%	-14.4%	-32.5%	BISG Continuous
6%	-7.6%	-37.8%	-15.3%	-33.3%	BISG Continuous
7%	-9.1%	-38.5%	-16.1%	-34.0%	BISG Continuous
8%	-10.6%	-39.3%	-17.0%	-34.8%	BISG Continuous
9%	-12.1%	-40.1%	-18.0%	-35.6%	BISG Continuous
10%	-13.6%	-40.9%	-18.9%	-36.5%	BISG Continuous
11%	-15.2%	-41.7%	-19.9%	-37.3%	BISG Continuous
12%	-16.7%	-42.6%	-20.9%	-38.1%	BISG Continuous
13%	-18.3%	-43.4%	-21.8%	-39.0%	BISG Continuous
14%	-19.8%	-44.2%	-22.8%	-39.9%	BISG Continuous
15%	-21.4%	-45.1%	-23.8%	-40.7%	BISG Continuous

Appendix J: Comparative Characteristics of Actual vs. Predicted Hispanics and APIs Under BISG 80% Threshold Rule

Hispanics	Total Actuals	–	False Negatives	=	True Positives	+	False Positives	=	Total Predicted
Average Median HH Income	$54,823		$55,702		$53,874		$58,171		$54,223
Average $ Fee Amount	$59.10		$58.53		$59.72		$55.63		$59.39
Sample Counts	1,589,902		824,535		765,367		67,586		832,953
% of Actual Hispanics			-51.9%		48.1%				
% of Predicted Hispanics					91.9%		8.1%		

APIs	Total Actuals	–	False Negatives	=	True Positives	+	False Positives	=	Total Predicted
Average Median HH Income	$65,667		$63,837		$68,139		$64,030		$67,810
Average $ Fee Amount	$48.65		$50.53		$46.12		$49.50		$46.39
Sample Counts	414,671		238,301		176,370		15,358		191,728
% of Actual APIs			-57.5%		42.5%				
% of Predicted APIs					92.0%		8.0%		

Appendix K: Comparative Characteristics of Actual vs. Predicted Hispanics and APIs Under BISG Max Rule

Hispanics	Total Actuals	–	False Negatives	=	True Positives	+	False Positives	=	Total Predicted
Average Median HH Income	$54,823		$52,861		$55,370		$56,561		$55,630
Average $ Fee Amount	$59.10		$60.54		$58.70		$58.05		$58.56
Sample Counts	1,589,902		346,636		1,243,266		346,750		1,590,016
% of Actual Hispanics			-21.8%		78.2%				
% of Predicted Hispanics					78.2%		21.8%		

APIs	Total Actuals	–	False Negatives	=	True Positives	+	False Positives	=	Total Predicted
Average Median HH Income	$65,667		$68,038		$64,455		$58,900		$63,197
Average $ Fee Amount	$48.65		$46.84		$49.58		$54.89		$50.78
Sample Counts	414,671		140,197		274,474		80,365		354,839
% of Actual APIs			-33.8%		66.2%				
% of Predicted APIs					77.4%		22.6%		

Appendix L: Comparative Characteristics of Actual vs. Predicted Groups Under BISG 50% Threshold Rule

Whites	Total Actuals	–	False Negatives	=	True Positives	+	False Positives	=	Total Predicted
Average Median HH Income	$61,331		$53,504		$62,054		$57,901		$61,536
Average $ Fee Amount	$52.17		$60.61		$51.39		$55.47		$51.89
Sample Counts	6,592,038		559,101		6,032,937		860,833		6,893,770
% of Actual Whites			-8.5%		91.5%				
% of Predicted Whites					87.5%		12.5%		

Blacks	Total Actuals	–	False Negatives	=	True Positives	+	False Positives	=	Total Predicted
Average Median HH Income	$47,221		$54,138		$41,034		$44,044		$41,767
Average $ Fee Amount	$67.25		$59.43		$74.25		$70.71		$73.39
Sample Counts	1,084,853		510,922		573,931		185,238		759,169
% of Actual Blacks			-47.1%		52.9%				
% of Predicted Blacks					75.6%		24.4%		

Hispanics	Total Actuals	–	False Negatives	=	True Positives	+	False Positives	=	Total Predicted
Average Median HH Income	$54,823		$52,585		$55,558		$57,856		$56,006
Average $ Fee Amount	$59.10		$61.03		$58.47		$56.63		$58.11
Sample Counts	1,589,902		392,948		1,196,954		289,589		1,486,543
% of Actual Hispanics			-24.7%		75.3%				
% of Predicted Hispanics					80.5%		19.5%		

APIs	Total Actuals	–	False Negatives	=	True Positives	+	False Positives	=	Total Predicted
Average Median HH Income	$65,667		$66,809		$64,969		$59,912		$64,049
Average $ Fee Amount	$48.65		$47.99		$49.06		$53.75		$49.91
Sample Counts	414,671		157,297		257,374		57,220		314,594
% of Actual APIs			-37.9%		62.1%				
% of Predicted APIs					81.8%		18.2%		

Appendix M: State-Level Disparate Impact Disparity Biases Under BISG Continuous Approach

State	Actual Number of Blacks	BISG Continuous Black Disparate Impact $ Disparity	Actual Black Disparate Impact $ Disparity	Estimated Black Disparate Impact Disparity Bias	Actual Number of Hispanics	BISG Continuous Hispanic Disparate Impact $ Disparity	Actual Hispanic Disparate Impact $ Disparity	Estimated Hispanic Disparate Impact Disparity Bias	Actual Number of APIs	BISG Continuous API Disparate Impact $ Disparity	Actual API Disparate Impact $ Disparity	Estimated API Disparate Impact Disparity Bias
AK	755	$28.48	$8.43	238.0%	2,258	$5.85	$2.39	144.8%	929	$1.56	$1.51	3.6%
AL	30,579	$28.88	$14.13	104.5%	14,879	$10.84	$5.11	112.1%	3,695	($9.33)	($3.32)	-180.7%
AR	10,643	$23.54	$11.11	111.8%	10,304	$5.80	$2.37	144.6%	2,354	($7.29)	($1.97)	-270.4%
AZ	10,593	$27.69	$9.22	200.2%	42,139	$15.93	$9.23	72.5%	7,051	($4.30)	($1.58)	-172.4%
CA	101,320	$38.34	$15.65	144.9%	308,274	$25.64	$13.29	93.0%	88,906	$2.69	$1.67	60.7%
CO	6,964	$36.92	$13.74	168.8%	28,389	$13.33	$6.98	91.0%	5,670	($5.04)	($0.10)	-4728.4%
CT	10,954	$55.97	$26.49	111.3%	17,262	$21.83	$12.37	76.5%	4,951	$1.04	$2.92	-64.5%
DC	7,457	$52.27	$30.75	70.0%	3,231	$27.77	$14.25	94.9%	798	$12.82	$8.33	53.9%
DE	4,786	$27.06	$10.78	151.0%	3,818	$14.35	$6.59	117.8%	1,138	($5.54)	($2.64)	-109.6%
FL	82,262	$27.13	$12.40	118.8%	128,531	$10.55	$5.63	87.4%	21,588	($2.39)	($0.97)	-147.6%
GA	69,521	$25.96	$12.45	108.5%	41,077	$10.51	$4.75	121.3%	11,025	($5.85)	($2.72)	-115.2%
HI	1,074	$28.44	$3.56	698.2%	5,661	$0.71	$1.03	-30.9%	9,347	($2.84)	($0.87)	-225.7%
IA	2,715	$43.19	$12.89	235.1%	9,437	$7.89	$3.76	109.8%	2,840	($3.85)	($0.19)	-1948.8%
ID	390	$17.49	$3.27	435.1%	6,798	$3.53	$1.42	148.7%	1,365	($4.99)	$0.09	-5642.3%
IL	50,332	$37.01	$21.88	69.2%	64,117	$13.79	$7.99	72.6%	16,463	($5.54)	($1.49)	-273.0%
IN	15,197	$32.34	$15.77	105.1%	22,062	$9.50	$5.01	89.7%	5,789	($7.88)	($0.96)	-722.2%
KS	5,014	$34.61	$12.41	178.8%	12,003	$10.05	$5.11	96.4%	2,829	($8.40)	($2.29)	-266.0%
KY	9,428	$20.08	$8.12	147.2%	11,852	$2.07	$0.67	208.0%	3,506	($15.12)	($6.20)	-143.7%
LA	33,767	$34.51	$17.16	101.1%	15,959	$10.31	$4.47	130.5%	4,025	($0.84)	$0.48	-275.7%
MA	13,711	$47.17	$20.20	133.5%	27,215	$20.27	$11.28	79.7%	10,240	$0.42	$3.05	-86.2%
MD	42,118	$22.92	$12.84	78.6%	24,220	$9.13	$4.58	99.1%	8,704	($2.48)	($0.49)	-408.3%
ME	467	$30.65	$7.71	297.6%	2,765	$2.86	$1.24	130.0%	1,210	($7.48)	($1.21)	-516.5%
MI	35,127	$33.84	$20.02	69.0%	32,465	$10.81	$5.53	95.7%	9,830	($11.90)	($3.40)	-250.1%
MN	7,856	$36.84	$13.25	178.0%	16,355	$8.11	$3.69	119.4%	6,810	($4.04)	($0.60)	-571.5%
MO	17,355	$26.30	$13.87	89.6%	18,305	$6.38	$2.98	114.2%	5,682	($10.52)	($2.93)	-259.2%
MS	25,317	$27.69	$12.96	113.6%	8,548	$9.94	$4.18	137.6%	2,068	($5.06)	($2.78)	-81.9%
MT	179	$28.15	$7.05	299.5%	2,912	$4.19	$2.39	75.1%	706	($6.04)	($0.68)	-781.7%
NC	52,191	$26.62	$11.60	129.6%	40,248	$9.75	$4.59	112.2%	9,507	($6.07)	($2.46)	-146.8%
ND	267	$41.79	$10.49	298.2%	1,697	$2.21	$1.43	54.6%	512	($0.87)	$1.98	-143.9%
NE	2,301	$35.76	$14.03	154.9%	7,161	$9.29	$5.01	85.5%	1,697	($5.07)	($0.53)	-864.1%
NH	597	$34.37	$8.72	294.3%	3,634	$3.42	$1.40	143.4%	1,494	($3.21)	$0.04	-7245.0%
NJ	35,339	$43.19	$22.54	91.6%	50,359	$23.19	$13.63	70.1%	15,483	($1.02)	$0.91	-212.0%
NM	2,040	$36.03	$7.51	379.7%	18,106	$17.58	$8.44	108.2%	1,570	($3.55)	($2.36)	-50.9%
NV	8,054	$28.88	$9.52	203.3%	16,574	$13.14	$6.33	107.5%	4,692	$0.24	($0.56)	142.1%
NY	87,557	$32.26	$20.06	60.8%	112,644	$19.26	$12.52	53.9%	33,914	$6.20	$6.02	3.0%
OH	35,457	$35.95	$18.67	92.5%	31,858	$10.58	$5.13	106.1%	10,734	($12.62)	($3.67)	-243.6%
OK	7,612	$35.50	$11.58	206.5%	14,425	$9.90	$3.95	150.6%	2,825	($11.96)	($3.97)	-201.1%
OR	2,545	$21.59	$7.24	198.3%	17,584	$6.72	$3.17	112.0%	4,915	($7.99)	($2.24)	-256.7%
PA	36,519	$35.05	$19.40	80.7%	41,510	$11.09	$5.80	91.4%	14,608	($8.31)	($1.97)	-322.5%
RI	2,100	$67.96	$26.33	158.1%	4,561	$22.85	$13.49	69.5%	1,388	$2.73	$4.92	-44.4%
SC	31,567	$31.95	$13.64	134.2%	16,796	$10.61	$4.27	148.2%	4,163	($5.49)	($2.25)	-144.1%
SD	315	$32.71	$6.69	388.8%	1,971	$5.64	$3.33	69.3%	614	($2.40)	$1.32	-282.1%
TN	26,266	$18.85	$9.61	96.2%	21,056	$5.65	$2.43	132.8%	5,557	($14.08)	($5.36)	-162.6%
TX	95,748	$31.83	$13.51	135.6%	208,005	$27.65	$15.48	78.6%	27,624	($4.49)	($1.73)	-159.6%
UT	1,102	$71.52	$12.82	457.8%	12,278	$7.80	$4.30	81.4%	3,045	$1.94	$2.45	-20.8%
VA	41,488	$27.46	$11.26	143.7%	32,033	$2.21	($0.54)	509.4%	11,189	($10.91)	($7.12)	-53.1%
VT	244	$34.08	$6.40	432.3%	1,525	$2.53	$0.45	456.5%	590	($7.03)	($0.73)	-861.4%
WA	8,598	$26.02	$8.99	189.3%	30,022	$10.64	$5.07	109.9%	11,375	($6.88)	($2.60)	-164.1%
WI	8,740	$42.78	$22.52	89.9%	19,072	$9.75	$5.50	77.4%	5,974	($1.15)	$1.11	-203.8%
WV	2,125	$6.97	$2.64	163.4%	3,567	($3.70)	($1.43)	-158.7%	1,206	($12.69)	($5.10)	-148.9%
WY	200	$38.77	$7.42	422.6%	2,380	$2.39	$1.07	122.6%	476	($6.53)	($1.78)	-266.9%

Appendix N: Impact of Proportional Regression Approach on State-Level Disparate Impact Bias Under BISG Continuous Approach

State	Actual Number of Blacks	Estimated Black Disparate Impact Disparity Bias	Black Disparity Bias Under Proportional Regression Approach	Actual Number of Hispanics	Estimated Hispanic Disparate Impact Disparity Bias	Hispanic Disparity Bias Under Proportional Regression Approach	Actual Number of APIs	Estimated API Disparate Impact Disparity Bias	API Disparity Bias Under Proportional Regression Approach
AK	755	238.0%	-5.2%	2,258	144.8%	-16.5%	929	3.6%	0.6%
AL	30,579	104.5%	-0.2%	14,879	112.1%	-1.1%	3,695	-180.7%	-5.9%
AR	10,643	111.8%	-0.5%	10,304	144.6%	9.7%	2,354	-270.4%	-15.9%
AZ	10,593	200.2%	-5.2%	42,139	72.5%	-0.4%	7,051	-172.4%	-15.1%
CA	101,320	144.9%	-0.1%	308,274	93.0%	0.1%	88,906	60.7%	4.0%
CO	6,964	168.8%	0.9%	28,389	91.0%	1.8%	5,670	-4728.4%	-74.3%
CT	10,954	111.3%	-0.5%	17,262	76.5%	1.0%	4,951	-64.5%	8.4%
DC	7,457	70.0%	0.0%	3,231	94.9%	0.1%	798	53.9%	1.2%
DE	4,786	151.0%	0.8%	3,818	117.8%	6.4%	1,138	-109.6%	37.4%
FL	82,262	118.8%	-0.7%	128,531	87.4%	-0.7%	21,588	-147.6%	-0.7%
GA	69,521	108.5%	0.4%	41,077	121.3%	1.0%	11,025	-115.2%	-3.5%
HI	1,074	698.2%	-12.4%	5,661	-30.9%	15.3%	9,347	-225.7%	2.7%
IA	2,715	235.1%	-2.0%	9,437	109.8%	2.7%	2,840	-1948.8%	-39.1%
ID	390	435.1%	42.7%	6,798	148.7%	30.3%	1,365	-5642.3%	-637.5%
IL	50,332	69.2%	0.1%	64,117	72.6%	-0.7%	16,463	-273.0%	-21.0%
IN	15,197	105.1%	1.0%	22,062	89.7%	-0.7%	5,789	-722.2%	-12.3%
KS	5,014	178.8%	1.7%	12,003	96.4%	-2.9%	2,829	-266.0%	-1.6%
KY	9,428	147.2%	0.7%	11,852	208.0%	-24.3%	3,506	-143.7%	-4.1%
LA	33,767	101.1%	-0.2%	15,959	130.5%	0.4%	4,025	-275.7%	7.9%
MA	13,711	133.5%	0.1%	27,215	79.7%	-0.3%	10,240	-86.2%	-9.7%
MD	42,118	78.6%	-2.8%	24,220	99.1%	-3.7%	8,704	-408.3%	-39.6%
ME	467	297.6%	-1.8%	2,765	130.0%	0.2%	1,210	-516.5%	32.7%
MI	35,127	69.0%	-0.3%	32,465	95.7%	1.3%	9,830	-250.1%	-2.8%
MN	7,856	178.0%	2.1%	16,355	119.4%	2.0%	6,810	-571.5%	29.5%
MO	17,355	89.6%	0.2%	18,305	114.2%	0.3%	5,682	-259.2%	14.5%
MS	25,317	113.6%	0.7%	8,548	137.6%	-1.0%	2,068	-81.9%	2.7%
MT	179	299.5%	-18.1%	2,912	75.1%	6.2%	706	-781.7%	24.1%
NC	52,191	129.6%	0.1%	40,248	112.2%	0.0%	9,507	-146.8%	-11.6%
ND	267	298.2%	-20.9%	1,697	54.6%	10.5%	512	-143.9%	-29.4%
NE	2,301	154.9%	-1.5%	7,161	85.5%	0.2%	1,697	-864.1%	-57.2%
NH	597	294.3%	-10.2%	3,634	143.4%	-3.8%	1,494	-7245.0%	73.9%
NJ	35,339	91.6%	0.1%	50,359	70.1%	0.6%	15,483	-212.0%	-1.7%
NM	2,040	379.7%	4.5%	18,106	108.2%	0.3%	1,570	-50.9%	24.7%
NV	8,054	203.3%	-0.2%	16,574	107.5%	0.2%	4,692	142.1%	64.3%
NY	87,557	60.8%	-0.1%	112,644	53.9%	0.1%	33,914	3.0%	-2.0%
OH	35,457	92.5%	-0.3%	31,858	106.1%	-1.9%	10,734	-243.6%	4.3%
OK	7,612	206.5%	3.1%	14,425	150.6%	2.2%	2,825	-201.1%	-0.7%
OR	2,545	198.3%	-6.6%	17,584	112.0%	-1.0%	4,915	-256.7%	-12.1%
PA	36,519	80.7%	0.0%	41,510	91.4%	3.3%	14,608	-322.5%	-3.1%
RI	2,100	158.1%	-0.8%	4,561	69.5%	-1.7%	1,388	-44.4%	-4.1%
SC	31,567	134.2%	-0.3%	16,796	148.2%	-0.6%	4,163	-144.1%	-14.1%
SD	315	388.8%	18.6%	1,971	69.3%	10.4%	614	-282.1%	-18.5%
TN	26,266	96.2%	2.2%	21,056	132.8%	6.2%	5,557	-162.6%	5.8%
TX	95,748	135.6%	0.3%	208,005	78.6%	-0.2%	27,624	-159.6%	-10.4%
UT	1,102	457.8%	4.9%	12,278	81.4%	0.0%	3,045	-20.8%	4.3%
VA	41,488	143.7%	-0.8%	32,033	509.4%	14.5%	11,189	-53.1%	-4.4%
VT	244	432.3%	16.1%	1,525	456.5%	100.6%	590	-861.4%	-87.4%
WA	8,598	189.3%	-2.6%	30,022	109.9%	-1.3%	11,375	-164.1%	0.6%
WI	8,740	89.9%	-2.0%	19,072	77.4%	0.1%	5,974	-203.8%	28.9%
WV	2,125	163.4%	-6.0%	3,567	-158.7%	-19.2%	1,206	-148.9%	4.5%
WY	200	422.6%	1.3%	2,380	122.6%	3.1%	476	-266.9%	30.6%

About the Author

Ric has over twenty years' experience assisting US consumer lenders of all sizes with the application of statistically-based methods to identify potential disparate treatment and disparate impact risks in loan pricing, credit decisioning, redlining, marketing, and product selection – including identification and remediation of algorithmic bias in decision models. Ric has also aided his clients in responding to federal- and state-level fair lending investigations and enforcement actions – including matters before the Civil Rights Division of the US Department of Justice, the Consumer Financial Protection Bureau, the Office of the Comptroller of the Currency, the FDIC, and the Federal Reserve Board.

Ric earned a PhD in Economics from the University of Rochester and a B.S. in Finance and Economics, summa cum laude, from the State University of New York College at Oswego.

Richard R. Pace
ricpace@gmail.com

Also Available