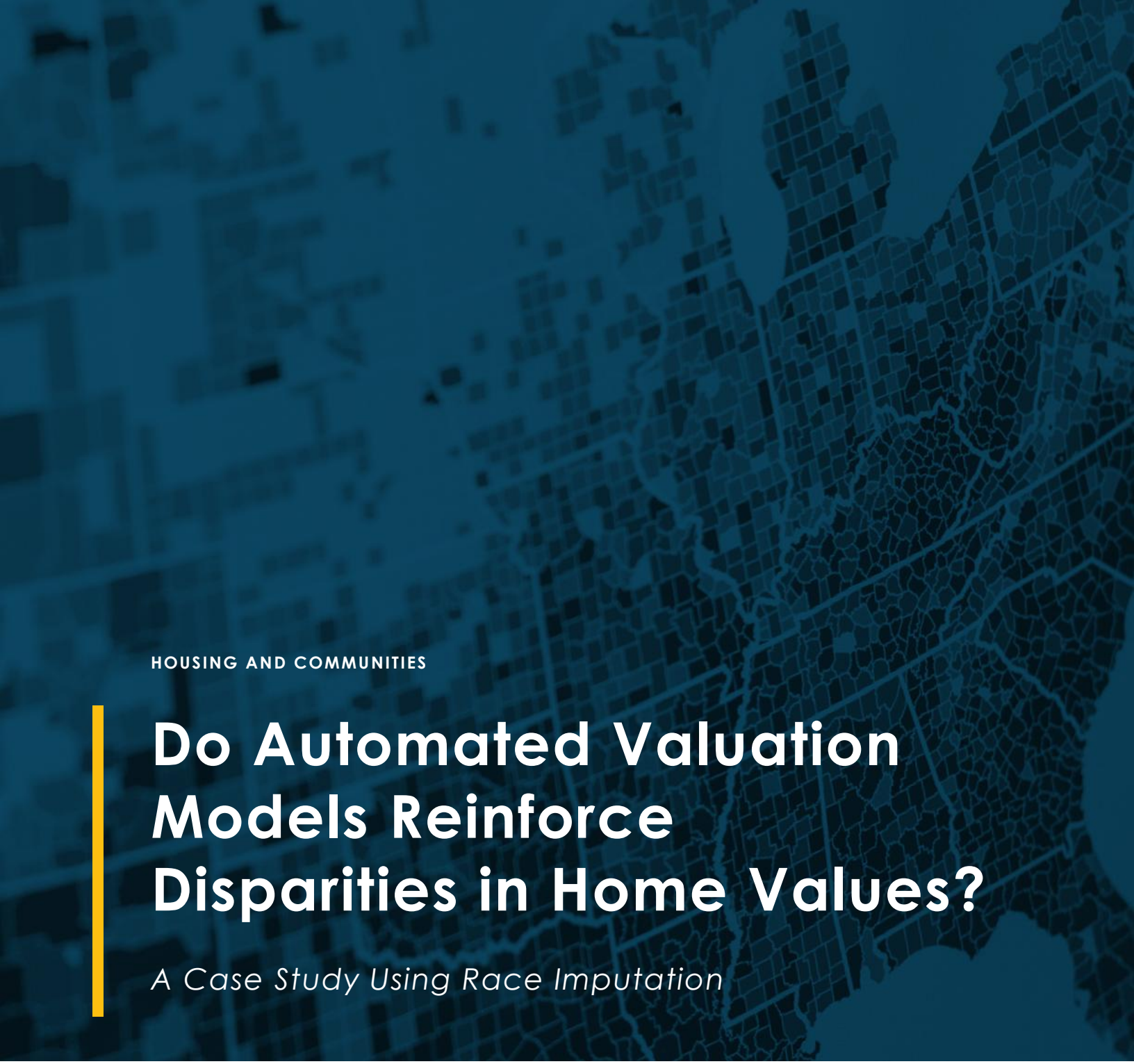


HOUSING AND COMMUNITIES

Do Automated Valuation Models Reinforce Disparities in Home Values?

A Case Study Using Race Imputation

Linna Zhu **Judah Axelrod** **Amalie Zinn**

URBAN INSTITUTE

RESEARCH REPORT

October 2025



ABOUT THE URBAN INSTITUTE

The Urban Institute is a nonprofit research organization founded on one simple idea: To improve lives and strengthen communities, we need practices and policies that work. For more than 50 years, that has been our charge. By equipping changemakers with evidence and solutions, together we can create a future where every person and community has the opportunity and power to thrive.

Copyright © October 2025. Urban Institute. Permission is granted for reproduction of this file, with attribution to the Urban Institute.

Contents

Acknowledgments iv

Executive Summary v

AVMs Perpetuate Long-Standing Disparities in Home Values 1

- Methodology and Data 4
- Key Findings 10
- Policy Implications 13
- Conclusion 16

Appendix. Regression Analysis 18

Notes 22

References 24

About the Authors 27

Statement of Independence 28

Acknowledgments

This report was funded by the Urban Institute.

The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, or its funders. Funders do not determine research findings or the insights and recommendations of Urban experts. Further information on the Urban Institute's funding principles is available at urban.org/fundingprinciples.

We would like to thank Chris Davis for his contributions and Alena Stern, Rekha Balu, Sue Popkin, and Michael Neal for their feedback.

Executive Summary

Automated valuation models (AVMs) are increasingly used in the mortgage industry to estimate home values, offering speed, efficiency, and cost savings over traditional in-person appraisals. But these algorithmic tools might unintentionally replicate or even amplify long-standing racial disparities in the housing market. Previous research has identified racial bias in AVM performance at the neighborhood level, but this study provides the first empirical evidence of racial disparities in AVM valuation errors at the individual level, using private property records data.

We ask whether AVMs systematically produce larger valuation errors for homeowners of color, even after controlling for property characteristics and neighborhood context. Because administrative property data lack self-reported race information, we apply an imputation technique—fully Bayesian Improved Surname Geocoding (fBISG)—to estimate each homeowner's racial or ethnic background based on surname and geographic location, using our estimates to assess AVM performance by race or ethnicity. We validate this method using Home Mortgage Disclosure Act (HMDA) data and find the method performs reliably, especially for Black and white homeowners, who make up the majority of our sample.

Our analysis reveals that the AVM associated with our property records generates significantly larger valuation errors for Black homeowners compared with white homeowners, even within the same neighborhood. We also find strong evidence of systematic undervaluation of homes owned by Black households in the Memphis and Atlanta metropolitan areas. These disparities persist even though AVMs do not explicitly incorporate race, suggesting that both biased training data and the structure of the algorithms themselves contribute to unequal outcomes. Inputs such as comparable sales, property condition, and neighborhood characteristics—factors historically shaped by redlining and segregation—can unintentionally encode discrimination into model predictions. Moreover, because AVMs are typically optimized to maximize accuracy across the entire dataset, this implicitly means the model will prioritize optimization for the largest racial group—white homeowners—at the expense of less represented groups (Hardt, Price, and Srebro 2016).

These findings have important policy implications. Solutions must be tailored to the source of the bias. When bias stems from incomplete or discriminatory data, state and local policymakers and internal developer policies should promote improvements in data coverage, timeliness, and transparency. For example, incorporating more detailed property-level data—including interior condition—can help correct inaccuracies, and regulators should encourage more scrutiny of the use of biased comparable

sales. Industry players have taken steps to improve AVM performance and data practices, and the growing use of desktop appraisals—a hybrid of AVMs and in-person assessments—reflects a broader shift toward addressing these concerns. Still, these developments do not fully eliminate the potential for algorithmic or data-driven bias. When bias originates from algorithmic design, interventions are needed at the federal, state, local, or company level. These include adding fairness as an explicit constraint during model training, mandating public reporting of AVM performance metrics by race and geography, and creating clear regulatory guidelines for evaluation and oversight. Formal collaboration between regulators and model developers will also be essential to ensure that innovation in automated valuation is both responsible and equitable.

Ultimately, our research underscores that race-neutral design is not enough to ensure fair outcomes. Without deliberate safeguards, AVMs risk reinforcing the disparities they aim to bypass. This study offers a new empirical framework for identifying AVM bias at the individual level and lays the groundwork for reforms that advance both equity and accuracy in automated home valuation.

AVMs Perpetuate Long-Standing Disparities in Home Values

The adoption of artificial intelligence (AI) technologies in mortgage lending has accelerated, promising to reshape nearly every stage of the lending process—from marketing and underwriting to fraud detection, risk management, and servicing (Neal, Perry, and Zhu 2023; Van Zetten, Ramackers, and Hoos 2022; Emad Azhar Ali et al. 2021). With its ability to analyze massive amounts of data and identify complex patterns, AI can improve speed, efficiency, and predictive accuracy.

In property valuation, AI-powered tools have emerged as an alternative to traditional in-person home appraisals. Home appraisals—the assessment of a property's value required when buying, selling, or refinancing a home—have historically been conducted in person, with a certified appraiser coming into the home to examine its condition and characteristics, evaluate the current real estate market and recent comparable sales, and ultimately, determine its value.[1] But within the past decade, the appraisal industry has changed considerably. The traditional industry, composed of certified appraisers who skew older than the general workforce, has shrunk,[2] and lenders have begun to leverage new technology to make the appraisal process more efficient. This technology includes automated valuation models (AVMs), which were first approved as appraisal waivers by the government-sponsored enterprises following the Great Recession, when traditional appraisal methods received scrutiny.[3] AVMs leverage large datasets on existing properties and transactions to model real estate values using machine learning methods.[4] AVMs became more popular during the COVID-19 pandemic, when the technology enabled lenders to continue with home purchases and refinances, as they eliminated the need for human contact during the appraisal process.[5]

Despite these advantages of offering time and labor cost savings to lenders and borrowers, AVMs—like traditional appraisals—are not immune to bias. Although they might reduce individual discretion, AVMs can replicate or even amplify historical inequities embedded in housing data (Neal et al. 2020). Research has shown that appraisal values can vary systematically by neighborhood racial composition (Narragon et al. 2021), and our prior work has documented greater AVM errors in majority-Black neighborhoods compared with majority-white ones, even after accounting for differences in property characteristics and market conditions (Zhu, Neal, and Young 2022). Although these errors occur in both directions (i.e., over- and undervaluation), disparities in the magnitude of errors raise questions about the fairness and accuracy of algorithmic valuations and their potential to perpetuate past discrimination in the housing market.

Although past research on appraisals has analyzed race and ethnicity at the property level, previous research investigating racial disparities in AVMs, including our own research, has focused mostly on AVM fairness and accuracy at the neighborhood level. Far less research has examined AVM performance at the person level, a gap in the field that has motivated our work. This gap is at least partly attributable to an empirical challenge: There is a lack of publicly available data on the race of individual homebuyers in datasets that measure home sales and AVM estimates.

This gap in the literature has important consequences. Systematic undervaluation of homes owned by people of color can undermine wealth accumulation, constrain access to credit, and reduce financial security. Overvaluation, on the other hand, inflates loan-to-value ratios, exposing borrowers to greater financial risk and potentially more expensive credit. In both cases, repeated errors—whether under- or overvaluation—can propagate over time through the use of comparable sales in subsequent valuations. The compounding nature of these errors means that bias today can shape valuations tomorrow.

Critically, solutions to neighborhood-level disparities are not the same as those needed to address individual-level bias. Neighborhood-level gaps might call for systemic reforms in housing finance, antidiscrimination enforcement, or data quality improvements. But if disparities persist at the individual level—holding neighborhood constant—they point more squarely to problems in the design and implementation of algorithmic models themselves. Disentangling these layers of bias is key to targeting the appropriate policy responses.

In this report, we pose the following question: Are homes owned by people of color more likely to experience AVM valuation errors, even when controlling for neighborhood-level factors and other observable property-level factors?

To analyze AVM bias at the individual level, we apply a widely used race imputation technique known as Bayesian Improved Surname Geocoding (BISG). One of the main reasons this type of analysis has been limited is that datasets containing AVM estimates and home sales typically lack self-reported race data. BISG addresses this gap by combining an individual's surname with the racial and ethnic composition of their residential area to assign probabilities across major racial and ethnic groups (Adjaye-Gbewonyo et al. 2014). This approach draws on the fact that because of historical patterns of residential segregation, an individual's last name and address often provide strong signals about their likely race or ethnicity (DeLuca and Curiel 2023).

BISG has been widely adopted by federal agencies, including the Equal Employment Opportunity Commission (Harris 2020) and the Consumer Financial Protection Bureau (2014), and has been used extensively in public health and housing research. In the health field, researchers have used BISG to examine disparities in disease outcomes and treatment access (Branham et al. 2022; Derose et al. 2013;

Zhang et al. 2022). In the housing and credit domains, researchers have applied BISG to assess fair lending risks (Zhang 2018), evaluate racial disparities in mortgage credit (Ambrose et al. 2021), study missing race data in Home Mortgage Disclosure Act (HMDA) filings (Voicu 2018), and impute race in credit bureau and eviction records (Stern, Narayanan, et al. 2021; Stern, Kovalski, et al. 2025).

By using BISG in this study, we can probabilistically infer homeowners' race and ethnicity, allowing us to conduct a robust analysis of AVM error patterns across racial groups while protecting privacy and working within the constraints of available data.

With race imputation in place, we analyze individual-level AVM valuation errors using regression analysis. This approach allows us to assess whether homes owned by people of color are more likely to experience over- or undervaluation, even when controlling for neighborhood characteristics and observable property factors. By focusing on individuals within the same neighborhood context, we can better isolate whether racial disparities in AVM performance are driven by differences in neighborhood characteristics or by individual homeowner identity.

To further investigate systemic bias, we then focus on undervaluation as a key indicator. Undervaluation directly undermines a homeowner's ability to build equity, refinance, or access credit in times of need. We employ a regression model with the log AVM value as our key outcome variable to identify undervaluation magnitude across racial groups, revealing persistent patterns that disproportionately affect Black homeowners. These findings suggest that even when AVMs are agnostic to a homeowner's race or a neighborhood's racial makeup, they might still reproduce historical inequities through data or algorithmic structures. By highlighting these risks, this report contributes a framework for evaluating and mitigating racial bias in algorithmic property valuation systems.

The remainder of the report proceeds as follows: The next section describes our data and methodology, including the race imputation process and our measures of AVM error. The section after that presents findings on the prevalence and magnitude of individual-level AVM bias by race. The final section concludes with implications for policy and industry practice and outlines steps toward fairer, more transparent use of artificial intelligence (AI) in the mortgage lending ecosystem.

Methodology and Data

Data Overview

Our analysis leverages proprietary property records data in 2018, which provide detailed property-level information on historical transactions, sales prices, AVM estimates, locations, and homebuyers' full names. Our analysis focuses on Atlanta, Georgia, and Memphis, Tennessee. Both cities have a high proportion of residents of color and offer a robust set of property-level pairings between AVM estimates and actual sales prices. Instead of using the entire core-based statistical area (CBSA) for each city, our study targeted counties with comprehensive historical deeds data that could be aligned with the AVM estimates. Although these counties represent only a fraction of the total counties within each CBSA, they account for a significant majority of the CBSA's overall population. Specifically, the counties we study in Atlanta and Memphis constitute 17 percent and 22 percent of the total counties in their respective CBSAs but are home to 63 percent and 74 percent of the respective city populations. To characterize the accuracy of AVM estimates, we compare AVM values with the sales price by calculating the percentage magnitude of inaccuracy, or the absolute difference between the AVM estimate and corresponding sales price divided by the sales price. We use a percentage-based approach because a $20,000 undervaluation means more for a home worth $200,000 than for one worth $300,000.

Following the work of Zhu, Neal, and Young (2022), we control for variables at both neighborhood and property levels that influence property values to examine the role of homeowners' race in algorithmic bias of AVMs. At the neighborhood level, we control for heterogeneity in properties within a neighborhood, neighborhood conditions, and turnover rates. Specifically, to capture within-neighborhood heterogeneity, we constructed two variables: the standard deviation of neighborhood property ages and the percentage deviation of neighborhood home values. Standard deviation and percentage deviation measure the dispersion of properties by age and home value, respectively. To capture neighborhood conditions, we included measures of gentrification, distressed sales, and median household income. We consider a neighborhood to be gentrified if it meets two criteria (Gould Ellen and O'Regan 2008): (1) the tract-level income is less than 70 percent of the income in the metropolitan statistical area and (2) the neighborhood (identified at the census tract level) experienced at least a 10 percentage-point increase in the ratio of tract-level income to metropolitan statistical area–level income over the year. In addition, we define turnover rate as the number of home sales per year divided by the number of homes. As AVM algorithms are based on comparable sales, higher turnover rates

would provide a larger sample of comparable sales for AVM algorithms, which would likely translate to more accurate estimates.

At the property level, we control for lot size, building age, sales price, number of bedrooms, and exterior property condition. Property condition, in particular, is a key contributor to home value that is absent from most administrative property records (Neal et al. 2021). Again, following from Zhu, Neal, and Young (2022), we use a measure called the exterior condition rating (ECR), provided at the property level by property intelligence firm CAPE Analytics. CAPE Analytics creates and applies computer vision algorithms to high-resolution images captured from airplanes to create structured data that include the ECR**.** The ECR covers all of a parcel's visible external features, including roofs, yards, driveways, and debris. In this analysis, we match our property records data for the Atlanta and Memphis metro areas with the ECRs from CAPE Analytics based on property latitudes and longitudes, parcel lot assessor parcel numbers, and transaction dates. Although the ECR excludes internal condition, newer AVMs are increasingly integrating more advanced data and techniques to capture interior property features. But Zhu, Neal, and Young (2022) found that ECR was a reasonable proxy for both external and internal condition of a property. The match rates are 98 percent for Atlanta and 90 percent for Memphis. The small share of unmatched properties was proportionately distributed between majority-Black and majority-white neighborhoods and thus do not skew the overall distribution.

Finally, our current analysis draws on additional information that supports our imputation strategy: homeowners' last names and addresses, as well as key inputs to estimate the likelihood that a homeowner belongs to a given racial category. Table 1 presents the summary statistics for all our variables in 2018.

TABLE 1

Descriptive Statistics of Neighborhood- and Property-Level Characteristics in 2018

N = 43,024

Variable	Mean	SD
Neighborhood-level characteristics		
Percentage deviation of neighborhood property ages	12.7%	6.8%
Percentage deviation of neighborhood property values	33.5%	10.8%
Gentrified neighborhood	4.0%	19.0%
Share of neighborhood distressed home sales	22.0%	41.0%
Neighborhood median household income	$79,287	$34,336
Turnover rate	8.8%	4.1%
Property-level characteristics		
Percentage magnitude of AVM inaccuracy	16.9%	32.5%
Sales price	$284,430	$198,562
Lot size	16,715	16,880
Number of bedrooms	3.42	0.47
ECR		
Good	13.0%	33.0%
Fair	50.0%	50.0%
Poor	37.0%	48.0%

Source: Authors' calculations using data from the American Community Survey, CAPE Analytics, and a major property records provider.

Note: AVM = automated valuation model; ECR = exterior condition rating; SD = standard deviation.

Methodology: A Three-Step Approach

To examine whether AVMs generate racially biased property valuations at the individual level, we used a three-step methodological approach: (1) imputing homeowners' race and ethnicity, (2) measuring AVM valuation error, and (3) identifying undervaluation. Our goal was to combine statistical rigor with transparency, making it possible to trace patterns of racial disparities in AVM performance even without access to self-reported race data.

Step 1: Imputing Homeowners' Race and Ethnicity

Because property records typically do not include information on homeowners' race or ethnicity, we employed statistical imputation techniques to infer this information. We tested two methods, both building upon the standard BISG. BISG uses a person's last name and residential location to estimate the probability that the person belongs to a particular racial or ethnic group, based on census data.

The first method, fully Bayesian Improved Surname Geocoding (fBISG), enhances traditional BISG by addressing two known data limitations: missing surname coverage and population undercounts in small geographic areas. The method incorporates voter registration records from six Southern states—where individuals self-identify their race—to improve accuracy, particularly for less common surnames

and smaller population groups (Imai, Olivella, and Rosenman 2022). We applied this method at the census block group level, which offers more geographic detail than broader aggregations such as zip codes.

The second method, Bayesian Instrumental Regression for Disparity Estimation (BIRDiE), takes the output from fBISG and further adjusts it using a statistical technique called instrumental variables regression. BIRDiE was designed to reduce bias that might arise in BISG-based approaches when imputed race estimates are correlated with the outcome being studied—in this case, property valuation. It treats surname as an instrument for race to correct for that correlation. BIRDiE has demonstrated improved accuracy in predicting the race of North Carolina voters (McCartan et al. 2024), and while it has been used by our colleagues to study eviction disparities (Stern et al. 2025), this is the first known use of BIRDiE in housing finance research. Because of computational demands, we apply this method at the zip code level. We evaluate two versions of BIRDiE: a base model using surname and zip code and a more detailed model incorporating building-level characteristics (e.g., year built, number of bedrooms) that are highly correlated with sales price.

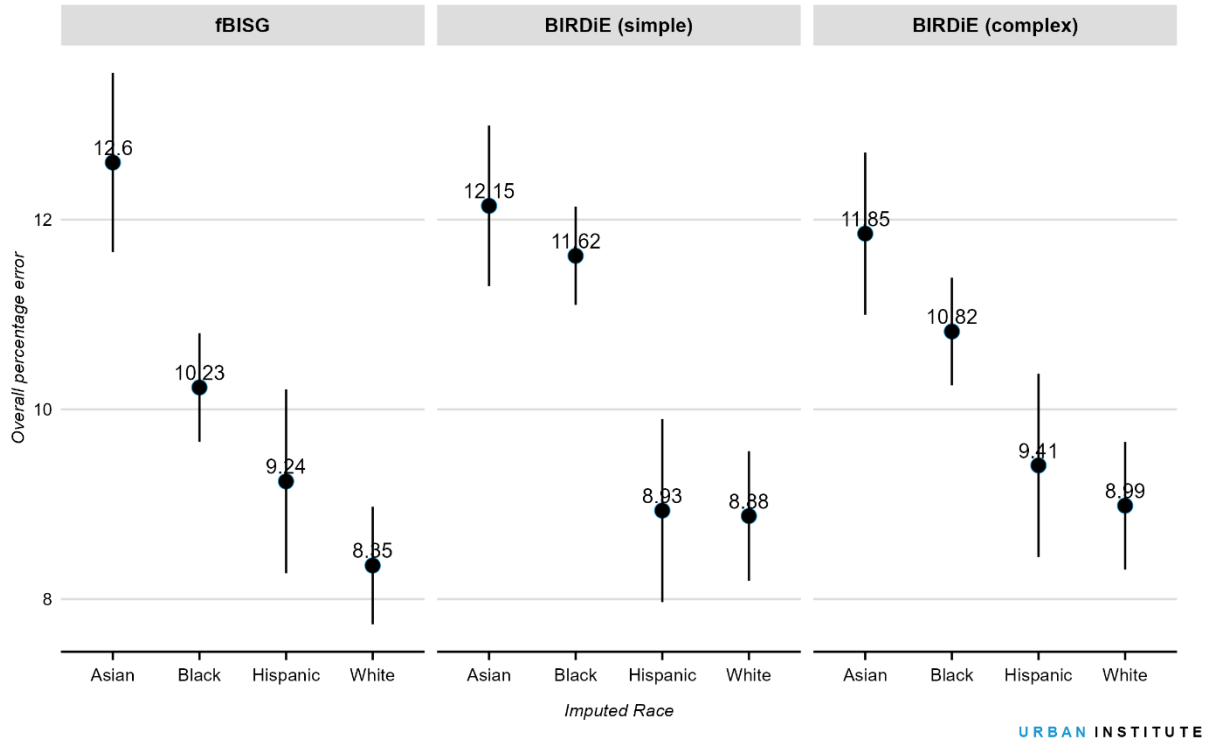
Although both methods are promising, we needed a way to determine which one produced more accurate results for our specific application. Because we lacked self-reported race at the individual level, we used loan-level HMDA data as a proxy. HMDA includes racial and ethnic data for mortgage applicants and serves a population similar to the homeowners in the property records we use.[6]

We compared the median loan amounts by race across census tracts in HMDA with those in our property dataset and calculated the percentage error between them for every census tract in our data.[7] About 99 percent of the census tracts (703 of 712) covered in our relevant property records were matched to census tracts in HMDA. Then, to obtain an overall error for each racial group, we average together all the tract-level errors, weighting each tract by the number of observations we have for that tract in the property records.

Figure 1 displays the overall percentage errors for each racial group, with error bars that represent 1 standard error in each direction to provide a measure of uncertainty. fBISG produces the lowest percentage errors for the Black and white homeowners who make up a majority of our sample. Based on this validation, we selected fBISG as the preferred imputation method for the rest of our analysis.

FIGURE 1

Percentage Errors, by Race or Ethnicity



Source: Urban Institute calculations using 2018 Home Mortgage Disclosure Act (HMDA) data and data from a major property records provider.
Notes: Individuals assigned race based on highest predicted probability from imputation (i.e., thresholding). Percentage errors are differences between median loan amount by race in the HMDA data and median loan amount by imputed race in property records, as a percentage of median loan amount by race in HMDA. Percentage errors are computed for each census tract and weighted by number of observations to compute overall percentage error for each race group. Error bars represent 1 standard error of each estimate. American Indian/Alaska Native and other racial groups are removed from validation because of insufficient sample size.

Step 2: Measuring AVM Valuation Error

Once we imputed homeowners' race and ethnicity, we analyzed whether AVM estimates systematically differ by race. We draw on ordinary least squares regression analysis implemented by Zhu, Neal, and Young (2022) to compare predicted home values with actual sales prices, controlling for various property and neighborhood characteristics.

An important feature of fBISG is that it produces a probability distribution for each racial or ethnic category rather than assigning each person to a single group. The empirical best practice is to use these probabilities directly in analysis, as they increase accuracy and reduce statistical bias (Chen et al. 2019; Elliott, Fremont, et al. 2008; Elliott, Morrison, et al. 2009). Although this improves statistical accuracy, it

makes interpreting regression results more difficult because in a regression setting, race *probabilities* do not have a practical interpretation as coefficients, while actual race assignments do.

To strike a balance between statistical rigor and interpretability, we adopt a multiple imputation approach. We simulate 100 regressions, each time randomly assigning homeowners to a racial or ethnic group by drawing from their fBISG probabilities. This allows us to estimate group-level effects while preserving interpretability. Appendix A further describes our multiple imputation process. The result of this process is a set of race assignments for each homeowner that allows us to interpret regression coefficients as the systematic difference in outcome variable for Black, Hispanic, or Asian homeowners relative to white homeowners as the reference group.

To measure AVM valuation error, we run ordinary least squares regression analysis on the percentage magnitude of error, defined as the absolute difference between a home's AVM-predicted and actual sales price divided by the actual price. This captures errors of both under- and overvaluation across racial groups. We control for the property attributes and neighborhood characteristics defined above to isolate the impact of homeowner race from other factors that affect home prices. We also include county fixed effects to control for any other unobservable differences in home value at the county level. Appendix A specifies our regression equations.

Step 3: Identifying Undervaluation

Although overall AVM error is important, undervaluation poses a particular risk to homeowners of color. Homes that are consistently valued below market can limit a household's ability to build equity, refinance, or access credit. Undervaluation can even lead to mortgage denial attributable to insufficient collateral. It is also worth noting, however, that undervaluation might also increase affordability for prospective homebuyers. Hamilton and coauthors (2023) find that low appraisals often lead to contract renegotiation, resulting in lower purchase prices.

To examine this more closely, we regress the log of AVM estimates on imputed homeowner race to examine the magnitude and direction of valuation error. This step allows us to directly assess how bias in AVM systems could suppress housing wealth accumulation, especially for Black homeowners. Identifying undervaluation as a systemic pattern strengthens the case for policy interventions focused on transparency, fairness, and accountability in AI-based property valuation.

Key Findings

The final specifications for all our regression models include property- and neighborhood-level variables defined to isolate the impact of homeowner race from other factors that affect home price. We also include county fixed effects to control for any other unobservable differences in home value at the county level. Appendix A includes our specific regression equations.

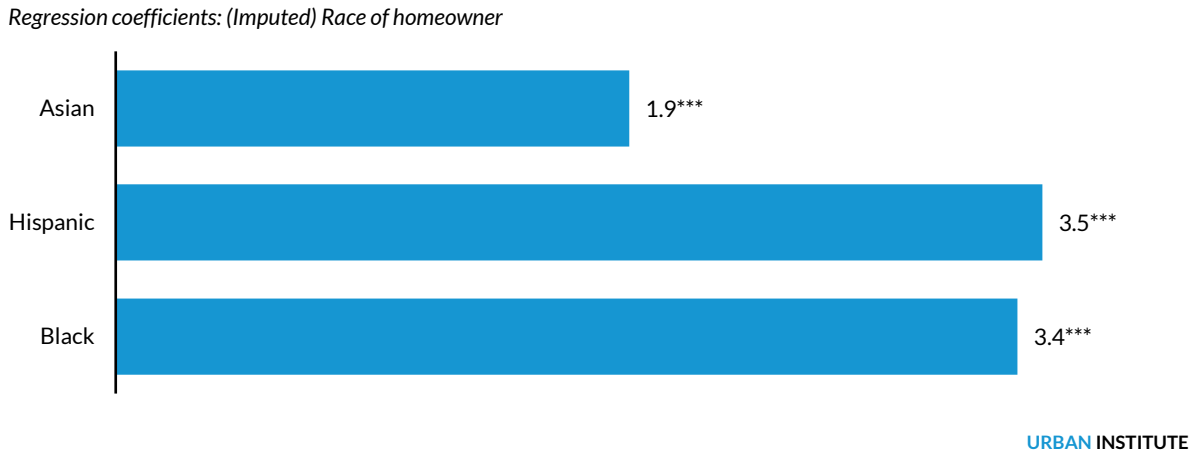
AVMs Tend to Yield Greater Valuation Errors for Black Homeowners

The goal of AVMs is to produce accurate and timely home value estimates for various purposes. These include mortgage underwriting—where valuation informs risk via loan-to-value ratios—as well as home equity lending, refinancing, and consumer-facing uses, such as online real estate platforms that help homeowners or buyers assess listing prices. Our first regression model examined whether the percentage magnitude of AVM errors differed by homeowner race or ethnicity. We found that the percentage magnitude of AVM error is greater for Black homeowners. Specifically, for homes owned by Black people, we found an AVM percentage error that is statistically significant and 3.4 percentage points greater than for homes owned by white people, all else equal (figure 2). Practically, this means at a median sales price of $420,000, a Black homeowner would encounter, on average, an additional AVM error of $14,960 relative to a white homeowner with a similarly priced and characterized property. Our analysis also found greater magnitudes of AVM error among Hispanic and Asian homeowners, but the majority of the sampled homeowners were Black or white. Future research should do robustness checks of these findings in places with greater counts of Asian and Hispanic homeowners.

FIGURE 2

Regression Coefficients of Percentage Magnitude of AVM Error on Homeowner Race (Relative to Non-Hispanic White Homeowners), Controlling for Property and Neighborhood Characteristics

Black and Hispanic homeowners have greater AVM error than white homeowners



Source: Urban Institute regression analysis.
Notes: AVM = automated valuation model. Not statistically significant. Full regression results are in appendix table A.1, column 4.
* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Black-Owned Homes Are Valued Lower by AVMs Than Comparable White-Owned Homes

The above results showed that Black-owned homes experience greater valuation error than those owned by white people. But the direction of error matters. In our second model, we examined the magnitude of AVM undervaluation. This is crucial information, as undervaluation artificially limits a homeowner's ability to build equity, refinance, and access credit.

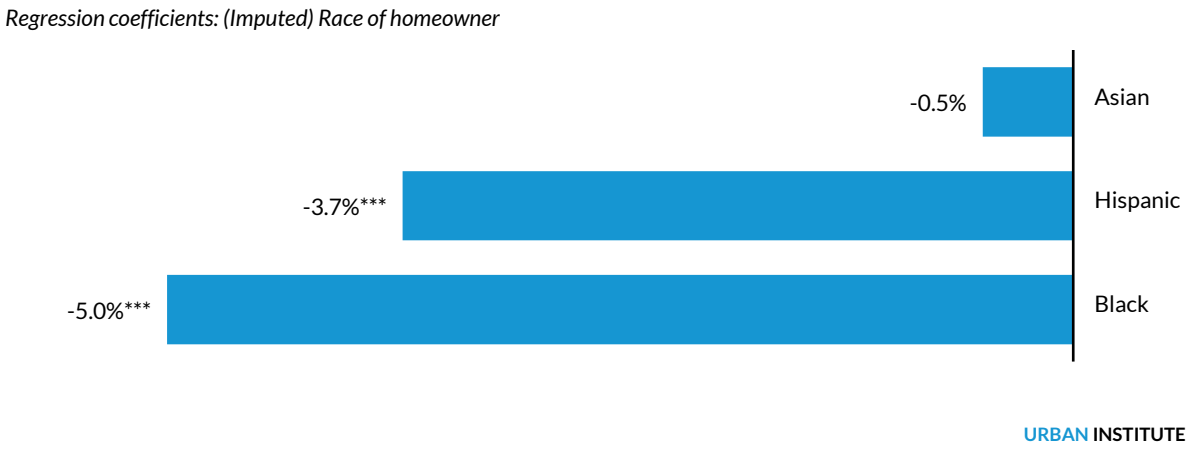
We used an additional regression model that looks at the log of the actual AVM estimate (not only its magnitude error) to further examine racial biases within AVMs. Even holding constant home value and property- and neighborhood-level variations, AVMs for properties owned by Black homeowners are undervalued by 5 percent relative to properties owned by white homeowners (figure 3). This disparity indicates a widening of existing racial gaps in housing wealth that contributes to reduced housing wealth accumulation for Black homeowners. Previous research finds that home price increases from 1984 to 2021 explain most of the increase in the gap in median wealth between Black and white homeowners (LaBriola 2024). The work finds that this is mostly attributable to the difference in homeownership rates between Black and white households (only existing homeowners fully benefit from home price appreciation), but it also finds that white homeowners tend to own more expensive

homes than Black homeowners, which increases their absolute gains in home equity. If Black homeowners face persistent undervaluation as they try to sell their homes or refinance, this could be contributing to the increasing gap in housing wealth, a key driver of the overall increase in the Black-white wealth gap.

FIGURE 3

Regression Coefficients of Log(AVM Estimate) on Homeowner Race (Relative to Non-Hispanic White Homeowners), Controlling for Property and Neighborhood Characteristics

Black and Hispanic homeowners receive systematically lower AVM estimates than white homeowners on comparative homes



Source: Urban Institute regression analysis.
Notes: AVM = automated valuation model. Not statistically significant. Full regression results are in appendix table A.2, column 4.
* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Limitations

Our analysis has several limitations. First, our imputation methods rely on proxy indicators such as surname and location, which might be less accurate for smaller groups, such as American Indian and Alaska Native and multiracial populations. Studies have shown that although BISG performs well for larger groups, such as Black, Hispanic, white, and Asian households, its accuracy is significantly lower for other groups (Haas et al. 2019).[8] Imputation will also fail to capture homeowners whose names change when they get married or cohouseholders whose names do not appear on property deeds. As the number of dual-race coborrowers and cohouseholders grows, these limitations become increasingly relevant.[9]

Second, although BIRDiE has shown promise in other applications, our validation exercise found fBISG to be more accurate for our context. But the comparison is not perfect: We applied the two

methods at different geographic scales, and HMDA data might not fully match the population in our property records—for example, non-arm's-length transactions and title issues might be underreported or missing from one source but present in another.

Third, this analysis is limited to a single year (2018) in only Atlanta and Memphis, both of which have predominantly Black and white populations. Future research should expand the time horizon and geographic scope to test whether similar patterns hold in more diverse regions with larger Hispanic or Asian American and Pacific Islander populations.

Policy Implications

Our findings reveal that AVMs can produce systematically different valuation errors across racial groups, even when race is not a direct input into the model. This result raises important questions about how such disparities arise and what it would take to build a more equitable property valuation system.

Although AVMs are often perceived as neutral tools—operating without human discretion or knowledge of homeowner race—the data they rely on are not free from bias. Inputs such as location, neighborhood characteristics, and sales prices of comparable homes are deeply intertwined with the legacy of racial segregation and structural inequality in the housing market (Howell and Korver-Glenn 2022). Properties in higher-income, more homogenous neighborhoods with fewer distressed sales tend to yield more accurate AVM estimates (Neal et al. 2020), and white homeowners are more likely to live in these areas. In contrast, AVM performance is systematically worse in neighborhoods with more diversity, lower median income, or higher levels of housing distress—factors that disproportionately affect communities of color (Neal et al. 2020).

This situation creates a central tension: Maximizing predictive accuracy without regard for race can perpetuate unequal outcomes, but pursuing fairness—such as by ensuring similar error rates across groups—might require sacrificing overall model accuracy. One hypothetical solution might involve building separate models tailored to different racial or demographic groups. But doing so would require incorporating race as an explicit variable, undermining the race neutrality that underlies most AVM design. Recent innovations such as desktop appraisals—which blend AVMs with human oversight—can address both efficiency and fairness concerns. But these hybrid models are not a panacea. They might still reflect biased data inputs or human subjectivity, and their effectiveness in reducing racial disparities remains understudied. These limitations underscore the importance of future research and

regulatory guidance to evaluate the trade-offs between predictive performance and fairness in AVM deployment.

To further assess these trade-offs, we must understand where the disparities originate. There are two primary sources of racial disparities in AVMs:

- First, biased data can embed historical and structural discrimination into model predictions. AVMs are trained on past sales, but those sales reflect long-standing patterns of housing discrimination, including buyer preferences shaped by race and neighborhood, as well as appraisal practices that have undervalued homes in communities of color. These biased patterns become part of the algorithm's logic when they are encoded in the training data.
- Second, algorithmic bias can arise from how models are optimized. AVMs are typically trained on datasets where certain groups are more represented than others, which can cause models to prioritize accuracy for majority populations. Because most home sales in most datasets come from white homeowners—because of racial gaps in homeownership—AVMs prioritize accuracy for white homeowners. Unless explicitly corrected, model performance will be skewed toward the group with the most data, often at the expense of underrepresented populations. Without built-in fairness constraints, predictive accuracy tends to come at the cost of equity.

In our analysis, we found that these forces combined to produce larger AVM errors for Black homeowners, especially in the form of undervaluation. This suggests that simply omitting race from models is insufficient to ensure fairness. Addressing these disparities will require both better data practices and intentional algorithmic design.

Since the second Trump administration began, we have seen a shift in federal policy geared toward studying bias in AI. On Inauguration Day, President Trump revoked a 2023 executive order that required AI developers whose products and systems pose risk to national security, the economy, public health, or safety to conduct and share results of safety tests with the government. Still, state and local governments and private-sector actors have several levers at their disposal to improve AVM rigor, transparency, and fairness.

Multiple states have shown an appetite to take action on housing appraisal bias. Maryland's Department of Housing and Community Development established a Task Force on Property Appraisal and Valuation Equity, which recommended steps to improve transparency and documentation and includes appraisers in the loop of any automated analysis (Task Force, n.d.). Although not specific to automated appraisals, California's Civil Rights Department secured financial settlements and new

antibias training from the company Clear Capital after demonstrating appraisal bias against a Black and Latino family.[10] Other litigation and regulatory regimes at the state and local level can follow the lead of Maryland and California in this space, and action around automated valuation can be encompassed under long-standing best practices for traditional appraisal bias. Finally, at the local level, city chief data officers, who are increasingly tasked with thinking about AI training and guidelines, can step in when AVMs have a disparate impact on homeowners in their cities or can evaluate disparities in AVMs themselves.[11] Similarly, real estate organizations and private-sector users of AVMs who want to advance the state of play can voluntarily take action to demonstrate the benefits of automation when used transparently and responsibly.

Policy Recommendations to Reduce Data Bias

- **High-quality, up-to-date data are critical to reducing data bias.** Developers should publish information on the accuracy of their underlying data and should update that information regularly. Standardizing the transparency and type of information published with AVMs will allow for better comparison by regulators and consumers. State and local policymakers, such as chief data officers, can encourage or require developers to publish or share this information. Across states, the State Chief Data Officers Network[12] could share best practices and promote data-sharing incentives that work.

- **But focusing only on data accuracy and reflection of market outcomes could perpetuate biases that are true *within housing market data*.** To ensure historical discrimination is not repeated, state and local policymakers could expand consumer protections against the exacerbation of discrimination, incorporate fair lending principles, strengthen compliance monitoring, and require regulators to mandate accounting of input data used and existing biases, such as disparities in comparable sales by race. They could also require developers to identify "less discriminatory alternatives," which are methods that achieve the same accuracy objectives but with improved fairness metrics, for compliance with fair lending laws. More research can be done to consider the potential for reduced bias by using synthetic data in training.

Policy Recommendations to Reduce Algorithmic Bias

For AVMs that exhibit algorithmic bias, the policy solutions differ, but policymakers have a role in developing fairer and more accountable AVMs.

- **To inform the public and better enable AVM comparisons across providers, housing market regulators could require lenders to publicly report the same statistics on a regular basis.** Regulators such as the Office of the Comptroller of the Currency or the Federal Housing Finance Agency could require AVM developers to publish model cards, short documents accompanying models that present evaluations of the model across different demographic groups (Mitchell et al. 2018). Brown and Engler (2023) suggest public reporting of mean absolute error and mean percentage error by geography and bands of property sales price, error magnitude and directionality, and relative accuracy on low- and high-value properties. This gives regulators and consumers better information and would align appraisals with the rest of the mortgage industry, where guidance requires lenders to provide precise reasons for adverse actions, such as loan denials and changes in credit conditions, when using AI or other complex models. Today, however, AVM algorithms are opaque, complex models fed with large datasets, and lenders often describe their models as "proprietary," with no further details.
- **Because the most accurate housing market data are likely to reflect biased outcomes, AVM algorithm developers can deliberately integrate fairness metrics into their design.** Additionally, developers might consider adopting new techniques to select the right algorithms in the first place. For example, Freddie Mac adopted machine learning techniques to search across a wider set of possible models and find less discriminatory alternatives that did not sacrifice predictive accuracy, leading to larger improvements in fairness than were previously possible (Kiefer 2023). Policymakers could require these types of techniques among model developers.

Conclusion

AI is being used in many corners of the housing market, from AVMs, to underwriting decisionmaking and risk projections, and even in construction project planning and management.[13] And while technological advancements as sophisticated as AI present a massive opportunity to increase housing market efficiencies and open credit and capital markets to more consumers, responsible use of these tools is crucial. Today's housing market cannot produce unbiased outcomes without explicitly focusing on and correcting for discrimination, in both the input data and algorithms themselves.

AVMs have the potential to transform the housing market and make the appraisal process more efficient and simpler for everyone, but the perpetuation of errors or undervaluation for individual homeowners of color, as we find in our research, comes at a real cost to millions of homeowners.

Moreover, racial disparities in home appraisals contribute to the racial wealth gap, which costs the United States trillions of dollars and reinforces long-standing inequities rooted in historical housing policies.[14] Rooting out bias in home appraisals is required to truly move forward into a society in which opportunity and mobility are determined by one's decisions and hard work, as opposed to their race, class, or social identity. But federal policymakers and regulators, model providers, developers, lenders, and independent researchers need to study and be transparent about bias and risk, not ignore it, to achieve this goal.

Today, there are few guidelines for evaluating and mitigating risks related to AVM bias. In our research, we do not have access to the models that generate AVM estimates, making it impossible to fundamentally determine the source of AVM bias. Despite this limitation, this study provides an empirical framework for testing bias in AVMs across individual estimates. Our results highlight the importance of encoding nondiscriminatory factors into algorithms to account for existing racial biases in the baseline data. Moreover, creating clear guidelines for risk evaluation and mitigation reduce the risk of innovation and product testing, as model developers are aware of clear bright lines and legal obligations (Bogen 2024). To ensure responsible use of AI in mortgage lending, policymakers can also encourage and create opportunities for partnerships between regulators and industry stakeholders that allow for both parties to share lessons, concerns, and opportunities. The research around AVMs, and AI more broadly, is still developing, and more work is needed to understand the origination of bias in AVMs and the best path forward for policy and regulation to maximize efficiency without expanding bias and historical discrimination. For this reason, continued study of AVMs by independent researchers and by federal and state agencies, regulators, and lawmakers remains critical, as do requirements or incentives that models and their output be made public and transparent. Policymakers at the federal, state, and local levels and industry leaders need to study and address bias in AI to avoid an increase in bias and disparities in housing, which would hurt the entire economy.

Appendix. Regression Analysis

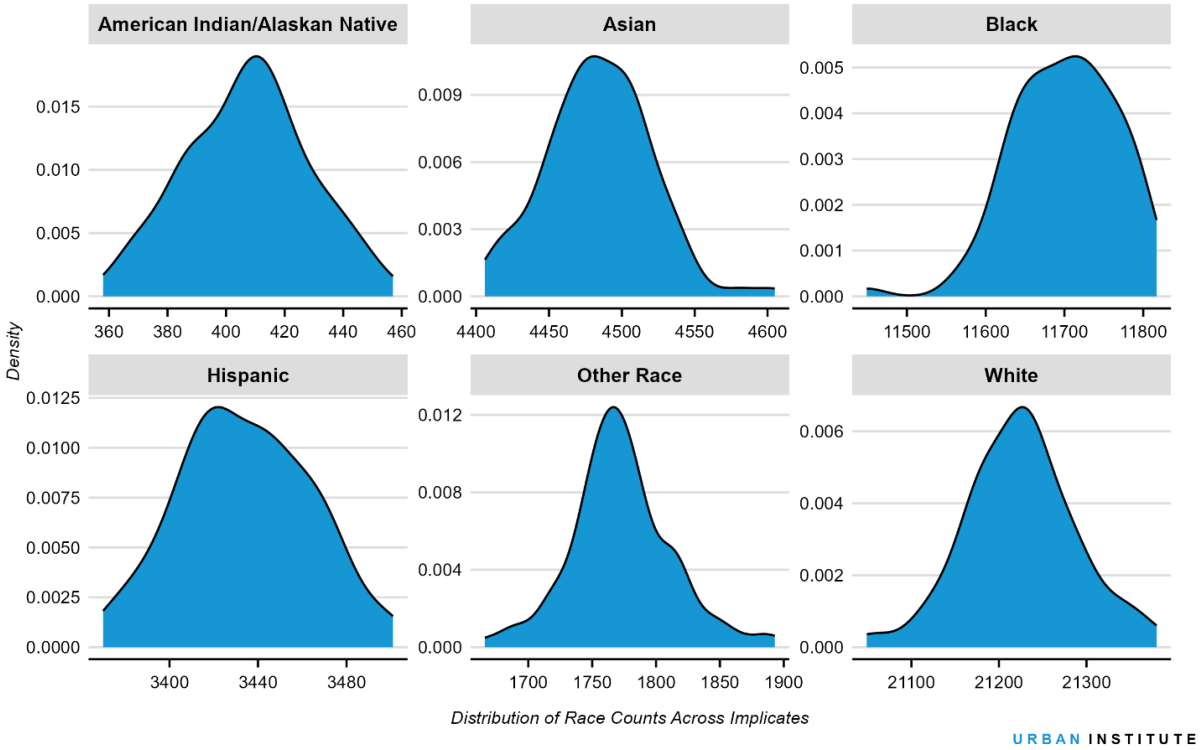
To run our regression analysis after computing race probabilities for each homeowner using fBISG, we use a multiple imputation framework. This approach is more statistically robust than the "thresholding" method of simply assigning the highest-probability racial group to each individual. Multiple imputation entails creating multiple copies, or *implicates*, of imputed race. It is a standard procedure used in many public data products, such as the Survey of Income and Program Participation Synthetic Beta, the National Survey of Children's Health, and the Survey of Consumer Finances. The benefit of this approach is that we can retain the interpretability of assigned race in our regressions while directly analyzing the variation that comes from uncertainty contained in fBISG's probabilities.

We create 100 implicates of race across all homeowners whose race we impute using fBISG. To do this, we sample 100 times from the fBISG probabilities. For instance, in a simple case where fBISG predicts a person as 90 percent likely to be Black and 10 percent likely to be white, we would draw repeatedly from that distribution and expect around 90 racial assignments to be "Black" and around 10 to be "white" (taking sampling variation into account). Replicating this process for every homeowner, we construct our 100 race implicates. We run a different regression model for every race implicate and average all coefficients together to arrive at the final results presented in figures 2 and 3. Standard errors are similarly pooled across the 100 implicates. We use the *mitools* package in R to pool our 100 regression results together and compute coefficients. These calculations, as well as our calculations for pooled R^2 and adjusted R^2 follow directly from Rubin's rules (Rubin 1987). The specific equations can be viewed in chapters 9 and 11 of Heymans and Eeekhout (2019).

Figure A.1 illustrates how much variation in race assignment results from our multiple imputation procedure. Each density plot shows the random variation in the total number of homeowners assigned to each race across the 100 draws from fBISG probabilities. For example, we can see that the total number of Black-assigned homeowners varies from a little under 11,500 to around 11,800, depending on the draw. The ranges are modest, indicating that fBISG tends to produce high-confidence predictions that are not subject to much random sampling variation.

FIGURE A.1

Variation in Race Assignment Results



Source: Urban Institute calculations from a major property records provider.
Note: Distributions of counts created by drawing 100 times from fully Bayesian Improved Surname Geocoding–imputed race probabilities for each homeowner.

We also detail the regression specifications discussed above.

Following equation 1, our first chosen outcome variable is percentage magnitude of AVM error, or the absolute difference between the AVM predicted sales price and the actual sales price of a home as a percentage of the actual sales price. Intuitively, we might think of racial discrimination in this setting manifesting in the form of Black- and Hispanic-owned homes being systematically undervalued compared with white-owned homes. But we also see bias in the other direction being important, as overvaluation might cause credit risk holders to underestimate risk and might speed up irrational inflation of property values, potentially resulting in a future home price correction (PAVE 2022). As a result, we consider the overall percentage magnitude of AVM error to be the outcome most of interest here.

$$Pct_{Diff_{i,2018}} = \alpha_0 + \beta Race_{i,2018} + \gamma N_{i,2018} + \delta P_{i,2018} + \pi_c + \varepsilon_{i,2018} \quad (1)$$

$Pct_{Diff_{i,2018}}$ is the percentage magnitude of AVM inaccuracy for property i in year 2018, measured by the absolute difference between the sales price and the AVM value, divided by the sales price. $N_{i,2018}$ and $P_{i,2018}$ represent a range of neighborhood-level and property-level indicators that are likely to directly affect the percentage AVM error. $N_{i,2018}$ includes percentage deviation of property age, percentage deviation of property values, turnover rates, gentrified neighborhood status, share of distressed sales, number of households, and median household income. $P_{i,2018}$ includes lot size, property age, number of bedrooms, and exterior property condition. We also include county fixed effects π_c to control for other idiosyncratic differences at the county level. We incorporate race into the regression by sampling from the individual probabilities as discussed above. β is the coefficient of interest that estimates the role of homebuyer race in AVM error. Statistically significant positive coefficients on the race variables would indicate that, after controlling for relevant housing characteristics, the magnitude of percentage AVM error is higher for homeowners of a certain imputed race or ethnicity, relative to imputed-white homeowners.

In addition to the percentage magnitude of AVM error, we also estimate the impact of individual race on AVM estimates, following equation 2. Statistically significant negative coefficients on the race variables would indicate that after controlling for sales price and neighborhood- and property-level characteristics, homeowners of a certain imputed race or ethnicity tend to experience underestimation in their AVM estimates, relative to imputed-white homeowners. $V_{i,2018}$ is the sales price of property i in 2018.

$$\ln AVM_{i,2018} = \alpha_0 + \beta Race_{i,2018} + \eta V_{i,2018} + \gamma N_{i,2018} + \delta P_{i,2018} + \pi_c + \varepsilon_{i,2018} \quad (2)$$

The results of our analyses are below in tables A.1 and A.2.

TABLE A.1

Dependent Variable: Percentage Magnitude of AVM Error

Variables	(1)	(2)	(3)	(4)
Black	9.614***	8.768***	3.225***	3.409***
	(0.477)	(0.471)	(0.516)	(0.515)
Hispanic	7.682***	6.277***	3.158***	3.503***
	(0.696)	(0.67)	(0.663)	(0.664)
Asian	0.757	2.964***	1.551***	1.941***
	(0.585)	(0.57)	(0.561)	(0.563)
Property-level controls	No	Yes	Yes	Yes
Neighborhood-level controls	No	No	Yes	Yes
County fixed effects	No	No	No	Yes
Constant	13.366***	10.999***	172.15***	160.582***
	(0.244)	(4.106)	(6.601)	(6.737)
Observations	43,024	43,024	43,024	43,024
R^2	0.018	0.088	0.132	0.134
Adjusted R^2	0.018	0.087	0.132	0.133

Source: Authors' calculations.

Notes: AVM = automated valuation model. Property-level controls include building age, lot size, number of bedrooms, and exterior condition rating. Neighborhood-level controls include an indicator of gentrification at the neighborhood level, the log of median household income in the tract, the log of the number of households in the tract, the turnover rate in the tract, variation in building age at the tract level, variation of home values at the tract level, and the share of distressed home sales.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

TABLE A.2

Dependent Variable: Log(AVM Estimate)

Variables	(1)	(2)	(3)	(4)
Black	-0.083***	-0.091***	-0.048***	-0.050***
	(0.004)	(0.004)	(0.004)	(0.004)
Hispanic	-0.041***	-0.047***	-0.034***	-0.037***
	(0.006)	(0.006)	(0.005)	(0.005)
Asian	-0.001	-0.013***	-0.003	-0.005
	(0.005)	(0.005)	(0.004)	(0.005)
Log(home value)	0.881***	0.830***	0.735***	0.731***
	(0.002)	(0.002)	(0.003)	(0.003)
Property-level controls	No	Yes	Yes	Yes
Neighborhood-level controls	No	No	Yes	Yes
County fixed effects	No	No	No	Yes
Constant	1.466***	1.998***	0.396***	0.468***
	(0.025)	(0.042)	(0.052)	(0.054)
Observations	43,024	43,024	43,024	43,024
R^2	0.165	0.873	0.885	0.885
Adjusted R^2	0.165	0.873	0.885	0.885

Source: Authors' calculations.

Notes: AVM = automated valuation model. Property-level controls include building age, lot size, number of bedrooms, exterior condition rating, and the share of distressed home sales. Neighborhood-level controls include an indicator of gentrification at the neighborhood level, the log of median household income in the tract, the log of the number of households in the tract, the turnover rate in the tract, variation in building age at the tract level, and variation of home values at the tract level.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Notes

[1] Angela Colley, "What Does an Appraiser Do?" Realtor.com, August 22, 2022, https://www.realtor.com/advice/buy/what-does-an-appraiser-do/.

[2] Michael Neal and Peter J. Mattingly, "Increasing Diversity in the Appraisal Profession Combined with Short-Term Solutions Can Help Address Valuation Bias for Homeowners of Color," *Urban Wire* (blog), Urban Institute, July 1, 2021, https://www.urban.org/urban-wire/increasing-diversity-appraisal-profession-combined-short-term-solutions-can-help-address-valuation-bias-homeowners-color; and Isaac Peck, "The Decline of Appraisers," Working RE, accessed August 4, 2025, https://www.workingre.com/the-decline-of-appraisers/.

[3] Laurie Goodman and Jun Zhu, "A Significantly Improved Appraisal Process Has Reduced the Riskiness of Refinance Mortgages," *Urban Wire* (blog), Urban Institute, March 7, 2019, https://www.urban.org/urban-wire/significantly-improved-appraisal-process-has-reduced-riskiness-refinance-mortgages.

[4] Adam Hayes, "Automated Valuation Model (AVM): Definition and How It Works," Investopedia, last updated March 28, 2025, https://www.investopedia.com/terms/a/automated-valuation-model.asp.

[5] Michael Neal and Laurie Goodman, "Appraisal Waivers Have Helped Homeowners Find Payment Flexibility amid Pandemic-Induced Economic Struggles," *Urban Wire* (blog), Urban Institute, October 16, 2020, https://www.urban.org/urban-wire/appraisal-waivers-have-helped-homeowners-find-payment-flexibility-amid-pandemic-induced-economic-struggles.

[6] Importantly, more than 20 percent of race and ethnicity information is missing from HMDA, likely in nonrandom ways. The populations might also differ slightly between the two datasets. Both these factors introduce uncertainty into our benchmarking process.

[7] For the property records data, we assign each individual to the highest-probability race category that fBISG or BIRDiE predicts. This approach, "thresholding," was used for ease of interpretation. We use the probabilities directly (i.e., "weighting") in our analysis of AVM disparities, after selecting a preferred imputation method.

[8] Lisa Lines and Jamie Humphrey, "Imputing Race and Ethnicity: Part1," *The Medical Care Blog*, January 1, 2022, https://www.themedicalcareblog.com/imputing-race-ethnicity-1/.

[9] John Walsh and Amalie Zinn, "Housing Market Researchers Tend to Ignore Coborrower Race. Differences in Denial Rates Suggest They Shouldn't," *Urban Wire* (blog), Urban Institute, October 16, 2023, https://www.urban.org/urban-wire/housing-market-researchers-tend-ignore-coborrower-race-differences-denial-rates-suggest.

[10] California Civil Rights Department, "Civil Rights Department Announces Settlements over Alleged Appraisal Discrimination against Black and Latino Family in the Bay Area," press release, April 14, 2025, https://calcivilrights.ca.gov/2025/04/14/civil-rights-department-announces-settlements-over-alleged-appraisal-discrimination-against-black-and-latino-family-in-the-bay-area/.

[11] Stephen Goldsmith, "The Growing Role of the City-Level Chief Data Officer," Harvard Kennedy School, Data-Smart City Solutions, July 10, 2024, https://datasmart.hks.harvard.edu/growing-role-city-level-chief-data-officer.

[12] "State Chief Data Officers Network," Georgetown University, Beeck Center, accessed August 4, 2025, https://beeckcenter.georgetown.edu/projects/state-cdo-network/.

[13] Sarah Rudge, "The Rise of Artificial Intelligence in Construction," Construction Today, September 17, 2024, https://construction-today.com/news/the-rise-of-artificial-intelligence-in-construction/.

[14] Nick Noel, Jason Wright, Duwain Pinder, and Shelley Stewart, "The Economic Impact of Closing the Racial Wealth Gap," McKinsey, August 13, 2019, https://www.mckinsey.com/industries/public-sector/our-insights/the-economic-impact-of-closing-the-racial-wealth-gap.

References

Adjaye-Gbewonyo, Dzifa, Robert A. Bednarczyk, Robert L. Davis, and Saad B. Omer. 2014. "Using the Bayesian Improved Surname Geocoding Method to Create a Working Classification of Race and Ethnicity in a Diverse Managed Care Population: A Validation Study." *Health Services Research* 49 (1): 268–83. https://doi.org/10.1111/1475-6773.12089.

Ambrose, Brent W., James Conklin, N. Edward Coulson, Moussa Diop, and Luis A. Lopez. 2021. *Does Appraiser and Borrower Race Affect Valuation?* New York: SSRN.

Bogen, Miranda. 2024. *Navigating Demographic Measurement for Fairness and Equity: AI Governance in Practice Guide*. Washington, DC: Center for Democracy and Technology.

Branham, D. Keith, Kenneth Finegold, Lucy Chen, Melony Sorbero, Roald Euller, Marc N. Elliott, and Benjamin D. Sommers. 2022. "Trends in Missing Race and Ethnicity Information after Imputation in HealthCare.Gov Marketplace Enrollment Data, 2015-2021." *JAMA Network Open* 5 (6): e2216715. https://doi.org/10.1001/jamanetworkopen.2022.16715.

Brown, Sylvia, and Alex Engler. 2023. *Governing the Ascendancy of Automated Valuation Models: Regulating AI in Residential Property Valuation*. Washington, DC: Brookings Institution.

Chen, Jiahao, Nathan Kallus, Xiaojie Mao, Geoffry Svacha, and Madeleine Udell. 2019. "Fairness under Unawareness: Assessing Disparity When Protected Class Is Unobserved." In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 339–48. Atlanta: Association for Computing Machinery. https://doi.org/10.1145/3287560.3287594.

Consumer Financial Protection Bureau. 2014. *Using Publicly Available Information to Proxy for Unidentified Race and Ethnicity: A Methodology and Assessment*. Washington, DC: Consumer Financial Protection Bureau.

DeLuca, Kevin, and John A. Curiel. 2023. "Validating the Applicability of Bayesian Inference with Surname and Geocoding to Congressional Redistricting." *Political Analysis* 31 (3): 465–71. https://doi.org/10.1017/pan.2022.14.

Derose, Stephen F., Richard Contreras, Karen J. Coleman, Corinna Koebnick, and Steven J. Jacobsen. 2013. "Race and Ethnicity Data Quality and Imputation Using U.S. Census Data in an Integrated Health System: The Kaiser Permanente Southern California Experience." *Medical Care Research and Review* 70 (3): 330–45. https://doi.org/10.1177/1077558712466293.

Elliott, Marc N., Allen Fremont, Peter A. Morrison, Philip Pantoja, and Nicole Lurie. 2008. "A New Method for Estimating Race/Ethnicity and Associated Disparities Where Administrative Records Lack Self-Reported Race/Ethnicity." *Health Services Research* 43 (5p1): 1722–36. https://doi.org/10.1111/j.1475-6773.2008.00854.x.

Elliott, Marc N., Peter A. Morrison, Allen Fremont, Daniel F. McCaffrey, Philip Pantoja, and Nicole Lurie. 2009. "Using the Census Bureau's Surname List to Improve Estimates of Race/Ethnicity and Associated Disparities." *Health Services and Outcomes Research Methodology* 9 (4): 252–53. https://doi.org/10.1007/s10742-009-0055-1.

Emad Azhar Ali, Syed, Syed Sajjad Hussain Rizvi, Fong-Woon Lai, Rao Faizan Ali, and Ahmad Ali Jan. 2021. "Predicting Delinquency on Mortgage Loans: An Exhaustive Parametric Comparison of Machine Learning Techniques." *International Journal of Industrial Engineering and Management* 12 (1): 1–13.

Fout, Hamilton, Nuno Mota, and Eric Rosenblatt. 2022. "When Appraisers Go Low, Contracts Go Lower: The Impact of Expert Opinions on Transaction Prices." *Journal of Real Estate Finance and Economics* 65:451–91. https://doi.org/10.1007/s11146-020-09800-6.

Gould Ellen, Ingrid, and Katherine O'Regan. 2008. "Reversal of Fortunes? Lower-Income Urban Neighbourhoods in the US in the 1990s." *Urban Studies* 45 (4): 845–69. https://doi.org/10.1177/0042098007088471.

Haas, Ann, Marc N. Elliott, Jacob W. Dembosky, et al. 2019. "Imputation of Race/Ethnicity to Enable Measurement of HEDIS Performance by Race/Ethnicity." *Health Services Research* 54 (1): 13–23. https://doi.org/10.1111/1475-6773.13099.

Hardt, Moritz, Eric Price, and Nathan Srebro. 2016. "Equality of Opportunity in Supervised Learning." arXiv. https://doi.org/10.48550/ARXIV.1610.02413.

Harris, Ada. 2020. "Using Bayesian Improved Surname Geocoding (BISG) to Classify Race and Ethnicity in Administrative Employment Data by Industry: A Validation Study." Paper presented at the JSM Virtual Conference, American Statistical Association, August 6.

Heymans, Martijn W., and Iris Eekhout. 2019. *Applied Missing Data Analysis with SPSS and (R)Studio*. Amsterdam.

Howell, Junia, and Elizabeth Korver-Glenn. 2022. *Appraised: The Persistent Evaluation of White Neighborhoods as More Valuable Than Communities of Color*. St. Louis: Washington University, Weidenbaum Center on the Economy, Government, and Public Policy.

Imai, Kosuke, Santiago Olivella, and Evan T. R. Rosenman. 2022. "Addressing Census Data Problems in Race Imputation via Fully Bayesian Improved Surname Geocoding and Name Supplements." *Science Advances* 8 (49): eadc9824. https://doi.org/10.1126/sciadv.adc9824.

Kiefer, Len. 2023. "Economic and Housing Research: Financial Stress, Affordability, Migration." Freddie Mac presentation at Investor Day 2023, McLean, VA, April 20.

LaBriola, Joe. 2024. "Housing Market Appreciation and the White-Black Wealth Gap." *Social Problems* spae030. https://doi.org/10.1093/socpro/spae030.

McCartan, Cory, Robin Fisher, Jacob Goldin, Daniel E. Ho, and Kosuke Imai. 2024. *Estimating Racial Disparities When Race Is Not Observed*. Working Paper 32373. Cambridge, MA: National Bureau of Economic Research.

Mitchell, Margaret, Simone Wu, Andrew Zaldivar, et al. 2018. "Model Cards for Model Reporting." https://doi.org/10.48550/ARXIV.1810.03993.

Narragon, Melissa, Danny Wiley, Doug McManus, et al. 2021. "Racial and Ethnic Valuation Gaps in Home Purchase Appraisals." McLean, VA: Freddie Mac.

Neal, Michael, Jung Hyun Choi, Kathryn Reynolds, et al. 2021. *Why Do Households of Color Own Only a Quarter of the Nation's Housing Wealth When They Compose a Third of the Nation's Households?* Washington, DC: Urban Institute.

Neal, Michael, Vanessa G. Perry, and Linna Zhu. 2023. "To Err Is Automated: Have Technological Advances in the Mortgage Market Increased Opportunities for Black Homeownership?" New York: SSRN.

Neal, Michael, Sarah Strochak, Linna Zhu, and Caitlin Young. 2020. "How Automated Valuation Models Can Disproportionately Affect Majority-Black Neighborhoods." Washington, DC: Urban Institute.

PAVE (Interagency Task Force on Property Appraisal and Valuation Equity). 2022. *Action Plan to Advance Property Appraisal and Valuation Equity*. Washington, DC: PAVE

Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys*. 1st ed. New York: Wiley.

Stern, Alena, Manu Alcala Kovalski, Leah Hendey, Elizabeth Burton, Sandra Ambrozy, and Carlos A. Manjarrez. 2025. *Unveiling Disparities in Eviction: A Novel Approach to Demographic Imputation in Cleveland*. Washington, DC: Urban Institute.

Stern, Alena, Ajjit Narayanan, Steven Brown, Graham MacDonald, LesLeigh D. Ford, and Shena Ashley. 2021. *Ethics and Empathy in Using Imputation to Disaggregate Data for Racial Equity: A Case Study Imputing Credit Bureau Data*. Washington, DC: Urban Institute.

Task Force (Task Force on Property Appraisal and Valuation Equity). n.d. *Addressing the Persistent Misvaluation and Undervaluation of Property Owned by Minorities*. Baltimore: Maryland Department of Housing and Community Development.

US Treasury Department. 2024. *Artificial Intelligence in Financial Services: Report on the Uses, Opportunities, and Risks of Artificial Intelligence in the Financial Services Sector*. Washington, DC: US Treasury Department.

Van Zetten, W., G.J. Ramackers, and H.H. Hoos. 2022. "Increasing Trust and Fairness in Machine Learning Applications within the Mortgage Industry." *Machine Learning with Applications* 10 (December): 100406. https://doi.org/10.1016/j.mlwa.2022.100406.

Voicu, Ioan. 2018. "Using First Name Information to Improve Race and Ethnicity Classification." *Statistics and Public Policy* 5 (1): 1–13. https://doi.org/10.1080/2330443X.2018.1427012.

Zhang, Guangyu, Charles E. Rose, Yujia Zhang, et al. 2022. "Multiple Imputation of Missing Race and Ethnicity in CDC COVID-19 Case-Level Surveillance Data." *International Journal of Statistics in Medical Research* 11 (January): 1–11. https://doi.org/10.6000/1929-6029.2022.11.01.

Zhang, Yan. 2018. "Assessing Fair Lending Risks Using Race/Ethnicity Proxies." *Management Science* 64 (1). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3169831.

Zhu, Linna, Michael Neal, and Caitlin Young. 2022. "Revisiting Automated Valuation Model Disparities in Majority-Black Neighborhoods: New Evidence Using Property Condition and Artificial Intelligence." Washington, DC: Urban Institute.

About the Authors

Linna Zhu is a senior research associate in the Housing Finance Policy Center at the Urban Institute. She conducts data-driven quantitative research on policy issues related to US mortgage finance, housing supply, and racial inequality. Central to her current work is how to ensure individuals and communities can equally adapt to technological advancements, such as artificial intelligence, demographic shifts (e.g., a rapidly aging population), and the adversities posed by climate change. Zhu has published widely noted research articles on such topics as algorithmic bias in automated valuation models, artificial intelligence in mortgage lending, the Community Reinvestment Act, aging in place, housing security at older ages, intergenerational wealth transfer, Asian American and Pacific Islander homeownership, housing supply, and local planning responses. Her work has been cited in notable media outlets, including the *New York Times*, the *Washington Post*, *USA Today*, Marketplace, National Mortgage News, and CNBC. Her research has been published in leading academic and policy journals, such as *Regional Science and Urban Economics* and *Cityscape*. She serves on the advisory board of the Design Thinking Executive Program at the University of California, Riverside. Zhu holds a BA in economics from Renmin University of China, an MS in finance from Johns Hopkins University, and a PhD in public policy and management from the University of Southern California.

Judah Axelrod is an associate director of data science at the Urban Institute. He works in collaboration with Urban's Technology and Data Science team and Racial Equity Analytics Lab, with a particular focus on the equitable use of AI and automated systems. Axelrod holds a BA in economics from Rutgers University and an MSc in data science from the London School of Economics.

Amalie Zinn is a former research analyst in the Housing Finance Policy Center. Her research focuses on the intersection of racial equity and housing finance and how historically underserved people and communities access and preserve homeownership, credit, and capital. Before joining Urban, she interned for the Consumer Financial Protection Bureau, where she researched racial differences in mortgage denial rates across the United States. Zinn was also a research assistant for the Center for Healthy Minds' Mindful Policing project—a research initiative that investigated the effects of mindfulness on police officer stress levels—during which she spearheaded a cost-benefit analysis of the mindfulness training program. Zinn graduated Phi Beta Kappa from the University of Wisconsin–Madison with a BA in economics and French and a minor in public policy.

Statement of Independence

The Urban Institute strives to meet the highest standards of integrity and quality in its research and analyses and in the evidence-based policy recommendations offered by its researchers and experts. We believe that operating consistent with the values of independence, rigor, and transparency is essential to maintaining those standards. As an organization, the Urban Institute does not take positions on issues, but it does empower and support its experts in sharing their own evidence-based views and policy recommendations that have been shaped by scholarship. Funders do not determine our research findings or the insights and recommendations of our experts. Urban scholars and experts are expected to be objective and follow the evidence wherever it may lead.



500 L'Enfant Plaza SW
Washington, DC 20024

www.urban.org