Racial Disparities in Automated Valuation Models: New Evidence Using Property Condition and Machine Learning

Linna Zhu
Michael Neal
Urban Institute

Caitlin Young
Yale Law School

Abstract

Automated valuation models (AVMs), which exclude an appraiser's input in estimating a home's price, hold great promise for reducing costs and increasing the accuracy of home valuations. However, AVMs can manifest racial disparities, even when the algorithm remains agnostic to the neighborhood's majority race or the homebuyer's race. This study provides a quantifiable measure for auditing the performance of AVMs in majority-Black neighborhoods compared with their majority-White counterparts. The authors find that including data on property condition and employing more sophisticated machine learning techniques can help more accurately assess the percentage of the magnitude of AVM error and its underlying contributors. In addition, even with data improvement and machine learning, the authors still find evidence that AVMs yield larger valuation errors in majority-Black neighborhoods.

Introduction

The racial gap in homeownership rates is wide and persistent. Even when Black households do achieve homeownership, the value of their homes is typically less than that of White households. Although household, property, and community differences play key roles in explaining the home value gap, evidence also suggests that differences in appraised value can contribute to the racial disparity in home values.

One potential solution to appraisal bias is using automated valuation models (AVMs). An AVM is a computer-driven mathematical formula that uses property characteristics, local market information, and price trends to arrive at an estimated value for a property. With a surplus of data, an AVM can provide a standard approach and faster property value estimation.

Hypothetically, AVMs should help address instances of racial bias. By eliminating the appraiser's input, a potential source of racial inequity, the AVM should produce property value estimates agnostic to the community's racial makeup. In addition, because an AVM uses a standardized approach to estimate a property's value, the supposed elimination of racial differentiation should scale to communities more broadly.

The prospect for greater accuracy and standardization also indicates that AVMs can bring greater efficiency to the market. In addition, AVMs may quicken property valuation, because a task that may take a significant amount of time with an appraiser can be completed more quickly with an AVM. The AVM can also be centralized, eliminating the need for much of the appraisal profession.

The efficiency proposition AVMs pose is enhanced with the use of artificial intelligence (AI). AI, specifically machine learning (ML), can further speed up calculating a property's value. In addition, it can extend the data inputs to nontraditional data sources that could more accurately capture nuances in communities of color. By extension, the flexibility provided by AI and ML can make an AVM valuable at high levels of geography and across different regions of the country. On balance, AVMs may improve the functioning of the housing market because of their varied uses. Consumers—potential homebuyers or sellers—may assess a home's value on popular multiple listing service websites. These sites often use an AVM to estimate a property's value. In addition, an AVM may be used in the process of underwriting a mortgage loan.

Greater and varied use of AVMs throughout the housing industry has led to a need for auditing tools to assess the potential for racial inequity. This set of tools could improve the development of future AVMs, inform federal authorities governing these tools, and ensure that all share the benefits.

Overview of the Literature

Emerging literature suggests the potential for variation in property valuation due to race. Earlier research motivating the scholarly and policy analysis in the area of appraisals suggested that homes in Black neighborhoods were devalued by as much as $48,000, amounting to $156 billion in cumulative losses (Perry, Rothwell, and Harshbarger, 2018). Additional analysis affirmed the potential for racial bias, indicating that majority-Black and Hispanic neighborhoods were more likely to experience undervaluation compared with White neighborhoods (Freddie Mac, 2021). However, the appraiser's race may not inform the potential for bias, because research shows little difference in appraisal valuation discounts by the appraiser's race (Ambrose et.al., 2021). Although most appraisers are White, these results may also suggest that the potential for a racial issue is embedded in the process. Moreover, a lower valuation may provide additional benefits for the homebuyer if the property is negotiated to a lower contract price (Fout, Mota, and Rosenblatt, 2022).

The prospect of racial bias in appraisals coincides with policy analysis assessing AVMs. Research suggests that AVMs can help improve variation in future delinquencies (Bogin and Shui, 2020).

Although an AVM audit suggested the models may produce greater percent error in majority-Black communities than in majority-White ones (Neal et al., 2020), additional research suggests that despite inaccuracies—at least for refinances—AVMs may be more accurate than appraisers (Williamson and Palim, 2022).

The potential for appraisal bias has emerged as a key policy area for the Biden Administration, which created the Interagency Task Force on Property Appraisal and Valuation Equity (PAVE) to address the possibility of discrimination in home valuation. The first PAVE report discussed inequities in home valuation and highlighted the background and administrative steps the task force planned to take regarding appraisals and AVMs. Comments by Vice President Kamala Harris further amplified PAVE activities, highlighting the federal government's desire to prevent algorithmic bias in home valuation.[1] In response, six federal agencies—the National Credit Union Administration, Office of the Comptroller of the Currency, Federal Housing Finance Agency, Federal Reserve, Federal Deposit Insurance Corporation, and Consumer Financial Protection Bureau—issued a comment request on a proposed rule designed to ensure quality control standards for AVMs.[2]

The federal government also indicated the potential for discrimination in AI. Through its Blueprint for an AI Bill of Rights,[3] the Administration hopes to make automated systems work for the American people. The AI Bill of Rights advocates for algorithmic discrimination protections and seeks to ensure that automated systems—including AVMs—are used and designed in an equitable way (Engler, 2022). These steps support the activities of the National Institute of Standards and Technology, an agency that developed a voluntary AI risk management framework,[4] with a focus on fairness, equality, and equity that addresses issues such as harmful bias and discrimination.

This research proposes an AI-based method for assessing the potential for error in AVMs, building on earlier research auditing AVM accuracy (Neal et al., 2020). However, the team shows how the auditing technique may not be the best approach given the structure of the housing data. This research incorporates greater flexibility to better fit housing market data through the use of an AI-ML technique, which presents richer auditing results that can be useful to the broader housing ecosystem. This research sits at the intersection of appraisal bias, AI, and racial equity and is a part of the growing discussion on appraisal bias and finding solutions where this problem exists. It is also a part of the research on AI in the housing industry and has important policy implications. Addressing steps toward improving the appraisal system contributes to the policy discussion on appraisal bias. At the same time, by using AI in its analysis, the research also informs the AI policy area.

This inquiry is of great importance in that AVMs, like human appraisals, seek to answer the question, "What will this property sell for under current market conditions?" This question is

[1] https://www.whitehouse.gov/briefing-room/statements-releases/2023/06/01/fact-sheet-biden-harris-administration-takes-sweeping-action-to-address-racial-bias-in-home-valuations/.

[2] https://ncua.gov/newsroom/press-release/2023/agencies-request-comment-quality-control-standards-automated-valuation-models-proposed-rule.

[3] https://www.whitehouse.gov/ostp/ai-bill-of-rights/.

[4] https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf.

important, because it addresses the system in which homeownership transactions occur, the ways in which property values are currently determined, and the racial equity implications therein.

However, an overlapping but separate focus is on the value disparity of properties in Black communities compared with White communities. Although this article seeks to answer the preceding question, a robust assessment of the Black-White housing wealth gap answers a more fundamental question: "What is the true value of this property?" Although a property's selling price and its true value can be similar, certain conditions, including the history of structural racism, serve as a reminder of why they may be radically different as well. This article uses sales price as the benchmark to assess the accuracy of AVM estimates and recommends that, in the future, literature in this area move from the first question anchored to a property's sale price to the second question exploring fundamental value.

The next section of this article describes the AVM data, followed by a discussion of the research team's methodology, findings, conclusions, and implications.

Data Description

This study uses property records data to capture home sales information and produces property-level pairings with AVM estimates and property condition from private data vendors. In addition, this study relies on the American Community Survey to capture neighborhood-level characteristics.

Property Records Data

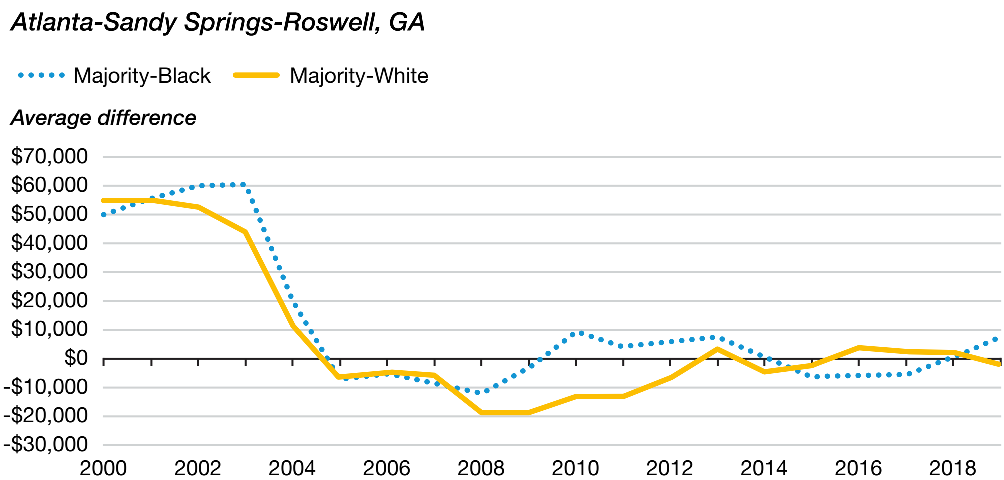
To examine whether AVM accuracy differs by race, this research compares AVM values with sales prices associated with arm's-length transactions at the property level between majority-Black and majority-White neighborhoods. The team analyzed Atlanta, Georgia, and Memphis, Tennessee. Each city had a significant Black population share and produced solid property-level pairings between AVM estimates and sales prices to analyze. In each city, instead of using the entire core-based statistical area (CBSA), the team used the counties with strong historical deeds data that could be matched with the AVM data. These counties are a small proportion of the total number of counties in each CBSA but account for most of the CBSA population. The Atlanta and Memphis counties account for 17 and 22 percent of the total counties in their CBSAs, respectively, and 63 and 74 percent of their CBSAs' respective populations.

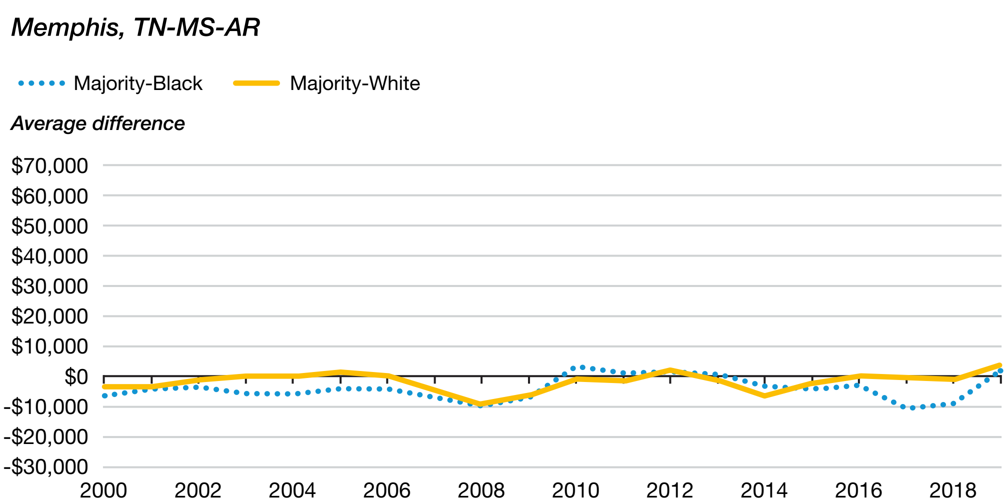
The research team employed property records data from a major data provider to combine information on AVM values, sales prices, and transaction dates for each traded property in 2018 for those selected counties within each city. The team then used 5-year (2014–18) American Community Survey data to extract the share of Black and White homeowners at the census tract level and merge the racial composition information with the property records data.

To characterize the differences between AVM values and sales prices, the research team first calculated directional inaccuracy, the difference between the AVM estimate and the corresponding sale price. For example, if one home is undervalued by $20,000 and a second is overvalued by $20,000, the average directional inaccuracy across these two properties is zero. As exhibit 1 illustrates, direction inaccuracy does not systematically differ according to neighborhood racial composition in Atlanta or Memphis and has not been significant between 2005 and 2018. In addition, the average inaccuracy across majority-Black neighborhoods fluctuated around zero during this period. In the Memphis CBSA, the average error across majority-Black neighborhoods has been systematically below zero over time but only to a modest degree. Exhibit 1 also shows that the average difference in both Atlanta and Memphis across majority-Black neighborhoods is neither consistently above nor consistently below that of majority-White neighborhoods.

Exhibit 1

Directional Automated Valuation Model Inaccuracy, by Majority Race in Neighborhood



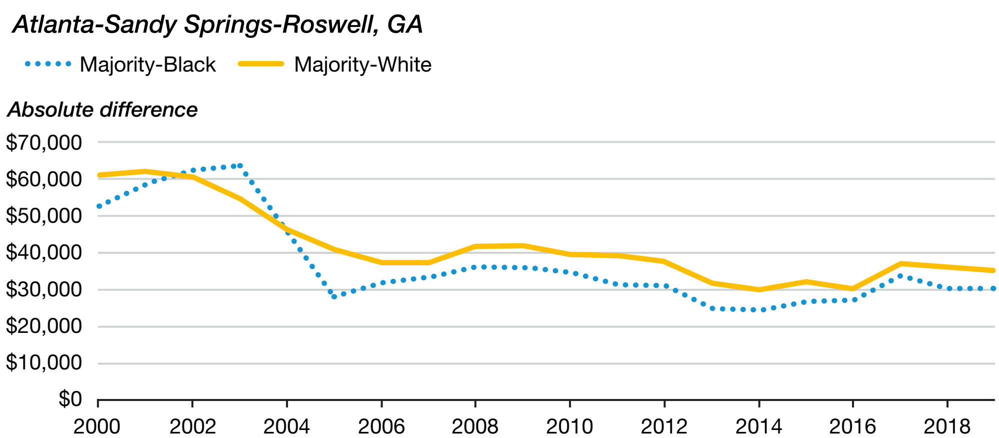


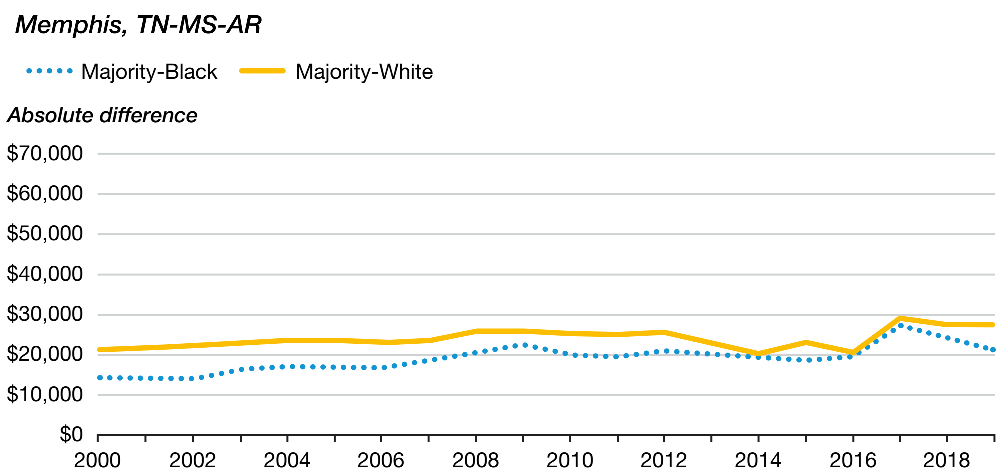
Sources: Authors' calculations of property records data; 2014–18 American Community Survey data

In contrast, the magnitude of inaccuracy, measured as the absolute difference between the AVM estimate and corresponding sale price, in majority-Black neighborhoods is consistently below the absolute inaccuracy in majority-White neighborhoods in the two CBSAs analyzed. For example, when one home is undervalued by $20,000 and a second home is overvalued by $20,000, the absolute value difference is $20,000, irrespective of whether the property is undervalued or overvalued by $20,000. Exhibit 2 shows that, except from 2002 to 2004, AVM inaccuracy in majority-Black neighborhoods in the Atlanta CBSA was smaller than in majority-White neighborhoods. Data from Memphis also reveal that the magnitude of inaccuracy in majority-Black neighborhoods was smaller than in majority-White neighborhoods.

Exhibit 2

Magnitude of Automated Valuation Model Inaccuracy, by Majority Race in Neighborhoods





Sources: Authors' calculations of property records data; 2014–18 American Community Survey data

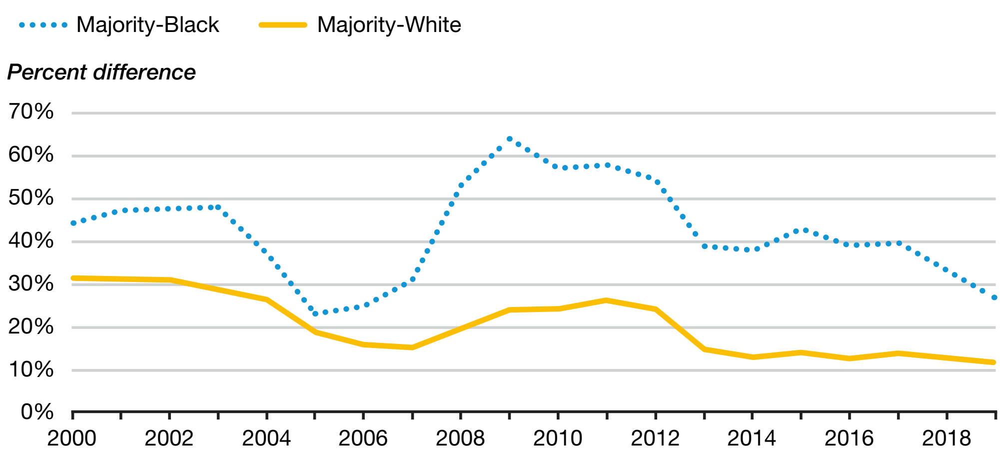
Finally, the research team calculated the percentage magnitude of inaccuracy, or the absolute difference between the AVM estimate and the corresponding sale price divided by the sale price. For example, if one home is undervalued by $20,000 and worth $200,000, the extent of error is greater than that of another home overvalued by $20,000 and worth $300,000. Exhibit 3 illustrates the percentage magnitude of inaccuracy in majority-White and majority-Black neighborhoods in

Atlanta and Memphis between 2000 and 2018. It demonstrates that the magnitude of inaccuracy is much higher in majority-Black neighborhoods in Atlanta and Memphis. The higher percentage magnitude of inaccuracy in majority-Black neighborhoods is attributable to lower average home values than in majority-White neighborhoods. The percentage magnitude of inaccuracy is roughly twice as large in majority-Black neighborhoods as in majority-White neighborhoods and is notably more volatile. For example, in 2009, the percentage magnitude of inaccuracy in majority-Black areas in Atlanta was 64 percent compared with 24 percent in majority-White neighborhoods. Although it has steadily improved since then, as of 2019, it was still more than twice the size in majority-Black neighborhoods than in majority-White neighborhoods. These results are consistent over time in both the Atlanta and Memphis CBSAs.

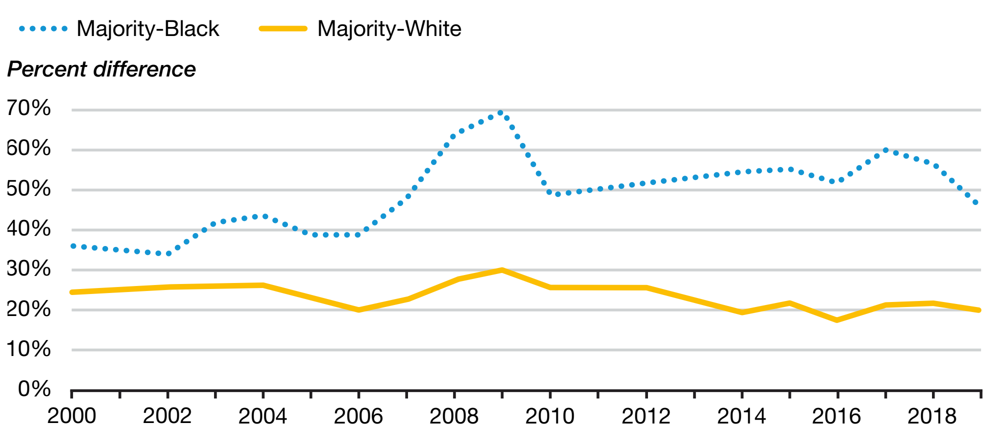
Exhibit 3

Percentage Magnitude of Automated Valuation Model Inaccuracy, by Majority Race in Neighborhood

Atlanta-Sandy Springs-Roswell, GA



Memphis, TN-MS-AR



Sources: Authors' calculations of property records data; 2014–18 American Community Survey data

These findings illustrate that AVMs could both undervalue and overvalue sales prices, both of which can be harmful. Undervaluation can limit wealth gains for homeowners seeking to refinance or sell their homes, and overvaluation may result in credit risk holders underestimating risk and may speed up irrational inflation of property values, potentially resulting in a future home price correction (PAVE, 2022). In addition, lower home values in majority-Black neighborhoods, partly reflecting historic discrimination, increase the risk of AVM error. Although the research team did not find systematic undervaluation bias in AVMs, the team observed that its AVM produced a racially disparate outcome in the form of a greater percentage magnitude of AVM error in majority-Black neighborhoods than in majority-White neighborhoods.

New Data on Property Condition

This research used the exterior condition rating (ECR) measure to capture property-level condition. The property intelligence firm CAPE Analytics provided the team with property-level ECRs. 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. Exhibit 4 provides the five-point scale definitions.

Exhibit 4

CAPE Analytics Exterior Condition Rating Scale Definitions

Rating	Definition
Excellent	Parcel condition falls within the best 5% of parcels
Good	Parcel condition falls within the best 20% but not the best 5% of parcels
Fair	Parcel condition is average (50% of parcels)
Poor	Parcel condition falls within the worst 23% but not the worst 2% of parcels
Severe	Parcel condition falls within the worst 2% of parcels
Unknown	Parcel could be assigned a property condition

Source: CAPE Analytics

In this analysis, the research team matched property records data for Atlanta and Memphis metropolitan areas with the ECRs from CAPE Analytics based on property latitudes and longitudes, parcel lot assessor parcel numbers, and transaction dates. 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, did not skew the overall distribution.

For this analysis, the team collapsed the five-point ECR scale from CAPE Analytics into three categories: poor (includes poor and severe), fair, and good (includes good and excellent). Exhibit 5 presents the ECR distributions based on the grouped categories for the matched sample within the Atlanta and Memphis CBSAs.

Exhibit 5

ECR Distribution in the Atlanta and Memphis CBSAs

CBSA	ECR	Majority-Black Neighborhoods	Majority-White Neighborhoods
Atlanta-Sandy Springs-Roswell, GA	Good	9%	13%
Atlanta-Sandy Springs-Roswell, GA	Fair	45%	52%
Atlanta-Sandy Springs-Roswell, GA	Poor	46%	34%
Memphis, TN-MS-AR	Good	10%	14%
Memphis, TN-MS-AR	Fair	46%	52%
Memphis, TN-MS-AR	Poor	44%	34%

CBSA = core-based statistical area. ECR = exterior condition rating.
Source: Authors' calculations using data from the American Community Survey; CAPE Analytics; and a major property records provider

In the Atlanta and Memphis CBSAs, single-family properties in majority-Black neighborhoods are more likely to have a poor rating and are less likely to have a fair or good rating than those in majority-White neighborhoods (exhibit 5). In Atlanta, 46 percent of single-family properties in majority-Black neighborhoods had a poor rating in 2018 compared with 34 percent in majority-White neighborhoods. In Memphis, 44 percent of single-family properties in majority-Black neighborhoods had a poor rating compared with 34 percent in majority-White neighborhoods.

Intuitively, a property condition assessment reflects both external and internal adequacy (Neal, Choi, and Walsh, 2020). Before examining the impact of the ECR measure on the percentage magnitude of AVM error, the team first established that external property condition is a reasonable proxy for the property condition overall, both inside and out. To do so, the team calculated the polychoric correlation—the correlation between two categorical variables—between exterior property conditions and interior structural conditions using American Housing Survey (AHS) data.[5,6]

The AHS is a recognized source of information on property condition, albeit with a limited suite of variables and geographic granularity. The team used the survey's information on roofs and outside walls across owner-occupied homes nationwide to assess exterior conditions and used its information on fundamental or structural problems, such as floors, windows, foundations, and peeling paint, to assess interior conditions. The team found a polychoric correlation of 0.67 between exterior and interior conditions.

This polychoric correlation should be regarded as a lower-bound estimate of the true strength of the correlation because of the AHS's limited variables to capture a property's exterior condition. The ECRs in this analysis cover all of a parcel's visible external features—including roofs, yards, driveways, and debris—compared with AHS variables that cover only roofs and outside walls. Because the ECR variable in this analysis is a more comprehensive measure of exterior condition, its correlation with interior condition would likely be greater than 0.67, suggesting that it should be a reasonable proxy for the property condition overall.

[5] The exterior condition categorical variable is a score variable derived by adding five dummy variables from the AHS—roofhole, roofshin, roofsag, wallside, and wallslope. These variables flag exterior condition problems with the roof and outside walls.

[6] The interior condition categorical variable is a score variable derived by adding six dummy variables from the AHS—floorhole, fndcrumb, paintpeel, wallcrack, winboard, and winbroke. These variables flag interior condition problems with floors, foundation, ceilings, windows, and interior paints.

Methodology

The Ordinary Least Squares (OLS) Approach

To determine how the ECR contributes to the percentage magnitude of AVM appraisal inaccuracy in the Atlanta and Memphis CBSAs, the research team first conducted the OLS regressions with 2018 as the analysis period and focused only on single-family home purchases. The team followed the model specification from previous research and controlled for key neighborhood characteristics affecting the percentage magnitude of AVM inaccuracy, as equation 1 shows (Neal et al., 2020).

$$Pct_{Diffi,2018} = \alpha_0 + \beta ECR_{i,2018} + \gamma Black_{n,2018} + \delta HP_{i,2018} + \theta NC_{n,2018} + \varepsilon_{i,2018} \quad (1)$$

$Pct_{Diffi,2018}$ is the percentage magnitude of automated valuation model inaccuracy measured by the absolute difference between the sales price and the AVM value divided by the sales price.

$ECR_{i,2018}$ captures property condition in the year 2018. $Black_{n,2018}$ represents the racial composition of the tract in which individual property i locates. This variable is a dummy, with 1 equal to majority-Black neighborhoods in which the share of Black households is greater than 50 percent and 0 equal to majority-White neighborhoods in which the share of White households is greater than 50 percent.[7] $HP_{i,2018}$ is the home value of property i. $NC_{n,2018}$ controls for key neighborhood characteristics affecting the percentage magnitude of AVM inaccuracy. These neighborhood characteristics are grouped along three dimensions: differences in properties within a neighborhood, neighborhood conditions, and turnover rates. Exhibit 6 presents summary statistics of those variables.

Exhibit 6

Summary Statistics

Variable	Majority-Black Neighborhood		Majority-White Neighborhood	
	Mean	SD	Mean	SD
Percent magnitude of AVM inaccuracy	36.3%	58.0%	13.8%	24.5%
Home value	127,756	80,969	329,443	204,256
Property age	46.4	24.3	37.5	22.1
Standard deviation of neighborhood property ages	14.0	7.1	12.5	6.5
Percentage deviation of neighborhood property values	43.2%	14.8%	34.4%	11.3%
Gentrified neighborhood	7.5%	26.3%	2.1%	14.3%
Share of neighborhood distressed home sales	15.7%	20.8%	5.0%	13.7%
Neighborhood median household income	46,198	16,657	92,312	30,955
Neighborhood number of households	2,320	1,328	2,630	1,214
Turnover rate at neighborhood level	8.8%	4.1%	7.5%	3.3%
Exterior Condition Rating				
Good	9.0%	29.0%	13.4%	34.0%
Fair	45.2%	50.0%	52.0%	50.0%
Poor	45.8%	50.0%	34.0%	47.0%

AVM = automated valuation model. SD = standard deviation.

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

[7] Census tracts in which the share of Black or White households is less than 50 percent of the total household population are excluded from the data. In the Atlanta, Memphis, and Washington, D.C. CBSAs, these tracts account for 63, 74, and 56 percent of their respective populations.

To capture property differences within neighborhoods, the research team constructed two variables: the standard deviation of neighborhood property ages and the percentage deviation of neighborhood property values.[8] Standard and percentage deviations measure the dispersion of properties by age and home value, respectively. Jiang and Zhang (2022) show a greater degree of house price dispersion in Black-dominant ZIP Codes.

To capture neighborhood conditions, the team included measures for gentrified neighborhoods, the share of distressed neighborhood sales, the neighborhoods' median household income, and the number of households within a neighborhood. Majority-Black neighborhoods are more likely to experience gentrification, which generally causes permanent and rapid home price increases as land values increase. AVMs cannot quickly pick up these house price shocks, contributing to greater AVM errors in gentrifying neighborhoods. The team considered a neighborhood to be gentrified if it met two criteria (Ellen and O'Regan, 2008): the tract-level income is less than 70 percent of the income in the metropolitan statistical area (MSA) and the neighborhood (identified at the census tract level) experienced at least a 10-percentage-point increase in the ratio of tract-level income to MSA-level income during the year. Under this definition, 7.3 percent of majority-Black neighborhoods in the United States were gentrified in 2018, which is almost five times the share of majority-White neighborhoods that were gentrified. In addition, majority-Black neighborhoods experienced significantly more distressed sales. Forced home sales, such as foreclosures, more often occur among low-price homes than among high-price homes (Campbell, Giglio, and Pathak, 2011). AVM accuracy will likely be compromised if a distressed sale results in a lower price than similar homes in the neighborhood. Among all home sales nationally in majority-Black neighborhoods in 2018, 16.0 percent were distressed home sales, almost four times the rate in majority-White neighborhoods (4.4 percent). The average household income in majority-Black neighborhoods is nearly one-half of that in majority-White neighborhoods. Lower incomes in majority-Black neighborhoods partly explain lower sales prices in these neighborhoods, which can increase the percentage magnitude of inaccuracy (Neal, Choi, and Walsh, 2020).

This analysis defines turnover rate as the number of home sales per year divided by the number of homes. Because AVM algorithms are based on comparable sales, greater turnover rates would provide a larger sample of comparable sales for AVM algorithms to provide more accurate estimates. The turnover rates in majority-Black neighborhoods are slightly higher than those of majority-White neighborhoods.

The OLS regressions usually make several assumptions about the underlying data: (1) a linear relationship between the dependent variable and the independent variables; (2) normality of the residuals—that is, the residual errors are assumed to be normally distributed; (3) homoscedasticity—that is, the residuals are assumed to have a constant variance; and (4) independence of residuals error terms. Given the complexity of the underlying data and the high dimensionality of the independent variables, the data structure in this analysis may not meet those linear assumptions, and thus, the OLS regression may not be the best audit approach to investigate the AVM error.

[8] For each property value, the percentage deviation of neighborhood home values subtracts that known value from the mean property value and divides the result by the property's value.

The LightGBM Approach

To address this issue, the team employed a nonparametric supervised ML approach, the LightGBM—a gradient boosting framework that uses tree-based learning algorithms. Nonparametric supervised ML is a highly innovative and effective vein in predictive data analysis and has several advantages over traditional linear parametric methods, such as OLS. First, ML methods fully use the available historical data. By repeatedly validating the model through training and prediction sets derived from existing data, the methods can map new data entries into specific dependent variables based on relevant independent variables used to train the model. Second, ML methods possess great capacities and effectiveness in handling interrelated variables (for example, collinearity; Aggarwal, 2015), thus boosting prediction accuracy from traditional regression methods. Third, ML methods do not assume linearity and can handle complex datasets that do not fulfill the requirements of traditional regression models.

LightGBM is among the most recent and efficient ML prediction algorithms (Ke et al., 2017). It provides more regularized model formalization and better overfitting control (Ashari, Paryudi, and Tjoa, 2013). It is also an algorithm that assumes no linearity, providing more appropriate handling to the complex dataset in this analysis. Thus, the team chose LightGBM as a nonparametric, tree-based machine learning counterpart to the OLS model, which helped the team explore the broader question of whether and how sophisticated artificial intelligence tools improve the analysis of automated systems.

The team first partitioned the entire dataset into a training set (70 percent) and a testing set (30 percent), then set up cross-validation through a stratified k-fold (k = 5) process. Next, the team entered all relevant independent variables into the LightGBM model as predictors and entered the outcome variable—the percentage magnitude of AVM inaccuracy—as the prediction target. Finally, the team employed a Bayesian optimization procedure to obtain the model parameters supporting the most accurate predictions of the target variable. The following discusses the detailed methodology.

Data Partitioning and Model Validation

This study divided the processed dataset for Memphis and Atlanta into two portions—the training and testing sets—to regulate the efficiency of the ML procedures. The LightGBM model is trained using only the training set and tested using only the testing set. This split is vital to demonstrate and tune the model's response to new data being processed for the first time. For the robustness of the division, the research team put 70 percent of the data into the training portion and the remaining 30 percent into the testing portion.

To enhance the model's validity, accuracy, and robustness, the team also employed a five-fold cross-validation procedure on the training set and adopted the k-fold (k = 5) cross-validation because of its efficiency and smoothness during the validation. Each dataset was randomly separated into k numbers of folds; k-1 folds were used for training purposes, and the remaining fold was simultaneously used for testing. The results over the k-testing folds were averaged at the end. Exhibit 7 lists the summary statistics on several key variables to show that the partition process does not distort the distribution in either the training or the testing datasets.

Exhibit 7

Summary Statistics: Training and Test Data

	Training Data (Mean)	Test Data (Mean)
Share of majority-Black neighborhoods	42.2%	42.3%
Home value	244,067	244,839
Exterior Condition Rating		
Good	11.6%	11.5%
Fair	49.2%	49.4%
Poor	39.3%	39.1%

Source: Authors

Model Parameters

To tune the hyperparameters of the LightGBM model in conjunction with the k-fold cross-validation procedure, the team employed a Bayesian optimization procedure to obtain the model parameters that would best predict the regression outcome. The parameter optimization boundaries are in exhibit 8. With those parameters, the team obtained its optimized LightGBM prediction model based on the 70-percent training set.

Exhibit 8

Model Parameters in the LightGBM Model

The parameter optimization boundaries:	
Learning rate	0–1
Number of leaves	5–40
Minimum gain to split	0–10
Minimum sum of hessian in leaf	0–20
The parameter value in the final optimized LightGBM model:	
Number of threads	6
Number of leaves	25
Learning rate	0.468
Minimum gain to split	1.823
Minimum sum of hessian in leaf	9.517

Source: Authors

Evaluation of Model Accuracy

Root mean square error (RMSE) is the standard deviation of the residuals (predicted errors) and is used to measure model prediction accuracy. The research team took advantage of its strong interpretability because it has the same unit as the regression target variable. The team tested the RMSE for the LightGBM prediction model and compared it with the RMSE for the OLS model to test whether the LightGBM model made more accurate predictions.

Identification of AVM Racial Disparity: Feature Importance

Shapley Additive Explanations (SHAP) is a novel way of computing feature contribution toward the prediction while preserving the sum of contributions being equal to the final outcome. It is especially well suited for tree-based models. SHAP values calculate a feature's importance by comparing what a model predicts with and without the feature. Given that the order in which a model sees a feature can affect its predictions, SHAP values account for all possible orders to make sure all features are fairly compared.

The team calculated the SHAP values for each predictor to determine the predictors' relative importance and effect on the model outcome. Their SHAP values allowed the team to delve deeper into the predictive model's complexity, partially unveil the ML black box, and evaluate the effect of neighborhood race and ECR on predicted AVM error.

Quantification of Feature Importance: Synthetic Control Method

Although the SHAP value could provide evidence of a specific feature's importance, it does not quantify the magnitude of the impact. To quantify the impact, the research team employed a synthetic control method to examine identified racial disparity in the AVM valuations, the ECR's impact, and the effect of the intersection of neighborhood majority race and the ECR.

The team extracted all properties in majority-Black neighborhoods from the test set and treated this selected group as the benchmark dataset. First, the team predicted the AVM error for the benchmark dataset using its constructed LightGBM model. Second, to create the corresponding synthetic datasets, the team changed the neighborhood majority race variable from Black to White, holding everything else constant. The team created 10 synthetic datasets and altered the neighborhood race in the share of properties, starting from 10 to 100 percent. For the tenth synthetic dataset, all properties' neighborhood majority race was randomly switched from Black to White. Third, the team predicted the AVM error for each synthetic dataset using its constructed LightGBM model and then obtained 10 corresponding predicted mean values of AVM error. Finally, the team compared the predicted target variable from the synthetic datasets with the benchmark dataset. This difference between the two predicted values measured the racial disparity. The properties in the synthetic groups were the same as those in the benchmark group, except that the neighborhood majority race was different. The results were expected to shed light on whether systemic racism is a key factor behind the AVM error.

Results

OLS Regressions

Exhibit 9 shows the OLS regression results, indicating that an ECR rating worse than good would raise the percentage magnitude of AVM error. Relative to an otherwise similar property with a good rating, a property with a fair rating would increase the AVM's percentage magnitude of error by 2.72 percentage points. Similarly, relative to a property with a good rating, a property with a poor rating would further increase AVM inaccuracy, increasing the percentage magnitude of error by 4.35 percentage points. In this case, the magnitude of the coefficient means that for a home with an average sales price of $250,000, having a poor rating is associated with a $10,875 greater percentage AVM error than a property with a good rating, holding all other attributes constant. Controlling for the ECR slightly reduces the magnitude of this Black neighborhood coefficient from 3.593 percentage points in column four to 3.499 percentage points in column five. This result indicates that even when controlling for property condition, location in a majority-Black neighborhood rather than a majority-White one still raises the percentage magnitude of error by 3.499 percentage points. The difference is a $4,549 greater percentage AVM error for a home with an average sales price of $130,000 in a majority-Black neighborhood compared with a property with the same attributes and sales price in a majority-White neighborhood.

Exhibit 9

Ordinary Lease Squares Regression Results

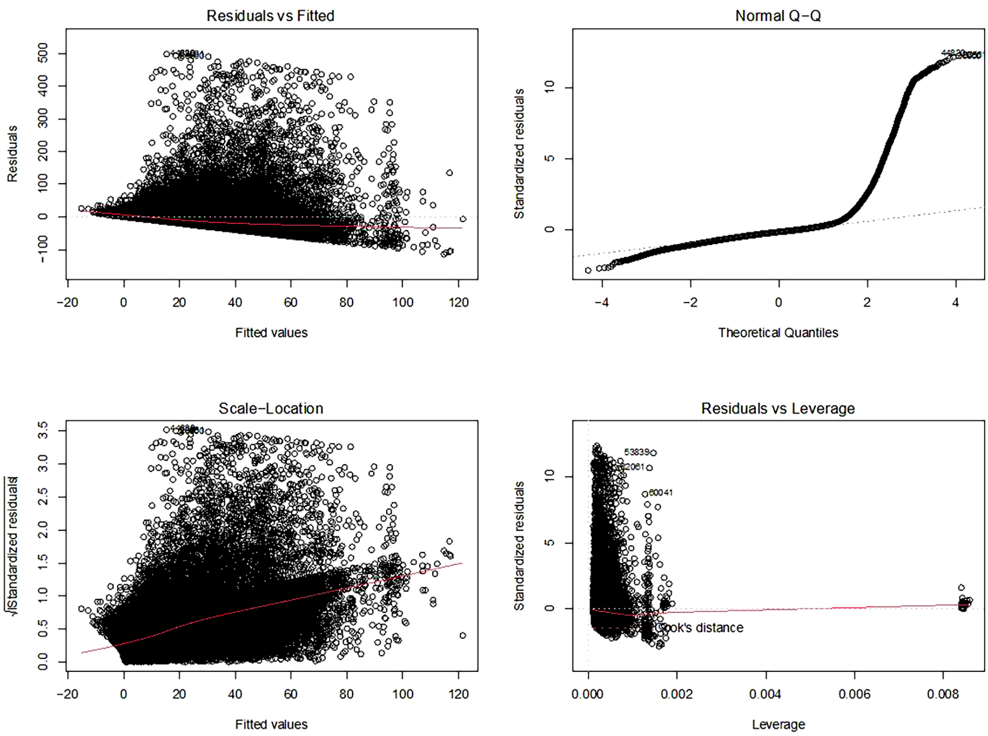
	Dependent Variable: Percentage Magnitude of Automated Valuation Model Inaccuracy				
	(1)	(2)	(3)	(4)	(5)
Black neighborhood	21.024*** (0.393)	4.816*** (0.504)	4.040*** (0.499)	3.593*** (0.542)	3.499*** (0.542)
Log (Home value)		– 15.785*** (0.316)	– 12.535*** (0.328)	– 10.358*** (0.402)	– 10.075*** (0.403)
Standard deviation of neighborhood property ages			0.155*** (0.028)	0.058** (0.029)	0.059** (0.028)
Percentage deviation of neighborhood property values (%)			0.453*** (0.014)	0.422*** (0.014)	0.422*** (0.014)
Share of neighborhood distressed home sales (%)				– 0.005 (0.010)	– 0.006 (0.010)
Gentrified neighborhood				2.155*** (0.817)	2.174*** (0.817)
Log (Neighborhood median household income)				– 4.116*** (0.652)	– 4.153*** (0.652)
Log (Number of households in neighborhood)				– 5.107*** (0.377)	– 5.016*** (0.377)
Neighborhood-level turnover rate (%)				– 0.246*** (0.049)	– 0.230*** (0.049)
Exterior condition rating (ECR): Fair					2.718*** (0.530)
ECR: Poor					4.350*** (0.546)
Constant	13.860*** (0.752)	213.240*** (4.063)	156.708*** (4.336)	219.975*** (6.708)	213.170*** (6.765)
County fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	62,609	62,609	62,606	62,606	62,606
R^2	0.086	0.121	0.138	0.142	0.143
Adjusted R^2	0.086	0.121	0.138	0.142	0.143
Residual standard error	41.587 (df = 62602)	40.784 (df = 62601)	40.383 (df = 62596)	40.297 (df = 62591)	40.276 (df = 62589)
F-statistics	981.818*** (df = 6; 62606)	1,230.664*** (df = 7; 62601)	1,115.744*** (df = 9; 62596)	739.866*** (df = 14; 62591)	652.256*** (df = 16; 62589)

$^{**}p < 0.05$. $^{***}p < 0.01$. df = degrees of freedom.
Note: The dependent variable is the percentage magnitude of automated valuation model error.
Source: Authors

The research team ran several diagnostic tests to confirm whether an OLS regression is the best approach for examining property condition's effect on AVM accuracy (exhibit 10). The residuals-versus-fitted plot indicates that the randomness of the error term was not met. The Normal Q–Q plot shows that the residuals from the OLS regressions (column five) are not normally distributed. In addition, the scale-location plot shows a severe heteroscedasticity problem. All these results suggest that OLS regression may not be the best approach.

Exhibit 10

Diagnostic Tests for Linear Regression Accuracy



Source: Authors

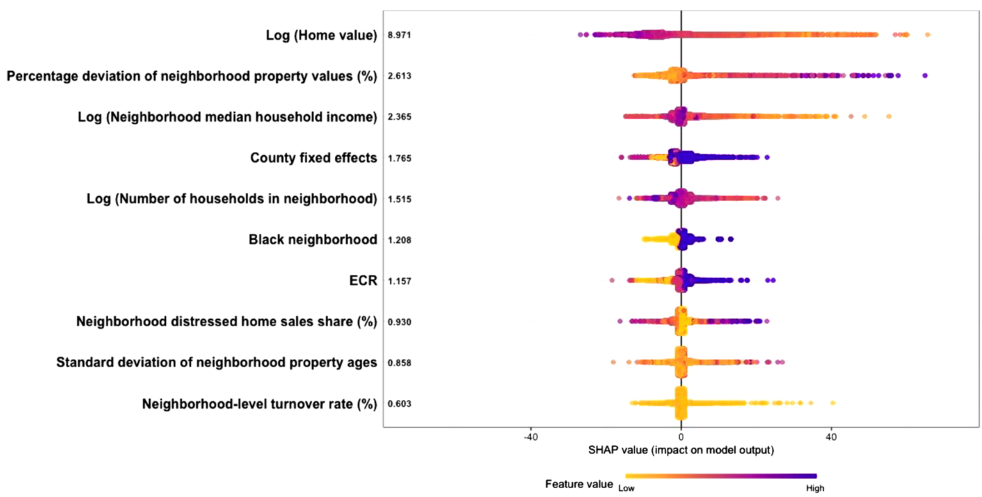
LightGBM

LightGBM has a greater predictive power than the OLS regressions. After completing data partitioning, model validation, and parameter tuning, the optimized LightGBM model produced an RMSE of 40.4. Applying the same data partitioning procedure to the OLS regression produced an RMSE of 46.2. This result suggests that LightGBM produces a 5.8-percentage-point improvement in the model's fit to explain the AVM inaccuracy. The magnitude of RMSE improvement does not differ between majority-Black and majority-White neighborhoods, with only around a 0.05-percentage-point difference. These results validate the team's selection of LightGBM over OLS regressions with respect to evaluating the identified AVM racial disparity. By relaxing the linear assumptions, this nonparametric, tree-based ML approach provides more appropriate handling of the complex dataset.

Majority-Black neighborhoods are associated with greater predicted AVM inaccuracy. Exhibit 11 illustrates the SHAP values for each feature. The y-axis displays the feature name in order of importance from top to bottom. The value next to the variable name is the mean SHAP value. The x-axis is the SHAP value. Each point represents a row from the training dataset. The gradient color represents that variable's original value. Continuous numerical variables, such as the log of home values, can contain the whole color spectrum. Dummy variables, such as majority-Black neighborhood, can take only two colors.

Exhibit 11

Shapley Additive Explanations Values



ECR = exterior condition rating. SHAP = Shapley Additive Explanations.
Source: Authors

For example, the percentage deviation of neighborhood property values is a key feature contributing to the AVM percentage error prediction, because it ranks as the second feature after the home values in log form. The percentage deviation of neighborhood property values is associated with high and positive values on the target (exhibit 11). The color distribution shows a high value in the color bar at the bottom of the figure, with purple indicating a high percentage deviation of neighborhood property values. The preponderance of purple observations indicates a positive value on the right side of the zero vertical line, signaling the SHAP value is above zero. Given that it is a continuous variable, its color scheme contains the full-color spectrum of the feature value from low (yellow) to high (purple). This result suggests that greater heterogeneity with respect to property values in a neighborhood contributes to a greater predicted percentage magnitude of AVM error.

Exhibit 11 illustrates that the majority-Black neighborhood variable is associated with high and positive values on the target. The value is high because of the figure value color bar and is positive from the x-axis SHAP value. Because this is a dummy variable, its color scheme contains only two colors for low (yellow) and high (purple). This high and positive relationship suggests that compared with majority-White neighborhoods, AVM error in majority-Black neighborhoods is greater.

Similarly, the ECR is associated with high and positive values on the target. The ratings are coded as 1 (good), 2 (fair), and 3 (poor), so high ECR values mean poor property conditions, indicating that properties in poor condition are associated with greater AVM inaccuracy. In addition, when looking at the order of importance from top to bottom, the majority-Black neighborhood variable ranks higher than the exterior condition rating, distressed sales share, and turnover rate variables. The SHAP values for neighborhood majority race combined with its ranking align with what was found in the OLS regressions, suggesting that even though an AVM algorithm does not have disparate input such as race, it still can produce racial disparities.

The Role of Historic Racism in AVM Estimates

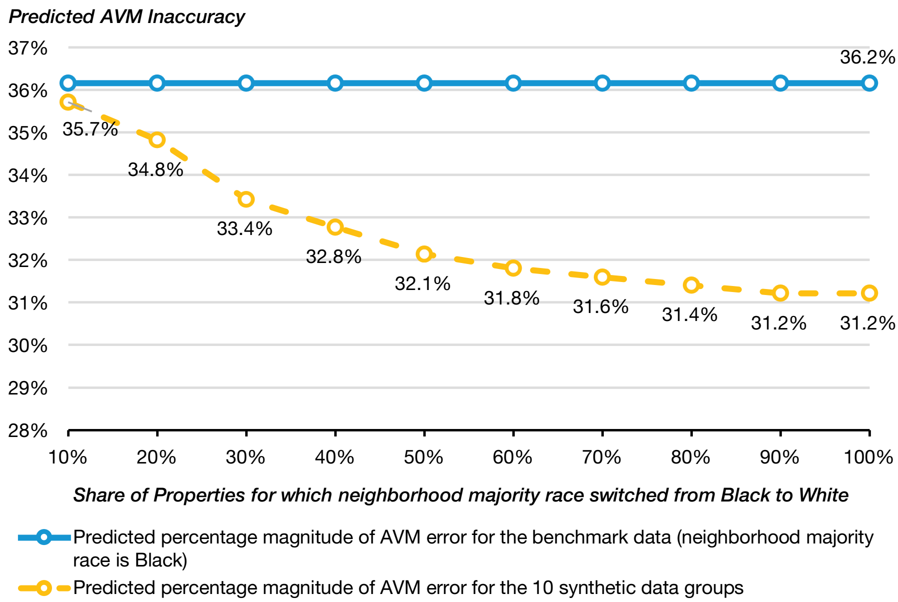
The blue line in exhibit 12 illustrates the predicted percentage magnitude of AVM error for the benchmark data, and the yellow line represents the predicted percentage magnitude of AVM error for the 10 synthetic data groups.[9] The x-axis represents the share of properties for which neighborhood majority race switched from Black to White. As the share increases, the gap between the two lines widens. This result indicates that, for example, if 60 percent of properties currently in majority-Black neighborhoods "move" to majority-White neighborhoods while keeping all other attributes constant, their associated predicted percentage magnitude of AVM error could decline from 36.2 to 31.8 percent, a 4.4-percentage-point difference. Further, if all properties currently in majority-Black neighborhoods "move" to majority-White neighborhoods, the predicted AVM error could decline by 5.0 percentage points, which is an upper-bound estimate of the racial disparity in AVMs. Such results suggest that historic racism could be a key factor behind greater AVM error in majority-Black neighborhoods.

A similar synthetic control approach was applied to examine the ECR's impact. The team used all properties with a poor ECR from the prediction set as the benchmark dataset (exhibit 13), then created 10 corresponding synthetic datasets by changing the ECR from poor to good, holding everything else constant. Again, the share that switched from poor to good ECR increased from 10 to 100 percent. For the tenth synthetic dataset, all properties' ECRs changed from poor to good. The team calculated the predicted AVM error for the benchmark group (the blue solid line) and the synthetic data groups (the yellow dotted line). Exhibit 13 demonstrates that if all properties currently in poor condition are upgraded to good condition, all else constant, their associated AVM error could decline from 26.1 to 21.8 percent, a 4.3 percentage-point difference. This decline in AVM error provides further evidence that policies to improve housing adequacy could reduce the adverse effect of the percentage magnitude of AVM error.

[9] The team ran the model 10 times against the same dataset, which is the 30-percent test set. Because the shares of properties in the test set were randomly selected using a with-replacement approach, results would differ slightly every time these 10 synthetic datasets were created. The 70–30 percent train-test split and the fivefold cross-validation ensured that the overall monotonically decreasing trend in exhibit 3 was statistically robust. In addition, the with-replacement selection caused slightly different results every time these 10 synthetic datasets were created.

Exhibit 12

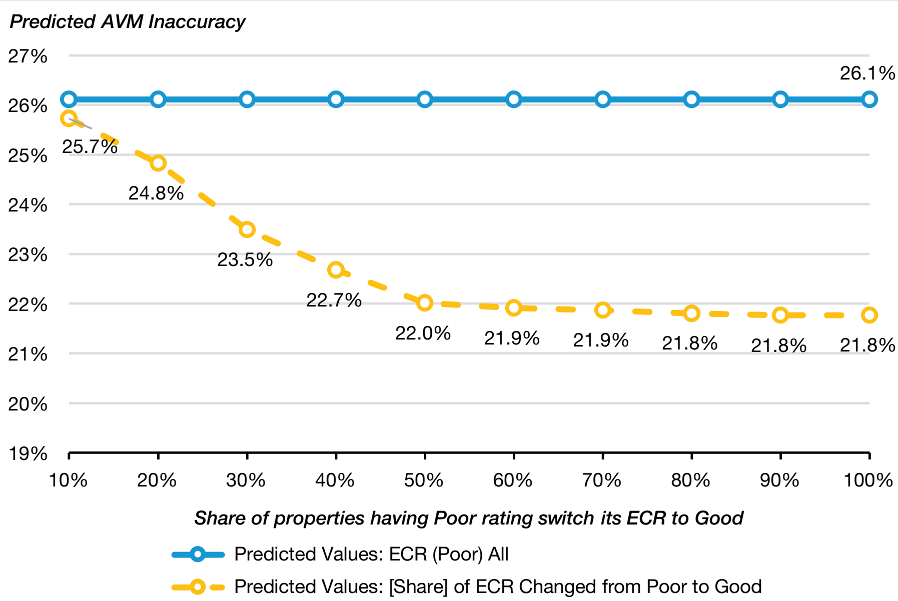
Racial Disparity: Neighborhood Majority Race



AVM = automated valuation model.
Source: Authors

Exhibit 13

Impact of ECR

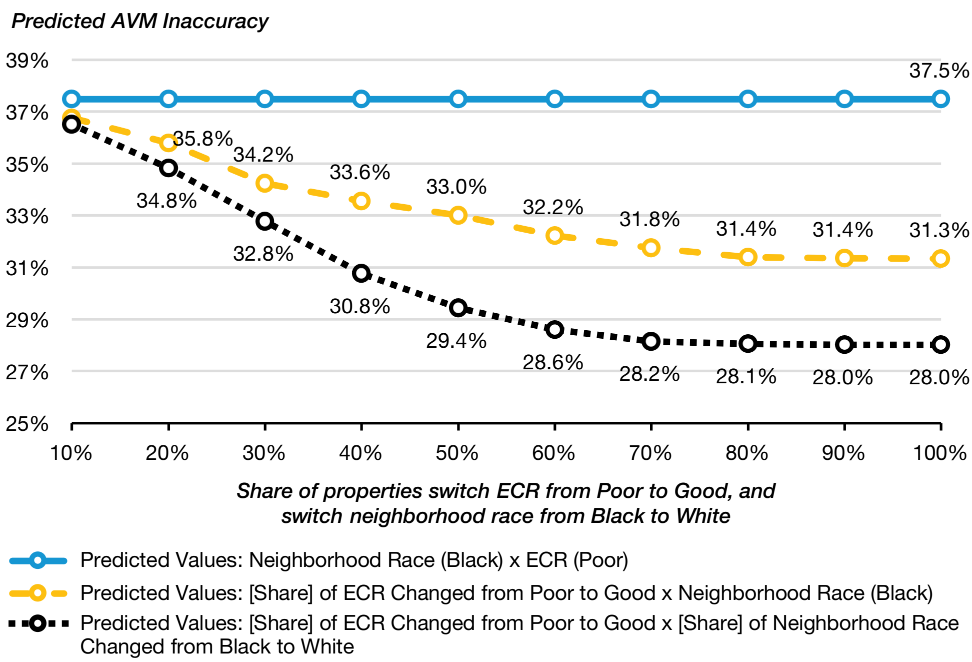


AVM = automated valuation model. ECR = exterior condition rating.
Source: Authors

Finally, the team examined the intersection of neighborhood majority race and ECR. The benchmark data consisted of properties in majority-Black neighborhoods rated as poor (the blue solid line in exhibit 14). The team changed the benchmark data by first altering the ECR from poor to good while keeping the neighborhood majority race and all other attributes constant. As before, for each synthetic data group, the team increased the share of change by 10 percent, ranging from 10 to 100 percent. Next, the team changed the neighborhood majority race from Black to White. When changing all the properties in the benchmark group from a poor to a good ECR, the predicted percentage magnitude of automated valuation model error fell from 37.5 to 31.3 percent (exhibit 14). The AVM error further declined to 28 percent once the neighborhood majority race flipped from Black to White. The gap between the two dash lines represents the effect of the intersection of neighborhood majority race and ECR. In other words, for two identical properties that have improved their ECR from poor to good, the home in a majority-Black neighborhood still experiences a 3.4-percentage-point greater percentage magnitude of AVM error, further suggesting that racial differences at the neighborhood level, which can reflect the effect of systemic discrimination, can play a role in producing percentage magnitude of AVM error.

Exhibit 14

Effect of the Intersection of Neighborhood Majority Race and ECR



AVM = automated valuation model. ECR = exterior condition rating.
Source: Authors

Conclusions and Implications

After including a direct control for property condition and employing more sophisticated ML techniques to examine the role of data omission and model selection, the research team finds that data on property condition and more sophisticated ML techniques can help more accurately assess the percentage magnitude of AVM error and its underlying contributors. Properties with poorer property conditions and in neighborhoods with more heterogeneous properties and a greater share of distressed sales are associated with greater predicted AVM error. In addition, despite data improvement and ML, evidence still shows that the percentage magnitude of AVM error is greater in majority-Black neighborhoods.

The evidence from this research contributes to the policy debate on appraisal bias and provides a quantifiable measure for auditing the performance of an AVM in majority-Black neighborhoods compared with majority-White ones. It serves as a starting point for developing a range of indicators that guard against a disproportionate effect on protected classes. If mortgage and housing providers are interested in determining the underlying contributors to the percentage magnitude of AVM error, Ordinary Least Squares regression is helpful, based on the findings of this research. However, if the goal is to assess the shortcomings of AVM models and their underlying contributors more accurately using large and complex data, the housing industry should consider exploring algorithms the AI community has developed. AI tools could enhance understanding of the complexity of predictive algorithms and partially unveil the AVM black box. The nonlinear regression results demonstrate that using a LightGBM model that includes ECR data could produce a 5.8-percentage-point improvement in the model's fit to assess the percentage magnitude of AVM error. Such results suggest that this nonparametric machine learning model more accurately copes with the complexity of variables of multiple dimensions. In addition, the SHAP values and the synthetic control approach shed light on the shortcomings of AVM algorithms that often get hidden in the black box.

Furthermore, the continued significance of a neighborhood's majority race suggests the need for integrating racial equity into the design of AVM algorithms. Researchers cannot yet reject the role of historic racism, which has persistently penetrated through home values, property conditions, and neighborhood conditions. Inequities in each of these dimensions can produce lower home values, less adequate housing, and lower household incomes across majority-Black neighborhoods. Encouraging regulatory oversight of AVMs and ML models will help ensure AVMs do not rely on biased data that could reinforce past discrimination.

This research illustrates the potential racial differences in one type of property valuation method, helping to shed light on current and standard valuation practices across the industry. However, the history of racial discrimination suggests conditions whereby the sale price may not be equivalent to a property's true value. Future research building relaxing assumptions embedded in current industry practice may represent steps toward a more fundamental assessment of differences between property values in Black communities compared with those in White communities.

Acknowledgments

The authors thank or acknowledge the Urban Institute for their financial support of this body of research, and particularly staff of the Housing Finance Policy Center, including Laurie Goodman and Janneke Ratcliffe for their input over the development of this research. The authors would also like to thank Cape Analytics, especially Sean Begley and Shane Lee (formerly of Cape Analytics) for their use of data and expert input. The authors would also like to thank the provider of their AVM data, who chooses to remain anonymous. This article was also enriched by many conversations over the years with subject matter experts around the housing industry. Finally, the authors would like to acknowledge the peer reviewers and the U.S. Department of Housing and Urban Development for their general support of policy research.

Authors

Linna Zhu is a senior research associate in the Housing Finance Policy Center at the Urban Institute. Michael Neal is a senior fellow in the Housing Finance Policy Center at the Urban Institute. Caitlin Young is a J.D. candidate at Yale Law School.

References

Aggarwal, Charu C. 2015. *Data Mining: The Textbook*, Vol. 1. New York: Springer.

Ambrose, Brent W., James Conklin, N. Edward Coulson, Moussa Diop, and Luis A. Lopez. 2021. *Do Appraiser and Borrower Race Affect Valuation?* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3951587.

Ashari, Ahmad, Iman Paryudi, and A. Min Tjoa. 2013. "Performance Comparison between Naïve Bayes, Decision Tree and k-Nearest Neighbor in Searching Alternative Design in an Energy Simulation Tool," *International Journal of Advanced Computer Science and Applications* 4 (11).

Bogin, Alexander N., and Jessica Shui. 2020. "Appraisal Accuracy and Automated Valuation Models in Rural Areas," *The Journal of Real Estate Finance and Economics* 60 (1–2): 40–52.

Campbell, John Y., Stefano Giglio, and Parag Pathak. 2011. "Forced Sales and House Prices," *American Economic Review* 101 (5): 2108–31.

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

Engler, Alex. 2022. "The AI Bill of Rights Makes Uneven Progress on Algorithmic Protections." Washington, DC: Brookings Institution. https://www.brookings.edu/articles/the-ai-bill-of-rights-makes-uneven-progress-on-algorithmic-protections/.

Fout, Hamilton, Nuno Mota, and Eric Rosenblatt. 2022. "When Appraisers Go Low, Contracts Go Lower: The Impact of Expert Opinions on Transaction Prices," *The Journal of Real Estate Finance and Economics* 65 (3): 451–491.

Freddie Mac. 2021. *Racial and Ethnic Valuation Gaps in Home Purchase Appraisals*. Freddie Mac Economic and Housing Research Note. McLean, VA: Freddie Mac. https://www.freddiemac.com/research/insight/20210920-home-appraisals.

Interagency Task Force on Property Appraisal and Valuation Equity (PAVE). 2022. *Action Plan to Advance Property Appraisal and Valuation Equity: Closing the Racial Wealth Gap by Addressing Mis-valuations for Families and Communities of Color*. Washington, DC: PAVE. https://pave.hud.gov/sites/pave.hud.gov/files/documents/PAVEActionPlan.pdf.

Jiang, Erica X., and Anthony L. Zhang. 2022. *Collateral Value Uncertainty and Mortgage Credit Provision*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4015452.

Ke, Guolin, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. "LightGBM: A Highly Efficient Gradient Boosting Decision Tree." In *Advances in Neural Information Processing Systems 30*, edited by Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S.V.N. Viswanathan, and Roman Garnett. Long Beach, CA: Neural Information Processing Systems: 3149–57.

Neal, Michael, Jung Hyun Choi, and John Walsh. 2020. *Before the Pandemic, Homeowners of Color Faced Structural Barriers to the Benefits of Homeownership*. Washington, DC: Urban Institute. https://www.urban.org/sites/default/files/publication/102781/before-the-pandemic-homeowners-of-color-faced-structural-barriers-to-the-benefits-of-homeownership.pdf.

Neal, Michael, Sarah Strochak, Linna Zhu, and Caitlin Young. 2020. *How Automated Valuation Models Can Disproportionately Affect Majority-Black Neighborhoods*. Washington, DC: Urban Institute. https://www.urban.org/sites/default/files/publication/103429/how-automated-valuation-models-can-disproportionately-affect-majority-black-neighborhoods_1.pdf.

Perry, Andre, Jonathon Rothwell, and David Harshbarger. 2018. *The Devaluation of Assets in Black Neighborhoods: The Case of Residential Property*. Washington, DC: Brookings Institution.

Williamson, Jake, and Mark Palim. 2022. *Appraising the Appraisal: A Closer Look at Divergent Appraisal Values for Black and White Borrowers Refinancing Their Home*. Washington, DC: Fannie Mae. https://www.fanniemae.com/media/42541/display.

Additional Reading

Akinwumi, Michael, Lisa Rice, and Snigdha Sharma. 2022. *Purpose, Process, and Monitoring: A New Framework for Auditing Algorithmic Bias in Housing and Lending*. Washington, DC: National Fair Housing Alliance. https://nationalfairhousing.org/wp-content/uploads/2022/02/PPM_Framework_02_17_2022.pdf.

Akinwumi, Michael, John Merrill, Lisa Rice, Kareem Saleh, and Maureen Yap. 2021. *An AI Lending Policy Agenda for the Federal Financial Regulators*. Washington, DC: Brookings Institution. https://www.brookings.edu/articles/an-ai-fair-lending-policy-agenda-for-the-federal-financial-regulators/.

Axelrod, Judah, Alena Stern, Michael Neal, and Linna Zhu. 2022. "Comment Letter on NIST Initial Draft of Artificial Intelligence Risk Management Framework." Washington, DC: Urban

Institute. https://www.urban.org/sites/default/files/2022-05/Comment%20Letter%20on%20NIST%20Initial%20Draft%20of%20Artificial%20Intelligence%20Risk%20Management%20Framework_0.pdf.

Buchak, Greg, Gregor Matvos, Tomasz Piskorski, and Amit Seru. 2020. *iBuyers: Liquidity in Real Estate Markets?* National University of Singapore, Institute of Real Estate and Urban Studies. https://ireus.nus.edu.sg/wp-content/uploads/2020/10/iBuyers-Liquidity-in-Real-Estate-Markets-by-Tomasz-Piskorski-.pdf.

Conklin, James, N. Edward Coulson, and Moussa Diop. 2022. *Distressed Comps*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4008585.

Fout, Hamilton, and Vincent Yao. 2016. *Housing Market Effects of Appraising Below Contract*. Washington, DC: Fannie Mae; Atlanta: Georgia State University. https://www.researchgate.net/profile/Vincent-Yao-2/publication/298807852_Housing_Market_Effects_of_Appraising_Below_Contract/links/57c04c7508ae2f5eb3321d07/Housing-Market-Effects-of-Appraising-Below-Contract.pdf.

Howell, Junia, and Elizabeth Korver-Glenn. 2018. "Neighborhoods, Race, and the Twenty-First Century Housing Appraisal Industry," *Sociology of Race and Ethnicity* 4 (4): 473–90. https://doi.org/10.1177/2332649218755178.

Narragon, Melissa, Danny Wiley, Doug McManus, Vivian Li, Kangli Li, Xue Wu, and Kadiri Karamon. 2021. *Racial and Ethnic Valuation Gaps in Home Purchase Appraisals*. Freddie Mac Economic and Housing Research Note. McLean, VA: Freddie Mac. https://www.freddiemac.com/research/insight/20210920-home-appraisals.

Neal, Michael, Jung Hyun Choi, Kathryn Reynolds, Joe Schilling, Gideon Berger, Elizabeth Champion, and Caitlin Young. 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. https://www.urban.org/sites/default/files/publication/105020/why-do-households-of-color-own-only-a-quarter-of-the-nations-housing-wealth-when-they-compose-a-third-of-the-nations-households.pdf.

Salkind, Neil J., ed. 2010. *Encyclopedia of Research Design*. New York: Sage Publications.

Shrestha, Noora. 2020. "Detecting Multicollinearity in Regression Analysis," *American Journal of Applied Mathematics and Statistics* 8 (2): 39–42.