GOOD SYSTEMS
A UT Grand Challenge

Modeling and Quantifying Group Disparities: A Case Study of Homeownership Racial Disparity

Inez Khan[1], Arya Farahi[1], Prachi Ingle[2]
[1]Department of Statistics and Data Science, [2]Department of Computer Science, The University of Texas at Austin

The University of Texas at Austin
Department of Statistics and Data Sciences
College of Natural Sciences

Abstract

Based on the explainable machine learning (ML) literature, we have created a racial disparity index (RDI), which measures the importance of race in the homeownership gap, after accounting for other household characteristics. We then develop an inference model that allows us to study the spatiotemporal trend of RDI and its dependence on different household characteristics. Our findings reveal that

1) less than half of the homeownership gap is explained by household characteristics, with the remainder being attributable to risk factors associated with race (Figure 1B),
2) the contribution of race to the homeownership gap has been increasing over time for all states, with Midwestern and Northern states experiencing more significant increases than Southern states (Figure 2).

Procedure

We developed the following procedure to analyze group disparities between white and black homeowners. We use data from the American Community Survey (ACS) for the years 2005 – 2019 at the Public Use Microdata Area (PUMA) region level. PUMA regions are statistical geographical regions developed by the U.S. Census Bureau that divide the U.S. population to groups of 100,000 – 200,000 inhabitants.

1) Selecting an appropriate ML model by regressing the probability of homeownership on demographic factors including race. We chose XGBoost.
2) Extracting SHapley Additive exPlanations (SHAP) values from ML model predictions to get feature importance values
3) Calculating the Racial Disparity Index (RDI), where RDI is defined as below. We define S_0 as the group of white householders and S_1 is the group of black householders. In addition, $\Phi_{Race}(x_j)$ is the SHAP value of feature Race for the j^{th} household.

$$\mathrm{RDI}(S) = \frac{1}{n_0}\sum_{\mathbf{x}_j \in S_0} \phi_{\mathrm{Race}}(\mathbf{x}_j) - \frac{1}{n_1}\sum_{\mathbf{x}_j \in S_1} \phi_{\mathrm{Race}}(\mathbf{x}_j)$$

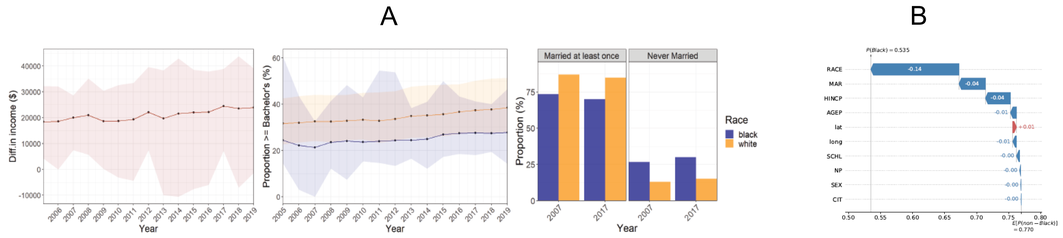
4) Fitting the following linear hierarchical model to explore RDI over time. We account for factors such as the proportion of black and white bachelor's degreeholders in a PUMA region in addition to the median income.

$$\mathrm{RDI}(t,p) = \beta_{0,k} + \beta_{1,k} x_{t,p}^{\mathrm{prop-black-bach}} + \beta_{2,k} x_{t,p}^{\mathrm{prop-white-bach}} + \beta_{3,k} x_{t,p}^{\mathrm{log-med-income}} + \beta_4 s(t) + \epsilon_t,$$

We use a Bayesian partial pooling model where the coefficients for each state are drawn from a Normal distribution.

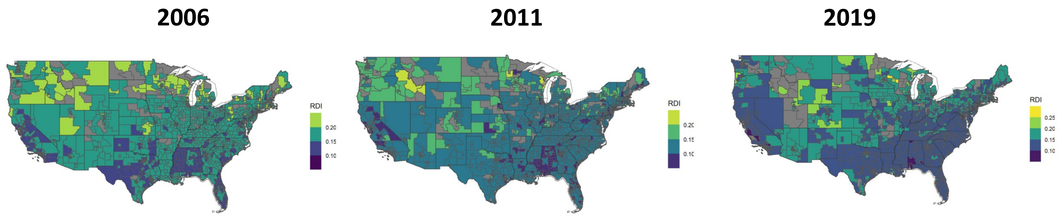
Figures

Figure 1. Comparing Various Statistics between White and Black Householders Nationally, 2005 - 2019



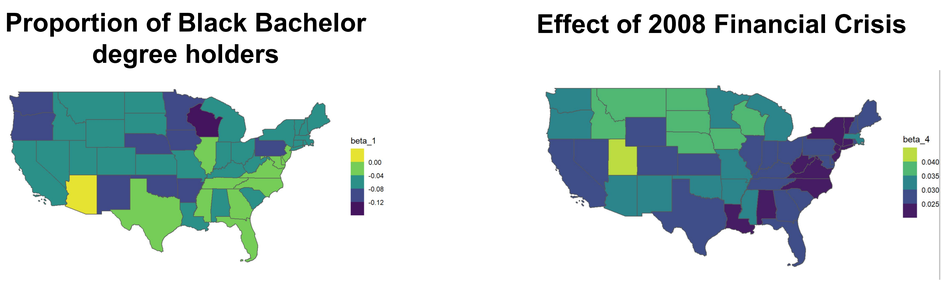
A) Comparing various statistics for white and black householders: income difference (left), bachelor's attainment (center), and marital status (right). Dark orange is for white householders, and dark blue is for Black householders. Differences are calculated by subtracting the statistic for Black from the statistic for white. B) The contribution of household characteristics to the homeownership gap for 2019 on average nationally.

Figure 2. Racial Disparity Index (RDI) for U.S. PUMA Regions



Comparing the RDI over time between white and black householders on the probability of homeownership. Higher RDI values are associated with race having a higher contribution towards the gap in homeownership between white and black householders. We notice greater geographic clustering over time as well.

Figure 3. Looking at RDI Over time: Model Coefficients

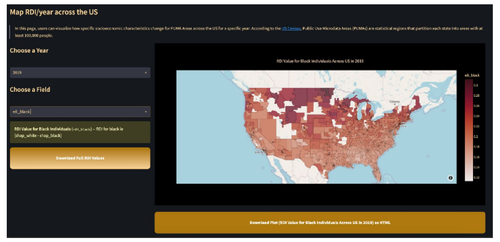


Selected model coefficients from the inference model explain how RDI is changed by these factors over time. We see over time having a higher proportion of black bachelors degree holders across states helps decrease the effect of race on homeownership. Meanwhile, the financial crisis has overall caused an increase in the effect of race on homeownership.

Dashboard

We developed Exploring Racial Disparity Index (ERDI) to allow users to :

1) Visualize and download data used and produced in our work
2) Upload their own household-level data into our models to make new predictions.



The dashboard can be accessed here:

Acknowledgments

This work was supported by Good Systems, a research grand challenge at the University of Texas at Austin, and the National Science Foundation (NSF-1916454, sub-award 5127276) through the South Big Data Innovation Hub SEEDS, Southern Engagement and Enichment in Data Science. The authors acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing HPC resources that have contributed to the research results reported within this paper.

References

Inez Khan, Arya Farahi, and Prachi Ingle. 2023. Modeling and Quantifying Group Disparities A Case Study of Homeownership Racial Disparity.

The University of Texas at Austin
Research Development
Vice President for Research, Scholarship and Creative Endeavors