Measuring Race and Ethnicity Gaps in U.S. Mortgage Approvals

Ray Carpenter

Advisor Chris Handy, PhD

September 2022

Abstract

Are there racial and ethnic gaps in mortgage approval rates throughout the United States mortgage market? This study seeks to measure these gaps using the open Home Mortgage Disclosure Act (HMDA) data from 2019 through 2021. I used a linear probability model to estimate whether a gap exists while controlling for other variables included in the dataset. The model produced estimates between White and Black applicants to be 4.6 percentage points, White and Asian to be 1.2, White and Hispanic to be 1.9, and White and Other applicants to be 3.2 percentage points. Other literature is also reviewed to further explore these gaps, given the closed dataset. This study suggests that these disparities exist in today's U.S. mortgage market.

Acknowledgments

This research has been made possible by a grant through The Guest Family Fund for Excellence in Economics. The support and guidance of Dr. Chris Handy has been essential to the success of this project. The University of Chapel Hill's access to their Longleaf Computing Cluster is an integral part of carrying out this project, and their education has allowed me to carry out this study.

I want to thank the Guest Family, Dr. Handy, and the University of Chapel Hill for providing this opportunity to learn a tremendous amount and be engaged in research.

Introduction

Owning a home helps individuals to invest in their future by building wealth and creating a stable environment in which to grow (Rappaport, 2010). Home ownership is considered an integral part of the American Dream. The ability to own a home is a right that should be afforded to everyone who is able. As the modern world moves to an equal society, the fairness of the steps to prosperity is a great necessity. The U.S. Census Bureau reported that the homeownership rates were 74.6 percent for non-Hispanic Whites, 61.1 percent for Asian households, 48.3 percent for Hispanic households, and 45.3 percent for Black households (United States Census Bureau, 2022). The current paper explores these gaps further by delving into what might have caused these gaps through a literature review and then measuring these gaps using mortgage data approval rates.

This paper seeks to measure differences in mortgage approval rates between different races. Open Home Mortgage Disclosure Act, HMDA, data was used to do this. This dataset allows the differences in access to homeownership to be further explored by controlling for aspects that may influence an approval. Some of these variables that lenders can look at are included in the model, such as Loan-to-Value Ratio, Debt-to-Income Ratio, and an individual's income. This study focuses specifically on home purchase loans, as opposed to other forms of mortgage lending.

The models used to estimate mortgage acceptance rates are similar to those used to measure the same gaps in previous literature. Different studies have tracked this from the mid-20th century to 2020 (Ky & Lim, 2022). This paper looks to add the open 2021 HMDA data to the current literature. It also provides a framework to estimate this gap in future years. The data cleaning process and model codes were built to be iterative. This allows for future studies to be performed, and it provides transparency. They can be found at https://github.com/RayCarpenterIII/Measuring-Race-and-Ethnicity-Gaps-in-U.S-Mortgage-Approvals.

Regression models are the primary estimation vehicle this paper uses to measure such gaps. Different multiple regression and fixed effects models were used. However, the fixed effects model, FE 4, is the primary model used when referring to results from this paper. The fixed effects model allows time-invariant unobserved factors to be accounted for in the estimations. These variables might affect approval rates and are not measured in the data.

A preliminary model regresses approval rates solely on Race and gender. This model is measured by regressing approved on the race variables. This can be found in the Appendix in Table 1 as OLS 1. This shows the average difference in approval rates between races and gender. Our initial findings show similar gaps to previous literature (Desliver & Bialik, 2017). I estimate the gaps to be 4.6 percentage points between White and Black Applicants, 1.2 between White and Asian applicants, White and Hispanic applicants to be 1.9, and White and Other applicants to be 3.2 These highlight the significant gaps between races that are so apparent.

This study measures the gap between White and Black applicants, White and Asian applicants, White and Hispanic applicants, and White and Other applicants. The results suggest statistically significant gaps in all four tests using a 1 percent significance level.

Literature Review

The Fair Housing Act (FHA) was passed on April 10, 1968, which prohibits discrimination in the rental or sale of housing on the grounds of Race, color, religion, sex, handicap, or national origin (Schill & Friedman, 1998). It was passed six days after the assassination of the Reverend Martin Luther King Junior. Since then, the United States has had a history of discrimination in mortgage acceptances by Race. This has been documented as far back as 1978 by the Comptroller of Currency – FDIC (Black et al., 1978) and continues into 2022. The most recent literature estimates that the gap between White and Black applicants is 2.9 percentage points, 1.5 percent between Asian and White applicants, and 2.2 percentage points between Latinx and White applicants (Ky & Lim, 2022).

The Home Mortgage Disclosure Act was another critical step to improving homeownership and mortgage lending gaps. It was passed in 1975 and compelled mortgage lenders to keep records on applications (Kolar & Jerison, 2006). This led to the creation of the HMDA data sets, the primary dataset used by this paper. The publicly available data is used for this paper. However, there is a confidential version of the HMDA dataset that researchers from the Federal Reserve and other institutions use that includes credit scores and other variables. The confidential datasets are what most other scholars in the field have used for similar studies.

The Federal Reserve began using similar data and methodology to this paper in 1992 (Browne et al., 1992). They showed a gap between White and minority races of 1.6 to 1 percent in the Boston area. They found that when a minority applicant has the credentials that lenders are looking for, they are 97 percent certain to be approved. The paper also states that even though racial discrimination gaps are small, there is still a significant divide in human capital between races, leading to a denial rate of 17 percent for minorities and 11 percent for White applicants. This is because minorities tend to have higher debt-to-income ratios, higher loan-to-value ratios, and higher credit scores. These three characteristics explain much of the gap.

A 2012 study shows statistically significant discrimination between 2004 and 2008 using HMDA data (Hirasuna & Allen, 2012). It provides evidence that, in some cases, mortgage application denial rate gaps may increase over time. Even though it does not claim that they are increasing over time, it focuses more on the volatility of decision-making in this process between company and time.

Other studies look at mortgage approval or denial race gaps in different markets and types of loans. Bartlett et al. observe lending differences in FinTech and non-FinTech lenders between 2009-2015 (Barlett et al., 2022). They found that FinTech leaders showed no difference in discrimination rates, except they showed a lower gap for FHA mortgages. They also showed that racial gaps persist regardless of whether a loan is a GSE security or is considered an FHA-insured loan. Interest rate gaps were used for that paper instead of approval rates. They estimated that the gap in interest rates cost minority borrowers over 450 million U.S. dollars yearly.

A more recent study by members of the Federal Reserve Bank of Minneapolis looked at mortgage denial rates in the U.S. (Ky & Lim, 2022). They use a similar model to the one used in this paper. They look at differences in mortgage denial rates between races conditional on other variables and find statistically significant gaps. Data between 2018 and 2020 is used. The study was also limited to conventional 30-year term mortgage applications. All the controls used for that paper are the same as the controls used for this paper. However, this current paper uses data from 2019-2021 and does not have access to credit scores. The lack of access to credit score data is one of the main reasons the results from this paper may vary from the results of the 2022 Fed working paper.

An August 2022 draft from three Federal Reserve Board of Governors, Neil Bhutta, Aurel Hizmo, and Daniel Ringo, measures racial gaps in the Automated Underwriting System (AUS) denials and in Lender Denials (Bhutta, N., Hizmo, A., & Ringo, D. 2022). They use the confidential HMDA data set also. They find "significant" progress in fair lending over the last 30 years in the U.S. mortgage market. Their regressions estimate the denial gap between White and Black to be between 1.5 for AUS Denials and 2 percentage points for Lender denials, White and Asian to be between .2 and 1.4 percentage points, respectively, White and Hispanic to be 0 to 1 percentage points, White and Other to be between .9 and 1.8 percentage points, and finally White and Joint Race to be between 0 and 1.7 percentage points. They show that the AUS system does an excellent job of closing the gaps. However, Lenders can skip the AUS's recommendation. This is where the difference came in between the AUS and Lender's denials.

In summary, current literature shows a statistically significant gap between races in mortgage denials. Estimates do vary between studies. However, they all estimate the gap to be in roughly the same range. This means that, most likely, the gaps exist for mortgage approvals.

Methodology

Datasets Overview

Home Mortgage Disclosure Act data is the source of data used to produce the results of this study. HMDA data provides mortgage data using loan-level information. Each observation consists of a mortgage application containing lender information, loan information, demographic information, and individual socio-economic information. The act went into effect in 1975. This study used nationwide data from 2021, 2020, and 2019. The three raw datasets used can be downloaded from the HMDA Data Browser at ffiec.cfpb.gov/data-browser/.

The total size of the three raw datasets was 25.21 gigabytes. Cleaning and analyzing data of this size required more power than my personal computer could provide. Access to the University of North Carolina at Chapel Hill's Longleaf computing cluster was applied for and granted to this project. The programs were managed using Linux through the SLURM job scheduler. This allowed for the more memory-intensive workloads that this project needed.

The Longleaf cluster had versions of Python and Stata pre-installed. A Python script was used to clean each year's data into cleaned chunks. The term "chunk" refers to a slice of the data. The chunks from each year were then merged into one dataset. This produced the final dataset that was used to run the regressions. It consists of nearly 12.8 million observations, 20 columns, and has a size of 2.17 gigabytes.

A snapshot of the finalized clean data is shown in Figure 1 in the appendix.

Data Cleaning Decisions

The open HMDA datasets used contains over 25 million observations and 99 columns of information. Only some information contained was pertinent to this study. These were trimmed and transformed into around 12.8 million observations and 20 columns. This section of the paper gives a more detailed analysis of the decisions made surrounding the data cleaning process for this study. This process can be found in the file "HMDA_Cleaner.py." It was built to be used on past and future open HMDA datasets for analysis down the road.

This study is currently interested in purely home purchase loans. All other loan types were removed from the dataset, including home improvements, refinancing, cash-out refinancing, other purposes, and not applicable.

The Race category was grouped into five categories: White, Asian, Black, Other, and LatinX. This is in keeping with current literature, including the paper by Neil Bhutta and Aurel Hizmo in 2020 and an additional paper by Kim-Eng Ky and Katherine Lim in 2022. This is assigned by the primary Race and ethnicity of the first applicant for this study. The Asian indicator consists of Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, and Other Asian races. The Hispanic indicator consists of any ethnicity considered to be Hispanic. Mexican, Puerto Rican, Cuban, Latino, and Other are all included with the Hispanic variable. The Other race variable consists of anyone that did not fall into the above categories. Any instances where Race or sex were not recorded and were removed.

Most current literature focuses on mortgage denials. However, this study focuses on mortgage approvals. The "Approval" variable includes approvals and approvals that were not accepted. Cases where denials occurred, were kept. All other cases were dropped. They included withdrawals, incompleteness, and pre-approval requests. This means that the approved and denied indicator columns explain the outcome entirely. Therefore, in the cleaned dataset, the approved variable will pick up the same information as denial when looking for gaps.

The size of the observations affects the hypothesis test's confidence interval. This study uses between 12.5 and 12.8 million observations. This means that most estimates will be statistically significant due to the size of the dataset. This may account for the statistical significance of variables that may not have an economic significance. In addition to the economic significance of a variable, variables used in previous studies were considered. When selecting variables to add to the model, special attention was given to the economic importance of each variable.

The log income variable was created from the income column in the HMDA dataset. Any instances where income was not reported, 0, or less than 0 were removed from the data, as these were most likely errors. The logarithm of income was taken because it better explains the relationship with approvals than just the income variable. The log loan amount was created using the same process, just concerning the loan amount.

The raw Debt-to-Income (DTI) variable consists of categorical and continuous data. Some of the data is in a range format, 0-20%, and the other half was in a numeric format between the more common DTI ratio percentages. To combine the categorical and numeric data, indicator variables were created for each categorical instance of DTI and each continuous value of DTI. This allowed all observations of DTI to be read as categorical data. Instances where DTI was null, were dropped. Loan-to-Value (LTV) ratios were already in a continuous format and only needed null values to be dropped. After all these calculations were completed, the cleaned datasets were combined and ready for the regression to be run.

Summary Statistics

Summary statistics can be found in the appendix under summary statistics. The final dataset consisted of around 12.8 million mortgage applications between 2019 and 2021. The average approval rate was 91.4 percent. Average approval rates by Race are explained under results using the OLS 1 regression from Table 1. The population comprised 69.6 percent White, 7.74 percent Asian, 8.83 percent Black, 13.01 percent Hispanic, and .81 percent Other. The average approval rates for White individuals were 93 percent, 91.1 percent for Asian applicants, 84.2 percent for Black applicants, 88.4 percent for Hispanic individuals, and 87.9 percent for other individuals. More in-depth tables can be found under Summary Statistics 1 and Summary Statistics 2 in the appendix.

Econometric Model

This study aims to measure race gaps in U.S. mortgage approvals. To do this, a series of different regressions were utilized where the approval variable is regressed on Race and a set of control variables. Multiple regressions are utilized to measure these gaps with different types of standard errors and to see how the variables of interest change when controls are added. The Race variable consists of White, Black, Asian, Other, and Hispanic. This process can be found in the file "Reg_Final.do" on the earlier linked GitHub.

The dependent variable, Approval, is a binary indicator of whether that case ended in an approval. Regressing a binary dependent variable on independent variables produces a linear probability model. The outcome is the probability that the dependent variable equals one, given the independent variables. In this case, the model's outcome describes the probability of Approval, given the race and control variables. The race variables can then be described as the percent difference in probability that Approval is equal to one between that Race and another chosen race, given the other control and race variables. The White variable was left out of the model to use as the baseline category. For example, the coefficient on Black is the difference in approval probability between Black and White, given the other coefficients.

In total, seven regressions were produced to be examined in this paper: Two linear models and five fixed-effects models. The primary model used can be described as follows:

$$P(Approval_{irt} = 1 \mid) = \beta_0 + \delta_r * 1(race = r) + \rho * X_{it} + \alpha_t + \varepsilon_{irt}$$

Where " δ_r" is the intercept, " δ_r" is a vector of coefficients for the different race variables of interest, " δ_r" is a vector of coefficients for the control variables, " δ_r" represents the fixed effects by year, and "δ_r" is the error term. The subscript represent represents each application, and the subscript represents the year.

The controls include Loan to Value Ratios, Debt to Income ratios, lenders, states, sex, log income, log loan amount, a pre-approval indicator, loan type, and occupancy type. The LTV and DTI ratios interacted. This is because LTV or DTI's effect may differ depending on how the other variable is presented. For example, someone's LTV might be looked at differently if they have a very low DTI. LTV and DTI account for many of the non-prohibited variables a lender can use to decide based on the Fair Housing Act and the Equal Credit Opportunity Act (Fair Lending, 2019). Race, Ethnicity, and Sex are prohibited under both acts. Race and Ethnicity are combined under the Race variable. The sex variable is accounted for also.

Different types of standard errors are utilized. In the first three models, standard errors are typical. In model "F.E. 2", robust standard errors are used. Two models were also produced where the only difference is clustering standard errors by State and Lender. For the remaining models, standard errors are clustered by Lender. Lender was chosen because lenders vary in how they make decisions. Thus, it is believed that the standard errors will deviate by Lender. In the model "F.E. 5" lender and county interact. Even though county effects are not accounted for, it only interacts with Lender. This accounts for how lenders might approve loans differently from one county to the next.

The "α_t" term is only used in the fixed-effects model. It represents variables that are constant over the set space and time. In this instance, space is the State, and time is the year. The model allows these effects to be accounted for, even though they are not observed. The "reghdfe" package was used to implement the fixed effects regressions. The package "reghdfe" uses the Correia estimator to produce the linear estimates. Further detail on this estimator can be found in a 2016 paper by Sergio Correia of Duke University (Correia, 2016).

These models were then analyzed to determine whether the coefficients on the Race variables were statistically significant or not. The null hypothesis is that each " δ_r" coefficient is equal to 0. The alternative hypothesis is that the " δ_r" coefficients are not equal to 0. Due to the size of the dataset, an alpha value of .01 is used.

$$H_0: \delta_r = 0$$

$$H_a: \delta_r \neq 0$$

$$\alpha = .01$$

$$Where\ \delta_r\ represents\ each\ Race\ coefficient\ to\ be\ tested.$$

Results

Table 1 supports the alternative hypothesis for all "δ_r" coefficients. Coefficients for Black, Asian, Hispanic, and Other are all statistically significant: the coefficients on the race variables are not equal to zero. The F.E. 4 model estimates the coefficient on Black to be -.046, -.012 for Asian, -.019 for Hispanic, and -.032 for Other. A 1 percent significance was used to test the coefficients.

Measuring Race Gaps in Mortgage Approvals

	OLS 1	OLS 2	FE 1	FE 2	FE 3	FE 4	FE5
Asian	-0.019^{***}	-0.011^{***}	-0.012^{***}	-0.012^{***}	-0.012^{***}	-0.012^{***}	-0.012^{***}
Black	-0.088^{***}	-0.048^{***}	-0.046^{***}	-0.046^{***}	-0.046^{***}	-0.046^{***}	-0.043^{***}
Hispanic	-0.046^{***}	-0.019^{***}	-0.019^{***}	-0.019^{***}	-0.019^{***}	-0.019^{***}	-0.019^{***}
Other	-0.051^{***}	-0.035^{***}	-0.032^{***}	-0.032^{***}	-0.032^{***}	-0.032^{***}	-0.027^{***}
Controls	-	Yes	Yes	Yes	Yes	Yes	Yes
State or Lender FE	-	-	Yes	Yes	Yes	Yes	Yes
Standard Errors	-	-	-	Robust	State	Lender	Lender
LEI * County	-	-	-	-	-	-	Yes
Adjusted R2	0.010	0.195	0.213	0.213	0.213	0.213	0.227
Observations	12802383	12742520	12742464	12742464	12742464	12742464	12553815

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

The table shows multiple regressions where another control variable is added, or part of the model changed. OLS 1 is the Approved variable regressed on Race without any controls. This gives the difference in mortgage approval probability averages between races. However, in OLS 1, the constant variable represents the average probability of mortgage approval for those who are White. This model shows an average probability of 93 percent for those who are White. This is shown in Table 1. The average probabilities are also shown in a graph in Figure 3. OLS 1 shows that those who are Black are 8.8 percentage points less likely to be approved for a mortgage loan than those who are white, Asians are 1.9 percentage points less likely, those who are Hispanic are 4.6 percentage points less likely, and those who fell in the other category are 5.1 percentage points less likely to be approved for a mortgage loan than those who are White.

These regressions seek to determine whether these gaps may be explained through differences in other characteristics between races or if other unobserved factors are causing the gap in OLS 1. OLS 2 adds the controls to the model. Adding the controls alone shows how they explain much of the gaps between the groups. The following regressions add fixed effects to account for variables that are not observed but constant across states in the U.S. and throughout 2019, 2020, and 2021.

F.E. 2, F.E. 3, and F.E. 4 all use different types of standard errors to estimate the models. F.E. 4 uses robust standard errors. F.E. 3 clusters standard errors around states. This is to account for any differences between states that may affect the heteroskedasticity between states. However, the decision of an approval is being made at the lender level. That is why the standard errors are clustered around Lender for F.E. 4 and F.E. 5. F.E. 7 adds a Lender and County interaction. This is to account for Lender's making different decisions based on a more granular level than the State. This model was run separately because it may improperly control approvals. This is because there are many counties with few or no observations.

The "adjusted r-squared" is added to the table due to many independent variables. This adjusted r-squared considers how many independent variables are added to a model. The adjusted r-squared is .01 for OLS 1 and .217 for most other models once the control variables and fixed effects were added.

Discussion

This study suggests that Black, Asian, Hispanic, and those who fell into the other category all have lower approval ratings relative to White applicants. The study suggested the same even after controlling for fixed effects and other variables. These include income, demographic, lender, and loan characteristics. The gap between White applicants and Black is estimated to be 4.6 percentage points, 1.2 percentage points between White applicants and Asian applicants, 1.9 percentage points between White and Hispanic applicants, and 3.2 percentage points between White and Other applicants. Our study supports other literature in that there is still work to be done to improve these gaps.

The coefficient on Race might suggest that racial discrimination is still occurring in the mortgage approval process. Recent studies, including data that this study does not have access to, suggest that this coefficient is likely much smaller than what is shown in Table 1. The coefficients on Race are due to unobservable factors and may not be due solely to racial discrimination. However, this gap still exists, given current data. It certainly does not rule out racial discrimination in these coefficients. On the other hand, other control variables can be sources of discrimination in themselves also, such as income, Lender, or loan-to-value ratios. There is still much to be done in working to fix these inequalities.

Our models estimate more significant coefficients for all tested variables than others have in more recent literature. This is most likely due to the inability to account for credit scores. The confidential HMDA dataset does include these variables, as the open HMDA data used in this report does not. Other researchers at the Federal Reserve do have access to this data and have shown that a person's credit rating is one of the main factors in deciding on whether a loan is denied (Ky & Lim, 2022).

This study tests four racial and ethnic gaps. These concepts are much more complex in reality. Many people fit into other categories not addressed here, and others may fit into multiple racial or ethnic groups. HMDA data allows more complex models to be estimated on the dimension of Race and ethnicity. Future studies might include a more in-depth grouping of races, secondary race variables, observed race variables, different ethnicity groupings, or co-applicant race variables. Someone who is Black is different from someone who is Black and Hispanic. This interaction should also be accounted for in future studies.

Another dimension I would like to study in the future is the term length of the loans. To be used as a future control variable that may be added to the F.E. 4 model and to study race and ethnicity gaps in different loan term lengths. This may be done by interacting with the loan length variables and the race variables.

The differences in Table 1's OLS 1 and F.E. 4 estimations suggest that many mortgage approval rating gaps are due to observable factors. This means that a large portion of the gap comes from differences that can be addressed. This allows policymakers looking to help close this gap to focus on factors that may be addressed. Many statistically significant control variables, such as LTV and DTI, suggest that policymakers can focus on building wealth for this marginalized group. This refers to the group that makes up the difference between these two estimations.

The years 2020 and 2021 have been different from any market's normal years. The mortgage market included. HMDA has open data on its website going back to 2007. Future studies might include a more in-depth time series analysis of racial and ethnic gaps in the mortgage market.

Work has also been done to study how these racial gaps might vary by State and Metropolitan Statistical Area. The Ky & Lim paper mentioned earlier measures these gaps by running the same regression used to measure Race and ethnic gaps. However, the state and race variables are interacted to understand how each Race or ethnic gap varies by State. (Ky & Lim, 2022) They found that the racial gaps varied not only by State but also that the different groups discriminated within a state might be different. I also ran a regression to test this. I mapped the state and race interaction coefficients onto the maps found in Figure 4. The results were similar to that of Ky & Lim's paper.

Conclusion

This study sought to explore racial and ethnic gaps in mortgage approval ratings. Using a linear probability model to control for other variables that may determine approval rating, we found that statistically significant gaps exist between White applicants and Black, Hispanic, and other applicants. Meaningful work has been done in recent literature outside this paper to measure these gaps using a more detailed and confidential version of the same dataset. They show that there have been significant improvements in recent years and that using Automated Underwriting Systems will help to slim these gaps. (Bhutta, N., Hizmo, A., & Ringo, D. 2022) However, work still needs to be done to bridge these divides both in and outside of the mortgage market.

References

Asante-Muhammad, D., Buell, J., & Devine, J. (2021, March 2). *60% Black Homeownership: A Radical Goal For Black Wealth Development*. National Community Reinvestment Coalition. Retrieved from https://ncrc.org/60-black-homeownership-a-radical-goal-for-black-wealth-development/

Bartlett, R., Morse, A., Stanton, R., & Wallace, N. (2021). Consumer-lending discrimination in the FinTech Era. *Journal of Financial Economics*, *163*(1), 30–56. Retrieved from https://www.sciencedirect.com/science/article/abs/pii/S0304405X21002403?casa_token=xre_SGOn2egAAAAA:KpivJHNRRgm6ZQ4KkbZjAev17KVvCTcRrNqHnyD5NF0wmcd0zn3BffD90XDz0ZIruomOFsI#bib0009.

Bayer, P., Ferreira, F., & Ross, S. L. (2016). What Drives Racial and Ethnic Differences in High Cost Mortgages? The Role of High Risk Lenders. *The Review of Financial Studies*, *31*(1), 175–205. https://doi.org/10.3386/w22004

Bhutta, N., & Hizmo, A. (2020). Do Minorities Pay More for Mortgages? *The Review of Financial Studies*, *34*(2), 763–789. Retrieved from https://doi.org/10.1093/rfs/hhaa047.

Bhutta, N., Hizmo, A., & Ringo, D. (2022). How much does racial bias affect mortgage lending? Evidence from human and algorithmic credit decisions. *Evidence from Human and Algorithmic Credit Decisions (August 2, 2022)*.

Black, H., Schweitzer, R. L., & Mandell, L. (1978). Discrimination in Mortgage Lending. *The American Economic Review*, *68*(2), 186–191. Retrieved from https://www.jstor.org/stable/1816686?seq=5#metadata_info_tab_contents.

Browne, L. E., McEneaney, J., Munnell, A. H., & Tootell, G. M. B. (1992). Mortgage Lending in Boston: Interpreting HMDA Data. *Federal Reserve Bank of Boston*, *86*(1), 25–53. Retrieved from https://ideas.repec.org/p/fip/fedbwp/92-7.html?ref=pislikcs.com.

Cheng, P., Lin, Z., & Liu, Y. (2015). Racial Discrepancy in Mortgage Interest Rates. *The Journal of Real Estate Finance and Economics*, *51*(1). https://doi.org/10.1007/s11146-014-9473-0

Correia, S. (2016). A Feasible Estimator for Linear Models with Multi-Way Fixed Effects. *Preprint*. Retrieved from http://scorreia.com/research/hdfe.pdf.

Desliver, D., & Bialik, K. (2017, January 10). *Blacks and Hispanics face extra challenges in getting home loans*. Pew Research Center. Retrieved from http://pewrsr.ch/2iYR1hC

Faber, J. W. (2018). Segregation and the Geography of Creditworthiness: Racial Inequality in a Recovered Mortgage Market. *Housing Policy Debate*, *28*(2), 215–247. Retrieved from https://doi.org/10.1080/10511482.2017.1341944.

Fair Lending. OCC. (2019, April 6). Retrieved from https://www.occ.treas.gov/topics/consumers-and-communities/consumer-protection/fair-lending/index-fair-lending.html

Federal Financial Institutions Examination Council's. (2018, September 6). *History of HMDA*. Home Mortgage Disclosure Act. Retrieved from https://www.ffiec.gov/hmda/history2.htm

Federal Financial Institutions Examination Council's. (2022). *HMDA Data Browser*. The Home Mortgage Disclosure Act. Retrieved from https://ffiec.cfpb.gov/data-browser/

Frame, W. S., Huang, R., Mayer, E. J., & Sunderam, A. (2021, October 19). *The Impact of Minority Representation at Mortgage Lenders*. SSRN.

Ghent, A. C., Hernández-Murillo, R., & Owyang, M. T. (2014). Differences in subprime loan pricing across races and neighborhoods. *Regional Science and Urban Economics*, *48*, 199–215. Retrieved from https://doi.org/10.1016/j.regsciurbeco.2014.07.006.

Hirasuna, D. P., & Allen, R. P. (2012). The resurgence of denial rates for home loans: an examination of disparate effects on groups of applicants in the upper Midwest. *Housing Policy Debate*, *22*, 573–603. https://doi.org/10.1080/10511482.2012.697910

Jerath, S. (2020). *Racial Discrimination in the Atlanta Mortgage Market, 2012-2017*. Georgia Tech Library. Retrieved from 2. https://smartech.gatech.edu/bitstream/handle/1853/62590/Jerath,%20Shikha.pdf

Kolar, J. M., & Jerison, J. D. (2006, February). *The Home Mortgage Disclosure Act: Its History, Evolution, and Limitation*. Retrieved from https://buckleyfirm.com/uploads/36/doc/HistoryofHMDAapr06.pdf

Ky, K.-E., & Lim, K. (2022, January 5). *The Role of Race in Mortgage Application Denials*. Federal Reserve Bank of Minneapolis. Retrieved from https://www.minneapolisfed.org/research/community-development-papers/the-role-of-race-in-mortgage-application-denials

Rappaport, J. (2010). The Effectiveness of Homeownership in Building Household Wealth. *Economic Review, Federal Reserve Bank of Kansas City*, *95*, 35–65. Retrieved from https://ideas.repec.org/a/fip/fedker/y2010iqivp35-65nv.95no.4.html.

Schill, M. H., & Friedman, S. (1999). The Fair Housing Amendments Act of 1988: The First Decade. *US Department of Housing and Urban Development*, *4*(3), 57–78. Retrieved from https://www.jstor.org/stable/20868475.

United States Census Bureau. (2022, August 2). *Quarterly Residential Vacancies and Homeownership, Second Quarter 2022*. Census.gov. Retrieved from https://www.census.gov/housing/hvs/files/currenthvspress.pdf

Appendix

Summary Statistics 1

Variable	Mean	Standard Deviation	Median
approved	0.9141728	0.2801087	1
denied	0.0858272	0.2801087	0
income	124.244	4861.715	86
log_income	4.504026	0.6912525	4.454347
loan_amount	298597.4	292090	245,000
log_loan_amount	12.34603	0.7949265	12.40901
ltv	20734.75	1.47E+07	90

This table shows the summary statistics for the dataset that was used to produce the regressions in this study. Further analysis on cleaning decisions and descriptions of each variable are given earlier in this study.

Summary Statistics 2

Race	Not Approved	Approved	Total	Percent of Total
Asian	87,993	902,522	990,515	7.74%
Black	178,491	951,553	1,130,044	8.83%
Hispanic	192,958	1,472,895	1,665,853	13.01%
White	626,767	8,285,390	8,912,157	69.61%
Other	12,553	91,261	103,814	.81%
Total	1,098,762	11,703,621	12,802,383	100%

This shows approved and not approved sums by race. It also shows the total counts of race and the percent of the total that they make up.

Table 1

Measuring Race Gaps in Mortgage Approvals

	OLS 1	OLS 2	FE 1	FE 2	FE 3	FE 4	FE5
Asian	-0.019***	-0.011***	-0.012***	-0.012***	-0.012***	-0.012***	-0.012***
Black	-0.088***	-0.048***	-0.046***	-0.046***	-0.046***	-0.046***	-0.043***
Hispanic	-0.046***	-0.019***	-0.019***	-0.019***	-0.019***	-0.019***	-0.019***
Other	-0.051***	-0.035***	-0.032***	-0.032***	-0.032***	-0.032***	-0.027***
Controls	-	Yes	Yes	Yes	Yes	Yes	Yes
State or Lender FE	-	-	Yes	Yes	Yes	Yes	Yes
Standard Errors	-	-	-	Robust	State	Lender	Lender
LEI * County	-	-	-	-	-	-	Yes
Adjusted R2	0.010	0.195	0.213	0.213	0.213	0.213	0.227
Observations	12802383	12742520	12742464	12742464	12742464	12742464	12553815

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The above table shows different regressions of U.S mortgage approval rates on race, ethnicity, and other control variables. The list of control variables can be found under the methodology portion of this paper. However, OLS 1 does not contain any controls. It represents the average difference in mortgage approvals by race.

Table 2

Measuring Race Gaps in Mortgage Approvals - extended

	OLS 1	FE 1	FE 2	FE 3	FE 4	FE 5
White	0.000	0.000	0.000	0.000	0.000	0.000
Asian	-0.019***	-0.012***	-0.012***	-0.012***	-0.012***	-0.012***
Black	-0.088***	-0.046***	-0.046***	-0.046***	-0.046***	-0.043***
Hispanic	-0.046***	-0.019***	-0.019***	-0.019***	-0.019***	-0.019***
Other	-0.051***	-0.032***	-0.032***	-0.032***	-0.032***	-0.027***
Male		0.000	0.000	0.000	0.000	0.000
Female		0.002***	0.002***	0.002***	0.002***	0.002***
Log Income		0.012***	0.012***	0.012***	0.012***	0.011***
Log Loan Amount		0.013***	0.013***	0.013***	0.013***	0.012***
No Preapproval Request		0.000	0.000	0.000	0.000	0.000
Preapproval Requested		0.110***	0.110***	0.110***	0.110***	0.107***
LTV		0.000	0.000***	0.000***	0.000***	0.000***
DTI 0%-20%		0.000	0.000	0.000	0.000	0.000
DTI 20%-<30%		0.023***	0.023***	0.023***	0.023***	0.020***
DTI 30%-<36%		0.024***	0.024***	0.024***	0.024***	0.021***
DTI 36		0.023***	0.023***	0.023***	0.023***	0.021***
DTI 37		0.023***	0.023***	0.023***	0.023***	0.021***
DTI 38		0.022***	0.022***	0.022***	0.022***	0.020***
DTI 39		0.022***	0.022***	0.022***	0.022***	0.019***
DTI 40		0.022***	0.022***	0.022***	0.022***	0.019***
DTI 41		0.021***	0.021***	0.021***	0.021***	0.018***
DTI 42		0.020***	0.020***	0.020***	0.020***	0.018***
DTI 43		0.017***	0.017***	0.017***	0.017***	0.014***
DTI 44		0.017***	0.017***	0.017***	0.017***	0.014***
DTI 45		0.006***	0.006***	0.006**	0.006	0.004
DTI 46		0.001	0.001	0.001	0.001	-0.001
DTI 47		0.000	0.000	0.000	0.000	-0.002
DTI 48		0.001	0.001	0.001	0.001	-0.001
DTI 49		0.003***	0.003***	0.003	0.003	0.001
DTI 50%-60%		-0.108***	-0.108***	-0.108***	-0.108***	-0.108***
DTI >60%		-0.693***	-0.693***	-0.693***	-0.693***	-0.691***
DTI 0%-20% # LTV		0.000	0.000	0.000	0.000	0.000
DTI 20%-<30% # LTV		-0.000	-0.000***	-0.000***	-0.000***	-0.000***
DTI 30%-<36% # LTV		-0.000	-0.000***	-0.000***	-0.000***	-0.000***
DTI 36 # LTV		-0.000	-0.000***	-0.000***	-0.000***	-0.000***
DTI 37 # LTV		-0.000	-0.000***	-0.000***	-0.000***	-0.000***
DTI 38 # LTV		-0.000	-0.000***	-0.000***	-0.000***	-0.000***
DTI 39 # LTV		-0.000	-0.000***	-0.000***	-0.000***	-0.000***
DTI 40 # LTV		-0.000	-0.000***	-0.000***	-0.000***	-0.000***
DTI 41 # LTV		-0.000	-0.000***	-0.000***	-0.000***	-0.000***
DTI 42 # LTV		-0.000	-0.000***	-0.000***	-0.000***	-0.000***
DTI 43 # LTV		-0.000	-0.000***	-0.000***	-0.000***	-0.000***
DTI 44 # LTV		-0.000	-0.000***	-0.000***	-0.000***	-0.000***
DTI 45 # LTV		-0.000	-0.000***	-0.000***	-0.000***	-0.000***
DTI 46 # LTV		-0.000	-0.000***	-0.000***	-0.000***	-0.000***
DTI 47 # LTV		-0.000	-0.000***	-0.000***	-0.000***	-0.000***
DTI 48 # LTV		-0.000	-0.000***	-0.000***	-0.000***	-0.000***
DTI 49 # LTV		-0.000	-0.000***	-0.000***	-0.000***	-0.000***
DTI 50%-60% # LTV		-0.000	-0.000***	-0.000***	-0.000***	-0.000***
DTI >60% # LTV		0.000***	0.000**	0.000*	0.000	0.000
Conventional		0.000	0.000	0.000	0.000	0.000
FHA		-0.004***	-0.004***	-0.004	-0.004	-0.003
RHS or FSA		-0.034***	-0.034***	-0.034***	-0.034***	-0.033***
VA		0.029***	0.029***	0.029***	0.029***	0.028***
Occupancy Type 1		0.000	0.000	0.000	0.000	0.000
OT 2		-0.011***	-0.011***	-0.011***	-0.011***	-0.006***
OT 3		-0.015***	-0.015***	-0.015***	-0.015***	-0.015***
AK		0.000	0.000	0.000	0.000	
AL		-0.014***	-0.014***	-0.014***	-0.014	
AR		-0.019***	-0.019***	-0.019***	-0.019*	

Measuring Race Gaps in Mortgage Approvals – extended (Continued)

	OLS 1	FE 1	FE 2	FE 3	FE 4	FE 5
AZ		-0.017***	-0.017***	-0.017***	-0.017*	
CA		-0.013***	-0.013***	-0.013***	-0.013	
CO		-0.011***	-0.011***	-0.011***	-0.011	
CT		-0.014***	-0.014***	-0.014***	-0.014	
DC		-0.024***	-0.024***	-0.024***	-0.024**	
DE		-0.008***	-0.008***	-0.008***	-0.008	
FL		-0.032***	-0.032***	-0.032***	-0.032***	
GA		-0.019***	-0.019***	-0.019***	-0.019*	
GU		-0.043	-0.043	-0.043***	-0.043***	
HI		-0.018***	-0.018***	-0.018***	-0.018	
IA		-0.004	-0.004	-0.004	-0.004	
ID		-0.006**	-0.006**	-0.006*	-0.006	
IL		-0.016***	-0.016***	-0.016***	-0.016*	
IN		-0.015***	-0.015***	-0.015***	-0.015*	
KS		-0.005*	-0.005**	-0.005*	-0.005	
KY		-0.016***	-0.016***	-0.016***	-0.016*	
LA		-0.020***	-0.020***	-0.020***	-0.020*	
MA		-0.014***	-0.014***	-0.014***	-0.014*	
MD		-0.013***	-0.013***	-0.013***	-0.013	
ME		-0.024***	-0.024***	-0.024***	-0.024**	
MI		-0.020***	-0.020***	-0.020***	-0.020*	
MN		0.002	0.002	0.002	0.002	
MO		-0.013***	-0.013***	-0.013***	-0.013*	
MS		-0.025***	-0.025***	-0.025***	-0.025**	
MT		-0.020***	-0.020***	-0.020***	-0.020**	
NC		-0.008***	-0.008***	-0.008***	-0.008	
ND		-0.007**	-0.007**	-0.007*	-0.007	
NE		0.002	0.002	0.002	0.002	
NH		-0.021***	-0.021***	-0.021***	-0.021**	
NJ		-0.026***	-0.026***	-0.026***	-0.026***	
NM		-0.018***	-0.018***	-0.018***	-0.018*	
NV		-0.016***	-0.016***	-0.016***	-0.016*	
NY		-0.031***	-0.031***	-0.031***	-0.031***	
OH		-0.011***	-0.011***	-0.011***	-0.011	
OK		-0.018***	-0.018***	-0.018***	-0.018*	
OR		-0.011***	-0.011***	-0.011***	-0.011	
PA		-0.014***	-0.014***	-0.014***	-0.014	
PR		-0.097***	-0.097***	-0.097***	-0.097**	
RI		-0.012***	-0.012***	-0.012***	-0.012	
SC		-0.020***	-0.020***	-0.020***	-0.020**	
SD		0.007**	0.007**	0.007	0.007	
TN		-0.020***	-0.020***	-0.020***	-0.020**	
TX		-0.015***	-0.015***	-0.015***	-0.015	
UT		-0.009***	-0.009***	-0.009**	-0.009	
VA		-0.010***	-0.010***	-0.010***	-0.010	
VI		-0.099	-0.099	-0.099***	-0.099*	
VT		-0.029***	-0.029***	-0.029***	-0.029***	
WA		-0.012***	-0.012***	-0.012***	-0.012	
WI		-0.009***	-0.009***	-0.009***	-0.009	
WV		-0.028***	-0.028***	-0.028***	-0.028***	
WY		-0.012***	-0.012***	-0.012***	-0.012	
Constant	0.930***	0.725***	0.725***	0.725***	0.725***	0.725***
FE	-	Yes	Yes	Yes	Yes	Yes
Standard Errors	-	-	Robust	State	Lender	Lender
LEI * County	-	-	-	-	-	Yes
Adjusted R2	0.010	0.213	0.213	0.213	0.213	0.227
Observations	12802383.000	12742464.000	12742464.000	12742464.000	12742464.000	12553815.000

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

A similar table to table 1 including most of the control coefficients. However, OLS 2 was left off due to the length it would have added to the table. The Lender variables have also been left out due to the amount there are and how much space it would add to this chart.

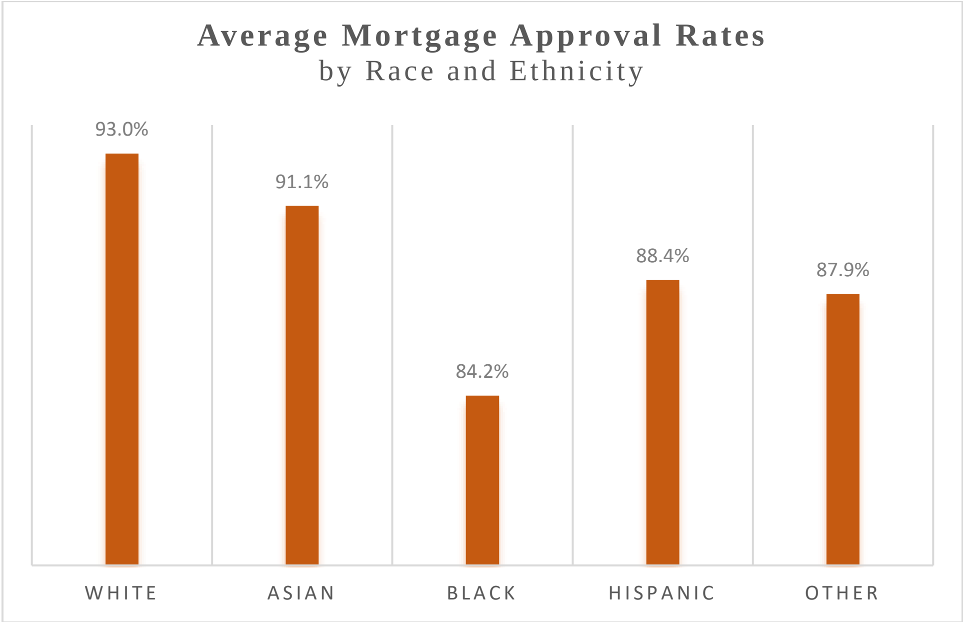
Figure 1

Snapshot from Data table used for Regressions

index	13	12	7	2	16
Year	2019	2019	2019	2019	2019
Lender_LEI	6B…02	6B.02	6B…M02	6B…02	6B…02
State	PA	MN	FL	AZ	IA
County_Code	42003.0	27003.0	12057.0	4013.0	19153.0
Census_Tract	4.2e+10	2.7e+10	1.2e+10	4.0e+09	1.915301e+10
Approved	1	1	1	1	1
Denied	0	0	0	0	0
Race	0_White	0_White	0_White	Asian	0_White
Sex	0_Male	Female	0_Male	Female	0_Male
Income	121.0	129.0	217.0	80.0	62.0
Log_Income	4.7	4.8	5.3	4.3	4.1
Loan_Amount	165000.0	205000.0	615000.0	385000.0	265000.0
Log_Loan_Amount	12.0	12.2	13.3	12.8	12.4
LTV	90.0	85.0	80.0	90.0	80.1
Loan_Type	Conv.	Conv.	Conv.	Conv.	VA
DTI_Ratio	39	30%-<36%	20%-<30%	30%-<36%	45
Preapproval	0 No…uest	0 No…uest	0 No…uest	0 No…uest	0 No…uest
Occupancy_Type	1	1	1	1	1
counts	136976	136976	136976	136976	136976

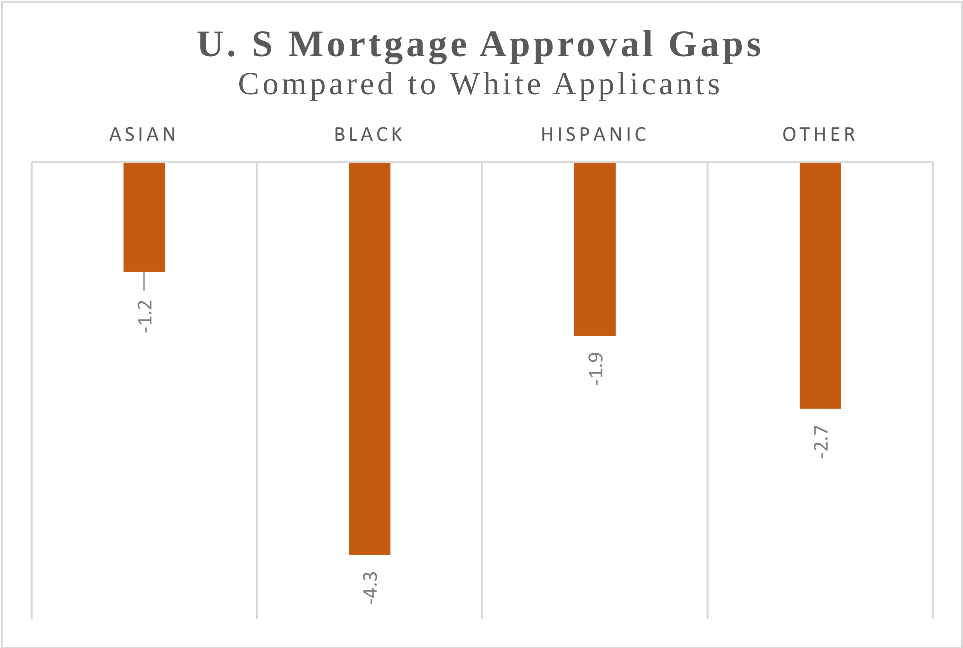
Above shows a snapshot from the data set used to run all regressions in this study. The datasets columns are in the first column in the figure above. The next columns show 5 rows from the dataset used.

Figure 2



This graph represents the average probability of an approval by Race. It was created in Excel using the coefficients for the "Race" variable from OLS 1. OLS 1 can be found in Table 1 in the appendix. The "Race" category includes those who are Hispanic for this study.

Figure 3



The graph represents U.S Mortgage Approval Gaps by Race estimated in this study. These are the gaps between each group listed and White applicants. They are based on the variables estimates from the regression FE 4. Which can be found under Table 1 in the appendix.

Figure 4

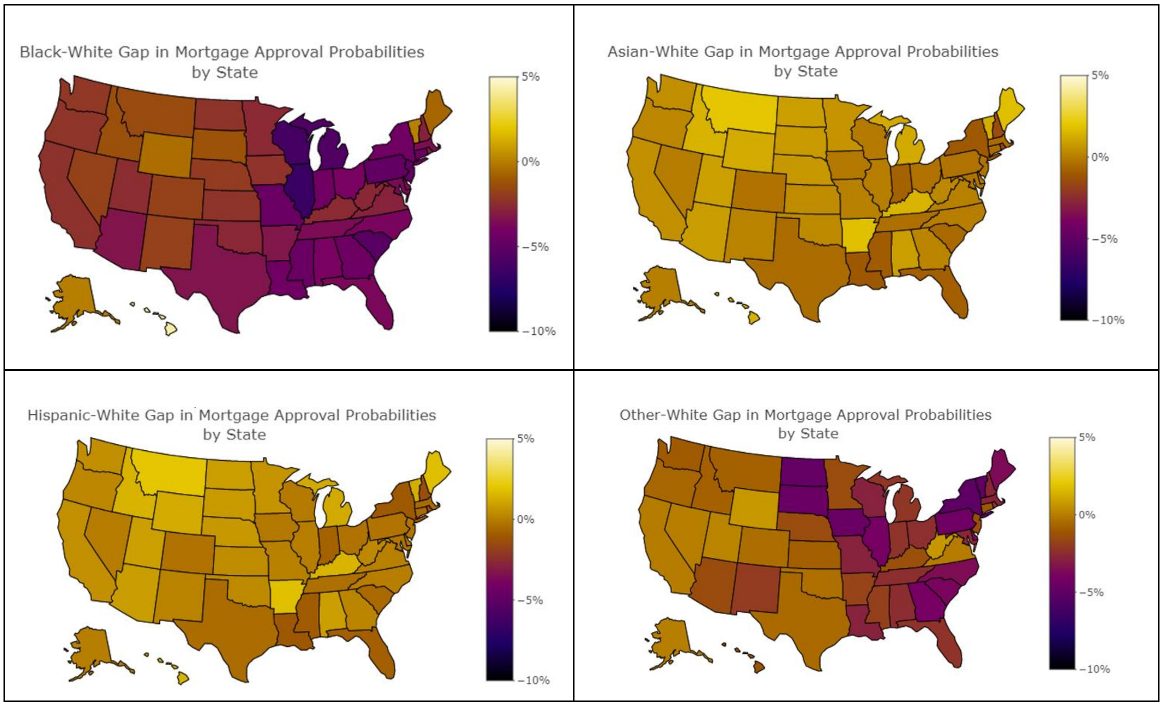


Figure 4 shows estimated gaps for mortgage approval rates by race mapped by U.S state. For example, the top left choropleth map shows the Black and White gaps in mortgage approval by state. These estimates were from a regression like FE 4. However, the race and state variables were interacted. The coefficients from the state and race interaction regression are what is mapped. Figure 5 shows a table of those coefficients.

Figure 5

Coefficients from Regression with State * Race Interactions

	Asian	Black	Hispanic	Other
AK	0	0	0	0
AL	0.00888	-0.0394	-0.0163	-0.0249
AR	0.0177	-0.0325	-0.0170	-0.0173
AZ	0.00816	-0.0327	-0.0150	-0.0146
CA	0.00472	-0.0234	-0.00151	-0.000758
CO	-0.00327	-0.0179	-0.00345	-0.00452
CT	-0.00523	-0.0450	-0.0238	-0.0103
DC	-0.00308	-0.0568	-0.0127	-0.0104
DE	-0.00161	-0.0367	-0.0185	-0.0361
FL	-0.00852	-0.0372	-0.0208	-0.0229
GA	0.00178	-0.0443	-0.0111	-0.0400
GU	0.0661	-0.185	0.131	0.108
HI	0.0144	0.0425	0.0229	-0.0121
IA	-0.00152	-0.0231	-0.0163	-0.0460
ID	0.0142	-0.0134	-0.00449	-0.00807
IL	-0.000273	-0.0657	-0.0214	-0.0404
IN	-0.00782	-0.0425	-0.0164	-0.0226
KS	0.00355	-0.0228	-0.0156	-0.00855
KY	0.0146	-0.0257	-0.0100	-0.0135
LA	-0.0106	-0.0431	-0.0147	-0.0285
MA	-0.00543	-0.0353	-0.0174	-0.0279
MD	-0.00260	-0.0368	-0.0187	-0.0267
ME	0.0174	-0.0067	-0.0195	-0.0363
MI	0.0118	-0.0564	-0.0182	-0.0217
MN	0.00584	-0.0259	-0.0138	-0.0145
MO	0.000578	-0.0457	-0.0146	-0.0284
MS	-0.00975	-0.0454	-0.0125	-0.0187
MT	0.0194	-0.0147	-0.0112	-0.00796
NC	-0.000104	-0.0384	-0.0156	-0.0358
ND	0.00875	-0.0227	-0.0287	-0.0478
NE	0.00682	-0.0199	-0.00923	-0.0143
NH	-0.0144	-0.0287	-0.0179	-0.0269
NJ	-0.00295	-0.0478	-0.0190	-0.0140
NM	0.00127	-0.0184	-0.00522	-0.0201
NV	-0.000478	-0.0181	-0.00186	-0.000309
NY	-0.0104	-0.0425	-0.0240	-0.0506
OH	-0.00324	-0.0393	-0.0144	-0.0240
OK	0.00302	-0.0268	-0.00797	-0.00402
OR	0.00233	-0.0224	-0.0102	-0.00638
PA	-0.00281	-0.0493	-0.0215	-0.0434
PR	-0.0589	-0.0865	0.0312	-0.0571
RI	-0.0141	-0.0284	-0.0325	-0.0214
SC	-0.00557	-0.0530	-0.0226	-0.0382
SD	0.00775	-0.0126	-0.0101	-0.0458
TN	-0.00392	-0.0361	-0.0154	-0.0242
TX	-0.00519	-0.0331	-0.00891	-0.00561
UT	0.00901	-0.0248	0.00204	0.00170
VA	0.000764	-0.0284	-0.0156	-0.000694
VT	0.0128	0.00073	-0.0311	-0.0493
WA	0.00445	-0.0215	-0.0130	-0.00997
WI	-0.000864	-0.0610	-0.0254	-0.0288
WV	0.00293	-0.0266	-0.00363	0.00664
WY	0.0122	-0.0044	-0.00686	0.00693

Above are the coefficients from a regression similar to that of FE 4. However, the state and race variables are interacted. The coefficients on those interactions are what is in the table above.