Race and Rates: Does Race Determine Mortgage Rates?

Mauricio Moreno

Economics Department, Middlebury College, Middlebury, Vermont 05753, USA

E-mail: mmoreno@middlebury.edu

Abstract

It goes without saying that race plays a large factor in everyday interactions and biases taught and observed towards minorities contribute towards unequal dynamics and treatment of individuals. The groups of interest in this research are Black and Hispanic/Latino-identifying groups of people. When it comes to the United States, not only do many hold biases in their personal lives, but many institutions have also demonstrated unfair treatment or racist-based policies and foundations. These biased and racist foundations tend to value white individuals more and make resource access unavailable to minorities in addition to standing as obstacles to social and financial mobility. Higher education, prison, and the housing market are examples of such institutions.

This research analyzes one aspect of the housing market – the mortgage loan application process. Past studies have shown that not only do, specifically, Black applicants have a lower probability of getting an approved mortgage loan, but other research has supported that checking off the "Black" race box also correlates with greater loan interest rates. Using 2019 data from the Home Mortgage Disclosure Act (HMDA) for applications originating from California, this research draws to see if those relationships are still observed in the present. The results from this study show that identifying as Black or Hispanic/Latino significantly decreases the odds of

getting a loan approved but surprisingly also decreases expected loan interest rates on approved loans when all other included economic variables are kept constant. We also discover that sex is not a significant determinant in multiple models.

Keywords: race; mortgage; loan interest rate; loan approval; Home Mortgage Disclosure Act

---

Introduction

In this paper, we will take a two-fold approach in order to examine the role of race as one of various causal factors for determining mortgage loan approvals, and the interest rates which accompany those loans. Our racial and ethnic groups of interest are Black and Hispanic/Latino. The reason for electing to analyze these groups of people is because they are underrepresented groups of homeowners and have also historically fallen victim to subprime, or predatory, lending (Been et al., 2009). These predatory lending practices have not only contributed to reverse redlining but have also concentrated low-quality credit in these minority-dominated areas (Haupert, 2019). Because of these subprime practices, there were loan modifications as part of federal government intervention, but by then, loan defaults were already prominent and high. Because of the nature of particular and racialized subprime targeting, we can use disclosed mortgage loan originations to further observe and critique the American housing system and how loans are distributed.

Recent studies have also identified disparities in mortgage lending with respect to applicant race, but even if statistical tests do support its significance, it can be difficult to determine if racial disparities in lending are an outcome of racial discrimination, or if it is just an unfortunate outcome of "efficiently functioning markets" (Leven et al., 1994). Although it can be

difficult to determine individual and personal biases, a quantitative analysis of mortgage lending data can surely shed light on common trends.

As mentioned above, the data analyzed is from California as opposed to aggregate national data. Minimizing our scope for the data sample was in part due to a lack of computational power to handle the millions of observations, but data from this particular state should be able to capture overall trends with our racial and ethnic groups of focus. Compared to other areas nationwide, California has prominent Black and Hispanic/Latino populations. Having such a large sample of representative data should allow us to extrapolate and normalize results and findings for the entirety of the US. That being said, California is a progressive state which does offer more intervention and assistance for minority, first-time homeowners.

In this particular study, we will follow previous research to demonstrate this disparity in mortgage loan approvals and loan interest rates, but we will also compare these racial effects on another aspect of the loan application process – total origination costs. I have not found prior studies that look at this so doing so may also help us understand if the mortgage loan application process encourages or deters Black and Hispanic/Latino individuals.

Background

What role has race played in mortgage lending? How has this racial discrimination been observed?

Despite government intervention to make resources more accessible and encourage minority first-time homeowners to buy homes, previous literature shows disparities in applicant

approvals with differing races, in particular, they found that there is a negative relationship between being a Black applicant, and getting your application approved. One paper found that "the racial composition of the neighborhood has a significant relationship with the volume of loans given to an area…even after controlling for economic variables" (Squires et al., 1991). Beyond that, Munnell et al. discovered in 1992 that holding for constant income, Black individuals were getting denied for mortgage loans twice as much as White individuals on the nationwide level, and denied up to three times more in metropolitan areas and big cities (1992)! Findings such as these only further support the notion that there are race-based loan denials, even when all other economic factors are the same across applicants.

Even with the basis established by other literature, Gotham brings up a key point of contention, if the data analyzing the effect of race in mortgage lending emphasizing disparate treatment (intentional discrimination) or disparate impact (institutional discrimination) (1998). As I have said before, I believe that the housing market is an institution that discriminates against minority applicants just by virtue of the application system and process. My goal is for continued research on this topic to create channels for reform and greater resource access to minority individuals.

Age and mortgage

There has surprisingly been very little prior research investigating the effects of age on mortgage loan approvals. Although age was not established as a continuous variable, but rather a categorical one, Epley et al. were able to uncover that financial institutions were more likely to reject loan applications from older applicants compared to younger applicants (1999). Although these finding did not make sense to me at first since I would expect older individuals to have greater financial security, the results do make sense because a financial institution would rather

give a loan to someone who has many years left to go compared to someone who is a senior citizen, for example.

Sex and mortgage

With respect to an applicant's sex, there appears to be conflicting data on whether or not being male or female will make an applicants loan application more likely to be approved versus denied. In the paper written by Delis et al., what they found was that for some time periods, sex (or in their case gender) was significant in determining loan approvals, but for others sex was not significant. Surprisingly, for the instances where sex was significant, it went in several different ways such that being male was beneficial in one instance and females were better off in the others. All in all however, the main effects by sex were not deemed all that significant or explanatory of approval rates.

Income and mortgage

Unfortunately, there did not appear to be any prior studies with respect to income effects on mortgage lending rate in the US. I am very happy to take this on even if it is in the shape of a short project such as this one. My reasons for selecting for income as an independent variable to test for determining approval and interest rates, is that I believe intuitively, that the more money and collateral you have, the more confidence a financial institution such as a bank has to lend you money. Upon selecting this variable for analysis, I foresaw that there may be a significant interaction between income and race/ethnicity since we have prevalent wealth gaps in the US.

Economic Models

Empirical model for loan approval:

$$\text{approved} = \alpha_0 + \underset{(-)}{\alpha_1}\text{black} + \underset{(-)}{\alpha_2}\text{hisp} + \underset{(+)}{\alpha_3}\text{sex} + \underset{(-)}{\alpha_4}\text{age} + \underset{(+)}{\alpha_5}\text{income} + \varepsilon$$

Empirical model for interest rate:

$$\text{interest\_rate} = \alpha_0 + \underset{(+)}{\alpha_1}\text{black} + \underset{(+)}{\alpha_2}\text{hisp} + \underset{(-)}{\alpha_3}\text{sex} + \underset{(+)}{\alpha_4}\text{age} + \underset{(-)}{\alpha_5}\text{income} + \varepsilon$$

Empirical model for origination fees:

$$\text{origination\_charges} = \alpha_0 + \underset{(+)}{\alpha_1}\text{black} + \underset{(+)}{\alpha_2}\text{hisp} + \underset{(-)}{\alpha_3}\text{sex} + \underset{(+)}{\alpha_4}\text{age} + \underset{(+)}{\alpha_5}\text{loan\_amount} + \varepsilon$$

Approved is coded as a dummy variable to represent if a mortgage loan application was approved (with value 1) or denied (with value 0).

Interest_rate is a continuous variable measured to the nearest hundredth of a percentage.

Origination_charges is a continuous variable measured in singular dollar units.

Black is a coded dummy variable which is used to identify an applicant as Black (coded as 1) or White (coded as 0). We dropped mixed/joint races so Black and White are unique in that sense. I suspect that Black is negatively correlated with approved since there is former literature that supports the fact that Black individuals experience lower approval rates than their white counterparts. Which that same logic, I would expect for Black individuals to receive higher interest rates if offered a loan and for the resulting origination charges to increase as well.

Hisp is another dummy variable which is used to identify applicants as Hispanic/Latino (coded as 1) or not (coded as 0). Any applicant who indicated a Hispanic/Latino ethnicity were coded as 1 REGARDLESS of race so that we can maintain unique and exclusive groupings. Just as with the Black variable, I also expect for Hispanic individuals to experience lower approval ratings while facing increased interest rates and originations costs.

Sex is a dummy variable which captures the applicant's sex. This binary variable codes for male (sex == 1) and female (sex == 0). Because of a general wealth gap and the overall more privilege that men experience, I expect for males to have higher rates of loan approvals, and to inversely receive lower interest rates and origination charges.

Age is a discrete variable that has been transformed so that it represents time intervals of ten years. Age == 1 is <25, Age == 2 is 25-34, Age == 3 is 35-44, …, Age ==7 is >74. Because the older you get, the greater your chance of old death, this is an increased liability. For that reason, I expect for age to be negatively correlated with approved, and positively correlated with loans interest rate and origination charges.

Income is a discrete variable measured in the nearest thousands of dollars. The data was already transformed such that incomes are in thousands of dollars (i.e., 78 = 78,000). I expect for individuals to not only be more likely to get their loan approved, but to also have lower interest rates and origination charges.

Loan_amount is a continuous variable measured in singular dollar amounts. I predict that this variable will be negatively correlated with origination costs since it's a greater financial risk for financial institutions.

All data was gathered from the Home Mortgage Disclosure Act with samples from 2019. Dummy variables such as those above were coded for using the HDMA data.

Data

The data for this study was mortgage loan application originations in California from the year 2019, made available via the Home Mortgage Disclosure Act. The HMDA does offer aggregate data at the National level, but due to excessive observations and increased computational power needed, I decided to limit the data set to only include the state of California due to its large populations and diverse demographics.

The original data set contained 1,715,924 observations. After dropping loan applications that were still in progress or that were closed for incompleteness, total observations lowered to approximately 475,000. We then continued to remove observations for which ethnicity was unavailable; this left a sample size of 314,593. We also dropped any observation for which race and sex were joint or unavailable, and observations that had an undefined age. Our final sample size was 144,798 total observations for each of the dummy variables. There was some variance in total observations amongst discrete and continuous variables.

Table 1 *Summary statistics of variables of interest.*

Variable	Obs	Mean	Std. Dev.	Min	Max
loan_amount	144,798	315981.1	384562.7	5000	2.80E+07
interest_r~e	42,477	4.480898	1.088985	0	39.06
originatio~s	19,201	2686.229	2634.383	0	37331.44
income	133,926	137.5792	2168.472	0	734000
approved	144,798	0.3014959	0.4589091	0	1
hisp	144,798	0.3586168	0.4795961	0	1
black	144,798	0.0748422	0.2631375	0	1
sex	144,798	0.6068316	0.4884554	0	1
age	144,798	4.136839	1.404523	1	7

Key Estimated Models

Loan approval

After completing a standard regression, we decided to create and utilize interaction terms between each of the independent variables and income since I believed that sex, race, ethnicity, and age all interact with income. Significant results were achieved but I decided to continue with a logit model in order to try to improve the R-squared value and address the non-linear relationship that is found with a naïve linear probability model. The results are shown in Table 2. All coefficients for main effects are significant at the 0.1% level with the exception of income which is significant at the 5% level. The interpretations for these results are that, controlling for all other variables, that Black applicants are 0.4138 log odds less likely to get their loan application approved compared to white counterparts; Hispanic/Latino applicants are 0.4575 log odds less likely to get their loan application approved compared to non-Hispanic counterparts;

males are 0.1352 log odds less likely to get their loan application approved compared to female counterparts; individuals are 0.2374 log odds less likely to get their loan application approved for every ten years they age; a one thousand dollar increase in income is expected to yield a 0.00017 log odds decrease in expected loan approval. The interaction terms were all significant so we can conclude that White people earn more than Black, non-Hispanics earn more than Hispanics, females earn less than males, and the older you are, the more you earn. Joint significance tested to be significant at 0.1% (see Figure 1).

Table 2 *Logistic regression on approved dependent variable with interaction terms.*

Logistic regression — Number of obs = 133,926; LR chi2(9) = 3303.50; Prob > chi2 = 0.0000; Log likelihood = -75511.006; Pseudo R2 = 0.0214

approved	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
black	-.4138159	.0304215	-13.60	0.000	-.473441	-.3541907
hisp	-.4574941	.0148208	-30.87	0.000	-.4865423	-.4284459
sex	-.1351571	.0140422	-9.63	0.000	-.1626793	-.107635
age	-.2374084	.0049982	-47.50	0.000	-.2472048	-.227612
income	-.0001697	.0000859	-1.98	0.048	-.000338	-1.38e-06
interact_bi	.000465	.0001411	3.30	0.001	.0001885	.0007416
interact_hi	.0001528	.0000437	3.50	0.000	.0000673	.0002384
interact_si	-.0000975	.0000491	-1.99	0.047	-.0001937	-1.33e-06
interact_ai	.0000562	.0000144	3.90	0.000	.0000279	.0000844
_cons	.1984445	.0256342	7.74	0.000	.1482024	.2486866

Figure 1 *Joint significance test*

```
. test black = hisp = sex = 0

 ( 1)  [approved]black - [approved]hisp = 0
 ( 2)  [approved]black - [approved]sex = 0
 ( 3)  [approved]black = 0

           chi2(  3) = 1117.03
         Prob > chi2 =    0.0000
```

Interest Rate

Results chart in Table 3 below. For this model, we regressed interest rate by black, hisp, sex, age, and income. With the first regression some significant coefficients were produced but the calculated R-squared value was low which meant that a small amount of variance of the dependent variable was being accounted for. I transformed the dependent variable to instead have it as the natural log of interest rate by that still yielded a small R-squared value. I suspected that the relationship was not too linear. I transformed the model again to take in the natural log of income as an independent variable, while also calculating new interaction terms to reflect the transformation. I then went back and conducted Ramsey's RESET tests to compare AIC and BIC values in order to choose the best model. The results for that in below in Table 3. Black, hisp, and age variables are all significant at 0.5%, while lnincome and sex are not significant. Although the lnincome result is shocking, the sex one is not considering prior literature did not find consistent data to show that sex was a significant determinant in mortgage rates. The main effects show that Black applicants are expected to experience a ((e^-0.1823)-1)*100 = 16.66% decrease in interest rate (NOT percentage point decrease) compared to white applicants; Hispanic applicants are expected to experience a ((e^0.0595)-1)*100 = 5.78% decrease in interest rate compared to non-Hispanic applicants; for every 10 years older you get, you are expected to experience a ((e^0.0281)-1)*100 = 2.85% increase in interest. Joint significance is significant and all but the sex interaction terms are significant (with same interpretation as previous model).

Table 3 *Regression of transformed interest rate dependent variable and income. Interaction terms included followed by Ramsey's RESET tests.*

Source	SS	df	MS	Number of obs	=	33,640
				F(9, 33630)	=	36.72
Model	18.3310106	9	2.03677895	Prob > F	=	0.0000
Residual	1865.346	33,630	.055466726	R-squared	=	0.0097
				Adj R-squared	=	0.0095
Total	1883.67701	33,639	.055996819	Root MSE	=	.23551

lninterest	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
black	-.1822989	.0403736	-4.52	0.000	-.2614325	-.1031653
hisp	-.0595129	.0211054	-2.82	0.005	-.1008802	-.0181455
sex	-.0227192	.0186325	-1.22	0.223	-.0592396	.0138012
age	.0280945	.0061922	4.54	0.000	.0159576	.0402314
lnincome	.0055913	.0069396	0.81	0.420	-.0080106	.0191932
interact_blni	.046281	.008683	5.33	0.000	.029262	.0632999
interact_hlni	.0200326	.0046216	4.33	0.000	.0109741	.029091
interact_slni	.0055229	.0039986	1.38	0.167	-.0023144	.0133602
interact_alni	-.003407	.0013483	-2.53	0.012	-.0060497	-.0007643
_cons	1.393737	.032282	43.17	0.000	1.330463	1.45701

```
.
. ovtest

Ramsey RESET test using powers of the fitted values of lninterest
       Ho:  model has no omitted variables
               F(3, 33627) =      9.84
                  Prob > F =      0.0000

.
. estat ic

Akaike's information criterion and Bayesian information criterion
```

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	33,640	750.3914	914.877	10	-1809.754	-1725.519

Note: BIC uses N = number of observations. See [R] BIC note.

Origination charges

This was a pretty straightforward test, especially since I did not need to account for activation terms. We settled on an independent variable transformed model since it yielded better AIC and BIC scores when compared to other tested models. See Table 4 below for results. Black applicants can expect to pay about $625 more in origination charges compared to White applicants; Hispanics can expect to pay approximately $553 more than non-Hispanic applicants; applicants can expect to pay $104 more for every ten years they age; a 1% increase in loan amount can expect to yield about $7.55 more in origination charges. These increased origination charges may very well dissuade Hispanic and Black applicants.

Table 3 *Regression of transformed interest rate dependent variable and income. Interaction terms included followed by Ramsey's RESET tests*

```
. reg origination_charges black hisp sex age lnloan
```

Source	SS	df	MS	Number of obs	=	19,201
				F(5, 19195)	=	124.81
Model	4.1955e+09	5	839090742	Prob > F	=	0.0000
Residual	1.2905e+11	19,195	6723209.86	R-squared	=	0.0315
				Adj R-squared	=	0.0312
Total	1.3325e+11	19,200	6939972.24	Root MSE	=	2592.9

origination~s	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
black	625.1713	83.91965	7.45	0.000	460.6814	789.6612
hisp	552.9218	42.77377	12.93	0.000	469.0814	636.7621
sex	-62.42022	38.70331	-1.61	0.107	-138.2821	13.44166
age	103.7379	14.09686	7.36	0.000	76.1068	131.3689
lnloan	755.5627	33.40132	22.62	0.000	690.0932	821.0322
_cons	-7497.293	437.3612	-17.14	0.000	-8354.559	-6640.026

```
.
. ovtest

Ramsey RESET test using powers of the fitted values of origination_charges
       Ho:  model has no omitted variables
                F(3, 19192) =     30.01
                   Prob > F =     0.0000

.
. estat ic

Akaike's information criterion and Bayesian information criterion
```

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	19,201	-178479.4	-178172.2	6	356356.5	356403.6

Note: BIC uses N = number of observations. See [R] BIC note.

Conclusions

The findings from this research offer substantial support for the position that there are racial biases when determining mortgage lending. As stated before, the housing market, as a lending market specifically, are perpetuators of disparate impact and build upon institutional discrimination. This cycle limits minorities, Black individuals specifically, and is one mechanism that prevents social mobility.

Looking at this data and reading previous works, I am curious to see if this is a trend that will continue for decades to come, or if we will be able to see less disparities, or perhaps complete equality, in the near future. Taking a moment to reflect on the sex independent variable, we noticed that in two of the three models carried out, sex did not have significant main effects or even significant interaction terms. The reason I bring this up is because women were once disenfranchised and still face a wage gap today, but in our latter two models, sex had no impact, and in the first, an applicant is more likely to get approved if they were female! I understand that race plays a far greater role in everyday life, but it makes me a little hopeful that we can work towards decreasing inequalities and remedying biases.

With that being said, however, it is still important to reflect on implications that can be experienced today. As we have seen through our economic models, being Black or Hispanic lowers chances of getting approved for a mortgage loan and yields subsequent higher origination fees. To address these issues, I believe that a suite of actions can take place – firstly addressing financial institutions that lend money. They should be mandated to receive implicit bias training, as well as comprehensive instruction to appreciate diverse backgrounds and perhaps have more of an instructional role for the application process for first time homeowners, especially minority one. Secondly, I think that the institutions should sponsor applicants by waiving origination costs

since those may very well deter applicants. Lastly, I propose that local, state, and federal governments subsidize loan costs or offer a grace period before adding the interest rate on a loan which can help applicants prepare for payments and feel some sense of financial security.

There were many aspects of my data sample that I could not address. In the future, next steps for this research can include looking at mixed applicant/co-applicant races, credit scores, zip code area property values, and debt-ratio to see their effect on interest rates or loan approvals. It would also be valuable to compare these values over a period of say 15 years.

References

Been, Vicki, Ingrid Ellen & Josiah Madar, THE HIGH COST OF SEGREGATION: EXPLORING RACIAL DISPARITIES IN HIGHCOST LENDING, 36 Fordham Urb. L.J. 361 (2009). Available at: https://ir.lawnet.fordham.edu/ulj/vol36/iss3/1

Boehm, Paul D. Thistle & Alan Schlottmann (2006) Rates and race: An analysis of racial disparities in mortgage rates, Housing Policy Debate, 17:1, 109-149, DOI: 10.1080/10511482.2006.9521563

Delis, Manthos D., and Panagiotis Papadopoulos. 2019. "Mortgage Lending Discrimination Across the U.S.: New Methodology and New Evidence." Journal of Financial Services Research 56 (3): 341–68. https://doi.org/10.1007/s10693-018-0290-0.

Epley, Donald R., and Kartono Liano. 1999. "Borrower Age Differences in Residential Mortgage Lending by Commercial Banks." Social Science Quarterly 80 (3): 604–13.

Faber, Jacob (2013) Racial Dynamics of Subprime Mortgage Lending at the Peak, Housing Policy Debate, 23:2, 328-349, DOI: 10.1080/10511482.2013.771788

Gotham, Kevin (1998) Race, Mortgage Lending and Loan Rejections in a U.S. City, Sociological Focus, 31:4, 391-405, DOI: 10.1080/00380237.1998.10571116

Hanson, R. A., Hawley, Z., Martin, H., & Liu, B. (2016). Discrimination in mortgage lending: Evidence from a correspondence experiment. Journal of Urban Economics, 92, 48–65.

Haughwout, Christopher Mayer, Joseph Tracy, Dwight M. Jaffee & Tomasz Piskorski (2009) Subprime Mortgage Pricing: The Impact of Race, Ethnicity, and Gender on the Cost of Borrowing

Haupert, Tyler (2019) Racial Patterns in Mortgage Lending Outcomes During and After the Subprime Boom, Housing Policy Debate, 29:6, 947-976, DOI: 10.1080/10511482.2019.1636286

Ladd, H. F. (1998). Evidence on discrimination in mortgage lending. Journal of Economics Perspectives, 12(2), 41–62. Retrieved from https://www.jstor.org/stable/2646961?seq=1#page_scan_tab_contents

Leven, Charles L. & Michael E. Sykuta (1994) "The Importance of Race in Home Mortgage Approvals." Urban Affairs Quarterly 29(3): 479-489.

Munnell, Alicia H., L. E. Brown, J. McEneaney and G.M.B. Tootel. 1992. Mortgage Lending in Boston: Interpreting HMDA Data. Boston: Boston Federal Reserve Board. Working Paper No. 92-97.

Rugh, J. S., Albright, L., & Massey, D. S. (2015). Race, space, and cumulative disadvantage: A case study of the subprime lending collapse. Social Problems, 62(2), 186–218.

Squires, Gregory D., William Velez & Karl E. Taeuber. 1991. "Insurance Redlining, Agency Location and the Process of Urban Disinvestment." Urban Affairs Quarterly 26(4): 567-588.