Financing the American Dream:

Examining Racial Inequality and The Mortgage Lending Industry

Mark Russeff, Renato Albolea, Shuai Ma

Introduction

Home ownership has long been an important part of the American dream. Owning a home provides psychological security, financial stability and it fosters a deeper connection to the community. Homeownership has shown to generate numerous positive effects including lower crime rates, increased civic engagement[1] and higher childhood educational attainment[2]. Nonetheless, for many this integral part of the American dream is still out of reach. According to the 1990 housing census[3] 64.2% of Americans owned a home, while only 43.4% of African Americans and 42.4% of Hispanics were homeowners. The divide is even more significant in Massachusetts where the total homeownership rate is 59.3%, while only 26.4% for African Americans and 18.7% for Hispanics. It is imperative that these home ownership disparities be examined so that policy makers and financial institutions can work together to bridge the gap. In most cases, access to credit is the most important factor influencing whether an individual can purchase a home. In fact, minorities have faced a long history of lending discrimination and limited access to credit. That was why, in 1976, congress enacted the Home Mortgage Disclosure Act (HMDA) in order to monitor minority and low-income individual's access to the mortgage market. The HMDA required that all financial institutions publicly disclose the data involved in mortgage lending decisions. Although the HMDA was an important piece of legislation it did fail to report several important decision factors such as loan-to-value and credit history. The HMDA data showed evidence of racial discrimination resulting in significantly higher denial rates for minorities than for whites. Lending institutions argued that that this data did not represent sufficient evidence for discrimination because it lacked many important financial factors. In 1990, responding to criticism of the HMDA data,

[1] DiPasquale, Denise, and Edward L. Glaeser. "Incentives and social capital: Are homeowners better citizens?." Journal of urban Economics 45.2 (1999): 354-384.

[2] Aaronson, Daniel. "A Note on the Benefits of Homeownership." Journal of Urban Economics 47.3 (2000): 356-369.

[3] U.S. Census Bureau, "Historical Census of Housing Ownership", 1990, https://www.census.gov/hhes/www/housing/census/historic/ownrate.html

the Federal Reserve Board of Boston requested that lending institutions provide additional information related to the mortgage lending process. The data provided by these lending institutions now contained additional information about the borrower's credit history as well as the loan to value ratio for their loan. This data will be the used for our analysis so that we can account for these additional factors and get a better idea of how race is related to the probability of loan approval. With our analysis we aim to answer the question: Do mortgage lending institutions discriminate against minorities when making lending decisions?

In order to address our research question, we estimated the parameters of a Logit and a Probit model to determine whether the probability of mortgage approval differs among racial groups in Boston. Based on the estimated parameter of our Logit model, we found that White applicants face roughly 2.4 times greater odds of approval than their Black and Hispanic counterparts. In addition, we were able to construct a set of 12 prototypical individuals and calculate their predicted probability of approval based on both models. These predicted probabilities showed that White individuals had a higher predicted probability of loan approval than their Black and Hispanic counterparts for every demographic combination. The results of our analysis suggest that there is racial discrimination among the mortgage lending decisions in our sample.

Data

The data for this study was provided by the Federal Reserve Board of Boston and contains a sample of 1,989 loan applications filed in 1990. The intention of this data was to monitor minorities access to the mortgage market. The variables being used for our analysis are: (1) The dependent variable "Approved Loan" – which is a binary variable representing the final decision of the loan application 1-approved or 0-denied; (2) "Meets credit guidelines" – a binary variable representing whether the applicant's credit history meets guidelines1-yes or 0-no; (3) "Loan to purchase price" – the ratio of loan amount to purchase price as a percentage; (4) "Other obligations"– the applicant's other financial obligations as a percent of total income; (5) "Married" – a binary variable that represents the applicant's marital status 1-married or 0-single; (6) "Black" – whether or not the applicant is black1-yes or 0-no; (7) "Hispanic" – whether or not the applicant is Hispanic1-yes or 0-no. (8) "White" – is the reference category for race.

Based on the preview of the raw data, the following steps were taken in order to clean the data and remove any unusual values. First, were noticed the presence of the "." symbol in "Married" and several "666" values in the binary variable "Meets credit guidelines"; these values were removed from the data set. Second, there were some values in "Loan to purchase price" greater than 100. The value greater than 100 in "Loan to purchase price" may indicate that the application was approved with more money than was required to purchase the property. This is not unusual because people often apply for extra money in order to refurbish or repair the property after purchase. However, this might also represent a typo or an error of unity (125 may represent 12.5%), given this uncertainty we limited the "Loan to purchase price" to values that were less than or equal to 100.

Table 1 represents the descriptive statistics for the full data set which contains 1,937 observations. We can observe that the average loan approval rate for the sample is 88% even though 91% of the sample meets the credit guidelines. This suggests that meeting the credit guidelines does not necessarily guarantee that the applicant will be approved for a loan. The average financial obligation rate is 32.37% and the average loan to purchase price is 76.08%, meaning the average down payment is 23.92%. Lastly, we can see that 66% of the applicants in this sample are married.

Table 1: Descriptive statistics - **All**

Statistic	Mean	St. Dev.	Min	Max
Approved	0.88	0.33	0	1
Meets credit guidelines	0.91	0.28	0	1
Other obligations (% of income)	32.37	8.25	0	95
Loan to purchase price (%)	76.08	16.76	2	100
Married	0.66	0.47	0	1

Table 2 presents the descriptive statistics for each race within the sample. The sample is comprised of 84.7% White, 9.9% Black and 5.4% Hispanic individuals. When we compare the average approval rate across race, we can see that the rate for Whites is 91% which is significantly higher than the average approval rate for both Blacks and Hispanics at 67% and 78% respectively. However, we can also see that among Whites 94% meet the credit guidelines while only 72% of Blacks and 87% of Hispanics meet the credit guidelines. Meeting the credit guidelines is likely the most important element in the credit decision, this may explain the lower approval rate for minorities. The average financial obligations across all the races is very similar, the largest difference is only 3.04 percentage points between Black and White. Lastly, the data shows us the average loan to purchase price for White applicants is 74.78% which is considerably lower than the averages for Blacks and Hispanic at 82.89% and 83.91% respectively. On average Whites are putting more money down than minorities when applying for a mortgage. Another interesting observation of the loan to purchase price is that the min value for White applicants is 2%, corresponding to a 98% down payment which is much lower than the min loan to purchase price for minorities.

Table 2: Descriptive statistics - **Race**

Statistic	Mean	St. Dev.	Min	Max
White (n = 1641)				
Approved	0.91	0.29	0	1
Meets credit guidelines	0.94	0.24	0	1
Other obligations (% of income)	31.99	8.18	0	95
Loan to purchase price (%)	74.78	17.17	2	100
Married	0.66	0.47	0	1
Black (n = 192)				
Approved	0.67	0.47	0	1
Meets credit guidelines	0.72	0.45	0	1
Other obligations (% of income)	35.03	8.13	5.60	63
Loan to purchase price (%)	82.89	12.63	28.99	100
Married	0.61	0.49	0	1
Hispanic (n = 104)				
Approved	0.78	0.42	0	1
Meets credit guidelines	0.87	0.34	0	1
Other obligations (% of income)	33.32	8.56	14.60	62
Loan to purchase price (%)	83.91	10.92	40	100
Married	0.71	0.46	0	1

Modeling Approach

Since the mortgage decisions are only limited to “Approved” or “Denied”, which is “1” or “0” when translating into mathematical language, we chose to construct two multiple regression analyses using Probit and Logit Models. The parameters of the model were estimated using the maximal likelihood estimation method. To address our research question of whether mortgage lending institutions discriminate against minorities when making lending decisions, the variables of the most interest were included as the categorical variables of “Black” and “Hispanic” with “White” as the reference category. The other factors influencing mortgage lending decisions were “Meet credit guideline”, “Other obligations”, “Loan to purchase price”, and “Marriage”. Marital status was included because most married couples have two sources of income which can mitigate the default risk and indirectly lead to a higher probability of loan approval. Although gender could affect lending decisions, we excluded gender in this study because the sample size of female applicants was very small. Especially for female Hispanic applicants, the applications from this group only accounts for 1% of all applications.

Probit Model

$$P(y=1|x) = G(\beta_0+\beta_1 MeetCreditGuidelines+\beta_2 OtherObligations+\beta_3 LoanToPurchasePrice+ \beta_4 Married+\beta_5 Black+\beta_6 Hispanic+\mu)$$

Logit Model

$$ln(\frac{P}{1-P}) = \beta_0+\beta_1 MeetCreditGuidelines+\beta_2 OtherObligations+\beta_3 LoanToPurchasePrice+ \beta_4 Married+\beta_5 Black+\beta_6 Hispanic+\mu$$

Empirical Results

Table 3 presents the parameter estimates for our Logit Model and the corresponding odds ratios. Based on these estimates we can observe that the coefficients of “Meets credit guidelines” and “Married” are both positive as was expected. The coefficient of 3.766 for “Meets credit guidelines” is statistically significant at the 1% level and corresponds to an odds ratio of 43.227. This indicates that, controlling for all other variables, an applicant who meets the credit guidelines faces, roughly, 43.227 better odds of approval than someone who does not meet the guidelines. This demonstrates that meeting the credit guidelines is the most important factor for improving your odds of approval in our model. The coefficient of

0.482 for "Married" is statistically significant at the 1% level and corresponds to an odds ratio of 1.619. This indicates that, controlling for all other variables, an applicant who is married faces roughly 1.619 greater odds of approval than someone who is single. This intuitively makes sense; a married individual will likely be more financial stable and would likely be two incomes contributing to the paying of the mortgage. Based on the estimate we can also observe that the coefficients of "Loan to purchase price" and "Other obligations" are both negative as was expected. The coefficient of -0.016 for "Loan to purchase price" is statistically significant at the 5% level and corresponds to an odds ratio of 0.984. This indicates that, controlling for all other variables, an applicant with a 1 percentage point higher loan to purchase price ratio will face roughly 1.6% lower odds of approval than a similar applicant. Loans with a high loan to purchase price indicate a lower down payment and a higher risk of default for the lender, so it makes sense that as this percentage gets higher the odds of approval are diminished. For the variable "Other obligations" we have a coefficient of -0.034 which is statistically significant at the 1% level and corresponds to an odds ratio of 0.967. This indicates that, controlling for all other variables, an applicant with other obligations 1 percentage point higher than another applicant will face roughly 3.3% lower odds of approval. Similar to the case with high loan to purchase ratio, a high percentage of other obligations indicates a borrower who may be a higher risk for the lender. When we examine the estimated parameter for the race variables, we can observe that both "Black" and "Hispanic" are negative. The coefficient of -0.869 for "Black" corresponds to an odds ratio of 0.419 and the coefficient of -0.860 for "Hispanic" corresponds to an odds ratio of 0.423. The odds ratios associated with those parameters indicates that a White applicant faces roughly 2.39 greater odds of approval than a Black applicant and roughly 2.36 greater odds of approval than a Hispanic applicant. This finding seems to indicate that there is evidence of racial discrimination in the lending market. Additionally, we can observe that both Blacks and Hispanics have similar odds of approval in relation to Whites, suggesting that that the racial discrimination is not specific to one minority group. Also, the size of the odds ratio indicates that this finding is economically significant.

Table 4 presents the predicted probabilities for 12 prototypical individuals that were estimated based on our Logit model. For the sake of comparison, the prototypical values for "Other obligations" and "Loan to purchase price" were set to their sample means of 32.37% and

76.08%. When we examine the predicted probability of prototypical individuals across race, we can see that there is evidence of a racial disparity. For prototypical individuals who meet the credit guidelines we can observe a roughly 5 percentage point difference between the predicted probability of approval for married Whites over married Blacks and Hispanics. However, the largest difference is for prototypical individuals who are married and do not meet the credit guidelines. In that case, the predicted probability of approval for a white prototype is 35.6%, but for a Black prototype, in the same situation, we observe a predicted probability of only 18.8% which results in roughly a 16.8 percentage point difference. This result demonstrates a very strong case for racial discrimination within our sample.

Table 3: Estimated Logit Model

	Dependent variable:	
	Approved Loan	
	Estimate	Odds Ratio
Meets credit guidelines	3.766***	43.227
	(0.221)	
Other obligations (% of income)	−0.034***	0.967
	(0.011)	
Loan to purchase price (%)	−0.016**	0.984
	(0.007)	
Married	0.482***	1.619
	(0.185)	
Black	−0.869***	0.419
	(0.243)	
Hispanic	−0.860***	0.423
	(0.323)	
Constant	1.233*	3.431
	(0.685)	
Observations	1,937	
Log Likelihood	−462.546	
Akaike Inf. Crit.	939.091	

Notes: *p<0.1; **p<0.05; ***p<0.01

Reference category: white

Standar Errors in parenthesis

Table 4: Predicted Probability of Loan Approval - **Logit**

	White	Black	Hispanic
Meets credit guidelines, Married	96.0%	90.9%	91.0%
Meets credit guidelines, Single	93.7%	86.1%	86.2%
Does not meet credit guidelines, Married	35.6%	18.8%	19.0%
Does not meet credit guidelines, Single	25.5%	12.5%	12.6%

Notes: Other obligations set at sample mean of 32.37%

Loan to purchase price set at sample mean of 76.08%

Table 5 presents the parameter estimates for our Probit Model. The interpretation of the estimated coefficients in the Probit model are not as straightforward as with a linear or Logit regression. So, for the sake of this analysis we will just be examining the sign of the coefficients to determine if they align with the Logit model in Table 3. We can observe that the "Meets credit guidelines" and "Married" coefficients are both positive while the "Other obligations" and "Loan to purchase price" coefficients are both negative. The signs align with our intuition and the results we observed using the Logit model in Table 3. For the variables of interest, "Black" and "Hispanic", both coefficients are negative. Once again, the signs align with what we have previously observed in Table 3.

Table 6 presents the predicted probabilities for prototypical individuals estimated based on our Probit model. For the sake of comparison, the prototypical values for "Other obligations" and "Loan to purchase price" were set to their sample means of 32.37% and 76.08%. We can observe that married white individuals who meet the credit guidelines have a predicted probability of approval of 96.1%. Whereas married Black and Hispanic prototypical individuals who meet the credit guidelines have a predicted probability of approval of 90.4% and 90.5% respectively. The gap between White applicants and their Black or Hispanic counterparts is roughly 5.5 percentage points. The lowest predicted probability of approval is observed among individuals that are single and do not meet the credit guidelines. The predicted probability of approval for these individuals is 25.8% for White applicants compared to 13.5% and 13.7% for Black and Hispanic applicants respectively. The results for the predicted probabilities based on the Probit model align with what we observed with the Logit model.

Table 5: Estimated Probit Model

Race	*Dependent variable:*
	Approved Loan
Meets credit guidelines	2.169***
	(0.123)
Other obligations (% of income)	−0.016***
	(0.005)
Loan to purchase price (%)	−0.007**
	(0.003)
Married	0.239***
	(0.092)
Black	−0.450***
	(0.128)
Hispanic	−0.444***
	(0.169)
Constant	0.435
	(0.337)
Observations	1,937
Log Likelihood	−462.491
Akaike Inf. Crit.	938.982
Notes:	*p<0.1; **p<0.05; ***p<0.01
Reference category: white	
Standar Errors in parenthesis	

Table 6: Predicted Probability of Loan Approval - **Probit**

	White	Black	Hispanic
Meets credit guidelines, Married	96.1%	90.4%	90.5%
Meets credit guidelines, Single	93.6%	85.7%	85.9%
Does not meet credit guidelines, Married	34.0%	19.4%	19.6%
Does not meet credit guidelines, Single	25.8%	13.5%	13.7%

Notes: Other obligations set at sample mean of 32.37%

Loan to purchase price set at sample mean of 76.08%

Conclusion

Table 5 presents the parameter estimates for our Probit Model. The interpretation of the estimated coefficients in the Probit model are not as straightforward as with a linear or Logit regression. So, for the sake of this analysis we will just be examining the sign of the coefficients to determine if they align with the Logit model in Table 3. We can observe that the "Meets credit guidelines" and "Married" coefficients are both positive while the "Other obligations" and "Loan to purchase price" coefficients are both negative. The signs align with our intuition and the results we observed using the Logit model in Table 3. For the variables of interest, "Black" and "Hispanic", both coefficients are negative. Once again, the signs align with what we have previously observed in Table 3.

Table 6 presents the predicted probabilities for prototypical individuals estimated based on our Probit model. For the sake of comparison, the prototypical values for "Other obligations" and "Loan to purchase price" were set to their sample means of 32.37% and 76.08%. We can observe that married white individuals who meet the credit guidelines have a predicted probability of approval of 96.1%. Whereas married Black and Hispanic prototypical individuals who meet the credit guidelines have a predicted probability of approval of 90.4% and 90.5% respectively. The gap between White applicants and their Black or Hispanic counterparts is roughly 5.5 percentage points. The lowest predicted probability of approval is observed among individuals that are single and do not meet the credit guidelines. The predicted probability of approval for these individuals is 25.8% for White applicants compared to 13.5% and 13.7% for Black and Hispanic applicants respectively. The results for the predicted probabilities based on the Probit model align with what we observed with the Logit model.

Contributions

All of us made similar contributions to the project across many areas and worked together on all elements. However, we each individually led for certain sections: Renato for programming/model specification, Shuai for EDA and Mark for final writing.