Mortgage Lending in Boston: Interpreting HMDA Data

Barkha Sharma
Sri Lakshmi Mallipudi
Amrapali Samanta
Jiayang Wu

Introduction:

There have been large racial/ethnic gaps that have formed in decades and it's worth mentioning it also has its effect on minority groups having less approval rate when applying for a mortgage than Whites. An official report from the National Association of Realtors released a fact about the disparity in homeownership among different race groups: Black and Hispanic Americans have much lower homeownership rates than whites (Luthi, 2022). While the mortgage is invented for those who need financial support to purchase assets and can repay the loan with interest on time, bias appeared in the mortgage market which is about colored people have less capability to afford loan expenditure and it leads to a stereotype that color people tend to get a lower average mortgage approval rate than whites. In the real world situation, mortgage lenders have the financial guidelines when assessing a borrower's qualification: expense/income ratio, loan-to-value ratio, and credit history (Day & Liebowitz, 1998).

In this paper, several additional factors were considered which might influence the result of whether a mortgage will be approved, such as marital status, gender, other obligations, race/ethnicity. Data involved in this study comes from lending institutions in the Boston area, and the sample size under observation is 1937. In this research paper, through logistic and probit regression analysis, we determine the relationship between race/ethnicity and mortgage lending decisions and we come to the conclusion that mortgage lending institutions are discriminating against minorities by analyzing the likelihood of loan approvals. This outcome might help to assess the effectiveness of policy interventions aimed at increasing loan approval rates for certain groups of people, especially minorities or low-income individuals.

Research question: **Controlling for relevant characteristics, is race/ethnicity associated with the outcome of a mortgage loan application?**

Econometric Model and Estimation Method:

For an **economic model** the loan approval can be derived as a function of various factors as below:
Logit:
Log(Odds(Approved)) = F (Married, CreditHistory, OtherObligations, Black, Hispanic, Loanamount)

Probit:
Pr(Approved=1) = F (Married, CreditHistory, OtherObligations, Black, Hispanic, Loanamount)

The below specific **econometric model** is considered based on different parameters determined by the economic theory and data considerations, with ‘ ε’ representing unobserved factors.

Logit:
$$\text{Log(Odds(Approved))} = \beta 0 + \beta 1\text{Married} + \beta 2\text{CreditHistory} + \beta 3\text{OtherObligations} + \beta 4\text{Black} + \beta 5\text{Hispanic} + \beta 6\text{Loanamount} + \varepsilon$$

Probit: $$\Pr(\text{Approved}=1) = \Phi(\beta 0 + \beta 1\text{Married} + \beta 2\text{CreditHistory} + \beta 3\text{OtherObligations} + \beta 4\text{Black} + \beta 5\text{Hispanic} + \beta 6\text{Loanamount} + \varepsilon)$$

Approved is a binary variable that takes the value of 1 if the customer mortgage application is approved and 0 otherwise, and f() is a function that represents the relationship between the independent variables and the probability of loan approval. The econometric logit and probit models expressed above are binary response models used to analyze the relationship between loan approval (a binary outcome) and several predictor variables. In this case, the predictor variables are marital status (Married), credit history (Credit history meets guidelines), other obligations (as % of total income), loan-to-value, and race/ethnicity (Black and Hispanic). The logit model uses a logistic function to estimate the probability of loan approval, while the probit model uses a standard normal cumulative distribution function. These models help us predict the likelihood of loan approval for different prototypical individuals across different races/ethnicity. The odds ratios of the predictor variables in the logit model represent the change in the odds of loan approval associated with a one-unit change in the predictor variable for continuous variables and for categorical variables, the comparison is between the group specified and the reference category, holding all other variables constant, whereas the marginal probability of the predictor variables in the probit model represents the change in the probability of the loan approval when all the predictor variables are set at their means.

Data:

The data is collected from Boston's mortgage lending institutions and HMDA, 1990, for all applications by Blacks and Hispanics and for a random sample of those by whites.

The number of observations under analysis is 1937.

Variable definitions:

Approved: a binary variable that takes the value of 1 if the loan application is approved and 0 otherwise.
Married: a binary variable that takes the value of 1 if the applicant is married and 0 otherwise.
Credit history: a binary variable that takes the value of 1 if the applicant's credit history meets the guidelines and 0 otherwise.
Other obligations: a continuous variable indicating the other obligations of applicants as % of their total income.
Male: a binary variable that takes the value of 1 if the applicant is male and 0 otherwise.
Loan amount: a continuous variable representing the loan amount as % of the mortgage value.
Black: a binary variable that takes the value of 1 if the applicant is black and 0 otherwise.
Hispanic: a binary variable that takes the value of 1 if the applicant is Hispanic and 0 otherwise.

Sample selection Criteria

From the initial exploratory data analysis of summary statistics, we noticed errors in observations of 'Gender,' 'Marital Status,' 'Credit history meets guidelines,' and 'Loan amount' variables in the data. As per the variable descriptions, gender =1 if male or =0 otherwise, but the gender column has character (.) in 15 observations (0.8%), marital status = 1 if married or =0 otherwise, but the marital status column has character (.) in 3 observations (0.2%), and guidelines =1 if credit history meets guidelines or =0 otherwise, but the data has 666 as the max value. The loan to value was presented as a proportion, but the data has 2.571 as the max value. We also observed no missing values in the data from the variable table. Hence, we have imposed sample selection criteria on the data, removed the errored observations, and restricted the loan amount to <=1 to perform data analysis.

Descriptive Statistics:

Summary statistics of overall data after imposing sample selection criteria

Number of Observations: 1937

Variables	Min	1st Quantile	Median	Mean	3rd Quantile	Max
Approved	0.00	1.00	1.00	0.88	1.00	1.00
Married	0.00	0.00	1.00	0.66	1.00	1.00
Credit history meets guidelines	0.00	1.00	1.00	0.91	1.00	1.00
Other Obligations	0.00	0.28	0.33	0.32	0.37	0.95
Male	0.00	1.00	1.00	0.81	1.00	1.00
Loan amount	0.02	0.70	0.80	0.76	0.90	1.00
Black	0.00	0.00	0.00	0.10	0.00	1.00
Hispanic	0.00	0.00	0.00	0.05	0.00	1.00

Summary Statistics by Race/Ethnicity after imposing sample selection criteria

Blacks

Number of Observations: 192

Variables	Min	1st Quantile	Median	Mean	3rd Quantile	Max
Approved	0.00	0.00	1.00	0.67	1.00	1.00
Married	0.00	0.00	1.00	0.61	1.00	1.00
Credit history meets guidelines	0.00	0.00	1.00	0.72	1.00	1.00
Other Obligations	0.06	0.31	0.35	0.35	0.39	0.63
Male	0.00	0.00	1.00	0.74	1.00	1.00
Loan amount	0.29	0.80	0.86	0.83	0.90	1.00

Hispanics

Number of Observations: 104

Variables	Min	1st Quantile	Median	Mean	3rd Quantile	Max
Approved	0.00	1.00	1.00	0.78	1.00	1.00
Married	0.00	0.00	1.00	0.71	1.00	1.00
Credit history meets guidelines	0.00	1.00	1.00	0.87	1.00	1.00
Other Obligations	0.15	0.29	0.33	0.33	0.38	0.62
Male	0.00	1.00	1.00	0.81	1.00	1.00
Loan amount	0.40	0.80	0.89	0.84	0.90	1.00

Whites

Number of Observations: 1641

Variables	Min	1st Quantile	Median	Mean	3rd Quantile	Max
Approved	0.00	1.00	1.00	0.91	1.00	1.00
Married	0.00	0.00	1.00	0.66	1.00	1.00
Credit history meets guidelines	0.00	1.00	1.00	0.94	1.00	1.00
Other Obligations	0.00	0.28	0.33	0.32	0.37	0.95
Male	0.00	1.00	1.00	0.82	1.00	1.00
Loan amount	0.02	0.68	0.80	0.75	0.89	1.00

From the above summary statistics by race/ethnicity, we observed that the average loan approval rate is 67% for Blacks, 78% for Hispanics, and 91% for whites. Though the average proportion of other obligations is approximately similar for all races, the approval rates for minorities are less than for whites.

About 94% of the white's credit history meets the guidelines set by the lending institutions. In contrast, only 72% of blacks and 87% of Hispanics' credit history meets the guidelines; this could be another reason for less loan approval rate for Blacks and Hispanics. Hence based on exploratory data analysis, mortgage lending institutions probably discriminate against minorities in mortgage lending decisions.

Correlation matrix:

Variables	Loan amt	Married	Credit history Guidelines	Obligations	Black	Hispanic	Male	Approved
Loan amount	1	-0.054	-0.135	0.209	0.135	0.111	-0.002	-0.141
Married	-0.054	1	0.023	-0.03	-0.034	0.027	0.346	0.066
Guidelines	-0.135	0.023	1	-0.154	-0.224	-0.041	-0.005	0.618
Obligations	0.209	-0.03	-0.154	1	0.107	0.027	-0.015	-0.169
Black	0.135	-0.034	-0.224	0.107	1	-0.079	-0.063	-0.215
Hispanic	0.111	0.027	-0.041	0.027	-0.079	1	-0.004	-0.073
Male	-0.002	0.346	-0.005	-0.015	-0.063	-0.004	1	0.017
Approved	-0.141	0.066	0.618	-0.169	-0.215	-0.073	0.017	1

From the above correlation matrix, we observed a negative relationship between Black/Hispanic and approved indicating that the loan approval rate is less for Blacks and Hispanics. The negative relationship between approved and loan amount indicates that as the loan to value increases approval rate decreases and vice versa. The positive correlation between married/male/guidelines indicates that the loan approval rate increases for persons who are male, married, and have a credit history that meets the guidelines set by the lending institutions.

Results:

Logit Model

Logit regression results (Unrestricted Model)

Dependent Variable: Mortgage Lending Decision (Loan approval response)

	Coefficients (β)		Std. Error of β	Odds Ratio (OR)	Std. error for OR
Intercept	1.247	.	0.701	3.48	2.439
Other Obligations (as % of total income)	-0.034	**	0.011	0.97	0.010
Credit history meeting guidelines	3.766	***	0.221	43.19	9.533
Loan amount (as % of mortgage value)	-1.589	*	0.701	0.20	0.143
Male	-0.023		0.239	0.98	0.233
Married	0.488	*	0.196	1.63	0.319
Black	-0.871	***	0.243	0.42	0.102
Hispanic	-0.861	**	0.323	0.42	0.137
Number of observations	1937				
Likelihood ratio chi-square (df)	-462.541 (df = 8)				
Pseudo R2	0.35				

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Likelihood ratio test for Male variable

LR statistic = 2(log-likelihood of the unrestricted model - log-likelihood of the restricted model)
= 2(-462.541 - (-462.546))
= **0.010** < 3.841 [Chi-square critical value at 5% (df = 1)]

Male has no statistical significance and as the LR statistic is less than the critical value, the independent variable, male is unimportant, hence dropping male variable from the model.

Logit regression results (Restricted Model - excluding male variable)

Dependent Variable: Mortgage Lending Decision (Loan approval response)				
	Coefficients (β)	**Std. Error of β**	**Odds Ratio (OR)**	**Std. error for OR**
Intercept	1.233 .	0.685	3.43	2.350
Other Obligations (as % of total income)	-0.034 **	0.011	0.97	0.010
Credit history meeting guidelines	3.766 ***	0.221	43.23	9.534
Loan amount (as % of mortgage value)	-1.590 *	0.701	0.20	0.143
Married	0.482 **	0.185	1.62	0.299
Black	-0.869 ***	0.243	0.42	0.102
Hispanic	-0.860 **	0.323	0.42	0.137
Number of observations	1937			
Likelihood ratio chi-square (df)	-462.546 (df = 7)			
Percent correctly predicted (Accuracy)	0.927			
Pseudo R2	0.35			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Interpretations:

Odds of loan approval is 3.43 when all independent variables, i.e., Other obligations, Credit history meets guidelines, Married, Black, and Hispanic, are equal to zero. In this model, the odds ratio of intercept is not meaningful, and also the result is not statistically significant at the 5% level. A one-percent increase in other obligations (as % of total income) decreases the odds of loan approval by approx. 3%, holding all other variables constant and the result is statistically significant at a 1% level. A credit history that meets guidelines had 43.23 times the odds of loan approval as to not having a credit history that meets guidelines, holding all other variables constant and the result is highly statistically significant at a 0.1% level. A one-percent increase in loan amount (as % of property value) decreases the odds of loan approval by approx. 80%, holding all other variables constant and the result is statistically significant at 5% level. Married people had 62% higher odds of loan approval than their unmarried counterparts, holding all other variables constant and the result is statistically significant at a 1% level. Blacks had 58% lower odds of loan approval than whites, holding all other variables constant and the result is highly statistically significant at a 0.1% level. Hispanics had 58% lower odds of loan approval than whites, holding all other variables constant and the result is statistically significant at a 1% level.

Probit Model

Probit regression results (Restricted Model - excluding male variable)

Dependent Variable: Mortgage Lending Decision (Loan approval response)				
	Coefficients (β)	**Std. Error of β**	**Marginal Probability**	**Std. Error of marg. prob.**
Intercept	0.435	0.337	NA	
Other Obligations (as % of total income)	-0.016 **	0.005	-0.002	0.001
Credit history meeting guidelines	2.169 ***	0.123	0.316	0.026
Loan amount (as % of mortgage value)	-0.739 *	0.328	-0.108	0.047
Married	0.239 **	0.092	0.035	0.013
Black	-0.450 ***	0.128	-0.066	0.019
Hispanic	-0.444 **	0.169	-0.065	0.025
Number of observations	1937			
Likelihood ratio chi-square (df)	-462.491 (df = 7)			
Percent correctly predicted (Accuracy)	0.927			
Pseudo R2	0.35			

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Interpretations:

The probability of loan approval decreases by approx. 0.2% for every one-percent increase in other obligations (as % of total income), when all variables are set at their means and the result is statistically significant at a 1% level. The probability of loan approval was approx. 31.6% higher for a person having a credit history that meets guidelines than not having a credit history that meets guidelines, when all variables are set at their means and the result is highly statistically significant at a 0.1% level. The probability of loan approval decreases by approx. 10.8% for every one-percent increase in loan to mortgage value, when all variables are set at their means and the result is statistically significant at a 5% level. The probability of loan approval was approx. 3.5% higher for married than unmarried people, when all variables are set at their means and the result is statistically significant at a 1% level. The probability of loan approval was approx. 6.6% lower for Blacks than Whites, when all variables are set at their means and the result is highly statistically significant at a 0.1% level. The probability of loan approval was approx.

6.5% lower for Hispanics than Whites, when all variables are set at their means and the result is statistically significant at a 1% level.

<u>Predicting probabilities of loan approval for prototypical individuals from Logit and Probit models</u>

Prototypes	Loan amount	Other Obligations	Credit history meeting guidelines	Married	Predicted Probability (Logit)	Predicted Probability (Probit)
Blacks						
Prototype - 1	0.76	32.37	1	1	0.91	0.90
Prototype - 2	0.76	32.37	1	0	0.86	0.86
Prototype - 3	0.76	32.37	0	1	0.19	0.19
Prototype - 4	0.76	32.37	0	0	0.13	0.14
Hispanics						
Prototype - 1	0.76	32.37	1	1	0.91	0.91
Prototype - 2	0.76	32.37	1	0	0.86	0.86
Prototype - 3	0.76	32.37	0	1	0.19	0.20
Prototype - 4	0.76	32.37	0	0	0.13	0.14
Whites						
Prototype - 1	0.76	32.37	1	1	**0.96**	**0.96**
Prototype - 2	0.76	32.37	1	0	**0.94**	**0.94**
Prototype - 3	0.76	32.37	0	1	**0.36**	**0.34**
Prototype - 4	0.76	32.37	0	0	**0.25**	**0.26**

We created four prototypes for each race group individuals (i.e., Blacks, Hispanics, and Whites) by setting the continuous variables: other obligations and loan amount at their mean and different combinations of categorical variables: ‘married & having a credit history that meets guidelines’, ‘unmarried & having a credit history that meets guidelines’, ‘married & not having a credit history that meets guidelines’, and ‘unmarried & not having a credit history that meets guidelines’. From the resulting probabilities, we observed that the combination of a credit history meeting the guidelines and being married has a higher chance of loan approval for all race groups. Whites have a higher logit probability(log(odds)) and higher probit probability of loan approval for every specified combination of individuals than Blacks and Hispanics.

Model comparison

Both logit restricted and probit models have the same signs for all the coefficients. Except for the intercept, all the other coefficients have statistical significance at a 5% level in both models. However, the coefficient of intercept has statistical significance at a 10% level in the logit model. The model's Goodness of Fit measures like Pseudo R2 and Percent correctly predicted are similar for both models. The predicted probabilities of loan approval for all the prototypical individuals are similar in both models, with a higher approval chance for whites, indicating discrimination against minorities concerning mortgage lending decisions.

Conclusion:

Both the logit and probit models provide similar regression results, hence one model cannot be labeled better over another. However, we choose the logit model as it is easier to interpret the coefficients and it is widely used in empirical research, which means there are more established methods for interpreting its coefficients and assessing goodness of fit.

From the results, we see the logit model shows that both Black and Hispanic individuals have significantly lower odds of loan approval than whites. For additional evidence, logit model results also show that the predicted probability of loan approval is significantly lower for Black and Hispanic individuals compared to whites.
The probit model is yielding similar results showing the probability of loan approvals is less for Blacks and Hispanics than Whites. In terms of policy implications, the choice between the logit and probit models affects the way policymakers interpret and act on the results. With both substantive and statistical significance of predictors, both models provide evidence to conclude that mortgage lending institutions do discriminate against minorities. It may be worthwhile to investigate potential policies such as if there are any possible high-interest rates or stricter rules that imply on minorities which lead to the issue of discrimination in mortgage lending. Addressing potential policy solutions can promote greater access to credit for all individuals.

The conclusions above are subject to some limitations. We cannot generalize these estimated results for cities other than Boston. There could still be some other variables in the error terms that influence the dependent variable and correlate to other independent variables leading to omitted variable bias. For example, there could be some cultural barriers or lack of familiarity with the lending process amongst minorities, or there could also be some economic conditions at the time when data has been collected, also to note is that Boston is a city with high living cost. In research, we lack data on the incomes of minorities which is an important factor, and also mortgage lending decisions can differ based on neighborhood locations. In future research on this topic, a larger sample size of minority people which covers more geographical areas is also expected and it would be interesting to further analyze the possible causes behind the obstacles that minority borrowers might be facing.

Contributions:

Worked as a team on establishing the logit and probit models with selected variables of interest, sample selection criteria, regression model analysis, and results aligning with the research question.
Specific Contributions are as follows:
Barkha Sharma worked on the conclusion of the regression results and model comparison.
Sri Lakshmi Mallipudi worked on the model interpretations, established the probabilities of the results, and proofread the final paper.
Jiayang Wu worked on the effective literature review focusing on the research question for the Introduction and worked on the limitations section for the final paper.
Amrapali Samanta ensured the regression analysis conformity with the research topic, worked on the Econometric model estimation and Data sections, and proofread the final paper.

References:

Luthi, Ben. "Mortgage Lenders Expanding Loan Programs for Minorities". Investopedia. Sep. 2022. https://www.investopedia.com/mortgage-lenders-expanding-loan-programs-for-minorities-6560676

Day, Theodore E., and S. J. Liebowitz. "Mortgage Lending to Minorities: Where's the Bias?" Economic Inquiry, vol. 36, no. 1, Jan. 1998, p. 3. EBSCOhost, https://doi-org.proxy.seattleu.edu/10.1111/j.1465-7295.1998.tb01692.x.

Word count: Excluding tables and references 2484