MORTGAGE LENDING DECISIONS

Exploratory Data Analysis, Logit Estimates

Date of Submission: 5th March 2023

Team: Shanmathi Chandaran | Pankhuri Taneja | Su Ting Chin | Rabiya Fatima

Introduction

The purpose of the Mortgage Disclosure Act (MDA), commonly known as the Truth in Lending Act (TILA), is to enhance transparency and accuracy in mortgage lending by forcing lenders to disclose the principal terms and expenses of a mortgage loan to the borrower. The law strives to empower borrowers by ensuring they have access to information about the complete cost of a mortgage, including interest rates, fees, and any other possible costs. By mandating lenders to give borrowers with clear and comprehensible information, the MDA intends to prohibit misleading or deceptive practices in the mortgage business, promote fair lending practices, and assist consumers in making educated mortgage decisions. Despite the government's assertions that its loans are nondiscriminatory, there is evidence to imply that racial bias may still be present. The availability of data has allowed for a more in-depth examination of racial disparities in mortgage lending, and studies have revealed that black and Hispanic borrowers are frequently charged higher interest rates and fees than their white counterparts, even when credit scores and other relevant factors are controlled for. Racial discrimination persists in mortgage approval, with minority mortgage applicants facing challenges such as poorer credit scores, more debt, and lower algorithmic approval rates from race-blind government automated underwriting systems (AUS) compared to their white counterparts. (Bhutta, Hizmo, & Ringo, 2022)

To determine if MDA exists discrimination, we must first collect data. The collected data revealed that blacks were more than twice as likely to be denied a mortgage as whites. Nevertheless, characteristics linked to race and creditworthiness are omitted from this data, therefore no inferences can be reached regarding the impact of race in mortgage lending. Given the modest number of mortgage applications filed by minorities, we used a collection of Black and Hispanic

applications for blacks and Hispanics as well as a random sample of whites to determine whether HMDA assessed mortgage applicants based on their race. Inquire as to whether there are any further variables that can influence the decision.

Econometric Model(s) and Estimation Methods

There are several theoretical debates on mortgage loan decisions, particularly on the role of discrimination and bias. The article by Munnell et al. (1996) titled "Mortgage lending in Boston: Interpreting HMDA data" argues that there was racial discrimination in mortgage lending decisions in Boston during the 1980s. The authors used data from the Home Mortgage Disclosure Act (HMDA) to examine mortgage lending patterns in Boston and found that minority applicants were less likely to be approved for mortgage loans than white applicants with similar financial characteristics.

In our analysis of loan approval, we are controlling for variables such as credit history, loan amount percentage, debt-to-income ratio, marital status, race, and gender. The sample size after cleaning and removing missing values from the original 1989 observations was reduced to 1936, including minority race groups such as Black and Hispanic. We removed irrelevant data values in credit history guidelines, married, male, and total income percentage, and considered only loan amount values less than or equal to 100%.

We used the Logit and Probit models to analyze a range of independent variables as we have a binary dependent variable and found that having a compliant credit history is a significant predictor of loan approval while being black or Hispanic decreases the likelihood of approval. Both the logistic and probit models have similar results. However, we preferred the Logit model in this case as it is generally preferred where there is a high degree of imbalance in the data, which is often the case in mortgage loan approval data where most applicants are approved. The log likelihood values

for both the logistic and probit regression models were computed and compared, and we found that the logistic regression model had a slightly higher log likelihood.

The logistic regression model estimates the log odds of loan approval as a function of several variables including credit history, loan percentage, other obligations, marital status, race (black and Hispanic), and gender (male). Given the importance of accurately modelling mortgage loan approval in economic research, logit is a suitable choice for this analysis due to its interpretability and ability to handle rare outcomes and imbalanced data.

Description of variables

TABLE 1 Variable Descriptions	
Loan Approval	= 1 if an applicant's mortgage loan was approved, = 0 otherwise
Complaint Credit History	= 1 if an applicant's credit history meets guidelines, = 0 otherwise
Loan Amount Percent	the percentage of loan amount/purchase price approved for loan
Obligation Percent Total Income	the percentage of an applicant's income that is dedicated to other obligations
Male	= 1 if an applicant is male, = 0 otherwise
Married	= 1 if an applicant is married, = 0 otherwise
Black	= 1 if an applicant is black, = 0 otherwise
Hispanic	= 1 if an applicant is Hispanic, = 0 otherwise

Overall, EDA Summary

TABLE 2 Summary Statistics				
Variable	Mean	Standard Deviation	Minimum	Maximum
Married	0.66	0.47	0	1
Complaint Credit History	0.91	0.28	0	1
Obligation Percent Total Income	32.38	8.22	4	95

Black	0.1	0.3	0	1
Hispanic	0.05	0.23	0	1
Male	0.81	0.39	0	1
Loan Approval	0.88	0.33	0	1
Loan Amount Percent	0.76	0.17	0.02	1
Non-Hispanic White	0.85	0.36	0	1
Non-Hispanic Black	0.1	0.3	0	1
Number Of Observations = 1,936				

Data

Our study utilizes data from the American Economic Review 86, published in 1996 by Alicia H. Munnell, Geoffrey M.B. Tootell, Lynn E. Browne, and James McEneaney, to analyze mortgage lending decisions based on race and credit worthiness. The study's EDA analysis included a summary of loan applicant demographics, such as marital status, credit history, race, and gender. Table 1 shows the descriptive statistics of each variable. It showed that most applicants were married, had compliant credit history, and were male. The study also provided summary statistics for non-Hispanic white, non-Hispanic black, and Hispanic loan applicants, which showed that most applicants were non-Hispanic white.

The table shows that out of 1936 loan applicants, 85% were non-Hispanic white, 10% were non-Hispanic black, and 5% were Hispanic. The mean and standard deviation for the variables of married, compliant credit history, male, and loan amount percentage are 0.66, 0.91, 0.81, and 0.76 respectively. The mean other obligation percentage for loan applicants is 32.38, with a standard deviation of 8.22.

The study's analysis suggests that race plays a significant role in mortgage lending decisions, even after controlling for other variables such as credit history and other obligation ratio. Overall, the

EDA suggests that further analysis and modeling are needed to fully understand the relationships between the variables in the dataset and loan approval. However, the data highlights potential disparities in lending practices based on race/ethnicity that require attention and intervention.

Table 3 - Regression Results from Logit and Probit Estimates

TABLE 3 Regression Results. Dependent Variable: Loan Approval				
Independent Variables	(1) Logit	(2) Odds Ratio	(3) Probit	(4) Odds Ratio
Obligation Percent Total Income	-0.034***	0.967	-0.016***	0.984
	(0.011)	(1.011)	(0.005)	(1.005)
Loan Amount Percent	-1.589**	0.204	-0.739**	0.478
	(0.701)	(2.016)	(0.328)	(1.389)
Complaint Credit History	3.766***	43.221***	2.169***	8.751***
	(0.221)	(1.247)	(0.123)	(1.131)
Married	0.481***	1.618	0.239***	1.269
	(0.185)	(1.203)	(0.092)	(1.096)
Black	-0.869***	0.419	-0.450***	0.638
	(0.243)	(1.275)	(0.128)	(1.137)
Hispanic	-0.860***	0.423	-0.444***	0.641
	(0.323)	(1.382)	(0.169)	(1.184)
Constant	1.231*	3.425*	0.434	1.544
	(0.685)	(1.984)	(0.337)	(1.401)
Observations	1,936	1,936	1,936	1,936
R-squared	0.354	NA	0.354	NA
*Note: *p<0.1; **p<0.05; ***p<0.01*				

Note: The quantities in parentheses below the estimates are the standard errors.

From the logit regression, we could discover that there are six significant variables namely Obligation Percent Total Income, Loan Amount Percent, Complaint Credit History, Married, Black and Hispanic in our regression result table which are highly statistically significant at 1% level. .The intercept of 1.231 indicates that on average log odds of Loan Approval is 1.23 when all other predictor variables are equal to zero . Keeping all other predictor variables fixed, on

average a one-unit increase in Obligation Percent Total Income is associated with a decrease in log odds of Loan_Approval by 0.034 which implies that as the percentage of total income contributed by the applicant decreases, the likelihood of loan approval decreases as well. And a one-unit increase in Loan Amount Percent is associated with a decrease in log odds of Loan Approval by 1.5894, on average, holding all the other variables constant. This indicates that as the percentage of loan amount requested by the applicant increases, the likelihood of loan approval decreases. Having a Compliant Credit History is associated with an increase of 3.766 in the log odds of loan approval compared to not having a Compliant Credit History, on average while keeping all other predictor variables constant. Specifically, it means that the odds of loan approval on average are 42.19 times higher for applicants with a Compliant Credit History compared to those without one. Applicants who are married on average had 61.81% greater log chances of loan approval than single applicants. Loan Approval probabilities for Black applicants on average are 58.07% less likely than for non-Black applicants. The log probabilities of loan approval on average are 57.7% lower for Hispanic applicants than for non-Hispanic applicants. We considered potential issues of multicollinearity in our analysis, and our VIF values suggested that it is not a significant concern in our model as all the values at below 5. Our Pseudo R-squared value of 0.354 from the logit model indicates that our model explains about 35% of the variation in loan approval decisions. After analyzing the dataset, we have found that there is a positive relationship between a good credit history, being married, and loan approval, whereas there is a negative relationship between the obligation percent of total income, loan amount percent, and belonging to the Black or Hispanic community, and loan approval. From the Table 3 we can also see that the variables have same statistical significance and the magnitudes of Probit model are similar to Logit model and there is a very slight difference in the magnitudes.

Further, we conducted a Likelihood ratio test by considering various combinations of interaction variables and independent variables. Interestingly, we found that being male is associated with a slightly lower probability of loan approval, but this effect is not statistically significant. Also, the interaction between being black and male is also not a significant predictor of loan approval. Based on the results of the test, we have decided to omit the variables of Male and the interaction terms. We used the techniques of coefficient analysis and odds ratio to arrive at the results of the analysis from the same dataset.

Conclusion

To sum up, the logit regression model yielded significant predictors of loan approval, including Obligation Percentage of Total Income, Loan Amount Percent, Compliant Credit History, marital status, and race. Our analysis found that a decrease in Obligation Percentage of Total Income and an increase in Loan Amount Percent corresponded with lower odds of loan approval. The findings suggest that lenders should pay close attention to applicants' credit history, marital status, and loan request amounts when making lending decisions. Additionally, lenders should be mindful of potential biases in lending decisions based on race/ethnicity. These results highlight the prevalence of racial and socioeconomic disparities in the mortgage lending process.

We hope that our findings will contribute to conversations on fair lending practices and inspire additional study on this crucial topic. However, it is crucial to highlight that our dataset may not be typical of the entire population, as it only contains applications filed in the Boston area and may not apply to other places. In addition, the sample used for the analysis consists of only 1936 records, which is a small dataset. We require a larger dataset that can more properly and comprehensively support our results. Moreover, past success is not indicative of future performance. Other non-financial elements, such as societal trends, demographic shifts, and

political issues, may also play a considerable effect in the mortgage lending market and should be taken into account when evaluating any analysis' conclusions.