ECON 5300
Mortgage Lending Decisions Paper
Dr. Bridget Hiedemann
Ziyu Jin, Florence Jiahui Yan, Wei Li

1. **Introduction**

Mortgage is a main resource for families to purchase real properties, especially in hotspot cities where the housing market barely cools down. It is reasonable that financial institutions establish rules for applicants who apply for mortgage lending to ensure the normal operation in their system. However, according to “Mortgage Lending in Boston: Interpreting HMDA Data,” the data obtained by The Home Mortgage Disclosure Act shows that minorities are more than twice as likely to be denied a mortgage as whites. Yet variables correlated with both race and creditworthiness were omitted from these data (*American Economic Review 86, 25-53*). In order to make the conclusion about race’s role in mortgage lending, the Federal Reserve Board of Boston requested lending institutions in the Boston area provided additional information relevant to mortgage lending decisions in 1990. For this paper, we will use the data collected in 1990 to analyze race’s role in mortgage lending start with the research question as below:

Is race/ethnicity associated with the outcome of a mortgage loan application?

The purpose of this paper is to conclude whether mortgage lending institutions discriminate against minorities by estimating the probability of mortgage loan approval for applicants of different characteristics, including race/ethnicity.

Understanding race’s character in mortgage lending is important. For example, if there are two applicants come with different races but similar financial conditions, the racial discrimination exists in the process of mortgage lending could result into a disparate result. The situation can be worse when a majority is approved in mortgage loan application even there are borrowers from other races with better qualifications, such as healthier credit histories and less loan amounts. Such consequence will intensify the residential segregation in the society and racial injustice in the country. Thus, it is indispensable to analyze the data and generate the issue.

Since the dependent variable as whether the applicant was approved for the loan is binary, we choose to use the probit and logit estimation method on the selected sample. And the results show that marriage status and the qualification of credit history display positive relationships with whether the application get approval or not. Loan amount, other obligations in total income, Hispanic and Black demonstrate negative relationships with being approved. Thus, race/ethnicity is associated with the outcome of a mortgage loan application.

2. Econometric Models and Estimation Methods

Since the dependent variable in our analysis is whether the mortgage loan is approved or not, it is classified as a binary variable. And we are aiming to obtain the probability of such variable in order to compare the results across different groups. As a result, we decide to use probit and logistic models instead of a linear probability model to ensure that the estimated probabilities fall between 0 and 1.

For the probit mode, we include marriage status (yes/no), credit meet (yes/no), other obligations as a percentage in total income, non-Hispanic Black (yes/no), Hispanic (yes/no), loan as a percentage in the purchase price. In this case, the non-Hispanic White is the reference category.

The dependent variable is the z score in a Standard Normal Distribution (SND). If we use G to represent the cumulative density function of SND, then we have the form:

$$P(y = 1 \mid x) = G(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k) = G(z),$$

where $z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k$. Thus we can use the z score in the SND to calculate the probability of being approved.

The parameters in the logit model are the same as those in the probit model for we want to observe the robustness of our results. For the dependent variable, we take natural logarithm of the ratio between the probability of being approved and the probability of being denied. The estimation model takes the form:

$$\ln(p/(1-p)) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k = z,$$

where p stands for the probability of approval, and therefore the left side of the equation represents the log odds of approval. For each observation with a vector of x's in all independent variables, we can calculate the z value from the equation, then we get the probability of being approved by the formula: probability $= 1 / (1 + e^{-z})$.

3. Data

3.1 Data Sources

We use the data collected by the lending institutions in the Boston Area under The Home Mortgage Disclosure Act (HMDA). As the Federal Reserve Board of Boston requested, data of all applications by Blacks and Hispanics and a random sample of those by Whites are collected. This aims to monitor minority and low-income access to the mortgage market, in light of the relatively small number of mortgage loan applications made by minorities. So we hope a conclusion about the possibility of the minorities' access to the mortgage market can be drawn based on the HMDA data.

3.2 Variables and Sample Selection Criteria

Approve

The values in the "Approve" column are either 0 or 1, where 0 represents this individual is not approved for the mortgage loan and 1 is for the ones who are approved. We use this column as the dependent variable. In the models we applied, the estimated value of "Approved" will help to calculate the probability of being approved.

Marriage Status

This is a categorical variable indicating whether the person is / isn't in a marriage status. We exclude the wrong values in this column so that the rest values of this parameter is either married, marked by 1 or unmarried by 0.

Credit Meet

This is also a categorical variable indicating whether the person's credit history meets guidelines or not. The wrong values are also excluded. The rest part is either meet, marked by 1, or not meet, marked by 0.

Other Obligations as a Percentage of Total Income

This is a continuous variable, indicating the percentage of the individual's other obligations in the total income. And we limited this variable by setting a lower bound of 0 and a upper bound of 100.

Black and Hispanic

These are two categorical variables, indicating whether the individual is a Black / Hispanic, marked as 1; isn't Blake / Hispanic, marked as 0. There are only three races/ethnicities in this data set: White, Black, and Hispanic.

Percentage of Loan Amount in Purchase Price

This is a continuous variable, showing the percentage of an applicant's loan amount to the purchase price. Supposedly, the values in this variable should be between 0 to 100. So we exclude the records whose percentages are larger than 100.

3.3 Descriptive Statistics

3.3.1 Overall Data Distribution

We present the overall data distribution in the two quantitative columns in Table 1, one is "Other Obligations as a Percentage of Total Income" and the other is "Percentage of Loan Amount in Purchase Price". There are 1937 observations in the data set. The highest obligation percentage is 95 and the lowest is 0. Intuitively, we can assume that the lower the obligation percentage, the higher probability that the individual would be approved for the mortgage loan. The average obligation percentage is 32.37. The highest loan percentage is 100 and the lowest is 2. The lower the number is, the stronger ability the individual has to pay for the purchase. The average loan percentage is 76.08, which seems to be quite high.

Table 1: Overall Data Distribution

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Obliga_TotIncome	1,937	32.366	8.247	0.000	28.000	37.000	95.000
Loan_Purchase	1,937	76.075	16.763	2	69.7	89.8	100

3.3.2 Obligation Percentage in Total Income by Races

Here in Table 2 we present the data distribution of obligation percentages in total income by races. Firstly, the number of observations in non-Hispanic White, non-Hispanic Black, and Hispanic are 1641, 192, and 104. The mean of the obligation percentages in the three groups are very close at around 33. The variations of the data in these groups are also close to 8. White group has the widest range of the obligation percentages from 0 to 95, and Black group has the narrowest range from 6 to 63. Generally speaking, the distributions seems to be quite close among the three groups.

Table 2: Percentages of Obligation to Total Income Data Distribution

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Black	192	35	8	6	31	39	63
Hispanic	104	33	9	15	29	38	62
White	1,641	32	8	0	28	36	95

3.3.3 Loan Percentage in Purchase Price by Races

In Table 3, we present the data distribution of loan percentages in purchase price by races. Among the three groups, the White group has the lowest mean and the widest range from 2 to 100. Hispanic group has the highest mean and the narrowest range from 40 to 100. From this perspective, the White group possibly has the better chance of being approved for the mortgage loan.

Table 3: Percentages of Loan to Purchase Data Distribution

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Black	192	83	13	29	80	90	100
Hispanic	104	84	11	40	80	90	100
White	1,641	75	17	2	68	89	100

4. Result

4.1 Probit Model

4.1.1 Model Results

Table 4: Probit Model Results	
Dependent Variable: approval on mortgage loan	
Variable	Parameter Estimates (Std. Error in parentheses)
Intercept	0.435
	(0.337)
Married	0.239**
	(0.092)
Credit History Meets Guidelines	2.169***
	(0.123)
Loan Amount	-0.007*
	(0.003)
Other Obligations (percentage)	-0.016**
	(0.005)
Black	-0.450**
	(0.128)
Hispanic	-0.444**
	(0.169)
Log-Likelihood	-462.491
Sample Size	1937

.$p < 0.10$, *$p < 0.01$, **$p < 0.001$, ***$p < 0.0001$

A. Non-racial/ethnic characteristics

According to the results of our probit regression model in Table 4, mortgage loan applicants who are married is associated with a higher probability of approval on average, when all variables are set at their means; and so are applicants who have a credit history that meets the guidelines. However, applicants whose loan amount is higher or other obligations constitute larger part of their income are estimated to face a lower probability of approval on average, controlling other characteristics. These findings are not only statistically significant (at the confidence level of 0.01) but also reasonable, for that mortgage lending organizations are, under most of the cases, unwilling to give loans to individuals when they are borrowing a large amount of money and/or having too much other obligations in comparison to their income.

B. Racial/ethnic characteristics

Table 4 shows that on average, Black and Hispanic mortgage loan applicants are both estimated to be less likely to obtain an approval, relative to White applicant, holding other variables constant. And this result is highly statistically significant with a confidence level of 0.001. We will discuss more about the magnitude of the difference in probabilities of approval for people of different races/ethnicities in the following section.

4.1.2 Predict probability of approval for prototypical loan applicants

To examine the relation between race/ethnicity with the outcome of a mortgage loan application, we defined several prototypical applicants by setting different values for their race/ethnicity, marriage status and credit history condition, and predicted the probability of approval for each one. In this paper, we set the loan amount and percentage of other obligations in income at their means across the whole sample for each prototype to simplify the case. Table 5 below presents the results of our prediction.

Table 5: Predicted Probabilities for Probit model			
	Mortgage Loan Application Approval		
	White	Black	Hispanic
Married Meet guidelines	***0.961***	0.904	0.905
Unmarried Meet guidelines	***0.936***	0.857	0.859
Married Meet no guidelines	***0.340***	0.194	0.196
Unmarried Meet no guidelines	***0.258***	0.135	0.137

A. Marriage status and credit condition

By comparing each row in Table 5, we discovered that whether the applicant's credit history meets guidelines affect the probability of mortgage loan approval most significantly. The probabilities of approval dropped by approximately two thirds under the same marriage status across all races/ethnicities. Although being married improves the probability of approval as well for each race/ethnicity, holding credit condition the same, the magnitude is not as significant.

B. Race and ethnicity

If we compare each column of the results within same rows in Table 5, it is obvious that White applicants are predicted to have the highest probability of receiving a mortgage loan, controlling the other two characteristics. However, Black and Hispanic applicants are both predicted to face a much lower probability for approval relative to white applicants (6~14 percentage points less).

4.2 Logit Model

4.2.1 Model Results

Table 6: Logit Model Results

Dependent Variable: approval on mortgage loan

Variable	Parameter Estimates (Std. Error in parentheses)	Odds Ratio
Intercept	1.233.	3.431
	(0.685)	
Married	0.482**	1.619
	(0.185)	
Credit History Meets Guidelines	3.766***	43.227
	(0.221)	
Loan Amount	-0.016*	0.984
	(0.007)	
Other Obligations (percentage)	-0.034**	0.967
	(0.011)	
Black	-0.869***	0.419
	(0.243)	
Hispanic	-0.860**	0.423
	(0.323)	
Log-Likelihood	-462.546	
Sample Size	1937	

.p < 0.10, *p < 0.01, **p < 0.001, ***p < 0.0001

I. Parameter Estimates

Table 6 reports the results of our logistic model, in which the signs of coefficients and significance levels for each variable are similar to those in the probit model. In other words, marriage, credit history that meets guidelines, lower loan amount, and less other obligations are all estimated to be indicators of higher average probabilities of approval, when controlling other characteristics, whereas being Black or Hispanic is associated with the opposite. And the findings are also statistically significant across all variables.

II. Odds Ratios

A. Non-racial/ethnic characteristics

Table 6 suggests that, setting other variables at the same value, the odds for applicants who are married and associated with a credit history that meets guidelines to receive an approval for mortgage loan is 1.619 and 43.227 as large as those who are not married and not associated with a credit history that meets guidelines, respectively. In other words, the estimated odds are 61.9% and 4222.7% higher among applicants who are married and whose credit history meets the guidelines, respectively.

In addition, we estimate that the odds of approval are expected to decrease respectively by 16% (0.984-1) and 33% (0.967-1) for a one-unit increase in loan amount and the percentage of other obligations in income, when controlling other variables.

B. Racial/ethnic characteristics

Tables 6 also indicates that, when other variables are fixed, the ratio between the odds for a Black or Hispanic applicant to be approved for mortgage loan and that for a White applicant is 0.419; and ratio between the same thing for a Hispanic applicant and a White applicant is 0.423. Therefore, Black applicants and Hispanic applicants are estimated to have respectively 58.1% and 57.7% lower odds for being approved for a mortgage loan than White applicants.

4.2.2 Predict probability of approval for prototypical loan applicants

We use our logistic model to predict the probability of approval for the same prototypes as the probit model and yield the following results in Table 7.

Table7: Predicted Probabilities for Logit model

	Mortgage Loan Application Approval		
	White	Black	Hispanic
Married Meet guidelines	***0.960***	0.909	0.910
Unmarried Meet guidelines	***0.937***	0.861	0.862
Married Meet no guidelines	***0.356***	0.188	0.190
Unmarried Meet no guidelines	***0.255***	0.125	0.126

A. Marriage status and credit condition

As it turns out, the predicted probabilities using logic model are quite similar to the results in Table 5. Namely, applicants whose credit history that meets guidelines are estimated to have a much greater chance to be approved for mortgage loan, whereas married applicants are estimated to have greater chance as well, but not as much.

B. Race and ethnicity

The results are also alike for those in Table 5 in terms of racial/ethnic characteristics that White applicants are associated with higher probabilities (5~16 percentage points difference) of receiving a mortgage loan than Black and Hispanic applicants at the same marriage status and credit condition.

4.2.3 Robustness of results (comparing to probit model)

As we mentioned in the previous two paragraphs, the predicted probabilities for each prototype we defined are quite close in Table 5 and 7 (difference ≤ 0.01). In this case, we can conclude that our estimation results are robust across the probit and logistic models.

5. Conclusion

The analysis in this paper shows that people of color receive more rejections of home loans compare to white people. The effect remains statistically significant after controlling all other variables. According to this finding, we can conclude that race/ethnicity is associated with the outcome of a mortgage loan application.

Marriage status holds a slight positive relationship with whether the loan can be approved in both models. The strong magnitude remains in odds ratio. It is intuitive since married individuals often are older and have been in workplace relatively longer than unmarried people. Thus, they have a longer credit history and more savings in their accounts. These are all critical qualifications to get approved of home loans from financial institutions. However, marriage status is not a predetermined condition to get approved. In other words, getting married doesn't always conclude a better credit history and a healthier financial situation. Therefore, it is logical that marriage status holds little correlation the result of mortgage lending. The variable of whether credit history meets guidelines remains significant positive relationship with the dependent variable across both models. This result is logical as a better credit history record indicates that the individual is more likely to repay the loan on time and thus, it is easier for them to be approved. The variables of loan amount and the percentage of other obligations show negative sign in both models. This result is also expected as less loan amount and less obligation of total income increase the likelihood for the applicant to repay the loan. For variables related with race/ethnicity, people of colors remain negative signs with the dependent variable across two models. Specifically, in the table of the predict probability of approval for prototypical loan applicants for two models, white continues receiving the highest probability among all four types. Overall, we believe that race/ethnicity is associated with the outcome of a mortgage loan application and white people get more approval than people of color, which includes black and Hispanic in this data set.

Although this paper provides the impact on mortgage loan approval brought by race and ethnicity, the conclusion above is subject to several limitations. First, considering relatively small number of mortgage loan applications made by minorities, the data we obtained were collected for all applications by blacks and Hispanics and for a random sample of those by whites. Thus, it may not include enough samples to represent every minority in Boston area. Secondly, due to the nature of regression models, our model is not a perfect fit for the entire dataset, and therefore variance and error are predictable.

Contribution:

Florence: Data Preparation and Cleaning, EDA, Sample Criteria, Paper Writing (Introduction, Econometric Models, and Conclusion)
Wei: Data Preparation and Cleaning, Sample Selection, Models Testing and Detection, Paper Writing (Econometric Models, Estimation models, Data)
Ziyu: Data Preparation and Cleaning, Sample Selection, Model Specification and Diagnosis, Code Modification and Finalization, Paper Writing (Results and Conclusion), Paper Editing