IS RACE/ETHNICITY ASSOCIATED WITH THE OUTCOME OF MORTGAGE LOAN APPLICATION?

Group 4: Sourabh Gupta, Huy Le, Han Li

1. INTRODUCTION

The ability to get a mortgage is often the key to an individual's ability to purchase a home. The United States has enacted a variety of laws making it illegal for lenders to discriminate against members of historically disadvantaged groups, particularly minorities. In this report, we are addressing the research question "*Controlling for relevant characteristics, is race/ethnicity associated with the outcome of mortgage loan application?*" Over the last several years, race and ethnic composition of Boston has changed a lot. In 1980, more than two-third people were white, and this proportion is decreasing with time. But employment and education are increasing with time. Hence, for this assessment we will use HMDA data from 1996: Mortgage Lending in Boston, containing marital status, if credit history meets guidelines, gender, race(black/white/hispanic), loan amount requested per purchase price, other obligations as a percentage of total income and approval status of each application. Our resultant/dependent variable is Loan approval status (1 = approved, 0 = declined). Because our dependent variable is binary(can take only two values), we are using Probit and Logit model to estimate the probability that an applicant with particular characteristics will get loan approval or not. Probit model uses the cumulative distribution function (CDF) of the standard normal distribution and Logit model uses the cumulative distribution function (CDF) of logarithmic function. We are using Maximum likelihood estimation method to obtain the parameter estimates.

We concluded that, the predicted probability of loan approval is slightly higher if White person apply for the loan as compared to other race but we can not conclude that mortgage lending decisions are discriminated against minorities because there are many other factors which can affect the result are not included in model like educational attainment, salary, etc.

2. ECONOMETRIC MODEL AND ESTIMATION METHOD

In this project, to address the question we created models to estimate the impact of race/ethnicity on the approval of mortgage loan application. In this specification, our dependent variable - the outcome of mortgage loan application is binary and takes on the values either zero(denied) or one(approved). We used estimated probit and logit models to represent the impact of independent variables on the probability of approval of a mortgage loan application P(outcome = "approved").

The set independent variables in both models are race/ethnicity, marital status, credit history meets guideline, other obligations as a percent of total income, loan amount/purchase prices. We used race/ethnicity as we want to address the impact of race and ethnicity on the outcome of mortgage loan application. We added marital status which can indicate the potential ability to meet repayment on the loan of applicant, since many applicants are co-applicant with their spouse and when controlling other factors, the total earnings of a married applicant is higher than a single applicant. To make our result is more reliable, we also included credit history (guideline), other obligations and loan to value ratio because those factors are considered in making mortgage decisions.

In the probit model, the dependent variable is the probability of approving P(outcome = approved). We applied a probit model to address the relationship between the probability of mortgage loan approval and race/ethnicity. Probit models are nonlinear regressions where coefficient are fitted with the maximum likelihood to the following function:

$$P(y = 1 \| x) = G(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k) = G(z)$$

Where G is the cumulative distribution of the standard normal distribution, x_j is the independent variable, β_j is the coefficients to be estimated.

In the logit model, we chose to model the natural logarithm odds of approval. The estimated coefficient in logit model represents the change in the natural logarithm of the odds ratio of the outcome which associates with one unit change in the independent variables.

To estimate the binary dependent variable models, we applied the maximum likelihood estimation (MLE) method for both probit and logit models. This method obtains the parameter estimates by finding the parameter values that maximize the likelihood function.

3. DATA

3.1 Data source

For this project, we used data from Mortgage Lending in Boston: HMDA data. As per the data, 66% mortgage loan applications made by married applicants. We assume that married applicants have high chances to get approval. Meeting credit history guidelines is the most important factor in the loan approval decision. In this data, 91% of applications had credit history guidelines met. 90% of applicants in this data are White people and 81% of applicants are male. Another important criteria in loan approval decision is obligations as a percent of income. Higher the obligations, less the chances. In HMDA data, on average applicants have 32% of their income for other obligations.

3.2 Variables, measurement, and selection criteria.

In the original data, it contains 1989 observations with 8 variables. Table 3.1 contains the name and description of these variables. All of the applicant in the dataset are either non-Hispanic white, non-Hispanic black or Hispanic.

After conducting basic analysis, we noticed that there are observations has wrong value in their features. Three observations have marital status as ".", two observations have Credit meet guidelines as "666". We exclude these 5 observations from the dataset for model estimation. We controlled for all of applicant characteristics in the dataset that may influence the probability of approval except gender of applicant. We believe that there are applications that the spouse may be a co-applicant. After making the restriction, our sample size is 1984 with 7 variables.

Table 3.1 Description of Variables

Variable	*Description*
APPROVE	= 1 If the mortgage loan was approved, 0 is denied
GDLIN	= 1 If the applicant credit history meets guidelines, 0 does not meet
LOANPRC	Loan amount/purchase price (%)
OBRAT	Other obligations as a percent of total income (%)
MALE	= 1 if male, = 0 female
MARRIED	= 1 if married, = 0 is otherwise
BLACK	= 1 if black, = 0 otherwise
HISPAN	= 1 if Hispanic, = 0 otherwise

3.3 Descriptive statistic

Table 3.2 and 3.3 give an overall descriptive statistic for each variable of the dataset. 84.6% of applicants in our sample are White people, followed by Black and Hispanic are 9.9% and 5.5% respectively. The proportion of Black and Hispanic people much smaller than white people. Table 3.4 and 3.5 further portrays the descriptive statistic by race/ethnicity. We noticed in table 3.4 that the percentage of white applicant was approved (90.8%) is higher than other race Black (67.5%) and Hispanic (76.4%).

Table 3.2 Overall descriptive statistic – categorical variable

	Number observation	Percentage of sample
Marital status		
Married	1306	65.8%
Unmarried	678	34.2%
Guideline		
Does not meet	171	8.6%
Meet guideline	1813	91.4%
Gender		
Female	368	18.5%
Male	1601	80.7%
Race/ethnicity		
Black	197	9.9%
Hispanic	110	5.5%
White	1677	84.6%
Denied	244	12.4%
Approved	1740	87.6%

Table 3.3 Overall descriptive statistic – continuous variable

	Min	Mean	Max	Median	Standard Deviation
Other Obligation	0.00	32.4%	95.00%	33.0%	8.26%
Loan amount/purchase price	2%	77%	257%	80%	19%

Table 3.4: Descriptive statistic by race/ethnicity - categorical variable

	Black		Hispanic		White	
Approve						
Approved	64	32.50%	26	23.60%	154	9.2%
Denied	133	67.50%	84	76.40%	1523	90.8%
Marital status						
Unmarried	76	38.60%	31	28.20%	571	34.00%
Married	121	61.40%	79	71.80%	1106	66.00%
Guideline						
Does not meet	53	26.90%	16	14.50%	102	6.10%
Meet	144	73.10%	94	85.50%	1575	93.90%
Gender						
Female	51	25.90%	22	20.00%	295	17.60%
Male	144	73.10%	86	78.20%	1371	81.80%

Table 3.5: Descriptive statistic by race/ethnicity - continuous variable

	Min	Mean	Max	Median	Standard Deviation
Other Obligation					
Black	5.6%	34.94%	63%	35%	8.16%
Hispan	14.6%	33.41%	62%	33%	8.45%
White	0%	32.03%	95%	32.5%	8.20%
Loan amount/purchase price					
Black	29%	84%	256%	87%	18%
Hispan	39%	85%	163%	89%	15%
White	2%	76%	257%	80%	19%

4. Result

In this paper, we would like to answer the question "Controlling for relevant characteristics, is race/ethnicity associated with the outcome of a mortgage loan application?" To address this question, we estimate logit and probit models of loan approval by using marital status, meeting guideline, other obligations, loan divided by purchase and races as variables.

4.1 Estimated Model

Table 4.1 Logit and Probit models of Loan Approval with Mean Loan to Value and Other Obligations

Dependent Variable	Logit Model		Probit Model
	Parameter Estimates (Standard Deviation)	Odds Ratios	Parameter Estimates (Standard Deviation)
Variable			
Married	0.458* (0.181)	1.580	0.227** (0.090)
Meet Guideline	3.731*** (0.217)	41.739	2.149*** (0.121)
Black	-0.804*** (0.239)	0.447	-0.418*** (0.126)
Hispan	-0.891** (0.310)	0.410	-0.457*** (0.163)
Other obligations (percent of total income)	-0.034*** (0.010)	0.967	-0.016*** (0.005)
Loan/purchase	-1.696*** (0.508)	0.183	-0.847*** (0.259)
Intercept	1.357		0.547
Log-likelihood	-480.808		
Sample Size	1984		1984
*p<0.10, **p<0.05, ***p<0.01			

When we discuss the approval vs one specific variable below, the assumption is when holding other factors constant.

4.1.1 Marriage

From Table 4.1, we could find that the estimated coefficient of the married is positive in both logit and probit models. It indicates that being married is an enhancement for an applicant to increase the chances of loan getting approved. It is statistically significant as the p-value is p-value is less than 0.1. The odds ratio is 1.580, which indicates that when holding other factors constant, the estimated odds of approval is 1.580 times higher for the married applicant than the unmarried applicant.

4.1.2 Guideline

The estimated coefficients of meeting guideline are positive in both logit and probit models, which indicates that if the applicant meets the guideline, he/she will have a higher chance to get approval for loan application. At the same time, it is statistically significant as the p-value is less than 0.01. The odds ratio is 41.739, which indicates that when holding other factors constant, the estimated odds of approval is 41.739 times higher for the applicants who meet guideline than the other whose credit history barely meets guideline.

4.1.3 Black

The estimated coefficient of Black is negative in both logit and probit models. It indicates that when holding other factors constant, the Black applicant will have a lower chance to get loan approval than the White applicant. It is statistically significant because the p-value is less than 0.01. The odds ratio in the logit model is 0.447. It could be interpreted as when holding other factors constant, the estimated odds of approval is 0.553 times lower for the Black applicant than the White applicant.

4.1.4 Hispan

Similar to the result of Black, the estimated coefficient of Hispan variable is negative in both logit and probit models. It indicates that when holding other factors constant, the Hispanic applicant will have a lower chance to get loan approval than the White applicant. It is statistically significant because the p-value is less than 0.01. The odds ratio in the logit model is 0.410. It could be interpreted as when holding other factors constant, the estimated odds of approval is 0.590 times lower for the Hispanic applicant than the White applicant.

4.1.5 Other Obligations

The estimated coefficients of other obligations are negative in both models, which indicates that if the applicant has a higher percentage of other obligations out of total income, the chance of loan approval will decrease. It is statistically significant as p-value is less than 0.01. When holding other factors constant, the odds ratio of getting loan approval changes by a factor of 0.967 for each additional percentage point of other obligations in the total income.

4.1.6 Loan/Purchase

The estimated coefficients of loan/purchase are negative in both logit and probit models. It indicates that if the applicant has a higher amount of loan and lower purchased assets, the chance for him/her to get loan approval will decrease. It is statistically significant only for the White because the p-value is less than 0.01. When holding other factors constant, the odds ratio of getting loan approval changes by a factor of 0.183 for each additional one hundred percentage points of loan amount / purchased price.

4.2 Predicted Probabilities of Loan Approval for Both Models

Table 4.2 Predicted Probabilities (%) for Applicant Approval with Mean Loan to Value and Other Obligations

		Married		Unmarried	
		Logit Model	**Probit Model**	**Logit Model**	**Probit Model**
Meeting Guideline	**White**	95.84	95.91	93.57	93.49
	Black	91.15	90.69	86.70	86.33
	Hispanic	90.42	90.03	85.66	85.46

Not Meeting Guideline	**White**	35.54	34.13	25.87	26.25
	Black	19.79	20.41	13.50	14.60
	Hispanic	18.45	19.33	12.52	13.73

Table 4.2 presents the predicted probabilities of loan approval for a few prototypical individuals by logit and probit models. We divided to individuals into 12 categories, who respectively are "Married White applicant meets guideline", "Unmarried White applicant meets guideline", "Married White applicant does not meet guideline", "Unmarried White applicant does not meet guideline", "Married Black applicant meets guideline", "Unmarried Black applicant meets guideline", "Married Black applicant does not meet guideline", "Unmarried Black applicant does not meet guideline", "Married Hispanic applicant meets guideline", "Unmarried Hispanic applicant meets guideline", "Married Hispanic applicant does not meet guideline", "Unmarried Hispanic applicant does not meet guideline".

From the result, the predicted probabilities by logit and probit models are similar for each category. The differences in predicted probabilities between two model are small. When comparing the married individuals who meets guideline by different races, we could find that among the married applicants who meet guideline, the probability of White applicants (96.11%, 96.09%) is higher than the Black(91.15%, 90.69%) and the Hispanic (90.42%, 90.03%) respectively by logit model and probit model. When comparing the unmarried individuals who meet guideline among the races, we find that the probability for While applicant (93.57%, 93.49%) is also higher than the Black(86.70%, 86.33%) and the Hispanic(85.66%, 85.46%) by logit model and probit model respectively.

When we compare the probability of approval for a married applicant who does not meet guideline among races, it is clear that White applicants (35.54%, 34.13%) have higher approval probabilities than the Black (19.79%, 20.41%) and Hispanic applicants (18.45%, 19.33%). While comparing the probability of approval for unmarried applicant who does not meet guideline among races, the White applicant (25.87%, 26.25%) has an higher probability of approval than the Black (13.50%, 14.60%) and the Hispanic (12.52%, 13.73%) by logit model and probit model respectively.

In general, when comparing the same category, the White applicants always have higher probabilities of approval than the Black and Hispanic applicants.

4.3 Robustness

Comparing the result of the logit and probit models, for each variable, the sign of coefficients are similar between the two models. At the same time, all the results from both models are statistically significant. Also, the predicted probabilities by the logit model are similar to the predicted probabilities by probit model. So we believe the results of models are robust.

5. CONCLUSION

Through the results from robust models, we find that the predicted probabilities seem to be higher for the White applicants than the Black and Hispanic applicants in the same condition when holding other factors constant. For the married applicants, the difference among the predicted probabilities is not big, but the unmarried applicant whose credit history does not meet guideline, the probabilities of the White applicant to get loan approval is almost twice more than the Hispanic applicants. However, we still cannot conclude that mortgage lending institutions discriminate against minorities. The reason is we are not holding some confounding factors when we interpret the relationship between dependent and independent variables. For example, educational attainment and the type of jobs would affect getting loan approval or not. Also, the market value and type of mortgage may also affect the decision of getting loan approval or not. Therefore, before we could take all the variables into consideration and minimize the omitted bias, we cannot say the mortgage lending institutions discriminate against minorities.

Disable to control all the relative variables is the limitation of this research. If possible, we would like to take more variable as independent variables into both models, such as educational attainment, type of job and the mortgage evaluation for each applicant to avoid violating the omitted bias.

We cooperate together to work on this research. Sourabh and Huy contribute to the introduction, model, and data, while Han works on the interpretation and conclusion.