**Project 2: MLD**

**Introduction**

Mortgages are very common in the US, especially in 2008 right before the credit crisis and recession. Are there discrimination biases of who financial institutions give mortgage loans out to? We will be answering the following question: can we conclude that mortgage lending institutions discriminate against minorities in the Boston area? According to Jay Fitzgerald, in “*Black, Latino mortgage rejection rates still high”*, “Some 21 percent of black home buyers in Boston were rejected for a mortgage in 2014, compared with just 6 percent of white loan applicants”, “even when minorities make as much money as the whites, the disparities remain”, based on studies done by Jim Campen, in his annual study on home-lending trends. Then he goes on to mention “Elizabeth Seymour, a spokeswoman for JPMorgan Chase & Co., ‘We lend to all individuals and families regardless of gender, race, or ethnicity and have a variety of loan options available to all qualified home buyers’”. Campen provided statistics and data concluding that mortgage lending institutions discriminate against minorities in the Boston area, yet the banks believe there are no biases. We will use this as motivation to conclude ourselves whether financial institutions discriminate against minorities in the Boston area. As discussed in Munnell et al (1996), the Home Mortgage Disclosure Act, or HMDA, data which is used in this paper, indicates approval or denial of an applicant’s mortgage request along with several demographic characteristics of the applicant. In 1990, approximately 94% of Boston residents were White, Black or Hispanic, therefore we will be limiting our data to those three races. We have presented and discussed the descriptive statistics for the entire sample. While controlling for relevant characteristics, we have presented probit and logit models of loan approval. For the logit model we have computed, presented, and interpreted the odds ratios associated with each independent variable. For both logit and probit we have presented predicted probabilities of loan approval using a few prototypical individuals. We will also present likelihood ratio tests and log-likelihood results. With the data acquired and the models presented, we will be able to conclude that mortgage lending institutions discriminate against minorities in the Boston area, at the time the census was collected.

**Econometric Model and Estimation Method**

For our model, we will analyze both probit and logit model. Probit model uses the cumulative distribution function of the standard normal distribution while logit model uses the cumulative distribution function of the logistic distribution. We use *Approved* as our dependent variable for both the models. As for our sample selection criteria, for both models, include *Guideline, Obligation, Black, Hispanic, and Loan to Value.*

There is, however, no simple interpretation of the models’ coefficients, or to simply say, the coefficients from a probit and logit model are not directly comparable. The reason why this is interesting is that both models are nonlinear in the parameters and thus cannot be estimated using OLS for meaning full interpretation. Therefore, one relies on a method called Maximum Likelihood Estimation (MLE), to instead, consider predicted probabilities and differences in predicted probabilities. MLE method is used to find the set of parameter estimates that yields the highest likelihood of observing the set of outcomes actually observed. To help answer our research question, we will proceed using the MLE method to find the parameter estimates of logit model. Using estimated logit model coefficients, we can compute for odd ratios which will then be interpreted to meaningful information.

**Descriptive Statistics**

On average, the amount of an individual’s income to other obligations is an average of 32.39% of their total income. The approval rate for a mortgage loan is 0.88 which infers that 88% of the mortgages that individuals request for are approved and financed. Black individuals for mortgage loans make up 0.1, which is 10% of the population sample. Whereas Hispanic individuals represent .06, which is 6% of the mortgage borrowers.

To better interpret the loan to value ratio, we will be presenting the values in percentage terms rather than as a proportion. The overall loan to value ratio from the original data is 2.57, which infers that the loan is 257% greater than the value for the maximum of the population. However, the maximum should be 100% which is a ratio of 1. Therefore, for our sample selection criteria, we will be removing any data that has a value greater than 100% loan to value percentage. Which, as a result, changes the maximum loan to value ratio of 257% to 100%. The minimum of the loan to value ratio is 2%, which is a great financial ratio for individuals in the population. This implies that only 2% of the value of the asset needed a loan to help that individual finance the asset. The independent variable guidelines are a dummy variable and must only receive the value of 0 and 1. Therefore, we will be removing any data that has a value greater than 100%.

**Table 1. Descriptive statistics Mortgage Loan Decisions**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Statistic** | **N** | **Mean** | **St. Dev.** | **Min** | **Max** |
| **Guideline** | 1,937 | 0.91 | 0.28 | 0 | 1 |
| **Obligation** | 1,937 | 32.37 | 8.25 | 0 | 95 |
| **Black** | 1,937 | 0.1 | 0.3 | 0 | 1 |
| **Hispanic** | 1,937 | 0.05 | 0.23 | 0 | 1 |
| **Approved** | 1,937 | 0.88 | 0.33 | 0 | 1 |
| **Loan to Value** | 1,937 | 76.08 | 16.76 | 2 | 100 |

**Results**

Overall, both models produce very similar estimates of the probability for loan applicants and the coefficients from a probit and logit model are not directly comparable (see Table 2 and Table 3). However, as mentioned in the econometric model and estimation method section, the coefficients from a probit and logit model are not directly comparable. Therefore, we will focus on interpreting the signs and significant levels of the coefficients rather than comparing them. For both probit and logit model, the independent variable, *Guidelines*, has a positive sign for its coefficient and is highly statistically significant. This is to be expected because a meeting the guideline implies that a loan applicant is more likely to get approved. The *Loan Amount* variable has a negative sign for its coefficient and is also highly statistically significant. This is also to be expected due to the fact that as the loan amount an application applied for increase; they are less likely to get approved for the loan.

Table 4. shows the computation of odds ratios associated with the logit coefficients. Following are the interpretation of the odd ratios: Controlling for all other characteristics in the model, on average, the estimated odds that a *Black*individual will receive an approval for a loan is 58.32% lower than a White individual. Controlling for all other characteristics in the model, on average, the estimated odds that a *Hispanic* individual will receive an approval for a loan is 55.89% lower than a white individual. Controlling for all other characteristics in the model,on average, one percentage point increase in *Other Obligations* is associated with 4% lower odds of being approved for a loan. Controlling for all other characteristics in the model, individuals who meet the *Guidelines* face 41.8657 greater odds of approval than individuals who did not meet the guidelines. Controlling for all other characteristics in the model, increasing the *Loan to Value* by one unit, will lower the odds of approval for a mortgage loan by 0.9836421 compared to individuals with one less unit of loan to value.

In the Predicted Probability (Table 5), the results demonstrate how the predicted probability of the probit model, White individuals who do not meet the guidelines have a higher chance than Black or Hispanic individuals of being approved for a loan, with White having around 31% chance and Black and Hispanic have around 17% and 18%. For the individuals who do meet the guidelines, White also have a higher probability of being approved for a mortgage loan, White have a 6-7 percentage points more likely chance to be approved for a mortgage loan. These differences are the similar in the logit model. These predicted probabilities demonstrate the discrimination between minorities and White individuals in favor of White individuals.

**Table 2. Probit Model of a Mortgage Loan Application**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Estimate** | **Std. Error** | **z value** | **Pr(>|z|)** | **Signif. codes** |
| **(Intercept)** | 0.6374 | 0.3290 | 1.937 | 0.0527 | . |
| **GDLIN** | 2.1584 | 0.1222 | 17.65 | <2e-16 | \*\*\* |
| **LOANPRC** | -0.0077 | 0.0032 | -2.366 | 0.0179 | \* |
| **OBRAT** | -0.0167 | 0.0054 | -3.077 | 0.0020 | \*\* |
| **BLACK** | -0.4534 | 0.1276 | -3.552 | 0.0003 | \*\*\* |
| **HISPAN** | -0.4183 | 0.1681 | -2.487 | 0.0128 | \* |

***Signif. Codes*:** 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1' '

**Table 3. Logit Model of a Mortgage Loan Application**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Estimate** | **Std. Error** | **z value** | **Pr(>|z|)** | **Signif. codes** |
| **(Intercept)** | 1.6312 | 0.6665 | 2.447 | 0.0143 | \* |
| **GDLIN** | 3.7344 | 0.2182 | 17.11 | <2e-16 | \*\*\* |
| **LOANPRC** | -0.016 | 0.0069 | -2.36 | 0.0182 | \* |
| **OBRAT** | -0.035 | 0.0105 | -3.291 | 0.001 | \*\* |
| **BLACK** | -0.875 | 0.2427 | -3.612 | 0.0003 | \*\*\* |
| **HISPAN** | -0.818 | 0.3221 | -2.541 | 0.0110 | \* |

***Signif. Codes*:** 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1' '  
***Sample Size*:** 1,937  
***Log-Likelihood***: -465.9147 ***DF*:** 6

**Table 4. Odds Ratio** **associated with the Logit Coefficients**

|  |  |
| --- | --- |
|  | **Odds Ratio** |
| **Intercept** | 5.112 |
| **GDLIN** | 41.865 |
| **LOANPRC** | 0.983 |
| **OBRAT** | 0.965 |
| **BLACK** | 0.416 |
| **HISPAN** | 0.441 |

**Table 5. Predicted Probabilities for Acceptance for Loans with Mean of Loan Price and Other Obligation**

|  |  |  |
| --- | --- | --- |
|  | **Probit** | **Logit** |
| **Black (met guidelines)** | 0.887122 | 0.891925 |
| **Hispanic (met guidelines)** | 0.893689 | 0.891925 |
| **White (met guidelines)** | 0.95202 | 0.951927 |
| **Black (didn’t meet guidelines)** | 0.171798 | 0.164666 |
| **Hispanic (didn’t meet guidelines)** | 0.180869 | 0.172604 |
| **White (didn’t meet guidelines)** | 0.310763 | 0.321106 |

\*Controlling that loan to value and other obligations are set to their mean.

**Conclusion:**

We conclude that there is discrimination against minorities in approval rate for mortgage loans from financial institutions. Can we conclude that mortgage lending institutions discriminate against minorities in the Boston area at the time of the census, was our initial question, which we can conclude is proven correct with our data and findings. In our findings there are limitations we must acknowledge. Our data is only valid if the individuals taking the census answered truthfully, and we cannot be certain that the census was given to every individual that applied for a mortgage loan. We also are limited to the Boston area and that these finding do not imply anything about the rest of the United States. We also are limited to only individuals who identify themselves as White, Black, or Hispanic races as they make up most the population but not the whole population. To conclude, we have noticed a correlation between minorities in the Boston area at the time of the census are discriminated in the approval rate for mortgage lending.

**Contribution:**

**Yahya** wrote the R code for this project and contributed to the writing of the Econometric Model and Estimation Method. He was also responsible of following up and with feedback and addressing comments provided by the professor.

**Dalaysone** imported outputs from R into Excel and organized them in the final paper. She was also keeping the group on track for meetings, following up with professor, and submitting every step of the paper on canvas.

**Allex** Wrote the introduction, the conclusion and other sections in the paper. Her analytical skills allowed us to make sense of the values we get from R. Additionally, Allex created tables for our results section and descriptive statistics.