Are Racial Differences a Determining Factor in Mortgage Lending Decisions?

Prepared for:
Professor Bridget Hiedemann
Applied Econometrics
ECON 5300

Prepared by:
Kyler Adams
Thomas del Zoppo
Ryan Loong

March 9th, 2017

Introduction

When Congress enacted the Home Mortgage Disclosure Act (HMDA) in 1975, they meant to provide greater transparency of mortgage lending practices by requiring certain lenders to publicize data on their mortgage lending decisions. The HMDA did not require mortgage lenders to alter their practices in ways that might benefit the public welfare. So, it remains incumbent upon the public itself to analyze this data and determine whether or not lending decisions are being made in the public interest.

This paper conducts an econometric analysis to determine if, all else being equal, an applicant's race is a significant factor in a lender's determination of mortgage approval decisions. After procuring data observations of mortgage lending decisions made in the city of Boston in 1990, we develop probit and logit models to estimate the average effect that an applicant's demographic characteristics will have on the probability and logarithmic odds of their mortgage loan application being approved. From this analysis, we conclude that, all else being equal, Black and Hispanic mortgage loan applicants have significantly lower odds of being approved for a mortgage loan than White applicants.

Econometric Model and Estimation Method

Our analysis utilizes maximum likelihood estimation (MLE) in order to estimate multivariate logit and probit models. The results are shown in Table 5. In the logit model we have estimated odds ratios for all independent variables, and in the probit model we estimated marginal probabilities of the independent variables. These marginal probabilities are shown in Table 6. Due to the binary nature of our dependent variable (mortgage "loan approved") we decided that an estimation derived by ordinary least squares (OLS) methodology would be inappropriate for this analysis. Although it is possible to model binary dependent variables through OLS estimations, doing so causes drawbacks that are otherwise mitigated by the MLE methodology of probit and logit models. In particular, when estimating binary dependent variables with OLS, the predicted probabilities may fall outside of the binary range, the error terms become heteroskedastic, and the error terms are not normally distributed. Thus, we decided to utilize MLE even though MLE-based probit and logit models are typically more difficult to interpret. In the descriptive statistics of Tables 1 through 4, the independent variables of 'other obligations' and 'loan amount' are expressed as percentages rather than proportions in order to ease their interpretability.

Data

Our initial data set included observations of mortgage lending decisions made in the city of Boston in 1990, obtained from various lending institutions. Included in the data were the maximum number of observations of lending decisions made about Black and Hispanic applicants, and a random sample of observations of lending decisions made about White applicants. Upon our preliminary exploration, we discovered some potential issues with these data. For example, in some categories, where the data entries were expected to be exclusively binary (i.e. only recorded as ‘0’ or ‘1’), there were other values present such as larger numbers or cryptic symbols. Since these alternative entries were present in negligible amounts (compared to our sample size) we decided to simply remove these misreported observations from the data set. Another potential issue was with the small number of loan percentage (LOANPRC) values recorded at over 100%, with some going higher than 250%. As the loan percentage represents the ratio of the loan amount to the purchase price, such high values seem incongruent with the typical purpose of mortgage lending. It would rarely make sense for anyone to obtain a loan valued at twice the price of their intended purchase. However, with a 203k mortgage loan, buyers may finance both the purchase price of a home, and the additional costs of constructing home improvements. This possibility led us to retain the abnormally large observations of loan percentages, especially when considering that they were relatively few in number. By filtering the initial data set with these criteria, we accordingly produced our sample data set of 1,969 observations for econometric analysis.

Tables 1 through 4 below include the summary statistics of our sample data. Table 1 includes a summary of all observations in the entire sample data set. These data include two types of variables: binary and interval. The binary variables are simple ‘yes’ or ‘no’ variables, characterized in the data sets as ‘1’s and ‘0’s respectively. The interval variables are ‘other financial obligations as a percentage of total income’ (OBRAT) and ‘loan amount as a percentage of the purchase price’ (LOANPRC). Since these variables are on an interval scale, and characterized by percentage values, their values typically range anywhere from 0 to 100%. However, as mentioned above, some LOANPRC values rise as high as 250%.

Since the binary variables are characterized by only ‘0’s and ‘1’s, the mean of each variable indicates the proportion of ‘1’s that were in the sample. For example, Line 1 in Table 1 shows that ‘Loan approved,’ our dependent variable, had a mean of 0.88. This indicates that 88% of the loan applications in the sample were approved.

Line 2 of Table 1 shows that applicant financial obligations as a percentage of total income had a mean of 32.4% and a median of 33.0%, suggesting this data has a somewhat normal distribution. Its value is consistent with what one might expect of typical loan applicants. A person with substantially higher financial obligations as a percentage of their income would not often consider purchasing a home, but might instead focus on paying down their current debts.

Line 3 shows the mean loan amount as a percentage of purchase price as 77.0% (indicating an average down payment of 23%). This is also consistent with what we expected. Mortgage lending standards in the United States conventionally dictate that an applicant should put down at least 20% of the cash needed to purchase a home, or else the applicant may be required to purchase mortgage insurance. Of note here are the maximum and minimum loan percentage values of 257.1% and 2.1%, respectively. Such large loan percentages may be explained by the reasoning above involving 203k loans. Very low percentages like 2% seem counterintuitive. Normally, it would not be expected for someone to save up and pay cash for 98% of a home purchase, but then stop short and pay for the closing costs of financing a meager 2%, all while incurring the administrative hassle of a loan application. However, there may be a rational basis for this, such as when applicants are refinancing their homes, or wishing to take a small loan simply to establish their own credit history.

Line 4 shows that the mean proportion of applicants who met the credit guidelines was 91%. While not implying causality, with an overall approval rate of 88%, this does suggest a high correlation between those meeting credit guidelines and those being approved for a loan. This would make intuitive sense.

Lines 5 through 8 show the means and other statistics that describe the remainder of the independent variables, all of which are binary. The mean of Male applicants was .81, meaning 81% of the applicants were male. This makes intuitive sense because men tend to earn more than women on average, placing them in better financial positions to make home purchases, and husbands probably complete more applications on behalf of their families than wives do.

Tables 2 through 4 depict the descriptive statistics of three individual subsets in our sample data, subsetted by applicant race; White, Black or Hispanic. Every applicant observed in our sample data was a member of one these three races. It should be noted here that while individual summary statistics like means and standard deviations were calculated for each of these subsets, our probit and logit models estimated the parameters of the full sample, and the resulting probability tables were calculated with the statistics of the full sample detailed in Table 1.

Table 1: Descriptive Statistics (All)

	Mean	SD	Median	Min	Max
Loan approved (dep. Var)	0.88	0.33	1	0	1
Other Obligations as % of Total Incon	32.39%	8.28%	33.00%	0.00%	95.00%
Loan Amount as % of Purchase Price	77.03%	18.95%	80.00%	2.11%	257.14%
Met credit guidelines	0.91	0.28	1	0	1
Married	0.66	0.47	1	0	1
Black	0.10	0.30	0	0	1
Hispanic	0.05	0.23	0	0	1
Male	0.81	0.39	1	0	1
Number of observations	1969				

Table 2: Descriptive Statistics (White)

	Mean	SD	Median	Min	Max
Loan approved (dep. Var)	0.91	0.29	1	0	1
Other Obligations as % of Total Incon	32.03%	8.23%	32.55%	0.00%	95.00%
Loan Amount as % of Purchase Price	75.65%	19.01%	79.88%	2.11%	257.14%
Met credit guidelines	0.94	0.24	1	0	1
Married	0.66	0.47	1	0	1
Male	0.82	0.38	1	0	1
Number of observations	1666				

Table 3: Descriptive Statistics (Black)

	Mean	SD	Median	Min	Max
Loan approved (dep. Var)	0.67	0.47	1	0	1
Other Obligations as % of Total Incon	34.90%	8.19%	35.00%	5.60%	63.00%
Loan Amount as % of Purchase Price	84.06%	17.84%	87.50%	28.99%	255.52%
Met credit guidelines	0.73	0.45	1	0	1
Married	0.62	0.49	1	0	1
Male	0.74	0.44	1	0	1
Number of observations	195				

Table 4: Descriptive Statistics (Hispanic)

	Mean	SD	Median	Min	Max
Loan approved (dep. Var)	0.76	0.43	1	0	1
Other Obligations as % of Total Incon	33.47%	8.46%	33.00%	14.60%	62.00%
Loan Amount as % of Purchase Price	85.63%	14.50%	89.63%	40.09%	162.63%
Met credit guidelines	0.85	0.36	1	0	1
Married	0.71	0.45	1	0	1
Male	0.80	0.40	1	0	1
Number of observations	108				

Empirical Results

In both our probit and logit econometric models, the dependent variable is mortgage loan approval, and the independent variables are whether or not the applicant is married, whether the applicant's credit history meets the mortgage requirement guidelines, the applicant's race (whether Black or Hispanic), other financial obligations the applicant is responsible for (expressed as a percentage of their income), and the amount of the mortgage loan applied for (expressed as a percentage of the purchase price of the home). The variable White was excluded from the models, allowing Whites to serve as a reference group, while Black and Hispanic variables were included. All dependent variables were statistically significant at the 1% level, except for married, which was statistically significant at the 5% level.

From our analysis, we can see that whether or not an applicant's credit history met the mortgage lending guidelines had, by far, the largest effect on whether or not they were approved for a mortgage. Applicants who met the mortgage requirements had 41.32 times the odds of being approved for a mortgage compared to those who did not meet the requirements, all else being equal. Additionally, a married applicant has about 59% higher odds of being approved for a mortgage compared to a non-married applicant.

Our results regarding the parameter estimates of our interval variables are interpreted with respect to a 1% increase in each of these intervals. A 1% increase in an applicant's 'other obligations' or 'loan percentage' leads to only slightly lower odds of their application being approved. However, these odds increase significantly as the percentage difference becomes greater. For example, consider two loan applicants, one with a 20% down payment and one with 0%. The applicant requiring a 100% mortgage has $(0.983)^{20}=0.71$, or 29% lower odds of being approved for a mortgage compared to the applicant requiring only an 80% mortgage, all else being equal. Similarly, when compared to Whites, Blacks and Hispanics respectively have roughly 56% and 59% lower odds of being approved for a mortgage, all else being equal.

In general, our analysis produces results that we would have expected in each model. It makes sense that married applicants have a higher rate of approval, as they are often perceived to be more financially stable and better able to complete the terms of a mortgage, especially when there are two potential incomes in the household. It also makes sense that if an applicant needs a larger loan or has more financial obligations that they are responsible for, the chance of approval is lowered. Finally, we did expect that an applicant's credit history meeting established guidelines would indeed be the most important factor in determining loan approval, but we were surprised that the estimated magnitude of this parameter was so extremely large. In regard to our research question, unfortunately we can see from the results that there may be some discrimination against Blacks and Hispanics, as their odds of approval are lower than Whites, all else being equal.

Table 5:

	Dependent variable:		
	APPROVE		
	probit	*logit*	*odds ratio*
Constant	0.542*	1.342**	
	(0.298)	(0.567)	
MARRIED	0.229**	0.461**	1.586
	(0.090)	(0.181)	
GUIDELINES	2.144***	3.721***	41.322
	(0.121)	(0.217)	
OTHER OBLIGATIONS	-0.016***	-0.034***	0.966
	(0.005)	(0.010)	
LOAN PERCENTAGE	-0.008***	-0.017***	0.983
	(0.003)	(0.005)	
BLACK	-0.423***	-0.811***	0.444
	(0.127)	(0.240)	
HISPANIC	-0.462***	-0.897***	0.408
	(0.163)	(0.310)	
Observations	1,969	1,969	
Log Likelihood	-479.46	-479.728	
Akaike Inf. Crit.	972.921	973.456	
Note:	*p<0.1; **p<0.05; ***p<0.01		

We translated the probit coefficients to marginal probabilities for several prototypical applicants. Specifically, we used ‘married’ or ‘unmarried,’ and ‘meets mortgage guidelines’ or ‘doesn’t meet mortgage guidelines’ for each of the three races. We chose these categorical variables because they were all statistically significant in our probit and logit models. We used the mean values of the ‘other obligations’ and ‘loan percentages’ from our full sample in order to calculate each probability. The marginal probabilities of the different prototypical applicants are presented in Table 6. In each cell, the numeric figure represents the probability of an applicant being approved for a mortgage who has the corresponding categorical traits . Again, the factor which contributes the most to whether or not an applicant’s loan is approved is whether or not the applicant’s credit history meets the mortgage guidelines. Out of all prototypical applicants, the White applicant was most likely to be approved, followed by the Black and Hispanic applicants. A married applicant who meets the guidelines is about 5% more likely to be approved if they are White than if they are Black, and about 6% more likely to be approved if they are White than if they are Hispanic. There is a larger discrepancy in applicants who do not meet the guidelines, in which case a married applicant is about 14% more likely to be approved if they are White than if they are Black, and about 15% more likely to be approved if they are White than if they are Hispanic.

Table 6 (Using Probit Model Coefficients):

Probabilities Predicted using Mean Values for OTHER OBLIGATIONS and LOAN PERCENTAGE

	Meets Guidelines		Does Not Meet Guidelines	
	Married	Not Married	Married	Not Married
Black	0.9056	0.8611	0.2034	0.1449
Hispanic	0.8989	0.8523	0.1925	0.1362
White	0.9588	0.9342	0.342	0.2624

Conclusion

With these results, we have strong statistical evidence to conclude that an applicant's race was a determining factor in whether or not they were approved for a mortgage loan in the city of Boston in 1990. Controlling for other relevant characteristics, Black and Hispanic applicants had significantly lower odds of being approved for a mortgage loan than White applicants. The marginal probabilities of White applicants being approved were greater in all cases of credit-guideline-qualified or unqualified, married or unmarried applicants. This implies that racial discrimination existed in the mortgage lending market. However, we cannot state whether or not racial discrimination appears to exist in Boston's lending market today, or whether it may exist in other markets of the United States.

Although our conclusions were highly statistically significant, our econometric modeling was limited in certain ways. With female applicants comprising only 19% of our 1,969 observations, we did not find gender to be of statistical significance to our analysis. Thus, we omitted gender variables. By choosing to include the observations with loan percentages greater than 100%, we may have inadvertently retained faulty data. Or we may have introduced omitted variable bias by including a special subset of real estate developers into the sample of applicants, developers far more likely to be approved on the basis of their professional qualifications or their longstanding relationships with mortgage lenders. We did not have any data on the wealth of our applicants. Nor did we have data on many other demographic characteristics that may have been relevant to our analysis, such as applicant ages, the number of children they have, or how many mortgages they had previously been approved for. So, our research was also subject to the limitations of our data.