Do Mortgage Lending Institutions Discriminate Against Minorities?

Garrison Hess and Mariah Harvey

May 2014

1 Introduction

Mortgage lending institutions are an integral part of both the American financial and housing systems. Fundamentally, mortgage lending decisions are made through quantitative assessment of a given individual's financial risk in repaying a substantial loan. Quantitative analysis naturally leads to either approval or disapproval of a loan and varying contractual terms for approved individuals. While variance in approval rates and contractual terms is rooted in risk management, there are still confounding trends. It is alleged that these trends are explained, in part, by racial and ethnic discrimination. The purported discrimination takes two general forms: redlining, and reverse-redlining. Accusations of redlining, a process by which banks discriminatorily avoid specific regions and demographics, have led to great controversy. Reverse-redlining involves the enticement of minorities into signing subprime loan agreements, despite their qualification for better loans with lower interest rates and fees. It is alleged that during the rapid proliferation of the housing market significant reverse-redlining was at play (Savage, 2011). Furthermore, Bank of America's Countrywide Financial unit paid $335 million to settle allegations of reverse-redlining; which lends itself to the veracity of these allegations. However, this issue becomes obfuscated when the government does not require mortgage lenders to report information about borrowers' credit scores, interest rates, down payments and other details used in risk assessment (Bajaj and Fessenden, 2007). In this paper we ask the question: can we conclude that mortgage lending institutions discriminate against minorities? We aim to contribute to the array of studies on racial and ethnic bias in mortgage lending decisions, and draw sound conclusions on this contentious subject. To assess these mortgage lending decisions, we estimate probit and logit models of loan approval by the method of maximum likelihood estimation (MLE) using data reported by the Home Mortgage Disclosure Act (HMDA).

This issue becomes increasingly important following the burst of the housing market bubble. Cities left in the wake of the crisis claim that vacated properties would not have become unoccupied if the predatory lending had not taken place (Martin, 2011). Vacated properties, in turn, bring other property values down, having a cascade effect on the economy. The existence of many vacated homes has deep implications for citizens across America and highlights the importance of recognizing discrimination in mortgage lending against minorities, if it is indeed a proven offense.

Previous research on mortgage loans suggests that discrimination is a true concern. Data collected in the 1990's through the HMDA showed that black or Hispanic applicants in the Boston area were more likely to be turned down for a mortgage, relative to a white applicant, even when controlling for credit scores or other observable individual risk factors (Munnell et al., 1996). Munnell used logit models and ordinary least squares regression; whereas, we use logit and probit models. Munnell also used a more comprehensive series of independent variables; which included: net worth, fixed rate of loans, age, and gender. We do not consider these variables in our study, as they are not included in our dataset. Other literature found that subprime loans were heavily concentrated in regions with relatively high black and Hispanic population proportions (Mayor and Pence, 2007; hereafter MP). While MP focused on subprime lending, we focus on lending aversion as measured by loan approval rate. Subsequent literature examining subprime loans showed that there was no evidence of interest rate bias by race, ethnicity, or gender (Haughwaut et al., 2009). Similar to MP, Haughtwaut examines the terms of the loan in an attempt to find discriminatorily unfavorable lending to specific demographics. As such, Haughwaut estimates initial mortgage interest rate using ordinary least squares (OLS) regression; as opposed to probability of mortgage approval using probit and logit models.

2 Econometric Models and Estimation Methods

2.1 Estimation Approach

In this paper we apply MLE to estimate probit and logit models. Each model predicts a given individual's probability of mortgage approval as a function of guideline fulfillment, loan to value, other obligations as a percentage of total income, marital status, and race/ethnicity. We include all these variables except race/ethnicity to control for omitted variable bias and examine approval discrepancies across different races and ethnicities. Furthermore, we use probit and logit models, rather than an OLS linear probability model, because we intend to predict probabilities for prototypical individuals with values between zero and one, obtain odds ratios, and avoid both heteroskedastic and non-normal error terms.

2.2 Measures

2.2.1 Guideline Fulfillment

We include guideline fulfillment because it represents an applicant's first step towards loan approval. This basic hurdle is a primary criterion in banks' loan-approving process.

2.2.2 Loan to Value

In the initial dataset, loan to value is reported as a proportion of purchase price. We multiply the initial figure by 100 so we can discuss the coefficient as a percent. Loan amount relative to purchase price is included because individuals who require a larger loan, relative to purchase price, have less equity in their home; thus, they represent riskier loans.

2.2.3 Other Obligations

Other obligations are already reported as a percentage of total income, so we do not transform the variable. Fundamentally, other obligations are relevant, in that, individuals with other significant financial obligations will have a lesser priority for making loan payments; thus, they are more risky applicants.

2.2.4 Marital Status

We include marital status, to serve as a proxy for the following factors: married men tend to work harder, and get paid more than unmarried men; married couples tend to be more financially stable, or are at least viewed as such, which may lead to preferential treatment by bankers; and loans to married couples have two individual co-signers, which is positive factor in banks' risk analysis process.

2.2.5 Race/Ethnicity

Lastly, we include our variables of interest: black, Hispanic, and white. These are dummy variables, for which white individuals are the reference category. All aforementioned regressors except black, Hispanic, and white are included as controls for banks' loan-approving process. Therefore, black, Hispanic, and white are included to capture discrepancies in loan approval rates among the three races.

3 Data

The data we use in this paper was collected according to the Home Mortgage Disclosure Act. Congress enacted the HMDA in 1975. Since then, lending institutions have been required to report public loan data to the Federal Reserve Board. The data assists the Fed in determining whether financial institutions, including banks, savings associations, credit unions, and other mortgage lending institutions, are serving the housing needs of their communities. It is also used to identify potential discriminatory lending patterns. The data we use was collected in 1990, and indicates whether an applicant's mortgage application was approved, given a set of demographic characteristics; which include: whether guidelines were met, loan amount as a proportion of purchase price, other financial obligation as a percentage of total income, race/ethnicity, and marital status. This randomized sample includes 909 loan applications from black, Hispanic, and white participants. One observation lacked a value for marital status. As such, we omit this observation, making our sample size 908.

Our descriptive statistics in Table 1 show that 87% of all applicants in the sample were approved for a mortgage loan. Approximately 91% of applicants met the credit guidelines, and the average loan to value was 76%. We also observe that, for the average applicant, other financial obligation as a percentage of total income was 32%. We also include in our analysis the variables white, black, Hispanic, and marital status. Table 1 shows that approximately 10% of applicants are black, and 6% Hispanic.

As such, small sample sizes of black and Hispanic applicants may prove somewhat problematic in model estimation.

Table 1: Descriptive Statistics

	Mean	**Median**	**Maximum**	**Minimum**	**Std. Dev**
Approve	0.8774	1	1	0	0.3315
Guidelines	0.9141	1	1	0	0.2804
Loan to Value	76.81	80	256	2.11	18.99
Married	0.6553	1	1	0	0.4755
Other Obligations	32.38	33	73	0	7.82
Black	0.1046	0	1	0	0.3062
Hispanic	0.0562	0	1	0	0.2304
White					
Approve	0.9042	1	1	0	0.2945
Guidelines	0.9209	1	1	0	0.2653
Loan to Value	75.39	79.87	149	2.11	18.67
Married	0.6653	1	1	0	0.4722
Other Obligations	32.08	32.65	73	0	7.72
Black					
Approve	0.6737	1	1	0	0.4714
Guidelines	0.8421	1	1	0	0.3666
Loan to Value	83.82	81.25	25.55	32.37	21.78
Married	0.5895	1	1	0	0.4945
Other Obligations	35.35	35.4	63	10.6	7.91
Hispanic					
Approve	0.8039	1	1	0	0.4009
Guidelines	0.9019	1	1	0	0.3003
Loan to Value	84.99	89.33	13.79	49.04	12.37
Married	0.6275	1	1	0	0.4883
Other Obligations	31.41	31.4	53.9	15	8.056

We have also stratified our descriptive statistics by race and ethnicity. Among the white applicants, 90% were approved for a mortgage loan; black applicants, 67%; and Hispanic applicants, 80%. Surprisingly, we find a 10 percentage point gap between approval rates for white applicants and Hispanic applicants, despite Hispanic applicants having otherwise similar characteristics to white applicants. We can see that 92% of white applicants met the credit guidelines while 90% of Hispanic applicants met the guidelines; which is only a 2 percentage point difference. When it comes to other obligations as a percentage of total income, Hispanic applicants have a favorable average to white applicants; with 32% for white applicants and 31% for Hispanic applicants. We also observe that black applicants have similar characteristics in this criterion, having 35% of total income as other financial obligations, on average. Despite observing generally similar characteristics for white, black, and Hispanic applicants, the outcome is still a drastically dispersed series of average approval rates:

90% for whites, 67% for blacks, and 80% for Hispanics. However, there is a variable with relatively dissimilar averages that may explain the approval rate dispersion, and that is loan to value. On average, the loan to value is 75% for white applicants, 84% for black applicants, and 85% for Hispanic applicants. While it is intuitive to look at the larger dissimilarities to explain approval rate variance, it may also be smaller discrepancies in more impactful variables that explain the difference in approval rates across races and ethnicities.

4 Results

Tables 2 through 5 present the main results of our analysis. First, we show the logit coefficient estimates and relevant odds ratios. Then, probit coefficient estimates. Lastly, Tables 4 and 5 display the predicted probabilities for all potential applicants. The following subsections provide a more thorough interpretation of our results.

4.1 Logit and Probit Models

All independent variable estimations have the same signs and similar statistical significance for both our probit and logit models; as such, we will jointly discuss them.

4.1.1 Guideline

Guideline fulfillment has a positive effect on loan approval probability, and is significant at the 1% level in both models. Its coefficient also has the greatest magnitude of all regressors. This is tested using 95% confidence intervals of 'guideline' and the second most magnitudinous variable, 'black', which do not overlap. The output makes sense, in that, meeting the guidelines is an important and basic factor of loan approval. Furthermore, the logit model shows that 'guideline' has an odds ratio of 15.43. Otherwise stated, individuals that meet the guidelines have 15.43 times greater likelihood of loan approval, than those who do not meet the guidelines.

4.1.2 Loan to Value

Loan to value has a negative effect on loan approval probability, and is significant at the 5% level in both probit and logit models. Intuitively this makes sense because as the ratio of loan to value rises, mortgage lenders are less inclined to grant approval. Our logit model shows that for each additional percent added to loan to value, an applicant is on average approximately 1.4% less likely to be approved for a loan.

Contrary to our prior notion that loan to value may the most impactful variable, it is not.

4.1.3 Other Obligations

Other obligations, as a percentage of total income, has a negative coefficient that is statistically significant at the 5% and 1% levels in both the probit and logit model, respectively. This negative relationship is intuitive because as applicants take on more financial obligations they have less uncommitted income. As an individual's uncommitted income decreases, their ability to make consistent payments on a long-term loan decreases, and their risk in holding a loan increases. Thus, as one's other obligations increases, banks become less likely to approve his or her mortgage. Accordingly, our logit model tells us that for each additional percent added to other obligations, an applicant is on average 3.9% less likely to be approved for a loan.

4.1.4 Married

The coefficient on 'married' is positive but not statistically significant in both models. The positive parameter estimate indicates that a married applicant is more likely to be approved for a loan. More specifically, the odds of a married applicant being approved are approximately 1.19 times greater than an unmarried applicant.

4.1.5 Black

The coefficient on 'black' is negative and statistically significant at the 1% level in both models. This negative relationship indicates that blacks are less likely to be approved than white applicants. More precisely, the odds of a black applicant being approved are approximately .257 times as great as those of a white applicant.

4.1.6 Hispanic

The coefficient on 'Hispanic' is negative and statistically significant in both the probit and logit model at the 10% and 5% levels, respectively. The negative parameter estimate shows that Hispanic applicants are less likely to be approved than white applicants. Looking at our logit model, we can conclude that the odds of a Hispanic applicant being approved are approximately .492 times as great as a white applicant.

Table 2: Mortgage Approval Probabilities (Probit Model)

Regressor	**Probit Coefficient**	**Standard Error**
Guideline	1.580***	(0.165)
Loan to Value (%)	-0.008**	(0.003)
Other Obligations (%)	-0.020**	(0.008)
Married	0.097	(0.125)
Black	-0.733***	(0.164)
Hispanic	-0.439*	(0.229)
Intercept	1.219***	(0.425)
Log Likelihood	-260.8	-
LR statistic	164.5	-
Sample Size	908	-

***, **, * denote statistically significant probit coefficients at the 1%, 5% and 10% level with standard errors in parentheses.

Table 3: Mortgage Approval Probabilities (Logit Model)

Regressor	**Logit Coefficient**	**Standard Error**	**Odds Ratio**
Guideline	2.736***	(0.284)	15.43
Loan to Value (%)	-0.014**	(0.007)	0.986
Other Obligations (%)	-0.040***	(0.015)	0.961
Married	0.181	(0.239)	1.198
Black	-1.358***	(0.295)	0.257
Hispanic	-0.846***	(0.430)	0.429
Intercept	2.310***	(0.814)	-
Log Likelihood	-261.2	-	-
LR statistic	163.7	-	-
Sample Size	908	-	-

***, **, * denote statistically significant logit coefficients at the 1%, 5%, and 10% level with standard errors in parentheses. Odds ratios are obtained by exponentiating the corresponding parameter estimate.

4.2 Predicted Probabilities

Whereas the logit model's odds ratios give an understanding of the impact of individual regressors, the predicted probabilities give a more encompassing view of given individuals' probabilities of approval. In Table 4 we display predicted probabilities based on the logit model. In Table 5 we display predicted probabilities based on the probit model. These probabilities are calculated for 12 prototypical individuals involving all potential combination of dummy variables. For the continuous variables, loan to value and other obligations, we use mean values from the entire distribution. Although it may be valuable to use mean values for these variables from each racial/ethnic subset, we must use them as controls to be able to make equal comparisons. Because the probit and logit predicted probabilities are very similar we will discuss them together.

From Table 4 and 5 we can discern certain trends. One trend we see across the two tables is that white applicants have the highest predicted probability of being approved for a loan across all categories. Hispanics have the next greatest predicted probability of being approved for a loan across all categories. As such, blacks have the lowest predicted probability of being approved for all categories.

More specifically, we can talk about the effects of meeting the guidelines. White applicants suffer the least from not meeting the guidelines. A Hispanic applicant meeting the guidelines nearly triples his or her predicted probability of receiving approval for a loan. Whereas, a black applicant that meets the guidelines multiplies his or her predicted probability by a factor of roughly 3.5.

When we focus specifically on married and non-married applicants, the aforementioned trends still apply. Black married applicants have lower predicted probabilities relative to Hispanic and white married applicants. Black non-married applicants also have lower predicted probabilities relative to Hispanic and white non-married applicants.

Overall, white married applicants that met the guidelines have the highest predicted probability of loan approval. Non-married blacks who did not meet the guidelines have the lowest predicted probability of loan approval. Meeting the guidelines consistently yielded the greatest increase in predicted probability of loan approval. Second in impact magnitude to meeting the guidelines is the racial/ethnic factor. The increase in predicted probability from being white instead of black, ranges from 12.8% to 30.6%. Lastly, marriage contributed to the predicted probability of receiving a loan, but did not have as large of an impact. Although Tables 4 and 5 show different probabilities for married and unmarried individuals, marital status is our only statistically insignificant variable. As such, we cannot conclude that whether an applicant is married makes a statistically significant impact on the probability of mortgage loan approval. However, we can still conclude from these predicted probabilities that white applicants are far more likely than Hispanic and black applicants to obtain loan approval, with all other variables held constant. Furthermore, Hispanic applicants are more likely than black applicants to be granted approval; and, accordingly, blacks are least likely to be granted approval.

Table 4: Predicted Probabilities (Logit Model)

	Met Guidelines		**Did Not Meet Guidelines**	
	Married	Not Married	Married	Not Married
Black	0.818	0.789	0.225	0.195
Hispanic	0.882	0.862	0.327	0.288
White	0.946	0.936	0.531	0.485

Values in table 4 represent predicted probabilities for prototypical individuals evaluated using the logit model with mean values of continuous variables. Coefficients are transformed into probabilities using the logistic function $1/(1+e^{-z})$ where z is the coefficient. Mean values for loan to value and other obligations are 32.38 and 76.81, respectively.

Table 5: Predicted Probabilities (Probit Model)

	Met Guidelines		**Did Not Meet Guidelines**	
	Married	Not Married	Married	Not Married
Black	0.811	0.782	0.246	0.213
Hispanic	0.880	0.860	0.346	0.288
White	0.947	0.936	0.518	0.475

Values in table 5 represent predicted probabilities for prototypical individuals evaluated using the probit model with mean values of continuous variables. Coefficients are transformed into probabilities according to the standard normal distribution. Mean values for loan to value and other obligations are 32.38 and 76.81, respectively.

5 Conclusion

Throughout our results, white applicants consistently have the highest probability of obtaining approval for a mortgage. Subsequent to whites are Hispanics, then blacks. These results were reached with several small caveats. First, we did not have as much data as we would have liked. This is because the Federal Reserve Board does not require financial institutions to report information pertaining to interest rates, fees, and payment terms of mortgages. Aforementioned studies used similar information to estimate subprime loan frequencies among different races and ethnicities, which enabled them to observe a highly relevant aspect of potentially discriminatory lending practices that we could not observe. As such, having deeper information would have been extremely valuable in drawing stronger conclusions. Furthermore, our data was from 1990, which disables us from observing any lending patterns directly preceding the burst of the housing market bubble. This would have been desirable, as it is alleged that with the rapid proliferation of the housing market, came the advent of reverse-redlining and other dubious lending practices. However, as previously stated, our study still would not have been ideal for capturing the effects of reverse-redlining, having only observed loan approval rates. Lastly, there were minor issues pertaining to sample selection. While we could have restricted outliers, and potentially obtained superior results, we only left out observations with missing criteria. We chose to do so because we did not code the initial dataset; thus, we could not determine whether outliers were coding errors, reporting errors, or genuine observations. Although these issues may have led to minor issues in our model estimations and inhibited our ability to draw stronger conclusions, we are confident in the veracity of our results.

While these results are tempting to call evidence of discrimination on the basis of correlation, we must account for other potential factors - we need causation, which is far more difficult to establish. In this case, we know from the inception of our research that we are not privy to all the information pertaining to these individuals' applications and mortgage-seeking processes. However, we also lack information regarding the unquantifiable human element of lending, which presents a far more convoluted problem. For example, when an individual enters a bank and applies for a mortgage, they are generally subject to the perception of the bank employee with whom they interact. If one enters a bank severely disheveled in appearance, should the banker accept their application? If individuals who present themselves in such a way are known to be unfit candidates for mortgages, surely the bank should reject the applications of said individuals. Yet, with our quantitative estimation methods, we would only see that this person did not get a loan, despite a potentially solid financial foundation. Therefore, contrary to what the data shows, the bank may have been entirely right to apply subjective analysis and reject an otherwise outstanding individual's application. This situation elucidates why we cannot definitively conclude that mortgage lending institutions discriminate by race and ethnicity.

Aforementioned rationale still in place, the human element also leaves room for discrimination. For example, if that person who entered the bank was white, rather than black, it is possible that they would be given the benefit of the doubt. One way to quash this potential discrimination would be to attempt to eradicate subjectivity as a whole. This becomes problematic, however, when it disables mortgage lending institutions from making efficient, non-discriminatory decisions regarding to whom they offer loans. As a whole, this would intuitively involve a government mandate that all mortgage lending institutions must follow strictly quantitative assessment processes, and report all numbers to the Federal Reserve Board. Due to impracticality and inefficiency, we do not call for such a mandate.

Idealistic measures aside, there are other more reasonable methods of dealing with potential discrimination by mortgage lending institutions; specifically, requiring more thorough data reporting by mortgage lending institutions, without infringing on their productivity and competitiveness. Just a few more pieces of vital information would be extremely effective in aiding econometricians and policymakers in finding the root of, and solving, problems with contemporary mortgage lending. As this information becomes more available, the United States' mortgage market should become more equitable, which may increase overall efficiency, because potentially discriminatory, inefficient incentives will be less prevalent in mortgage lending institution's decision-making processes. Ultimately, while we cannot definitively conclude that there is discrimination, we can conclude that policymakers must further investigate structural and institutional biases that contribute to the mortgage approval rate gap among racial and ethnic minorities.

References

[1] Bajaj, V., & Fessenden, F. (2007, November 4). *What's Behind the Race Gap.* New York Times, p. 16.

[2] Haughwout, A., Mayer, C., Tracy, J., Jaffee, D. M., & Piskorski, T. (2009). *Subprime Mortgage Pricing: The Impact of Race, Ethnicity, and Gender on the Cost of Borrowing*. Brookings-Wharton Papers on Urban Affairs, 33-63.

[3] Martin, A. (2011, May 6). *Judge Allows Redlining Suits to Proceed*. New York Times.

[4] Mayer, C., & Pence, K. (2008). *Subprime Mortgages: What, Where, and to Whom*?

[5] Munnell, A. H., Tootell, G. M., Browne, L. E., & McEneaney, J. (1996). *Mortgage Lending in Boston: Interpreting HMDA Data*. The American Economic Review , 86 No. 1, 25-53.

[6] Savage, C. (2011, Decemeber 22). *Countrywide Will Settle a Bias Suit*. New York Times.