Discrimination in Mortgage Lending Markets in US Metropolitan Area

Joon J. Park

Professor, Department of Business Administration
University of Arkansas at Pine Bluff
1200 North University Drive, Pine Bluff, Arkansas 71601, USA
E-mail:parkj@uapb.edu

870) 575-8083

Abstract

This paper investigates that denial rates are higher or lower based on characteristics of applicants holding constant the loan amount-to-income ratio. It is very interesting to see applicant characteristics are likely influence the probability of denial. All models indicate that denial rates are higher for African American applicants holding constant the loan amount-to-income ratio. The estimated differences in denial probabilities between two hypothetical applicants with the same loan amount-to-income ratio are 14.46% for Linear Model 2, 14.95% for Probit Model 2, and 15.08% for Logit Model 2. As for the all model coefficients are highly significant and we obtain positive estimates for the coefficients on A/I ratio and black.

Key words: HMDA, Mortgage Market, Discrimination, Probit Model, Logit Model

1. Introduction

The Home Mortgage Disclosure Act (HMDA) was enacted by Congress in 1975 and is implemented by the Federal Reserve Board's Regulation C. This regulation provides the public with loan data that can be used to assist: in determining whether financial institutions are serving the housing needs of their communities; public officials in distributing public-sector investments so as to attract private investment to areas where it is needed; and in identifying possible discriminatory lending patterns.

The US government collects and distributes an enormous database with information about US mortgages. The HMDA dataset contains the most comprehensive publicly available information on mortgage market activity. Each fall, new HMDA data are made available. In 2016, almost 7,000 institutions released over 16 million records, making HMDA an invaluable administrative dataset on housing and homeownership for policymakers, regulators, and researchers. The latest records were made available in September 2017 and used in this research.

Previous attempts have been made to analyze discrimination in the mortgage market by using HUMDA and other financial data. James Lindley et at., (1984), Alici Munnell et al., (1996), Harold Black et al., (1978), Fidel Ezeala-Harrison and Glenda Glover (2008), and Jerry Ingram and Emma Frazier (1982) all found that being a minority increased the probability of being rejected for a mortgage loan.

This paper investigates that denial rates are higher or lower based on characteristics of applicants holding constant the loan amount-to-income ratio for owner-occupied housing loans to minority applicants of Little Rock Metropolitan Statistical Area (MSA). It is very interesting to see applicant characteristics are likely influence the probability of denial.

2. Method: Econometric Models

This study will use three econometric models for Little Rock metropolitan area in Arkansas. A special class of regression models that aim to explain a limited dependent variable, where the dependent variable is binary. The regression function can be interpreted as a conditional probability function of the binary dependent variable.

The Linear Probability: Linear probability models are easily estimated in most statistical softwares. A binary dependent variable Y_i is called the linear probability model. In the linear probability model we have

$$E(Y|X_1,X_2,\ldots,X_k) = P(Y=1|\ X_1,X_2,\ldots,X_k)$$

where

$$P(Y=1|\ X_1,X_2,\ldots,X_k)=\beta_0+\beta_1X_{1i}+\ \beta_2X_{2i}+\cdots+\ \beta_kX_{ki}.$$

Thus, β_j can be interpreted as the change in the probability that Y_i=1, holding constant the other k−1 regressors. Just as in common multiple regression, the β_j can be estimated using OLS and the robust standard error formulas can be used for hypothesis testing and computation of confidence intervals.

In most linear probability models, R^2 has no meaningful interpretation since the regression line can never fit the data perfectly if the dependent variable is binary and the regressors are continuous.

Probit Model: In statistics, a *probit model* (binary dependent variable case) is a type of regression in which the dependent variable can take only two values (0 and 1), for example, denied or not denied. The name comes from probability and unit. The purpose of the model is to estimate the probability that an observation with particular characteristics will fall into a specific category.

A standard statistical textbook such as Greene (2011) would show that the estimator β^ could be calculated through maximizing the following log-likelihood function L(β):

$\hat{\beta} = \text{argmax}_\beta[\ln L(\beta)] = \text{argmax}_\beta[\sum_t(y_t \ln\Phi(x'_t\beta)+(1-y_t)\ln(1-\Phi(x'_t\beta)))]$.

In order to report standard regression outcomes such as t-statistic, p-value and others, we need the estimated co-variance matrix of the estimator β^, i.e., Vβ^, which is based on the inverse Hessian matrix according to Greene (2011),

$\hat{V}_{\hat{\beta}} = (\hat{H})^{-1}$,

Where $\hat{H}=\nabla^2 \ln L(\beta)|_{\hat{\beta}}$ is the estimated Hessian of the log-likelihood function lnL(β) at the solution point β^.

Logit Model: A *logit* (or logistic regression) model is a type of regression where the dependent variable is categorical. It could be binary or multinomial; in the latter case, the dependent variable of multinomial logit could either be ordered or unordered. On the other hand, the *logit* is different from the *probit* in several key assumptions. A standard statistical textbook such as Greene (2011) would show that the estimator β^ could be calculated through maximizing the following log-likelihood function lnL(β):

$\hat{\beta} = \text{argmax}_\beta[\ln L(\beta)]$

$= \text{argmax}_\beta[\sum_t (y_t \ln(\exp(x'_t\beta) / (1 + \exp(x'_t\beta)) + (1-y_t)\ln(1/(1+\exp(x'_t\beta))]$.

3. Data and Variables

3.1. Data

The HMDA data show geographic distribution of loans and applications; ethnicity, race, sex, age, and income of applicants and borrowers; and information about loan approvals and denials. HMDA data will be collected from the Consumer Financial Protection Bureau's Web site (wwww.consumerfinance.gov/hmda). The description of the data set given in Appendix I and Appendix II was supplied by the Consumer Financial Protection Bureau.

3.2. Variables: Selection of Dependent and Independent Variables

We use a subset of the 2017 HMDA data for Arkansas. To narrow the scope of the analysis, we use a subset of the data for conventional loan only (thereby excluding data on FHA, VA, and FSA/RHS loans), single-family residences only (thereby excluding data on multi-family homes), home purchase only (thereby excluding data on home improvement and refinancing), owner-occupied as a principal dwelling only (thereby excluding data on not owner-occupied), loan originated/application approved but not accepted/application denied by financial institution only, and black and white applicants only (thereby excluding data on applicants from other minority groups). This leaves 5207 observations, 538 observations for black and 4669 observations for white and 635 observations for application denied or not accepted and 4572 observations for loan originated, Appendix III.

4. Empirical Results

4.1. Linear Probability Model

The variable we are interested in modelling is deny, an indicator for whether an applicant's mortgage application has been accepted (Y=0; deny = no) or denied (Y=1; deny = yes). The regressors that ought to have power in explaining whether a mortgage application has been denied are the loan amount, the applicant's income, and/or the size of the loan amount relative to the applicant's income. It is straightforward to translate this into the simple regression model

$Y = \beta_0 + \beta_1$aount/income ratio + u. (Linear Model 1)

$Y = \beta_0 + \beta_1$aount/income ratio + $\beta_2$black + u. (Linear Model 2)

We estimate Linear Models 1 and 2 just as any other linear regression model by using EViews. Before we do so, the dependent variable must be converted to a numeric variable, Y=1; denied and Y = 0; originated.

Linear Model 1 indicates that there is a positive relation between the amount to income ratio and the probability of a denied mortgage application so individuals with a high ratio of loan amount to income are more likely to be rejected. According to the estimated model, in Table 1, the estimated regression line is

$\hat{Y} = 0.102 + 0.0089$ a/i ratio.

Table 1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.101911	0.009915	10.27882	0.0000
A/I	0.008889	0.003911	2.272755	0.0231

The true coefficient on a/i ratio is significant and can be interpreted as follows: a 1 percent point increase in a/i ratio leads to an increase in the probability of a loan denial by $0.0089 \times 0.01 = 0.000089 \approx 0.0089$ %.

We augment the simple model, Linear Model 1, by an additional regressor black which equals 1 if the applicant is an African American and equals 0 otherwise, Linear Model 2. Such a specification is the baseline for investigating if there is racial discrimination in the mortgage market: if being black has a significant (positive) influence on the probability of a loan denial when we control for factors that allow for an objective assessment of an applicant's credit worthiness, this is an indicator for discrimination.

The estimated regression function, in Table 2, is $\hat{Y} = 0.085 + 0.010$a/iratio$+0.145$black.

The coefficient on black is positive and significantly different from zero at the 0.01% level.

Table 2

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.085244	0.009972	8.548539	0.0000
A/I	0.009656	0.003877	2.490660	0.0128
B	0.144582	0.014763	9.793547	0.0000

The interpretation is that, holding constant the a/i ratio, being black increases the probability of a mortgage application denial by about 14.5%. Another indicator for discrimination, Linear Model 3, shows the coefficient on black in Table 3 is positive and significantly different from zero at the 0.01% level. Linear Model 3 indicates that there is a negative relation between the amount of loan and the probability of a denied mortgage application so individuals with a larger loan amount to income are less likely to be rejected. The relation between the income of applicants and the probability of a denied mortgage is negative but the coefficient is statistically insignificant.

$$Y = \beta_0 + \beta_1 \text{aount} + \beta_2 \text{income} + \beta_3 \text{black} + u. \qquad \text{(Linear Model 3)}$$

The estimated regression function, in Table 3, is

$$\hat{Y} = 0.140 - 0.0002\text{aount} - 1.25\text{E-}05\text{income} + 0.135\text{black}.$$

The interpretation is that, holding constant the loan amount and income, being black increases the probability of a mortgage application denial by about 13.5%.

Table 3

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.140425	0.008543	16.43729	0.0000
A	-0.000164	4.85E-05	-3.382290	0.0007
I	-1.25E-05	6.54E-05	-0.190781	0.8487
B	0.134614	0.014878	9.047863	0.0000

These findings are compatible with racial discrimination. However, it might be distorted by omitted variable bias so discrimination could be a premature conclusion.

Additional regressors that likely have power in explaining whether a mortgage application has been denied are the applicant's sex which equals 1 if the applicant is female and equals 0 otherwise and the percentage of minority population where the property is located. It is straightforward to translate this into the simple regression model

$$Y = \beta_0 + \beta_1 \text{aount} + \beta_2 \text{income} + \beta_3 \text{black} + \beta_4 \text{sex} + \beta_5 \text{minority population percentage} + u. \qquad \text{(Linear Model 4)}$$

The estimated regression function, in Table 4, is $\hat{Y} = 0.143 - 0.0002\text{amount} - 2.11\text{E-}05\text{income} + 0.132\text{black} - 0.017\text{sex} + 0.0002\text{percentage}$.

Table 4

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.142533	0.011734	12.14678	0.0000
A	-0.000167	4.90E-05	-3.407327	0.0007
I	-2.11E-05	6.56E-05	-0.321994	0.7475
B	0.131928	0.016223	8.132214	0.0000
S	-0.017316	0.009707	-1.783758	0.0745
P	0.000248	0.000282	0.878440	0.3797

The coefficient on sex in Table 4 is positive and insignificant. Linear Model 4 indicates that there is a negative relation between the sex of applicant and the probability of a denied mortgage application so female applicants are less likely to be rejected. The relation between the percentage of minority population where the property located and the probability of a denied mortgage is positive but the coefficient is statistically insignificant.

The linear probability model has a major flaw: it assumes the conditional probability function to be linear. This does not restrict $P(Y=1|X_1,\ldots,X_k)$ to lie between 0 and 1. It is possible the probability of a mortgage application denial to be bigger than 1. For some applications, the predicted probability of denial is even negative so that the model has no meaningful interpretation here. In order to obtain a more trustworthy estimate of the effect of being black on the probability of a mortgage application denial we estimate a linear probability model as well as several Probit and Logit models.

4.2. Probit Model

In Probit regression, the cumulative standard normal distribution function $\Phi(\cdot)$ is used to model the regression function when the dependent variable is binary, that is, we assume

$$E(Y|X)=P(Y=1|X)=\Phi(\beta_0+\beta_1X). \qquad (4.2.1)$$

$\beta_0+\beta_1X$ plays the role of a quantile z. Remember that

$$\Phi(z)=P(Z\leq z),\ Z\sim N(0,1) \qquad (4.2.2)$$

such that the Probit coefficient β_1 is the change in z associated with a one unit change in X. Although the effect on z of a change in X is linear, the link between z and the dependent variable Y is nonlinear since Φ is a nonlinear function of X. We estimate the following Probit models:

$$Y = \Phi\,(\beta_0 + \beta_1\text{aount/income ratio} + u\,) \qquad \text{(Probit Model 1)}$$

$$Y = \Phi\,(\beta_0 + \beta_1\text{aount/income ratio} + \beta_2\text{black} + u\,) \qquad \text{(Probit Model 2)}$$

$$Y = \Phi\,(\beta_0 + \beta_1\text{aount/income ratio} + \beta_2\text{black} + \beta_3\text{sex} + \beta_4\text{minority population percentage} + u\,). \qquad \text{(Probit Model 3)}$$

Since the dependent variable is a nonlinear function of the regressors, the coefficient on X has no simple interpretation. The expected change in the probability that Y=1 due to a change in A/I ratio can be computed as follows:

1. Compute the predicted probability that Y=1 for the original value of X.
2. Compute the predicted probability that Y=1 for $X+\Delta X$.
3. Compute the difference between both predicted probabilities.

Of course, we can generalize (4.2.1) to Probit regression with multiple regressors to mitigate the risk of facing omitted variable bias. We now estimate Probit Model 1, a simple Probit model of the probability of a mortgage denial. The estimated model is, in Table 5,

$$\hat{P}(\text{deny}|\text{A/I ratio}) = \Phi(-1.25+0.036\text{A/I ratio}).$$

Just as in the linear probability model we find that the relation between the probability of denial and the loan amount-to-income ratio is positive and that the corresponding coefficient is significant.

Table 5

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-1.247856	0.046139	-27.04561	0.0000
A/I	0.036194	0.017590	2.057704	0.0396

We continue by using an augmented Probit model to estimate the effect of race on the probability of a mortgage application denial.
The estimated model equation, Probit Model 2, is, in Table 6, ˆP(deny|A/Iratio,black) =Φ(−1.339+0.042A/Iratio+0.576black). We augment the simple model, Logit Model 1, by an additional regressor black which equals 1 if the applicant is an African American and equals 0 otherwise. Such a specification is the baseline for investigating if there is racial discrimination in the mortgage market: if being black has a significant (positive) influence on the probability of a loan denial when we control for factors that allow for an objective assessment of an applicant's credit worthiness, this is an indicator for discrimination.

Table 6

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-1.338553	0.047717	-28.05179	0.0000
AI	0.041840	0.017653	2.370112	0.0178
B	0.575997	0.063646	9.049959	0.0000

While all coefficients are significant, both the estimated coefficients on the loan amount-to-income ratio and the indicator for African American descent are positive and significant. Again, the coefficients are difficult to interpret but they indicate that, first, African Americans have a higher probability of denial than white applicants, holding constant the loan amount-to-income ratio and second, applicants with a high loan amount-to-income ratio face a higher risk of being rejected.

How big is the estimated difference in denial probabilities between two hypothetical applicants with the same loan amount-to-income ratio? As before, we may use predict Φ ( ) to compute this difference. In this case, the estimated difference in denial probabilities is about 14.95%.

Additional regressors, Probit Model 3, that likely have power in explaining whether a mortgage application has been denied are the applicant's sex which equals 1 if the applicant is female and equals 0 otherwise and the percentage of minority population where the property is located. The estimated regression function, in Table 7, is ˆY = Φ(−1.366 + 0.043A/Iratio + 0.54black −0.051sex + 0.002minority population).

Table 7

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-1.366039	0.056136	-24.33445	0.0000
AI	0.043485	0.017713	2.455057	0.0141
B	0.539700	0.071502	7.548089	0.0000
S	-0.051414	0.048340	-1.063580	0.2875
P	0.001966	0.001319	1.489786	0.1363

The Probit Model 3 indicates that there is a negative relation between the sex of applicant and the probability of a denied application so female applicants are less likely to be rejected. The relation between the percentage of minority population where the property located and the probability of a denied mortgage is positive. However, these coefficients are statistically insignificant.

4.3. Logit Model

As for Probit regression, there is no simple interpretation of the model coefficients and it is best to consider predicted probabilities or differences in predicted probabilities. Here again, t-statistics and confidence intervals based on large sample normal approximations can be computed as usual.
The population Logit regression function is

$$P(Y=1|X_1,X_2,\ldots,X_k) = F(\beta_0+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k)$$

$$=1 / (1+e^{-(\beta_0+\beta_1X_1+\beta_2X_2+\ldots+\beta_kX_k)}).$$

The idea is like Probit regression except that a different CDF (cumulative distribution function) is used: $F(x)=1/(1+e^{-x})$ is the CDF of a standard logistically distributed random variable. Both models produce very similar estimates of the probability that a mortgage application will be denied depending on the applicant's loan amount-to-income ratio. We extend the simple Logit model of mortgage denial with the additional regressor black, sex, and minority population percentage. We estimate the following Logit models:

$Y = F(\beta_0 + \beta_1$aount/income ratio + u) (Logit Model 1)

$Y = F(\beta_0 + \beta_1$aount/income ratio + $\beta_2$black + u) (Logit Model 2)

$Y = F(\beta_0 + \beta_1$aount/income ratio + $\beta_2$black + $\beta_3$sex + $\beta_4$minority population percentage + u). (Logit Model 3)

Logit Model 1 indicates that there is a positive relation between the loan amount to income ratio and the probability of a denied mortgage application so individuals with a high ratio of loan amount to income are more likely to be rejected. According to the estimated model, in Table 8, the estimated regression line is

$\hat{Y} = F(-2.16 + 0.079$aount/income ratio)

Table 8

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-2.155001	0.091535	-23.54297	0.0000
AI	0.078827	0.034730	2.269721	0.0232

The true coefficient on a/i ratio is significant. We augment the simple model, Logit Model 1, by an additional regressor black which equals 1 if the applicant is an African American and equals 0 otherwise, Logit Model 2. Such a specification is the baseline for investigating if there is racial discrimination in the mortgage market: if being black has a significant (positive) influence on the probability of a loan denial. Logit Model 2, shows the coefficient on black in Table 9 is positive and significantly different from zero at the 0.01% level.

Table 9

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-2.317911	0.094306	-24.57855	0.0000
AI	0.085412	0.034656	2.464608	0.0137
B	1.034629	0.110295	9.380572	0.0000

How big is the estimated difference in denial probabilities between two hypothetical applicants with the same loan amount-to-income ratio? As before, we may use predict F( ) to compute this difference. In this case, the estimated difference in denial probabilities is about 15.08%. As for the Probit Model 2, coefficients are highly significant and we obtain positive estimates for the coefficients on A/I ratio and black. For comparison we compute the predicted probability of denial for two hypothetical applicants that differ in race and have an A/I ration of 3. We find that the white applicant faces a denial probability of 11.29%, while the African American is rejected with a probability of 26.37%, a difference of 15.08%.

In Table 10, the Logit Model 3 indicates that there is a negative relation between the sex of applicant and the probability of a denied application so female applicants are less likely to be rejected. The relation between the percentage of minority population where the property located and the probability of a denied mortgage is positive. However, these coefficients are statistically insignificant.

Table 10

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-2.372184	0.109131	-21.73711	0.0000
AI	0.089728	0.034762	2.581175	0.0098
B	0.967008	0.125869	7.682628	0.0000
S	-0.104316	0.091435	-1.140873	0.2539
P	0.003772	0.002434	1.549813	0.1212

5. Conclusions

All models, Linear Model 2, Probit Model 2, and Logit Model 2, indicate that denial rates are higher for African American applicants holding constant the loan amount-to-income ratio. The estimated differences in denial probabilities between two hypothetical applicants with the same loan amount-to-income ratio are 14.46% for Linear Model 2, 14.95% for Probit Model 2, and 15.08% for Logit Model 2. As for the all model coefficients are highly significant and we obtain positive estimates for the coefficients on A/I ratio and black. Both results could be subject to omitted variable bias. We estimate a linear probability model as well as several Logit and Probit models. We thereby control for financial variables and additional applicant characteristics which are likely to influence the probability of denial and differ between black and white applicants. In the Linear Model 2, the coefficients have direct interpretation. Having a high loan amount-to-income ratio is estimated to face higher risk of denial than those with a low loan amount-to-income ratio, ceteris paribus. The estimated coefficient on the race dummy, which indicates the denial probability for African Americans, is 14.46% larger than for white applicants with the same characteristics except for race. Linear Models 3 and 4 provide similar evidence that there is racial discrimination in the mortgage lending market in Little Rock metropolitan area. Using this approach, the estimate for the effect on the denial probability of being African American of the Probit Model 2 and Logit model are 14.95% and 15.08%, respectively.

The estimates of the impact on the denial probability of being black are similar for all models. The results in this paper are consistent with those of Alici Munnell et al., (1996) and Fidel Ezeala-Harrison and Glenda Glover (2008) It is important to know that the magnitude of the estimated effects can be different with other models which include omitted variables, such as credit scores, ratio of total monthly debt payments to total monthly income, and ratio of size of loan to assessed value of property. This indicates that these simple models produce biased estimates due to omitted variables. All models provide evidence that there is an effect of being African American on the probability of a mortgage application denial: in all specifications, the effect is estimated to be positive and is significantly different from zero. While the linear probability model seems to slightly underestimate this effect, it still can be used as an approximation to an intrinsically nonlinear relationship.

References

Black, Harold, Robert L. Schweitzer and Lewis Mandell. "Discrimination in Mortgage Lending." American Economic Review, May 1978, 68 (2), pp. 186-191.

Ezeala-Harrison, Fidel and Glenda Glover. "Determinants of Housing Loan Patterns Toward Minority Bowwowers in Mississippi." Journal of Economic Issues, March 2008, 42 (1), pp. 75-96.

Greene, William. 2011. *Econometrics Analysis, 7th ed.* Prentice Hall, Upper Saddle River, NJ.

Hanck, Christoph, Martin Arnold, Alexander Gerber, and Martin Schmelzer, Introduction to Econmetrics with R, 2019.

Munnell, Alicia, Geoffrey Tootell, Lynn Browne and James McEneaney. "Mortgage Lending in Boston: Interpreting HMDA Data." American Economic Review, March 1996, 86 (1), pp. 25-53.

Startz, Richard, EViews Illustrated, 2019

Table 1
Dependent Variable: Y
Method: Least Squares
Sample: 1 5207
Included observations: 5207

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.101911	0.009915	10.27882	0.0000
A/I	0.008889	0.003911	2.272755	0.0231

R-squared	0.000991	Mean dependent var	0.121951
Adjusted R-squared	0.000799	S.D. dependent var	0.327261
S.E. of regression	0.327130	Akaike info criterion	0.603466
Sum squared resid	557.0082	Schwarz criterion	0.605985
Log likelihood	-1569.124	Hannan-Quinn criter.	0.604347
F-statistic	5.165413	Durbin-Watson stat	0.006536
Prob(F-statistic)	0.023082		

Table 2
Dependent Variable: Y
Method: Least Squares
Sample: 1 5207
Included observations: 5207

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.085244	0.009972	8.548539	0.0000
A/I	0.009656	0.003877	2.490660	0.0128
B	0.144582	0.014763	9.793547	0.0000

R-squared	0.019071	Mean dependent var	0.121951
Adjusted R-squared	0.018694	S.D. dependent var	0.327261
S.E. of regression	0.324188	Akaike info criterion	0.585587
Sum squared resid	546.9279	Schwarz criterion	0.589365
Log likelihood	-1521.576	Hannan-Quinn criter.	0.586909
F-statistic	50.58658	Durbin-Watson stat	0.007474
Prob(F-statistic)	0.000000		

Table 3
Dependent Variable: Y
Method: Least Squares
Sample: 1 5207
Included observations: 5207

Variable	Coefficient	Std. Error	t-Statistic	Prob.

C	0.140425	0.008543	16.43729	0.0000
A	-0.000164	4.85E-05	-3.382290	0.0007
I	-1.25E-05	6.54E-05	-0.190781	0.8487
B	0.134614	0.014878	9.047863	0.0000

R-squared	0.022040	Mean dependent var	0.121951
Adjusted R-squared	0.021476	S.D. dependent var	0.327261
S.E. of regression	0.323728	Akaike info criterion	0.582940
Sum squared resid	545.2726	Schwarz criterion	0.587978
Log likelihood	-1513.685	Hannan-Quinn criter.	0.584702
F-statistic	39.08530	Durbin-Watson stat	0.009422
Prob(F-statistic)	0.000000		

Table 4
Dependent Variable: Y
Method: Least Squares
Sample: 1 5207
Included observations: 5207

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.142533	0.011734	12.14678	0.0000
A	-0.000167	4.90E-05	-3.407327	0.0007
I	-2.11E-05	6.56E-05	-0.321994	0.7475
B	0.131928	0.016223	8.132214	0.0000
S	-0.017316	0.009707	-1.783758	0.0745
P	0.000248	0.000282	0.878440	0.3797

R-squared	0.022749	Mean dependent var	0.121951
Adjusted R-squared	0.021809	S.D. dependent var	0.327261
S.E. of regression	0.323673	Akaike info criterion	0.582983
Sum squared resid	544.8771	Schwarz criterion	0.590539
Log likelihood	-1511.795	Hannan-Quinn criter.	0.585626
F-statistic	24.21429	Durbin-Watson stat	0.010242
Prob(F-statistic)	0.000000		

Table 5
Dependent Variable: Y
Method: ML - Binary Probit (Newton-Raphson / Marquardt steps)
Sample: 1 5207
Included observations: 5207
Convergence achieved after 5 iterations
Coefficient covariance computed using observed Hessian

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-1.247856	0.046139	-27.04561	0.0000

A/I	0.036194	0.017590	2.057704	0.0396

McFadden R-squared	0.001100	Mean dependent var	0.121951
S.D. dependent var	0.327261	S.E. of regression	0.327036
Akaike info criterion	0.741542	Sum squared resid	556.6888
Schwarz criterion	0.744061	Log likelihood	-1928.605
Hannan-Quinn criter.	0.742423	Deviance	3857.209
Restr. deviance	3861.456	Restr. log likelihood	-1930.728
LR statistic	4.247070	Avg. log likelihood	-0.370387
Prob(LR statistic)	0.039318		
Obs with Dep=0	4572	Total obs	5207
Obs with Dep=1	635		

Table 6
Dependent Variable: Y
Method: ML - Binary Probit (Newton-Raphson / Marquardt steps)
Sample: 1 5207
Included observations: 5207
Convergence achieved after 5 iterations
Coefficient covariance computed using observed Hessian

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-1.338553	0.047717	-28.05179	0.0000
AI	0.041840	0.017653	2.370112	0.0178
B	0.575997	0.063646	9.049959	0.0000

McFadden R-squared	0.021339	Mean dependent var	0.121951
S.D. dependent var	0.327261	S.E. of regression	0.324167
Akaike info criterion	0.726917	Sum squared resid	546.8568
Schwarz criterion	0.730695	Log likelihood	-1889.529
Hannan-Quinn criter.	0.728239	Deviance	3779.058
Restr. deviance	3861.456	Restr. log likelihood	-1930.728
LR statistic	82.39846	Avg. log likelihood	-0.362882
Prob(LR statistic)	0.000000		
Obs with Dep=0	4572	Total obs	5207
Obs with Dep=1	635		

Table 7
Dependent Variable: Y
Method: ML - Binary Probit (Newton-Raphson / Marquardt steps)

Sample: 1 5207
Included observations: 5207
Convergence achieved after 5 iterations
Coefficient covariance computed using observed Hessian

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-1.366039	0.056136	-24.33445	0.0000
AI	0.043485	0.017713	2.455057	0.0141
B	0.539700	0.071502	7.548089	0.0000
S	-0.051414	0.048340	-1.063580	0.2875
P	0.001966	0.001319	1.489786	0.1363

McFadden R-squared	0.022135	Mean dependent var	0.121951
S.D. dependent var	0.327261	S.E. of regression	0.324058
Akaike info criterion	0.727095	Sum squared resid	546.2817
Schwarz criterion	0.733392	Log likelihood	-1887.991
Hannan-Quinn criter.	0.729297	Deviance	3775.982
Restr. deviance	3861.456	Restr. log likelihood	-1930.728
LR statistic	85.47413	Avg. log likelihood	-0.362587
Prob(LR statistic)	0.000000		
Obs with Dep=0	4572	Total obs	5207
Obs with Dep=1	635		

Table 8
Dependent Variable: Y
Method: ML - Binary Logit (Newton-Raphson / Marquardt steps)
Sample: 1 5207
Included observations: 5207
Convergence achieved after 5 iterations
Coefficient covariance computed using observed Hessian

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-2.155001	0.091535	-23.54297	0.0000
AI	0.078827	0.034730	2.269721	0.0232

McFadden R-squared	0.001291	Mean dependent var	0.121951
S.D. dependent var	0.327261	S.E. of regression	0.326964
Akaike info criterion	0.741400	Sum squared resid	556.4425
Schwarz criterion	0.743919	Log likelihood	-1928.235
Hannan-Quinn criter.	0.742281	Deviance	3856.471
Restr. deviance	3861.456	Restr. log likelihood	-1930.728

LR statistic	4.985492	Avg. log likelihood	-0.370316
Prob(LR statistic)	0.025561		

Obs with Dep=0	4572	Total obs	5207
Obs with Dep=1	635		

Table 9
Dependent Variable: Y
Method: ML - Binary Logit (Newton-Raphson / Marquardt steps)
Sample: 1 5207
Included observations: 5207
Convergence achieved after 5 iterations
Coefficient covariance computed using observed Hessian

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-2.317911	0.094306	-24.57855	0.0000
AI	0.085412	0.034656	2.464608	0.0137
B	1.034629	0.110295	9.380572	0.0000

McFadden R-squared	0.021399	Mean dependent var	0.121951
S.D. dependent var	0.327261	S.E. of regression	0.324150
Akaike info criterion	0.726873	Sum squared resid	546.8019
Schwarz criterion	0.730651	Log likelihood	-1889.413
Hannan-Quinn criter.	0.728194	Deviance	3778.826
Restr. deviance	3861.456	Restr. log likelihood	-1930.728
LR statistic	82.62991	Avg. log likelihood	-0.362860
Prob(LR statistic)	0.000000		

Obs with Dep=0	4572	Total obs	5207
Obs with Dep=1	635		

Table 10
Dependent Variable: Y
Method: ML - Binary Logit (Newton-Raphson / Marquardt steps)
Sample: 1 5207
Included observations: 5207
Convergence achieved after 5 iterations
Coefficient covariance computed using observed Hessian

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-2.372184	0.109131	-21.73711	0.0000
AI	0.089728	0.034762	2.581175	0.0098
B	0.967008	0.125869	7.682628	0.0000

S	-0.104316	0.091435	-1.140873	0.2539
P	0.003772	0.002434	1.549813	0.1212

McFadden R-squared	0.022280	Mean dependent var	0.121951
S.D. dependent var	0.327261	S.E. of regression	0.324020
Akaike info criterion	0.726987	Sum squared resid	546.1541
Schwarz criterion	0.733284	Log likelihood	-1887.711
Hannan-Quinn criter.	0.729190	Deviance	3775.421
Restr. deviance	3861.456	Restr. log likelihood	-1930.728
LR statistic	86.03468	Avg. log likelihood	-0.362533
Prob(LR statistic)	0.000000		

Obs with Dep=0	4572	Total obs	5207
Obs with Dep=1	635		

Appendix I

HMDA Loan Application Register Format
(File format is comma separated value)

Fields	**Maximum Length**	**Type**
As of Year	4	Numeric
Respondent ID	10	Alphanumeric
Agency Code	1	Alphanumeric
Loan Type	1	Numeric
Property Type	1	Alphanumeric
Loan Purpose	1	Numeric
Occupancy	1	Numeric
Loan Amount (000s)	5	Numeric
Preapproval	1	Alphanumeric
Action Type	1	Numeric
MSA/MD	5	Alphanumeric
State Code	2	Alphanumeric
County Code	3	Alphanumeric
Census Tract Number	7	Alphanumeric
Applicant Ethnicity	1	Alphanumeric
Co Applicant Ethnicity	1	Alphanumeric
Applicant Race 1	1	Alphanumeric
Applicant Race 2	1	Alphanumeric
Applicant Race 3	1	Alphanumeric
Applicant Race 4	1	Alphanumeric
Applicant Race 5	1	Alphanumeric
Co Applicant Race 1	1	Alphanumeric
Co Applicant Race 2	1	Alphanumeric
Co Applicant Race 3	1	Alphanumeric
Co Applicant Race 4	1	Alphanumeric
Co Applicant Race 5	1	Alphanumeric
Applicant Sex	1	Numeric
Co Applicant Sex	1	Numeric
Applicant Income (000s)	4	Alphanumeric
Purchaser Type	1	Alphanumeric
Denial Reason 1	1	Alphanumeric
Denial Reason 2	1	Alphanumeric
Denial Reason 3	1	Alphanumeric
Rate Spread	5	Alphanumeric
HOEPA Status	1	Alphanumeric
Lien Status	1	Alphanumeric
Edit Status	1	Alphanumeric
Sequence Number	7	Alphanumeric
Population	8	Alphanumeric
Minority Population %	6	Alphanumeric
FFIEC Median Family Income	8	Alphanumeric
Tract to MSA/MD Income %	6	Alphanumeric
Number of Owner-occupied units	8	Alphanumeric

Number of 1-to 4-Family units	8	Alphanumeric
Application Date Indicator	1	Numeric

Appendix II

HMDA LOAN APPLICATION REGISTER CODE SHEET

RESPONDENT INFORMATION
Respondent ID: 10 Character Identifier
Agency
: 1 -- Office of the Comptroller of the Currency (OCC)
2 -- Federal Reserve System (FRS)
3 -- Federal Deposit Insurance Corporation (FDIC)
5 -- National Credit Union Administration (NCUA)
7 -- Department of Housing and Urban Development (HUD)
9 -- Consumer Financial Protection Bureau (CFPB)

LOAN INFORMATION
Loan Type:
1 -- Conventional (any loan other than FHA, VA, FSA, or RHS loans)
2 -- FHA-insured (Federal Housing Administration)
3 -- VA-guaranteed (Veterans Administration)
4 -- FSA/RHS (Farm Service Agency or Rural Housing Service)
Property Type:
1 -- One to four-family (other than manufactured housing)
2 -- Manufactured housing
3 -- Multifamily
Loan Purpose:
1 -- Home purchase
2 -- Home improvement
3 -- Refinancing
Owner-Occupancy:
1 -- Owner-occupied as a principal dwelling
2 -- Not owner-occupied
3 -- Not applicable
Loan Amount: in thousands of dollars
Preapproval:
1 -- Preapproval was requested
2 -- Preapproval was not requested
3 -- Not applicable
Action Taken:
1 -- Loan originated
2 -- Application approved but not accepted
3 -- Application denied by financial institution
4 -- Application withdrawn by applicant
5 -- File closed for incompleteness
6 -- Loan purchased by the institution
7 -- Preapproval request denied by financial institution
8 -- Preapproval request approved but not accepted (optional reporting)

PROPERTY LOCATION
MSA/MD: Metropolitan Statistical Area/Metropolitan Division

State: Two-digit FIPS state identifier
County: Three-digit FIPS county identifier
Tract: Census tract number

APPLICANT INFORMATION

Ethnicity:
- 1 -- Hispanic or Latino
- 2 -- Not Hispanic or Latino
- 3 -- Information not provided by applicant in mail, Internet, or telephone application
- 4 -- Not applicable
- 5 -- No co-applicant

Race:
- 1 -- American Indian or Alaska Native
- 2 – Asian
- 3 -- Black or African American
- 4 -- Native Hawaiian or Other Pacific Islander
- 5 – White
- 6 -- Information not provided by applicant in mail, Internet, or telephone application
- 7 -- Not applicable
- 8 -- No co-applicant

Sex:
- 1 – Male
- 2 – Female
- 3 -- Information not provided by applicant in mail, Internet, or telephone application
- 4 -- Not applicable
- 5 -- No co-applicant

Gross Annual Income: in thousands of dollars

PURCHASER AND DENIAL INFORMATION

Type of Purchaser
- 0 -- Loan was not originated or was not sold in calendar year covered by register
- 1 -- Fannie Mae (FNMA)
- 2 -- Ginnie Mae (GNMA)
- 3 -- Freddie Mac (FHLMC)
- 4 -- Farmer Mac (FAMC)
- 5 -- Private securitization
- 6 -- Commercial bank, savings bank or savings association
- 7 -- Life insurance company, credit union, mortgage bank, or finance company
- 8 -- Affiliate institution
- 9 -- Other type of purchaser

Reasons for Denial:
- 1 -- Debt-to-income ratio
- 2 -- Employment history
- 3 -- Credit history
- 4 – Collateral
- 5 -- Insufficient cash (downpayment, closing costs)
- 6 -- Unverifiable information
- 7 -- Credit application incomplete
- 8 -- Mortgage insurance denied

9 -- Other

OTHER DATA
Rate Spread
HOEPA Status (only for loans originated or purchased):
1 -- HOEPA loan
2 -- Not a HOEPA loan
Lien Status (only for applications and originations):
1 -- Secured by a first lien
2 -- Secured by a subordinate lien
3 -- Not secured by a lien
4 -- Not applicable (purchased loans)
Edit Status:
Blank -- No edit failures
5 -- Validity edit failure only
6 -- Quality edit failure only
7 -- Validity and quality edit failures
Sequence Number: One-up number scheme for each respondent to make each loan unique

CENSUS INFORMATION
Population: total population in tract.
Minority Population %: percentage of minority population to total population for tract. (Carried to two decimal places)
FFIEC Median Family Income: FFIEC Median family income in dollars for the MSA/MD in which the tract is located (adjusted annually by FFIEC).
Tract to MSA/MD Median Family Income Percentage: % of tract median family income compared to MSA/MD median family income. (Carried to two decimal places)
Number of Owner Occupied Units: Number of dwellings, including individual condominiums, that are lived in by the owner.
Number of 1- to 4-Family units: Dwellings that are built to house fewer than 5 families.
Application Date Indicator
0 -- Application Date >= 01-01-2004
1 -- Application Date < 01-01-2004
2 -- Application Date = NA (Not Available)

Appendix III

Dependent Variable Frequencies

Dep. Value	Count	Percent	Cumulative Count	Percent
0	4572	87.80	4572	87.80
1	635	12.20	5207	100.00