Race, Gender & Class: Volume 21, Number 1-2, 2014 (48-63)

Gender & Class Website: www.rgc.uno.edu

Race in the Mortgage Market: An Empirical Investigation Using HMDA Data

Madhavi Cherian
Sociology
New York University

Abstract: This paper investigates the role of race in the mortgage market using data from the Home Mortgage Disclosure Act and the census from 1992 to 2003 for the Indianapolis Metropolitan Statistical Area. A multinomial logit model is used to isolate the effect of race on the final outcome of a loan application, after accounting for individual, loan, neighborhood, and property characteristics. The findings of this paper are consistent with those of previous works in this field, that is, being black or Hispanic lowers the probability of a successful loan application. Although data limitations prevent this paper from claiming explicit discrimination in the mortgage market, the project demonstrates that race is a statistically significant variable in predicting the outcome of a mortgage application.

Keywords: race; housing; mortgage; discrimination

Madhavi Cherian is a Doctoral Candidate in Sociology at New York University. She is currently working on the politics behind the Public Distribution System for subsidized foods in India and its effectiveness as an anti-poverty tool. Her background is in Economics with a BA (Honors) from Delhi University and an MA from Indiana University at Indianapolis. Her interests include political sociology, poverty and inequality.

Address: New York University, 295 Lafayette St., 4th Floor, New York, NY 10012. Ph.: (317) 457-2117. Email: mcherian@nyu.edu

The 2000 census data in the Indianapolis Metropolitan Statistical Area (MSA) reveal differential rates of home ownership among races. While the average homeownership rate is 62 percent, the homeownership rate for African Americans is 47 percent (Table 1). This could be indicative of differences in access to capital from the mortgage market, since mortgage financing is an important part of the home buying process. As the Home Mortgage Disclosure Act (HMDA) data show, the probability of loan origination, that is, an application resulting in the lending agency issuing a loan, is lower for nonwhites than whites at all income levels.

Table 1: *Population in Occupied and Rental Housing Units in the Indianapolis MSA (Figures in parentheses are percentages)*

	Total population in occupied housing units	Owner occupied	Renter occupied
All population	774,011	482,515 (62)	291,496 (38)
African-American	195,104	91,870 (47)	103,234 (53)

Source: US Census Bureau, 2000

Studies have linked home ownership with various lifetime benefits such as better health and psychological well-being. Home ownership is also associated with intergenerational benefits for children and other obvious benefits like increased home equity. These positive effects are amplified in the case of low-income households, which stand to benefit the most from the stability that an owned home can provide (Harkness & Newman, 2003). Fair access to mortgage credit is an important step towards ensuring that benefits of homeownership accrue to all races.

Using information from the HMDA data and the census for the Indianapolis MSA for the 1992-2003 period, this paper studies the relative weights of various factors in the lender's decision-making process with the aim of isolating the role of race. I use a multinomial logit model to determine the role of race in the outcome of a loan application. Race consistently shows up as a sizeable and statistically significant factor. The role of race cannot be attributed to lower income and other neighborhood level characteristics of nonwhites, since I control for the effects of these variables in my model. However, this paper cannot claim that explicit discrimination exists in the mortgage market, due to lack of information on credit scores and loan–to-value ratios which are needed to replicate the lender's decision making model more precisely.

Discrimination in the Housing Market

Discrimination in the housing and home loan markets can take several forms. Ross and Yinger (2002) describe the following forms of discrimination. Blatant discrimination, as the term suggests, is the overt practice of discrimination based on a particular characteristic of the applicant, such as race. This sort of discrimination would be easily noticed since loan originations to members of a particular racial group would be absent.

The second type of discrimination is disparate treatment. This means that factors that would not play a role in the lender's decision for applicants of a certain race are given importance for members of other races. The lender would attribute different weights to the factors considered (like income and debt) in the decision making process for different races. This would be the case if the lender believed that the race embodies certain unobserved characteristics that predict loan performance. This would be more difficult to spot in a dataset. With the data that are available, missing variables such as loan-to-value ratio and credit history could be used to explain away suggestions of disparate treatment on the basis of race.

The third type of discrimination is adverse impact, that is, when a seemingly innocuous measure adversely affects a particular group. Although the lender does not explicitly consider race, factors negatively correlated with a particular race would have a greater weight in the lender's decision-making framework. For example, a lender could decide that only people belonging to a certain club or institution will be approved for loans. If minorities were less likely to belong to that club, their chances for a successful loan application would be considerably diminished even though this factor has nothing to do with subsequent loan performance. Like disparate treatment, adverse impact in the mortgage market would be difficult to prove with the data used for this project. We would have to know the specific criteria employed by a particular lending institution and whether it had disparate impact on white and nonwhite borrowers, respectively.

Another form of discrimination prevalent in the mortgage market is redlining, that is, when the lender's decision is affected by the racial characteristics of the location of the property, rather than the applicant's individual characteristics. In this case, the probability of loan approval decreases when the house is in a nonwhite neighborhood, in spite of otherwise equivalent economic variables. The present study includes neighborhood level variables to account for the effect of neighborhood characteristics on the probability of loan origination.

Race, Mortgage and Homeownership: Prior Works

A number of studies have examined the role of race in the mortgage market to explain racial discrepancies in loan access as well as homeownership rates. All the studies find differences in mortgage origination rates for white and nonwhite applicants. Most of these works use HMDA data as their primary data. Due to shortcomings in HMDA data, researchers augment HMDA with the census or some other data. The studies below represent a variety of techniques that have been used to study the relationship between race and access to mortgage credit and consequently, the probability of homeownership.

Munnell, Tootell, Browne, and McEneaney (1996) supplemented HMDA data with additional information, such as financial, personal, and property characteristics for blacks, Hispanics, and a random sample of 3300 whites, collected by the Federal Reserve Bank of Boston from banks in the Boston MSA. The authors used logit and ordinary least squares to estimate and interpret differences in loan denial rates on the basis of race. The paper concluded that black and Hispanic applicants have a greater probability of denial, given the same characteristics as whites. The data set used here did not include the credit score of each applicant (but the authors did construct their own measure of credit history) and the loan-to-value ratio. The Federal Reserve Bank also collected information on each applicant's employment and the authors' constructed a measure to predict the stability of employment in the applicant's occupational field. They also knew whether the applicant purchased private mortgage insurance, since that would indicate the cost of default. They found that given equivalent characteristics, mortgage denial rates for nonwhites were 8 percentage points higher than that for whites. Additionally, separate white and nonwhite equations were used to predict denial rates for the different groups. Using the coefficients in the white equation, the authors predicted denial rates for nonwhites. The denial rates for nonwhites predicted using white coefficients are a function of the lower economic and neighborhood characteristics of nonwhites. The actual denial rates were a lot higher at 28 percent than the predicted 21 percent. The difference between the actual nonwhite denial rate and that based on the coefficients of the white equation is the effect of race on denial rates.

Collins and Margo (1999) use Integrated Public Use Microdata Series (IPUMS) to investigate the correlation between race and the probability of home ownership. Using IPUMS data enables the authors to track the changes in the relationship between race and home ownership since the beginning of the century. This study uses a linear probability model to estimate the effect of race on home ownership holding other characteristics like age, income, marital status, family size, and region of residence and migrant status constant. The paper posited mortgage discrimination as a possible reason for different rates of home ownership among different races. It concluded that blacks are less likely than whites to hold a mortgage. They also found that holding individual, loan and neighborhood characteristics constant, the probability of loan origination is much higher for whites than for blacks.

Avery, Beeson and Sniderman (1993) examined the prevalence and causes of racial discrimination in the mortgage market. This paper explores whether racial discrimination arises because of characteristics other than race associated with that group, neighborhood features (low income neighborhood with low property values) or some other reason. The authors use HMDA data from 1990 to construct a fixed effect linear probability model to decompose racial differences in application denial rates. It assumes that applicants' risk is represented through their economic characteristics. It concluded that widespread racial differences in denial rates exist across different loan products and geographical regions. Property location is advanced as a possible reason since nonwhite neighborhoods see less property appreciation and may be considered more risky by the lender. Appraisers, too, point out that valuing property in a low income or nonwhite neighborhood is more difficult since the housing stock tends to be older and more heterogeneous.

Weink (1992) explored shortcomings in the existing housing discrimination research. His paper concluded that discrimination in the housing credit market does exist. However, HMDA cannot be used to estimate the effects of discrimination since it lacks information on a number of key variables such as property rates and interest rate that are needed to replicate the lenders decision-making process. HMDA also does not include information on lenders' treatment of potential borrowers. Consequently, Weink's paper is hesitant to point to specific causes of differing outcomes and instead cites other studies that rule out explanations such as asset risk of the property and creditworthiness of applicants in explaining the lower mortgage rates among nonwhites.

Wyly and Holloway (1999) reconstructed the pathbreaking 1988 study, "The Color of Money: Mortgage Redlining in Atlanta," to examine whether there were any changes in the incidence of redlining in Atlanta since the original study. To examine whether mortgage origination rates in black neighborhood had changed, the authors reconstructed neighborhood taxonomies to make neighborhood racial composition and income levels comparable to those in the census tracts in the original study. Using HMDA data from 1992 to 1996, the authors found that there was a slight improvement in the later period. The ratio of number of loans originated per owner-occupied unit in middle income white neighborhoods to those originated in middle income black neighborhoods declined in the 1990s. However, the number of mortgage origination was still higher for white neighborhoods.

Buist, Megbolugbe and Trent (1994) wrote on the importance of the preapplication process using 1990 Census data and HMDA data. The authors found that if nonwhite applicants are not discouraged in the earlier stages of the application process, the numbers of originations increase. Specifically, the authors used HMDA data to calculate the percentage of mortgage applications for each race by dividing the number of mortgage applications by the number of rental units occupied in each MSA. This figure was the lowest for African Americans at about 3.1 percent. Additionally, the home ownership rates, as measured by the ratio of observed homeowners to renters for a particular race in an MSA, were the lowest for African Americans at 40.8 percent. The home ownership rate is regressed on a number of variables, including the percentage of mortgage applications of a particular race. They find that the mortgage application rate is statistically significant predictor of loan origination. It finds that if minorities were encouraged to apply instead of being prescreened, rejection rates would increase but so would the volume of home purchases. It also emphasized the importance of availability of affordable housing since the high mortgage expenditure-to- income ratio has a significant adverse effect on the home ownership rates amongst blacks and Hispanics. The availability of affordable housing would also help in decreasing loan default rates, as it would lower monthly mortgage payments.

An alternative way to test for race-based discrimination is paired testing. The National Fair Housing Alliance (NFHA) conducted a study of this type in 1993, wherein testers posing as first time homebuyers approach lenders with enquiries. At this stage, potential borrowers only want to collect information to determine the loan products that they can afford and do not make a formal application at this stage. The NFHA conducted these tests in seven cities and about two-thirds of the tests were concentrated in Chicago and Oakland. Smith and DeLair (1999) reconstructed the NFHA data in 1998, but restricted the pairing to an African American tester in a minority neighborhood and a white tester in a white neighborhood. The mortgage application contained information on total debt payment-to-income ratio and the prospective mortgage loan payment-to-income ratio, respectively. The actual purchase amounts, income, and debt were not used since property values tend to be different in African American and white neighborhoods. The two outcome measures were the information made available and the time the loan officer spent with each tester. In four of the five cities, African Americans were more likely to be denied a quote than whites and have loan officers spend less time with them. The differences in borrower treatment were statistically significant in Chicago and Atlanta, but were rare and insignificant in Oakland.

The above studies provide us with some of the approaches that have been taken to understand the role of race in obtaining mortgages and purchasing homes throughout the country as well as in specific regions up to 2003. All of these works find that race and homeownership rates are correlated and the likely culprit is differential access to mortgage credit between whites and nonwhites. Both regression techniques and paired testing reveal that whites and nonwhites with similar socio-economic characteristics face different outcomes in the mortgage market. Additionally, they shed light on mechanisms such as prescreening of nonwhite applicants, which lead to fewer successful loan applications among nonwhites. From a methodological perspective, these studies demonstrate the importance of including a number of non-HMDA variables in any study of race and mortgage outcomes. Like some of the previous studies, I use regression analysis to see if race explains any of the individual level variation in access to mortgage credit.

Framework

To isolate the effect of race, I construct a model that is an estimation of a bank's lending model (Equation 1). I include a number of economic factors that are expected to affect a bank's decision. However, race is included as an explicit factor in the model. A statistically significant race variable would mean that belonging to a particular race would alter the probability of an outcome of a mortgage application, after controlling for a number of other factors. In this model, the dependent variable is the final outcome of the application, described in Table 2. The independent variables consist of individual, loan, property, neighborhood characteristics, and factors affecting asset risk, presented in Table 3. Race and income are considered both at the individual and the neighborhood level. Since HMDA does not provide information on individual credit scores or the loan-to-value ratio, this model cannot be an exact replication of the lender's decision-making process. The credit score of an applicant is one of the most important and widely used tools available to the lender to evaluate an applicant's propensity to make future payments.

Equation 1

$$Outcome = \beta_0 + \beta_1(race) + \beta_2(income) + \beta_3(assetrisk) + \beta_4(year) + \beta_5(loan) + \beta_6(property) + \varepsilon_i$$

Data and Variable Description

In this paper, I use HMDA and census data for the Indianapolis MSA from 1992-2003 to isolate the effect of race on mortgage application outcomes. Congress enacted HMDA in 1975. HMDA requires certain lending institutions to submit information about the applicant, the loan, the type as well as the location of property, and the final outcome of the loan application. This regulation applies to certain financial institutions including banks, savings associations, credit unions, and other mortgage lending institutions. HMDA allows us to view data on lending practices of banks, determine whether a bank is serving the needs of its community, and look for indication of discrimination in lending patterns. Using loan data submitted by financial institutions, the Federal Financial Institutions Examination Council creates aggregate reports for each MSA that are available to the public at central data depositories (History and Background of HMDA, n.d.).

For this project, I consider 11 counties in central Indiana comprising the Indianapolis MSA. (The data from 1992 to 1999 does not contain information for Putnam and Brown County since they were not considered part of the MSA. However, the 2000 census included the two counties as part of the MSA and are therefore part of the final dataset.) Additionally, I use the 1990 and 2000 census by using census tract level variables as indicators of neighborhood characteristics.

In this paper, I restrict the study to loan applications meant for home purchase. I exclude loans meant for refinancing and home improvement because HMDA provides limited knowledge on loans meant as home equity lines of credit. Also, individuals applying for home improvement or refinancing loans are already own a home.

Lending institutions have an option of reporting six types of application outcomes. I excluded observations with an outcome of "loan purchased by your institution", because a loan purchased by an institution would perfectly predict loan origination. I classify loan outcomes as loan origination, loan denied and *other outcomes*, which includes loan applications with outcomes of application withdrawn, incomplete or application approved but not accepted. Consequently, the dependent variable, *Application Outcome*, is modified to consist of three types of outcomes (Table 2).

Table 2: *Description of Dependent Variable*

Application Outcome	Type of action	Observations
1	Loan originated	955,239
2	Application approved but not accepted Application withdrawn by applicant File closed for incompleteness	285,219
3	Loan denied	298,179

To explain the variation in loan application outcomes, I consider both economic and non-economic factors, as economic factors may be correlated with both the applicant's race and the lender's decision. Excluding relevant economic variables could lead to biased estimates of the impact of race on the final outcome of a loan application.

In this paper, I use individual level variables such as race and gender as well as economic measures of individuals, that is, applicant income and loan-to-income ratio as a measure of the potential indebtedness of the applicant. As HMDA does not require banks to report the total debt-to-income ratio, I use the loan-to-income ratio as an indicator of the mortgage burden of the applicant. For individual characteristics such as race and gender, I construct binary variables. (Binary variables can take two values, either 0 or 1 and are also known as dummy variables). In the case of race, I include binary variable for black, Asian, Native Americans, and Hispanic in the equation to see the effect of belonging to these races on the mortgage application outcome, relative to being white. For gender, I include a binary variable for male. Additionally, I include loan value and whether the property will be occupied by the owner or used for some other purpose once purchased. I also include a time variable in the form of year to capture any changes that happen with time.

A number of neighborhood level variables are used to approximate the lenders' decision framework. Lending institutions submit information on the census tract where the property is located. I use the census to obtain additional variables such as population, number of vacant homes, and housing units occupied by renters and owners in each census tract. To use data on neighborhood characteristics (which are not included in the HMDA data set), I include census data for 1990 and 2000. Neighborhood level variables determine property value and are important as they indicate the cost of default for the bank. Lenders will incorporate both the risk as well as the cost of default in their decision-making models (Munnell, Tootell, Browne, & McEneaney, 1996). I use the ratio of owned homes to total population as well as that of vacant homes to total homes in a census tract as a proxy for asset price risk. Additionally, I use minority population percentage and the median income of a census tract as measures of neighborhood characteristics and to see whether they explain any variation in loan application outcome. The number of loan applications as a proportion of total population of the census tract is used as an indicator of the economic activity of the area, which will enable the lending institution to generate more accurate appraisals. The summary statistics are presented in Table 3.

Multinomial Logit Model

Since there are three categorical outcomes, Equation 1 is estimated using a multinomial logit model. A multinomial logit model is an extension of the binomial logit model, with multiple instead of two possible outcomes. Both these models are based on the random utility model, which means that an agent will pick the outcome that will maximize their utility (Kennedy, 1998). The utility to the lender and the borrower of each alternative in the random utility model is specified as a linear function of characteristics such as the individual, neighborhood, and that of the property of the potential borrower, and the error term in Equation 1. The random utility model suggests that the lender and the borrower will pick the alternative that will lead to the highest utility to the respective agents.

Table 3: *Summary Statistics*

Variable	Observati ons	Mean	Std Dev	Min	Max
Loan Amount (in '000s)	1,538,637	87.23	74.44	1	1000
Applicant Income (in '000s)	1,436,699	61.82	54.27	1	1000
Loan to Income Ratio	1,436,699	1.58	.99	.001	5
Minority Population Percent	1,4947,19	12.8	22.37	0	100
Loan Applications/ Total Households	869,433	2.59	1	.14	27.06
Owned Home Ratio	869,433	. 71	.16	.06	.96
Vacancy Ratio	1,404,800	.06	.04	.01	.34
Median Income of Census Tract (in '000s)	1,497,903	44.5	17.3	0	140.2 1
Male*	1,483,859	.67	.46	0	1
Owner Occupied *	1,538,637	.94	.23	0	1
White*	1,252,288	.87	.33	0	1
Black*	1,252,288	.09	.29	0	1
Hispanic*	1,252,288	.01	.10	0	1
Native American*	1,252,288	.003	.06	0	1
Asian*	1,252,288	.009	.09	0	1

*Binary Variables
Year variables are not shown for reasons of space

In this case, the multiple outcomes are loan originated, loan denied and *other outcomes* (which comprises applications withdrawn, incomplete applications and applications approved but not accepted). With these three outcomes, a multinomial logit is roughly equivalent to running two binary logits comparing two outcomes to an arbitrarily chosen baseline outcome. This is possible while estimating a multinomial logit model since all of the logits are estimated simultaneously, which enforces the logical relationship among the parameters and uses the data more efficiently (Long 1997:150). In order to correct the standard errors for heteroskedasticity, the model is estimated using the robust standard error option. Additionally, I find that the Independence of Irrelevant Alternatives assumption is valid, which makes the multinomial logit model appropriate to use (see Appendix A).

To interpret the multinomial logit model, the paper uses marginal effects. The marginal effect of a given variable is the change in the dependent variable with respect to a unit change in the independent variable. These changes are calculated at the mean values of the independent variables. Marginal effects are estimated using the coefficients from the multinomial logit estimation. They are useful in interpreting the changes in predicted probability for both continuous and discrete independent variables. In the case of discrete independent variables, marginal effects tell us the changes in the probability of the dependent variable in case of a change from 0 to 1. For continuous variables, the marginal effects can be interpreted as the change in probability for a unit change of that variable.

Table 4: *Marginal Effects of Select Variables*
Number of Observations= 603,016

Variable Name	Loan Originated Pr(Y)=.68	Other Outcomes Pr(Y)=.15	Application Denied Pr(Y)=.15
Loan Amount	.0009***	-.0001***	-.0008***
Applicant Income	.0007***	.00006**	-.0007***
Male	.018***	-.002*	-.015***
Loan to Income	.009***	.008	-.017***
Owner occupied	-.026***	.006***	.02***
Black	-.09***	.039***	.05*
Black*Income	-.0004***	.00007*	.0003***
Hispanic	-.035***	.005	.029***
Minority Population %	-.0004***	.0003***	.0001***
Median Income of Census Tract	-.003***	-.0016***	-.002***
Owned Home Ratio	-.1109***	.034***	.066***
Vacant Home Ratio	-.303***	.15***	.22***
Loan Applications	.002***	-.002***	-.0002

Marginal effects for Years, Native Americans and Asians not shown.
*p < 0.1. **p < 0.01. ***p < 0.001. (Level of Statistical Significance)

Results

Table 4 presents the marginal effects for select variables for the three different categories of the dependent variable. Row 1 in Table 4 presents the predicted probabilities for each outcome at the mean values of the independent variables. The marginal effects in each column can be interpreted as the magnitude of change in the predicted probability with a unit change in the independent variable. (I do not present the marginal effects for the dummy variable for year or for Native American and Asian.)

Race

In this model, I exclude the dummy variable for white, which means the coefficient for race tells us the changes in probability for a particular outcome, relative to being white. The coefficient for race is statistically significant for all races, except Asian. The probability of a loan being originated is 9 percentage points lower for blacks than for whites. For Hispanics, it is 4 percentage point lower than for whites. This indicates that being black or Hispanic has a negative effect on the probability of loan origination. This difference suggests that even after taking other factors into account, lending institutions may view the race of an applicant as embodying certain characteristics that can impact the final outcome of the application. The interaction variable between *black* and *applicant income* tells us that if the applicant is black, the marginal effect of applicant income on the probability of loan origination decreases. Specifically, as income increases by a \$1000, the probability of loan origination increases by .0007 for an average applicant who is white, but only by .0003 for a black applicant. The probability of a loan being denied is 5 percentage points higher for blacks compared to whites and 3 percentage points higher for Hispanics. Blacks are also more likely to have outcomes such as withdrawn and incomplete applications and files being closed, relative to whites.

Applicant Income

The marginal effect of applicant income is small, but is statistically significant in deciding whether the loan application is rejected or originated. Higher income means that the applicant is likely to have a greater proportion of income available to pay towards the mortgage obligation. If applicant income increases by \$1000, the probability of loan origination will increase by .0007 percentage points. However, as noted earlier, being black reduces the marginal effect of income on loan origination. Interestingly, being male also increases the probability of loan origination by 1.8 percentage points.

Neighborhood Characteristics and Factors Affecting Asset Risk

If the buyer defaults, then the bank should be able to repossess the property and recover the value of the loan. In this model, I find several neighborhood (census tract) level characteristics to be statistically significant. Firstly, I find a positive statistically significant relationship between the numbers of loan applications in a census tract and the probability of loan origination. The number of loan applications relative to the number of total houses in a neighborhood is an indicator of the degree of economic activity in the area. If lenders do not make loans in a neighborhood, there will be inadequate information on the values of the home, thus increasing the risk of the loan to the lenders. If there is not too much market activity in the housing sector in a particular census tract, it would be difficult for the appraisers to evaluate the property. This would lead to lower quality appraisals, thus driving down the value of the home. This can lower the quality of the application, especially if the applicant lives in a poorer area (as indicated by the median income).

I use two additional neighborhood level factors to proxy asset risk: *vacant home ratio* and *owned home ratio*. The higher the proportion of vacant homes to total homes, the lower the chances of origination. This is not surprising as the proportion of vacant homes is indicative of asset price risk in the neighborhood. The ratio of owned homes to total homes has a negative effect on the probability of the loan being originated. This result seems rather puzzling, since one would expect that neighborhoods with high home ownership rates would be less risky in terms of property prices. A possible explanation could be that a unit increase in home ownership rate in a census tract leads to an increase in the property value of the housing units (Rohe & Stewart, 1996). The higher home ownership rates could be indicative of the high property rates and thus explain the lower probability of origination for properties located in tracts with higher home ownership rates.

The model also incorporates two other characteristics of neighborhoods: proportion of nonwhite population and median income. If the nonwhite population increases by a percentage point, the probability of loan origination would decrease by .0004 and that of denial increases by .0001. While the effect of nonwhite population percentage of the census tract is statistically significant, its magnitude is small. If the median income of the census tract increases by $10,000, the probability of loan origination will increase by .03 percentage points and the probability of denial will decrease by .02 percentage points, holding other factors constant.

Loan and Property Characteristics

Loan amount plays a statistically significant, but a small role in the lenders' decision making process. If the loan amount increases by $100,000, the probability of success increases by 9 percentage points to .76 and the probability of denial falls to .08, holding other factors constant. The variable of interest is the loan amount-to-income variable since it would give the lender an idea about the applicant's potential indebtedness. The marginal effect for the loan-to-income ratio is positive for loan origination and is negative for loan denial. Thus, as the loan-to-income ratio increases, the probability of denial decreases. This unexpected result could be attributed to not having information about the specific features of the loan product that would affect the monthly as well as the total payment to be made in service of the loan. HMDA data lack information on whether the loan is an adjustable or fixed rate mortgage and whether it is a 15 or a 30 year loan. Additionally, if the applicant intends to occupy the property upon purchase, the probability of loan origination decreases by 2.6 percentage points and that of denial increased by 2 percentage points.

Time

The probability of a loan being originated is lower in all years relative to 1992 (the base year), except for 1993 (results not shown). The probability of *other outcomes* and denial increases from 1994 onwards, relative to 1992. Examining the coefficients for the year variable closely, it is seen that the probability of loan origination changes dramatically every year. This could be due to the changes in the mortgage interest rate through this period. Including the time variable in the equation is necessary to account for unobserved changes every year.

Conclusion

Previous works have identified the role of race in the mortgage market on both a regional and a national scale. For example, Munnell et al. (1996) use data from the Boston MSA and Wyly and Holloway (1999) restrict themselves to Atlanta to understand the relation between race and credit availability. Collins and Margo (1999) use IPUMS data for the entire country to investigate the correlation of race and the probability of home ownership. I use HMDA and census data to extend the study of this topic to the Indianapolis MSA for the period from 1992 to 2003. Using a multinomial logit model to isolate the effect of race on outcomes for mortgage applications after controlling for a host of other factors such as individual, asset and neighborhood characteristics, I find that the probability of denial of loan applications is higher for blacks and Hispanics than for whites.

One shortcoming of HMDA is that it lacks information on key variables such as credit scores and loan-to-value ratio. The estimate of the effect of race on probability of a certain outcome could be biased by the lack of credit history to the extent that race is correlated with credit scores. As a result, this analysis cannot be used as definitive support for existence of discrimination in the home loan market. What can be said is that race consistently shows up as a statistically significant variable in the models for all outcomes. The effect of race cannot be explained away by factors such as lower income status of minorities or the neighborhoods where they live, since their effect has already been accounted for. An argument can be made that since blacks and Hispanics tend to be economically weaker, the chances of them having a worse credit score than a white applicant are higher. This could distort the estimate of the effect of race in the mortgage market. However, the data contain information on income, which is the one of the main determinants of economic status. Since income is included in my model, the only way credit history could impact the effect of race would be if credit scores and race were correlated, after taking income into account. Interestingly, Munnell et al. (1996) included a measure of credit score in their model and still found that blacks and Hispanics had higher rates of loan denial than white with similar characteristics. Also, no study has been able to conclusively prove the link between race and the probability of default (Munnell, Tootell, Browne, & McEneaney, 1996). Moreover, the lower income and other economic variables of minority applicants could be the function of discrimination in other markets, including the labor market, which is not considered in this paper.

As indicated in other works, a possible explanation for race being an important factor could be differential treatment of potential borrowers in the preapplication stage. When a potential borrower enquires about possible loan options, the lender's attitude can make a difference. The Smith and DeLair (1999) study showed that lenders are more encouraging with white applicants and offer them a wider option of products to meet their financial condition.

From a policy perspective, the various legislative acts that were passed to promote fair housing and equal access to the mortgage market have not been effective. Since the reasons for the significant role of race are not completely obvious, it would be difficult to formulate policies that would mitigate the effect of race. The ideal policy prescription would be a mix of steps to elevate the target population's socioeconomic characteristics and programs to improve lenders' attitudes and applicants' awareness.

Future Research

Future researchers can use HMDA data from 2005 onwards to investigate the effects of borrower treatment during the preapplication stage on mortgage applications outcome. This would be possible since HMDA now requires lenders to collect information on preliminary enquiries made by potential borrowers. Since information regarding borrower treatment in the preapplication stage was not available for the period until 2003, this paper could not use that information to understand the variation in mortgage outcomes among different races. Additionally, to effectively implement fair housing laws, it would be useful to isolate the effect of race for different institutions. Another aspect that needs to be addressed is the correlation between race and credit scores. Finding the extent of correlation, if any, would help in determining the validity of our results.

Appendix A

The multinomial logit model is based on the assumption of Independence of Irrelevant Alternatives (IIA). This means that the odds of choosing between two alternatives do not depend on the presence of other outcomes. That is, the relative probability of choosing two existing alternatives remains unaffected by the addition and deletion of other outcomes. In that sense, other alternatives are irrelevant. Suppose a new alternative almost identical to an existing alternative is added to the set of choices. One would expect that the probability of choosing the duplicated alternative would be halved and the probability of choosing other alternatives would be unaffected. If this happens, then the multinomial logit model would be inappropriate since IIA would be violated. In other words, the multinomial logit model cannot be used whenever two or more alternatives are close substitutes. This can be illustrated using the example of an individual who has a choice between walking, driving or taking a blue bus to her destination. If she were given an additional choice of taking a red bus, it would be natural to expect the relative odds of taking the blue bus to another choice to be halved. However, for IIA to hold, the relative probability of choosing the blue bus to an existing alternative should remain unchanged despite the introduction of a similar alternative (Kennedy, 1998:241).

The Hausman test is commonly used to test this assumption. The Hausman test can have multiple outcomes depending on the base category and is based on the idea that if a category were dropped, the IIA assumption would mean that the estimated coefficient would not change. For this test, a full model is estimated with all the outcomes and a restricted model is estimated by eliminating one or more outcome categories. Test statistics are computed using information from the full model and the restricted model. In this case, the model seems to satisfy the IIA assumption (Table A-1).

Table A-1: *Hausman test for IIA assumption*
Ho: Odds (Outcome-J vs Outcome-K) are independent of other alternatives.

Omitted Category	Chi 2	Degrees of freedom	P>chi2	Evidence
2	-1.4e+03	28	1.000	For Ho
3	-226.686	27	1.000	For Ho
1	-2.8e+03	28	1.000	For Ho

Acknowledgements

I would like to acknowledge the invaluable guidance of Todd Sears, formerly of the Indianapolis Neighborhood Housing Partnership and that of Anne Royalty at the Department of Economics at Indiana University-Purdue University at Indianapolis. Additionally, I would like to thank Glen Pine of the Department of Sociology at New York University for his comments. All errors, of course, remain mine.

References

Federal Financial Institutions Examination Council. www.ffiec.gov/hmda/history (accessed 2004 October).

Avery, R.B., Beeson, PE., & Sniderman, M.S. (1993). Acounting for racial differences in housing credit market (Working Paper 9130). *Federal Reserve Bank of Cleveland*, December.

Buist, H., Megbolugbe,I.F., & Trent, T.R. (1994). Racial home ownership patterns, the mortgage market and public policy. *Journal of Housing Research*, 5(1):91-116.

Collins, W.J. & Margo, R.A. (1999, August). Race and home ownership, 1900 to 1990 (NBER Working Paper 7277). *NBER Working Paper Series*.

Harkness, J. & Newman, S. (2003). Effects of homeownership on children: The role of neighborhood characteristics and family income. *FRBNY Economic Policy Review*, June.

Immergluck, D. (2004). *Credit to the community, community reinvestment and fair lending policy in the United States.* Armonk, New York: M.E. Sharp.

Kennedy, P. (1998). *A guide to econometrics.* Fourth. Cambridge: The MIT Press.

Long, S.J. (1997). *Regression models for categorical dependent variables.* Thousand Oaks, California: Sage Publishers.

Munnell, A., Tootell, M.B., Browne, L.E., & McEneaney, J. (1996). Mortgage Lending in Boston: Interpreting HMDA data. *American Economic Review,* 86(1, March):25-53.

Ross, S.L. & Yinger, J. (2002). *The color of credit: Mortgage discrimination, research methodology and fair lending enforcement.* Cambridge: The MIT Press.

Smith, R. & DeLair, M. (1999). New evidence from lender testing: Discrimination in the preapplication stage. In M.A. Turner & F. Skidmore (Eds.), *Mortgage lending discrimination: A review of existing evidence*, pp. 23-36. The Urban Institute.

Stewart, L. &. Rohe, W.M. (1996). Home ownership and neighborhood stability. *Housing Policy Debate,* 7(1):37-81.

Weink, R.E. (1992). Discrimination in the urban credit market: What we don't know and why we don't know it. *Housing Policy Debate,* 3(2):217-240.

Wyly, E.K. & Steven, R. (1999). Holloway. "The color of money revisited: Racial lending patterns in Atlanta neighborhoods." *Housing Policy Debate,* 10(3):555-600.