

# POLICY STUDIES

## **Home Mortgage Lending by Applicant Race/Ethnicity: Do HMDA Figures Provide a Distorted Picture?**

**Paul F. Huck**  
**Federal Reserve Bank of Chicago**

Consumer Issues Research Series  
Consumer and Community Affairs Division  
October 2000 (2000-3)

FEDERAL RESERVE BANK  
OF CHICAGO

October 25, 2000

**Home Mortgage Lending by Applicant Race/Ethnicity:  
Do HMDA Figures Provide a Distorted Picture?**

Paul Huck  
Senior Economist  
Federal Reserve Bank of Chicago  
Phone: (312) 322-2350

The views expressed are the author's and do not necessarily reflect the views of the Federal Reserve Bank of Chicago or the Federal Reserve System.

## **Abstract**

The Home Mortgage Disclosure Act of 1975 (HMDA) was designed to further the goal of fair access to mortgage credit and requires lenders to report information including the location, loan amount, income, and the race/ethnicity and sex of the applicant(s) for each application taken by the lender. However, the race/ethnicity of applications taken by mail or phone is not provided in a significant proportion of applications. Given the widespread use of HMDA data by lenders, community groups, researchers, and regulators and the importance of mortgage lending as a public policy issue, the strengths and shortcomings of the data must be clearly understood.

The main findings are that reported approval rates by race/ethnicity are significantly overstated for refinance and home improvement loans, whereas home purchase loans are little affected by this reporting problem. A review of trends in how race/ethnicity is reported and in the technology of mortgage lending leads to the reasonable conjecture that missing race/ethnicity will become a bigger and bigger problem in the near future.

Keywords: Mortgages, Data, Minorities.

## **Introduction and summary**

The Home Mortgage Disclosure Act of 1975 (HMDA) was designed by Congress to throw a spotlight on mortgage lending so that the public and regulators could better determine whether or not individuals or specific neighborhoods were being unfairly denied access to credit. HMDA requires lenders to report information including the

location, loan amount, income, and the race/ethnicity and sex of the applicant(s) for each application taken by the lender. Lenders also report whether the application resulted in an origination, denial, or some other action. The data generated by HMDA reporting are available publicly and provide a detailed picture of how geographic lending patterns vary depending on the income status or the racial/ethnic make-up of neighborhoods.<sup>1</sup>

The fact that denial rates are higher for minorities than for non-minorities is clearly evident in the HMDA reports, and this disparity has attracted much attention. Although HMDA data do not contain all of the variables that are relevant to the credit underwriting decision, the data have played an important role in research on racial discrimination and redlining--the denial of credit to neighborhoods on the basis of the income or the race or ethnic status of residents.<sup>2</sup> The HMDA reports are also an important source of information for examiners during the course of a Community Reinvestment Act (CRA) examination of a lending institution.

Given the widespread use of HMDA data by lenders, community groups, researchers, and regulators and the importance of mortgage lending as a public policy issue, the strengths and shortcomings of the data must be clearly understood. This paper explores the potential for a distorted interpretation of the HMDA data as a result of the way that race/ethnicity is reported. As discussed below, the race/ethnicity of applications taken by mail or phone is not provided in a significant proportion of applications. Since these applications really do belong in some racial/ethnic category, the number of applications for each race/ethnicity is understated by some amount. In addition,

---

<sup>1</sup> See Canner and Passmore (1995) for an example of an analysis of the national mortgage market using the HMDA data.

<sup>2</sup> See Ladd (1998) for a survey of the research on discrimination in mortgage lending. See Schill and Wachter (1993) for an example of the use of HMDA data for an analysis of redlining in mortgage markets.

applications for which race/ethnicity is not provided are less likely to be approved than applications with a known race/ethnicity. This suggests that the reported approval rates for the various racial/ethnic groups are overstated. The key question that this paper will address is how much the missing race/ethnicity distorts the interpretation of the HMDA reports.

The main findings are that significant potential exists for skewed interpretation of the HMDA data and that the magnitude of the distortion depends on the loan product under consideration. The problems raised by this issue are more pronounced for refinance and home improvement loans compared to home purchase loans. A review of trends in how race/ethnicity is reported and in the technology of mortgage lending leads to the reasonable conjecture that missing race/ethnicity will become a bigger and bigger problem in the near future.

### **Reporting race/ethnicity and action taken**

Lenders are required to gather information regarding the sex and race or ethnicity of applicants for a loan, usually by having applicants fill out a form.<sup>3</sup> The categories for reporting race or ethnicity are as follows: 1) American Indian or Alaskan Native; 2) Asian or Pacific Islander; 3) Black; 4) Hispanic; 5) White; 6) Other; 7) Information not provided by applicant in mail or telephone application; and 8) Not applicable. Applicants are not required to furnish this information, but if the applicant refuses to do so, the lender is required to put down the appropriate category based on visual appearance or surname. The exceptions to this procedure are the cases of applications taken entirely by mail or telephone. Lenders who take an application entirely by telephone are not

required to ask for this information. Forms for mail applications must include questions for this information, but if the applicant does not provide this information, the lender is not required to inquire further. “Not applicable” may be used for purchased loans, which are not originated by the reporter, and should be used for applications made by non-natural persons, such as partnerships or corporations.

The HMDA reports also provide information on the actions taken by lenders on the applications they receive. The categories for reporting the action taken are as follows: 1) Loan originated, 2) Application approved but not accepted, 3) Application denied, 4) Application withdrawn, 5) File closed for incompleteness, and 6) Loan purchased by your institution. For the purposes of this analysis, the first two categories are combined into a category for approvals, whether originated or not. Denials are left as a category, and categories four and five are combined into a category for withdrawn/incomplete application. The last category for purchased loans is excluded from the analysis in this paper.

Although attention is usually focused on ethnic/racial disparities in denial rates, it is equally valid to look at differences in approval rates. Approval and denial rates are not quite two sides to the same coin because of the presence of the third category of withdrawn/ incomplete applications. We shall see that applications for which racial/ethnic information is not reported are much more likely to fall in the denied or withdrawn/ incomplete categories as opposed to being approved. If we focus on denials, we will lose track of the withdrawn/incomplete applications. Thus, it will prove more illuminating to talk about approval rates in the following discussion.

---

<sup>3</sup> See FFIEC (1998) for a detailed discussion of the HMDA reporting process.

While the use of information not provided as a racial/ethnic category is allowed under the HMDA reporting rules, some information is obviously lost when this category is used. This raises the question: does the use of the information-not-provided category have an empirically important effect on the interpretation of the HMDA data? If information not provided were reported in a purely random fashion, then this reporting practice would not seriously distort our interpretation of the HMDA data. However, if the use of information not provided is systematically related to variables of interest, then distortion of the data will be a problem. To see whether or not this is the case, a first step is to look for systematic patterns in the use of information not provided, using 1997 HMDA data for ten major MSAs as a case study, as well as some national figures.<sup>4</sup> In order to sharpen the focus of this study, only loan applications for one-to-four-family houses are included in the analysis. These include applications for purchase, home improvement, and refinance loans. Loans that were reported as purchased from the originator were excluded from the sample, as were loans with race/ethnicity reported as not applicable.

### **Widespread use of information not provided**

Given the importance of HMDA data for policy discussions about lending in potentially underserved neighborhoods, variation in the use of information not provided by neighborhood income category is potentially interesting. The prevalence of the use of the information-not-provided category for the ten MSAs taken together is shown in Table 1. From this table, we can see that information not provided is reported for 19.9 percent

---

<sup>4</sup> The MSAs in the study are Atlanta, Chicago, Detroit, Houston, Los Angeles, Miami, Milwaukee, New York, Philadelphia, and Washington DC. These MSAs are clearly not a random sample and are not meant to be

of all loan applications. Although not reported in the table, this proportion for individual MSAs ranges from 11.9 percent in Houston to 30.3 percent in Philadelphia. The use of information not provided varies by neighborhood income and loan product, however. Across all census tracts, home purchase applications are the least likely to fall into the information-not-provided category (6.4 percent), whereas this category appears much more often for refinance applications (28.2 percent) and home improvement applications (33.4 percent). The use of information not provided ranges from 4.6 percent in Detroit to 9.5 percent in New York for purchase applications, from 17.9 percent in Detroit to 49.8 percent in New York for refinance applications, and from 19.4 percent in Detroit to 42.0 percent in Los Angeles for home improvement applications.

National figures provide a similar picture, as information not reported is used for 17.8 percent of the applications within MSAs in the nation for 1997. The corresponding figures for the various loan products are 5.8 percent for home purchase applications, 26.5 percent for refinance applications, and 29.7 percent for home improvement applications.

Table 1 also shows that information not provided is used for a higher proportion of applications for all loan products in low- and moderate-income tracts (25.7 percent) relative to middle (20.5 percent) and upper income tracts (16.2 percent) in the MSAs taken together. This pattern also holds for each loan product taken separately. If we consider the individual MSAs for each loan product, New York provides the most extreme example of the use of information not provided, as 63.8 percent of refinance loans in low- and moderate-income neighborhoods fall into this category. This figure is based on over 21,000 applications, so it is not a result of a small sample. Thus, the use of information not provided is more prevalent in low- and moderate-income neighborhoods

---

representative of the nation.

in part because this category is used relatively more for each loan product in these neighborhoods.

Race/ethnicity is also reported less for applications from low- and moderate-income neighborhoods because a lower proportion of applications in these neighborhoods are for home purchase loans, which have relatively low incidence of information not provided compared to refinance and home improvement loans. Table 2 shows the proportions of applications for each loan product by neighborhood income. We can see that low- and moderate- income neighborhoods have a lower proportion of purchase applications (32.7 percent) compared to middle- (41.1 percent) and upper-income neighborhoods (47.8 percent).

Another aspect of an overview of the incidence of missing race/ethnicity is that this incidence is increasing over time. I illustrate this increase using national figures for applications made within MSAs using the same criteria for excluding applications that apply to the ten MSA sample. Between 1993 and 1997, the proportion of applications for all loan products using information not provided in the nation increased from 6.7 percent to 17.8 percent. This proportion rose for purchase loans (from 3.7 to 5.8 percent), for refinance loans (from 6.9 to 26.5 percent), and for improvement loans (from 14.9 to 29.7 percent).

These results demonstrate that the use of information not reported is quite substantial in these selected MSAs and in the nation, particularly for refinance and home improvement applications, and becoming more common in recent years. It is also the case that applications from low- and moderate-income neighborhoods are more likely to not report race/ethnicity. These findings suggest that there is a systematic component to

the use of information not provided and that potential distortions in interpreting lending patterns for the various racial/ethnic groups are more likely to be a problem in low and moderate income neighborhoods relative to middle and upper income neighborhoods. One potential explanation for these findings is that lenders that target low- and moderate-income and minority neighborhoods market their services in such a way as to promote the use of mail or telephone applications. Thus, product delivery systems may vary systematically by neighborhood, perhaps because of the historical location of branch offices.

### **Approval rate distortion**

The potential for a distorted view of the actions taken on applications for the various racial/ethnic groups is compounded by the fact that applications in this category are much less likely to be approved than applications for which the race or ethnicity of the applicant are known. Approval rates by race/ethnicity and loan product for the aggregated MSAs are shown in Table 3.<sup>5</sup> The table illustrates that approval rates for minority applicants are lower than is the case for White applicants, with the exception of purchase applications by Asians.

The table also shows that approval rates for applications reporting information not provided are substantially lower than applications with known race/ethnicity for all loan products. For example, for all loan products taken together, the approval rate for applications with known race/ethnicity is 73.1 percent, whereas the approval rate for applications with information not provided is only 40.9 percent. This pattern holds for

---

<sup>5</sup> Due to the small number of applications in these groups, applications in the Native American and Other categories are omitted from the analysis of loan applications by race/ethnicity.

each of the various loan products taken separately. It is also the case that the approval rates for all of the ethnic/racial groups for each loan product are higher than the corresponding figure for the information-not-provided category. A potential explanation for these findings is that some of the people who initially make applications by telephone or mail at some time may have a face-to-face meeting with the lender when race/ethnicity can be observed. Applicants who are rejected, especially early in the process, are less likely to meet with lenders than those who are accepted. Thus, denied applications are more likely to fall into the information-not-provided category. Another possibility is that less credit-worthy applicants may select more anonymous lending channels, such as phone or mail, in order to reduce any embarrassment from rejection.

These findings suggest that ignoring applications for which race/ethnicity is not provided results in a distorted picture of approval rates. Table 3 allows us to calculate the magnitude of the distortion for the combined race/ethnic categories. For example, the table shows that the approval rate for purchase applications with a known race/ethnicity is 77.6 percent, whereas the rate for all purchase applications, including those using the information-not-provided category, is 76.5 percent. Thus, the actual approval rate for total purchase applications is overstated by the difference between these figures, 1.1 percent. This seems to be a minor distortion of the approval rate. This finding is the result of a relatively low incidence of the use of information not provided for purchase applications (as shown in Table 1) combined with the relatively small difference in approval rates for applications with and without a known race/ethnicity (as shown in Table 3).

However, if we calculate the overstatement in the approval rate for the other loan products, we see that the approval rate for refinance loans is overstated by 9.4 percentage points (70.2 percent minus 60.8 percent) and the rate for improvement loans is overstated by 9.1 percentage points (64.8 percent minus 55.7 percent). For all loan products together, the approval rate is overstated by 6.5 percentage points (73.1 percent minus 66.6 percent). These results also hold for the MSAs taken individually. The most extreme example is the case of New York where the refinance approval rate for applications with known race/ethnicity is 64.1 percent and the rate for all applications is 47.6 percent, an overstatement of 16.5 percent. Thus, the overstatements of approval rates for refinance and home improvement applications are fairly substantial. This finding is due to the relatively common use of information not provided for refinance and improvement applications (as shown in Table 1) combined with the relatively large difference in approval rates for applications with and without a known race/ethnicity (as shown in Table 3).

We can easily calculate the overstatement of the approval rate for all applications combined, but calculating the distortion in the approval rate for the separate racial/ethnic groups is not so straightforward. In order to make this calculation, the applications using the information-not-provided category must be allocated in some way to the known racial/ ethnic categories. In other words, the missing racial/ethnic information must be filled in, or imputed, by some method.

## **Imputing unknown race/ethnicity**

Short of tracking down and surveying applicants for whom race/ethnicity is missing, we can never be certain of the actual race/ethnicity of such applicants. However, it is possible to impute the missing race/ethnicity using established statistical methods. This imputation is made using information about the loan and the applicant from the HMDA data combined with census data for the tract in which the property is located.

Race/ethnic status is an example of a categorical variable in which the value the variable takes on is limited to a number of discrete outcomes, in this case the categories defined by HMDA. Race/ethnic status is also unordered in the sense that the various categories convey no ranking and could be arranged in any order without affecting the analysis. An example of the opposite case, an ordered categorical variable, would be a bond-rating system, for which the order or ranking is clearly important.

The multinomial logit is a standard statistical model for analyzing an unordered categorical variable, and it is used to generate estimated probabilities for the unobserved race or ethnic category using the known individual loan and neighborhood characteristics.<sup>6</sup> This is an example of a regression imputation, where the missing variable is estimated by the predicted value of a regression on the known variables.<sup>7</sup>

There is a good deal of information available for predicting race/ethnic status. HMDA reports provide information about the individual applicant, the lender, the loan

---

<sup>6</sup> See Greene (1997) for a technical description of the multinomial logit model.

<sup>7</sup> There is a large literature on the analysis of data with missing observations. See Little and Rubin (1987) for a useful introduction to the topic. Another way of filling in missing data is a hot-deck imputation procedure. This method works by matching an observation with a missing variable to another observation that is similar in terms of the variables that are thought to be useful for prediction. The missing value is then imputed using the variable value of the matched observation. Avery, Beeson, and Sniderman (1999) use a hot-deck procedure in their analysis of mortgage lending. One of the variables that they impute is race/ethnicity.

characteristics, and the census tract in which the property is located. Tract-level census data, such as information on housing, income, and demographics, can be combined with the individual loan data to build a statistical picture of the loan application and the neighborhood in which the property is located. To the extent that these variables are correlated with the racial/ethnic status of the applicants, they will be useful for prediction. Because of the small number of loan applicants who fall into the American Indian and Other categories, applications in these categories were omitted from the analysis.

In order to accommodate differences across the MSAs and loan products , separate regressions were run for each MSA and product. A fairly large number of applicant- and tract-level variables have been commonly used in the mortgage lending literature.<sup>8</sup> In order to conserve on computational resources, I began with these variables and then evaluated alternative specifications on the basis of goodness-of-fit measures, such as the log likelihood and cross tabulations of actual and predicted outcomes. It quickly became apparent that the extra predictive value of adding variables to a fairly parsimonious specification is small. The applicant-level variables used to estimate the model include the log of the loan amount, a dummy variable equal to one for lenders who report HMDA data to HUD, and a dummy variable equal to one if the loan was approved.<sup>9</sup> The tract-level variables include the tract median family income, the population proportions for Asians, Blacks, and Hispanics, and the square of these proportions. Given widespread housing segregation, it is not surprising that the key variables for predicting race/ethnicity are the minority proportion variables.

---

<sup>8</sup> See Schill and Wachter (1993) for a representative selection of variables used to describe mortgage lending markets.

<sup>9</sup> Applicant income and the ratio of loan amount to income were not used as an explanatory variables because a significant number of applications in some MSAs did not report income. In practice, income adds little to the results after including loan amount because of the correlation of these two variables.

Because the purpose of estimating the model is to impute unknown race/ethnicity, evaluating the models' goodness-of-fit is important. One method is a crosstabulation that compares the actual and predicted race/ethnicity, where the predicted race/ethnicity for each observation is taken to be the outcome with the maximum probability. Note that a feature of the logit model is that the estimated probabilities will on average equal the sample proportions for each outcome. Thus, if the predictive variables are utterly useless, the model would simply return the population proportion for each outcome and the predicted outcome would always be the most common sample outcome.

Although not reported here, the comparisons of actual and predicted outcomes for each MSA and loan product suggest that models overall fit the data reasonably well. Not surprisingly, given that a majority of the applicants are White, a high proportion of these observations are correctly identified. It is also not surprising that the models correctly identify relatively few observations when a racial/ethnic group makes up a small proportion of the sample and does not stand out in terms of the explanatory variables. This result holds for Asians in all of the MSAs except Los Angeles where about 20 percent of Asian applications are correctly identified. Where Blacks and Hispanics make up a more substantial proportion of the observations (and often live in segregated neighborhoods), the models do a reasonable job of correctly categorizing the actual race/ethnicity. For example, Black applicants make up about 20 percent of all applications and Hispanics make up about 10 percent in the Chicago MSA. The proportion of correctly identified observations for Black applicants ranges from 57.4 percent for purchase loans to 79.2 percent for refinance loans. For Hispanics, this proportion ranges from 27.1 percent for purchase loans to 66.9 for improvement loans.

## Results

The predicted probabilities of the racial/ethnic categories for the loan applications using information not provided are used as estimates of the unobserved race/ethnicity and provide the foundation for further analysis. At least two approaches could be taken in allocating information-not-provided applications to a particular racial/ethnic category. One is to allocate the entire application to the race/ethnicity having the highest predicted probability. Another is to, in effect, split the application and allocate a portion to each race/ethnicity equal to the predicted probability for each category. I chose the latter strategy and allocated portions of the unknown applications, on the grounds that information is lost by putting all the weight on the highest predicted probability. Now we are in a position to see if adjusting for omitted race/ethnicity makes a difference in our interpretation of the HMDA information. Specifically, we will pay particular attention to approval rates in light of the importance attached to differences in the outcomes of applications across the various racial/ethnic categories.

The results of the imputation exercise are used to allocate the applications using information not provided into the known racial/ethnic categories. The adjusted figures thus represent estimates of the actions taken on applications in the various racial/ethnic groups after accounting for missing race/ethnicity. The adjustment is calculated as the difference between the adjusted approval rate and the reported rate. Table 4 shows the reported and adjusted approval rates by loan product and race/ethnicity for the aggregated MSAs.

The table suggests that approval rates are overstated for each racial/ethnic group. For all loan products taken together, the estimated adjustments to the approval rates range

from -5.5 percentage points for Asians to -8.4 percentage points for Blacks. The estimated adjustments vary depending on the loan product in question. The adjustments for home purchase loans are quite small and range from -0.9 percentage points for White and Asian applicants to -2.0 percentage points for Black applicants. The adjustments for both refinance and improvement loans are more substantial and range from about -8 to -10 percentage points. The approval rate adjustments for Black applicants tend to be higher (in absolute values) than those of the other racial/ethnic groups, with the exception of improvement loans. However, the adjustments tend to be of similar magnitude across the various racial/ethnic groups.

The adjustments for some of the individual MSAs are larger than those reported in the table. The most extreme case is that of refinance loans in the New York MSA where the adjustments range from -13.5 percent for White applicants to -18.2 percent for Black applicants. These adjustments are quite substantial, as for example, the reported approval rate for Black applicants of 55.9 percent is reduced to an estimated value of 37.7 percent. In round figures, the reported approval rate of just over one-half is reduced to just over one-third.

Racial/ethnic disparities in the actions taken on applications have received a good deal of attention as indicators of discrimination in the home mortgage market.<sup>10</sup> Although disparities in denial rates have generally been reported, for the purposes of this paper we have been discussing approval rates. One way to measure racial/ethnic differences in outcomes is to calculate the differences between approval rates for White and for minority applicants. The adjustments to approval rates for the various loan

---

<sup>10</sup> For an example of a recent media discussion of discrimination and disparities in denial rates, see New York Times (1999).

products in Table 4 do not vary markedly across racial/ethnic groups, at most about two percentage points. Since a similar downward adjustment to the approval rate applies to each racial/ethnic group, the differences between Whites and minorities are not affected much by the adjustment process. The ratios formed by dividing the approval rate for Whites by the minority approval rates can also be used to measure differences in outcomes. Although, the approval ratio after adjustment tends to show an increase in disparities between White applicants and minority applicants, the increases in approval ratios tend to be fairly small, generally under five percentage points.<sup>11</sup> Again, this is a reflection of the fact that the adjustment to approval rates are fairly similar across the racial/ethnic groups.

Another aspect of the results gets at the issue of how the estimated incidence of information not provided varies with race/ethnicity. After allocating applications with unknown race/ethnicity to a category, we can calculate the estimated proportions of the total applications in the known racial/ethnic categories that are left out of the HMDA reports. These estimated proportions are shown in Table 5. The estimated incidence of omitted race/ethnicity reflects the overall use of the information-not-provided category, as we see relatively low figures for purchase loans and relatively high figures for improvement and refinance loans. These proportions for refinance and improvement applications are quite substantial, as they range from 24.6 percent for White applicants for refinance loans to 38.8 percent for Black applicants for refinance loans. If we compare the estimated proportion omitted across the racial/ethnic groups, we see that this proportion for Black applicants is higher than the corresponding figures for other groups

---

<sup>11</sup> The biggest change is an increase in the ratio of White to Black approval rate from a reported ratio of 1.27 to an adjusted ratio of 1.37.

for purchase and refinance loans. The estimated proportions for improvement loans are similar for the minority groups, which in turn are higher than the figure for White applicants. For all loan products taken together, the estimated proportion of applications missing race/ethnicity is highest for Black applicants at 29.4 percent. The largest estimated proportion across the ten MSAs is the case of refinance loans in the New York MSA, where this proportion ranges from 42.7 percent for White applicants to 65.9 percent for Black applicants.

In summary, the results show that reported approval rates generally are overstated for all the racial/ethnic groups. The estimated adjustments to approval rates are small for the purchase loans but more substantial for refinance and improvement loans. Since the approval rate adjustments are fairly similar across the racial/ethnic groups, the adjustments do not substantially increase disparities in approval rates relative to the reported rates. The estimated incidence of the use of information not provided in each racial/ethnic category is quite substantial, particularly for Black applicants.

## **Discussion**

One of the most important implications of these results is that the potential for distortion in the HMDA reports depends on the loan product in question. The findings indicate that the incidence of omitted race/ethnicity is relatively small for home purchase loans, both overall and for the individual racial/ethnic groups. Both because of the relatively low incidence and because of the relative similarity in approval rates for applications with and without a known race/ethnicity, the adjustments to approval rates are fairly small for purchase loans. Thus, at current levels of the use of the information-

not-provided category, the interpretation of HMDA reports for purchase lending seems little affected by this reporting issue. However, this is not the case for home improvement and refinance loans.

Another important finding is that approval rates are overstated for all racial/ethnic groups, substantially so for refinance and improvement loans. Racial/ethnic disparities in reported approval rates are quite substantial, and one issue is whether or not the adjustments to approval rates widen the disparities. There is some evidence that disparities in approval rates, as measured by differences in this rate between White and minority applicants as well as the White/minority ratio of approval rates, are somewhat increased by the estimated adjustments, especially for some individual MSAs. However, because the adjustments to approval rates are fairly similar across the racial/ethnic groups for the aggregated MSAs, disparities in approval rates are not dramatically affected by the estimated adjustments.

Although disparities in approval rates are not dramatically misstated, the relatively high incidence of the use of information not provided for the aggregated MSAs for all loan products (20.4 percent) raises a warning flag for some uses of the data. Table 5 shows that home purchase is the only loan product for which the estimated proportion of applications missing race/ethnicity is small. The proportions are much higher for refinance and improvement loans, and although not reported in the table, can be higher yet in selected MSAs.

As one example of problems raised when the incidence of omitted race/ethnicity is high, consider comparing the minority approval rates for an individual lender with the MSA aggregate approval rates, perhaps as part of a fair lending review. Suppose the

lender under review did not report many loans with omitted race/ethnicity. Since we have seen that the reported approval rates for an MSA aggregate are overstated, a comparison will be biased against a lender who fully reports race/ethnicity. A better comparison might be made with a control group of lenders that also fully report race/ethnicity.

Widespread use of the information-not-provided category also raises problems for the statistical analysis of whether or not a lender's approve/deny decision depends on the race/ethnicity of an applicant. If applications with missing race/ethnicity are simply dropped from the analysis, then the sample obviously will be a sub-sample of the actual population. This results in a statistical bias if the sample selection process is systematically related to the accept/deny process. This clearly is the case, because applications that are not approved are much more likely to be reported with missing race/ethnicity (Table 3). In other words, race/ethnicity is not missing at random. This sample selection problem is compounded if the selection process depends on applicant race/ethnicity. This would be the case, if for example, lenders that marketed phone and mail applications tended to target minority neighborhoods. This selection problem means that statistical tests of discrimination in the loan approval process may be untrustworthy unless selection bias is accounted for.<sup>12</sup>

Methods are available for dealing with sample selection. One approach is to statistically model the selection process and include this in the regression analysis. This requires a reasonable understanding of the selection process. It also requires access to data for the variables in the selection model. However, these data might not be readily

---

<sup>12</sup> See Phillips, Trost and Yezer (1994) for a discussion of selection problems in this context.

available. If, for example, the selection process depends on the marketing methods or product delivery systems of individual lenders, then lender-specific data would be required. Another approach is to impute, or fill in, the missing race/ethnicity and then proceed as usual with the regression analysis of the accept/deny decision. However, this is not a simple matter either. The simple regression imputation used in this paper, although adequate for my purposes, is not sufficient for statistical inference based on the filled in data set, and more advanced imputation methods would be required.<sup>13</sup> In any case, statistical analysis of discrimination in the accept/deny decision is considerably more difficult in the face of substantial non-reporting of race/ethnicity.

We have already noted some evidence that the proportion of applications reported with race/ethnicity not provided has been rising in recent years. This means that reported racial/ethnic trends in lending will be distorted. For example, if the number of reported home purchase originations by race/ethnicity are compared over time, then increasing use of information not provided means that the reported number of originations will be increasingly understated for each racial/ethnic group. Since the numbers of purchase originations generally have been increasing in recent years, originations by race/ethnicity have actually increased by more than the reported trend.<sup>14</sup> However, the results show that reported approval rates for known race/ethnicity are overstated. Thus, increasing use of information not provided means that approval rates are increasingly overstated.

We have seen that the current incidence of missing race/ethnicity raises significant difficulties for the interpretation of reported lending patterns by race/ethnicity.

---

<sup>13</sup> Multiple imputation is one approach to imputing missing data in a way that hopefully avoids biased statistical inferences. See Rubin (1987) for a discussion of how this approach might be implemented.

<sup>14</sup> See FFIEC (2000) for national lending trends by race/ethnicity in recent years.

Will these difficulties become pronounced in the future? While it is not possible to foretell the future, a reasonable projection of current mortgage industry trends makes it seem likely that this race/ethnicity reporting problem will become more serious.

There can be little doubt that rapidly evolving information technology is making it easier to complete mortgage transactions with less face-to-face contact between lenders and borrowers. Many financial institutions have Internet Web sites that allow customers to apply online for a variety of products and services, including mortgage loan products. Some online banks without any brick-and-mortar branches are open for business, although it is not clear how successful this strategy will be. A consequence of this investment in online banking is that Internet originations should become more common (American Banker 1999).

How the growing presence of online lending will affect the HMDA reports depends on the proportion of online applications that report race/ethnicity as information not provided. Internet application materials, like mail applications, should include a form that gives an applicant a chance to report race/ethnicity, but lenders are not required to follow up if this information is not provided. Thus, reporting of race/ethnicity for Internet applications depends on how potential borrowers fill out this portion of the application materials. While we still have little experience with the HMDA reporting of online mortgage lending, there is some evidence that applications taken over the Internet will provide little information about applicant race/ethnicity.

Although HMDA reports for wholly internet lenders are still rare, we do have at least one example of the reporting of race/ethnicity by such a lender. Overall, 90.9 percent of applications reported by this lender in 1998 used the information-not-provided

category. This high incidence of missing race/ethnicity is observed for both home purchase and refinance loans. Thus, if this example is at all representative, then the increasing use of online lending will entail increasingly less complete HMDA reporting of applicant race/ethnicity.

Another industry trend that results in the increased use of the information-not-provided category is the growth of the subprime lending market. It is estimated that firms that specialize in subprime lending take in about 10 percent of the conventional home purchase applications in the national HMDA figures in 1998 and about 30 percent of the conventional refinance applications.<sup>15</sup> Subprime applications are more likely to have missing race/ethnicity compared to prime applications. For example, subprime specialists in the Chicago MSA in 1998 used the information-not-provided category for 15.1 percent of home purchase applications compared to a corresponding figure of 7.7 percent for all lenders in the MSA. The same figure for refinance applications for subprime lenders is 40.0 percent compared to 17.8 percent for all lenders in the MSA. Thus, the increasing size of the subprime lending market has contributed to the rising incidence of missing applicant race/ethnicity in recent years. If the subprime market continues to expand, we can expect to see continued reduction in the completeness of HMDA reporting of applicant race/ethnicity.

Turning to policy options, easy fixes for this problem are not apparent for at least two reasons. First, the reporting of race/ethnicity for applications taken by phone, mail, or the internet ultimately depends on the willingness of the applicant to supply the

---

<sup>15</sup> These estimates are taken from Scheesele (1999). Subprime loans are not identified in the HMDA reports, but HUD estimates the extent of the subprime lending market by identifying lenders that specialize in this market. A list of subprime specialists can be found in Scheesele (1999).

information. Following up on applicants that refused to provide this information would be extremely intrusive and would negate part of the cost savings of these product delivery modes. Second, the HMDA reports are meant to provide an accurate picture of lending by individual lenders and for relatively small localities, such as neighborhoods. Thus, statistical measures, such as imputing missing race/ethnicity or surveying non-respondents, that might be acceptable for establishing a general picture would not work for specific cases involving a small number of applications.

### **Concluding remarks**

Imputing missing race/ethnicity for applications that report information not provided allows the calculation of the estimated adjustment to the approval rate and the estimated proportion of a given racial/ethnic group with missing information for the sample of ten large MSAs. The results suggest that missing race/ethnicity is not a serious problem for the interpretation of home purchase lending. The overstatement of the approval rate and the proportion of applications missing race/ethnicity for purchase loans are fairly small. However, the overstatement of the approval rate and the proportion of applications missing race/ethnicity for refinance and improvement loans are much more substantial, particularly for some individual MSAs. These findings suggest that the impact of missing information should not be ignored when using the HMDA reports to examine racial/ethnic patterns in lending, particularly for refinance and improvement loans.<sup>16</sup>

---

<sup>16</sup> Note that missing race/ethnicity is not a major problem for the enforcement of the CRA because a lender is evaluated on the basis of the distribution of activity by borrower and neighborhood income categories.

The rising incidence of missing race/ethnicity in recent years and emerging technological trends in the mortgage industry, which more easily allow lending without face-to-face contact, make it reasonable to forecast that this problem will become more serious in the not-to-distant future. One of the three stated purposes of the HMDA is to provide loan data that can be used “to assist in identifying possible discriminatory lending patterns and enforcing antidiscrimination statutes (Board of Governors 1995).” If current trends continue, the ability of the HMDA reports to further this purpose may be compromised.

## References

- American Banker*. July 27, 1999. Decisive Shift to On-Line Applications Predicted.
- Avery, Robert B., Patricia E. Beeson, and Mark S. Sniderman. 1999. Neighborhood Information and Home Mortgage Lending. *Journal of Urban Economics* 45: 287-310.
- Board of Governors of the Federal Reserve System. 1995. *Regulation C Home Mortgage Disclosure*. Washington, DC.
- Canner, Glenn B., and Wayne Passmore. 1995. Home Purchase Lending in Low-Income Neighborhoods and to Low-Income Borrowers. *Federal Reserve Bulletin* 81: 71-103.
- Federal Financial Institutions Examination Council. 1998. *A Guide to HMDA Reporting: Getting it Right!*, Washington, DC.
- Federal Financial Institutions Examination Council. 2000. August 8, 2000 Press Release. World Wide Web page <<http://www.ffiec.gov/hmcrpr/hm080800.htm>>.
- Greene, William H. 1997. *Econometric Analysis*, third edition. New Jersey: Prentice-Hall.
- Ladd, Helen. 1998. Evidence on Discrimination in Mortgage Lending. *Journal of Economic Perspectives* 12: 41-62.
- Little, Roderick A. 1987. *Statistical Analysis with Missing Data*. New York: John Wiley and Sons, Inc.
- New York Times*, September 16, 1999. Bias Worsens For Minorities Buying Homes.
- Phillips, Robert F., Robert P. Trost, and Anthony M. J. Yezer. 1994. Bias in Estimates of Discrimination and Default in Mortgage Lending: The Effects of Simultaneity and Self-Selection. *Journal of Real Estate Finance and Economics* 9: 197-215.
- Rubin, Donald B. 1987. *Multiple Imputation of Nonresponse in Surveys*. New York: John Wiley and Sons, Inc.
- Scheessele, Randall M. 1999. 1998 HMDA Highlights. Working paper No. HF-009. Office of Policy Development and Research, HUD.
- Schill, Michael H., and Susan M. Wachter. 1993. A Tale of Two Cities: Racial and Ethnic Geographic Disparities in Home Mortgage Lending in Boston and Philadelphia. *Journal of Housing Research* 4: 245-275.

Table 1

Incidence of Missing Race/Ethnicity, 10 MSAs (%)

	Purchase	Refinance	Improvement	All Loans
Low-mod income	7.1	34.3	35.9	25.7
Middle income	5.7	29.6	33.6	20.5
Upper income	6.9	22.9	30.7	16.2
All tracts	6.4	28.2	33.4	19.9

Notes: Neighborhood income categories are defined as follows: Low and moderate income is less than 80 percent of MSA median family income; middle income is at least 80 percent and less than 120 percent of MSA median family income; and upper income is equal to or greater than 120 percent of MSA median family income.

Source: FFIEC, 1997 HMDA data, and author's calculations.

Table 2

Distribution of Loan Products, by Tract Income, 10 MSAs (%)

	Purchase	Refinance	Improvement
Low-mod income	32.7	44.3	23.0
Middle income	41.1	41.5	17.4
Upper income	47.8	39.8	12.4

Notes: Neighborhood income categories are defined as follows: Low and moderate income is less than 80 percent of MSA median family income; middle income is at least 80 percent and less than 120 percent of MSA median family income; and upper income is equal to or greater than 120 percent of MSA median family income.

Source: FFIEC, 1997 HMDA data, and author's calculations.

Table 3

Reported Approval Rates by Race/Ethnicity and Loan Product, 10 MSAs (%)

	White	Asian	Black	Hispanic	Known Race/ Ethnicity	Information Not Provided	Total
Purchase	80.0	80.1	69.1	73.5	77.6	60.9	76.5
Refinance	75.3	68.8	59.1	58.8	70.2	37.5	60.8
Improvement	71.1	59.7	53.2	53.6	64.8	38.3	55.7
All Loans	77.1	74.8	62.1	66.3	73.1	40.9	66.6

Note: Applications in the American Indian and Other ethnic categories have been excluded from the analysis.

Source: FFIEC, 1997 HMDA data, and author's calculations.

Table 4

Adjustments and Approval Rates by Loan Products and Race/Ethnicity, 10 MSAs (%)

		White	Asian	Black	Hispanic
Purchase	Reported	80.0	80.1	69.1	73.5
	Adjustment	-0.9	-0.9	-2.0	-1.0
	Adjusted	79.1	79.2	67.1	72.5
Refinance	Reported	75.3	68.8	59.1	58.8
	Adjustment	-8.5	-8.0	-10.3	-8.0
	Adjusted	66.8	60.8	48.8	50.8
Improvement	Reported	71.1	59.7	53.2	53.6
	Adjustment	-9.0	-9.8	-7.7	-9.0
	Adjusted	62.1	49.9	45.5	44.6
All loans	Reported	77.1	74.8	62.1	66.3
	Adjustment	-5.7	-5.5	-8.4	-6.3
	Adjusted	71.4	69.3	53.7	60.0

Note: Applications in the American Indian and Other ethnic categories have been excluded from the analysis.

Source: FFIEC, 1997, HMDA data, and author's calculations.

Table 5

Estimated Proportion of Applications Missing Race/Ethnicity, 10 MSAs (%)

	White	Asian	Black	Hispanic	Total
Purchase	6.1	6.3	8.9	6.2	6.5
Refinance	25.5	24.6	38.8	28.9	28.9
Improvement	31.9	37.7	37.7	38.5	34.2
All Loans	18.1	16.4	29.4	19.8	20.4

Note: Applications in the American Indian and Other ethnic categories have been excluded from the analysis.

Source: FFIEC, 1997 HMDA data, and author's calculations.