

# Opportunities and Issues in Using HMDA Data

---

**Authors** Robert B. Avery, Kenneth P. Brevoort, and Glenn B. Canner

---

## Abstract

Since 1975, the Home Mortgage Disclosure Act (HMDA) has required most mortgage lending institutions to disclose to the public information about the home loans they originate or purchase during a calendar year. In using these data, however, researchers need to be aware of a number of issues and potential problems that characterize HMDA. This article provides a comprehensive enumeration of these issues, focusing on practical problems that can potentially influence choices researchers make in using the data or in interpreting the findings. The article also includes an illustrative example of how the data that is reported in HMDA can be used to gain a better understanding of trends and practices in the home mortgage market.

---

Since 1975, the Home Mortgage Disclosure Act (HMDA) has required most mortgage lending institutions with offices in metropolitan areas to disclose to the public information about the geographic location and other characteristics of the home loans they originate or purchase during a calendar year. Disclosure of home lending activity is intended to help the public determine whether institutions are adequately serving the housing finance needs of their local communities, to facilitate enforcement of the nation's fair lending laws, and to guide public- and private-sector investment activities. It is estimated that the more than 8,800 lenders currently covered by the law account for approximately 80% of all home lending nationwide. Because of its expansive coverage, the HMDA data likely provide a representative picture of most home lending in the United States. For a previous analysis of HMDA coverage, see Bercovec and Zorn (1996).

Over the years, the Congress has amended HMDA to extend the reach of the law to a broader range of institutions and to expand the types of information that must be reported and disclosed. The most sweeping legislative amendments occurred in 1989; these required the disclosure of application and loan-level information for home loans, including the disposition of applications and the income, sex, and race or ethnicity of the individuals applying for credit. Before that time, HMDA disclosures were limited to summary totals covering loan activity at the census tract level. Analysis of the loan-level information prompted widespread public

discussion about the fairness of mortgage lending decisions; the disclosures revealed wide disparities in the rates of approval of loan applications across racial and ethnic lines.<sup>1</sup> Since that time, the HMDA data have become an important component of fair lending enforcement reviews and public scrutiny of lender activities in this regard.

The Federal Reserve Board revised Regulation C, which implements HMDA, in 2002.<sup>2</sup> As a result of this review, a number of important changes were made to the reporting requirements beginning with the 2004 data. The changes substantially increase the types and the amount of information made available about home lending. The most important change was the requirement that lenders disclose pricing information for loans with prices (interest rates and fees) above designated thresholds. Loans with prices above the thresholds are referred to here as “higher-priced loans.” Other new information being reported include lien status (whether a loan is a first lien, a junior lien, or unsecured) and whether a loan is secured by a manufactured home or is subject to the protections of the Home Ownership and Equity Protection Act (HOEPA) of 1994.

The 2004 HMDA data, the first to reflect the recent Regulation C revisions, were released to the public by individual lending institutions in the spring of 2005. In September 2005 and again in September 2006, the Federal Financial Institutions Examination Council (FFIEC) made available to the public various summary reports (statistical tables) pertaining to each lender and lending activity in each metropolitan statistical area (MSA), along with a comprehensive data file that included all the reported information (except the dates of loan application and of credit decision).<sup>3</sup> With the release of the 2004 and 2005 HMDA data, the staff of the Federal Reserve Board prepared assessments of the expanded data, which were published in the *Federal Reserve Bulletin*.<sup>4</sup>

Although the HMDA data are quite comprehensive in some respects, in important ways the HMDA data are limited. The data include few items that pertain directly to the credit risks posed by applicants and borrowers and little detail about the types of loans extended. The data also do not include information about the standards lenders use to evaluate credit or about the methods used to establish loan prices or compensate individuals that take and process applications or underwrite loans. Nonetheless, the expanded HMDA data offer many opportunities for researchers and the general public to monitor mortgage market activities in general, as well as the performance of individual lenders. This article details a number of issues and potential problems that researchers need to be aware of in using the HMDA data. It also provides an illustrative example of how the data that are reported in HMDA can be used to gain a better understanding of trends and practices in the home mortgage market.

---

### Basic Reporting Requirements

HMDA applies to most depository institutions (commercial banks, savings associations, and credit unions) with home or branch offices in MSAs.

Depositories that are exempt are small (those with assets of less than \$35 million for the 2006 HMDA reporting year), or are not in the home-lending business, or have offices exclusively in rural (nonmetropolitan) areas. HMDA also extends to mortgage and consumer finance companies, whether such companies are independent or subsidiaries of banking institutions or affiliates of bank or thrift institution holding companies. Covered mortgage and consumer finance companies (referred to henceforth as “mortgage companies”) include those that extend 100 or more home purchase or home refinancing loans per year; such institutions are deemed to have an office in a metropolitan area if they receive five or more applications for properties in such areas.<sup>5</sup>

Depository institutions account for the bulk of the reporting institutions, but mortgage companies report the majority of the applications and loans. In 2005, for example, nearly 80% of the 8,850 reporting institutions were depository institutions but together they reported only 37% of all the lending-related activity. Mortgage companies accounted for 63% of all the reported lending; 70% of these institutions were independent and not related in any way to a depository institution.

Since the 1990 reporting year, the HMDA data have included application-level information about applicants and borrowers, the home-loan products they seek, the disposition of their requests for credit, and details about the location of the property that relate to the application. From 1990 to 2003, the reporting requirements were little changed; lenders were required to report on applications for home purchase, refinance, or home improvement loans. Historically, lenders were allowed some limited flexibility in determining which loans to report, but starting in the 2004 reporting year, requirements were standardized. Lenders are now required to report on all applications for home purchase loans secured by the home, loans with at least a partial purpose of home improvement secured by the home, refinanced mortgages secured by the home where the refinanced mortgage replaces another dwelling-secured mortgage, and certain unsecured home improvement loans. Open-ended home equity loans may be reported at the lender’s option. Also, if the lender has a formal home purchase pre-approval program, applications to the program must be reported.

Since 2004 limited information has also been collected on loan pricing. In particular, lenders are required to report the difference between the annual percentage rate (APR) on a loan and the rate on Treasury securities of comparable maturity for loans with spreads above designated thresholds. To calculate the rate spread, the lender uses the yield on Treasury securities as of the fifteenth day of a given month depending on when the interest rate was set on the loan.

---

### Non-Pricing Issues in HMDA

HMDA data are widely used in both academic and policy research and also to aid enforcement of consumer protection regulations, such as the fair lending laws

and the Community Reinvestment Act. However, there are a number of reporting issues that data users should be aware of when using the annual submissions.

### Timing Issues

HMDA requires that all applications with an *action taken date* within a calendar year must be reported in that year's annual HMDA filing. This rule has implications for analysis of denial rates and other lender actions. For loan denials, the action taken date is the date the loan was denied; for originated loans it is the date the loan was originated. Because lenders typically decide to deny a loan request in a shorter time frame than it takes to close a loan, the average time between application date and action date is generally longer for originated loans than for denials. This can lead to a distortion in the implied HMDA denial rates for loan applications at both the end and beginning of the year. End-of-year applications will appear to have very high denial rates since the denials will tend to be reported in the same HMDA reporting year. However, many of the approvals will carry over into the following reporting year if the closing takes place after January 1, making the approval rate for HMDA filings in the beginning of the year abnormally high.

If a lender is in "steady state" the excess denial rate at the end of the year will be offset by the excess approval rate at the beginning of the year. If a lender is growing rapidly (through merger or expanded business) or there was a large "application boom" at the end of the year, these effects will not balance out. Solving this problem is not straightforward. Knowledge of the application date would allow one to restrict the analysis to applications filed in a period less likely to be subject to this distortion, for example, January through October. Unfortunately, date information is not released to the public. Alternatively, if there was a temporary "boom" in applications, several years could be strung together for analysis, but otherwise there is not much that can be done. Fortunately, this timing issue has become less of a problem over time due to improvements in information processing in loan underwriting. The median number of days between application and closing for both home purchase and refinance loans in the 2005 HMDA data was under 30 days, down considerably from earlier years.<sup>6</sup>

Finally, there was a particular timing problem created by the significant changes in HMDA reporting requirements between the 2003 and 2004 filing years. To ease compliance burdens, the Federal Reserve established transition rules to cover applications taken in 2003, but filed during the 2004 HMDA reporting year. The rules affected a number of aspects of the filing. Lenders were allowed to use the old Regulation C rules covering loan purpose definitions and institutions were not required to designate manufactured home loans as such. Applications initiated through a pre-approval program similarly did not have to be so designated.

Transition problems were also created for the racial and pricing variables. Lenders were required to use the pre-2004 racial classifications for all applications received

in 2003 but were required to use the revised rules for all applications received in 2004 and later. As discussed below, this creates significant problems in analyzing racial patterns in the 2004 HMDA filing year, which includes both applications covered by the transition period rules and those that were not.

The transition rules provide that for loans with application dates before January 1, 2004, lenders need not report pricing information. As a consequence, some indeterminate proportion of higher-priced loans is reported in the same way as loans that did not meet the price reporting threshold requirements. The reported incidence of high-priced loans for loans covered by the transition rule is only about one-half that of other loans. The inability to distinguish higher-priced loans from others that were originated in 2004 and 2005 but with application dates before January 1, 2004, means that users of the data need to take this limitation into account when assessing the data. Generally, this means dropping such loans from pricing analysis (but not necessarily other types of analysis such as denial rates, where for reasons discussed above, it is important to keep in mind that the transition period applications are disproportionately approvals). To facilitate the segregation of transition period applications and loans from others, the Federal Financial Institutions Examination Council (FFIEC) placed a flag on the 2004 and 2005 public data release files identifying transactions initiated in the transition period.<sup>7</sup>

### Type of Lender

Information in HMDA can be used to identify the type of lender and affiliations among lenders. Commercial banks can generally be identified by their agency code as filing with one of the three commercial bank regulators, thrift institutions as filing with Office of Thrift Supervision (OTS), credit unions as filing with National Credit Union Administration (NCUA), and independent mortgage companies as filing with U.S. Department of Housing and Urban Development (HUD). Unfortunately, these classifications are not always accurate. Some misclassifications are systematic; for example, some thrift institutions (federal savings banks and some cooperative banks) file with the Federal Deposit Insurance Corporation (FDIC) not the OTS. In other cases, reporters just make errors in their filings. These errors (which happen relatively infrequently) are often associated with changes in status (e.g., change in charter or an acquisition by another institution).

Classification problems can also occur because of the rule that stipulates that applications are filed under the HMDA identification number of whatever entity owns the lender at the end of the filing year. Thus, if a lender acquires another lender in the middle of the year, one filing is made for the combined institutions. It is not possible to determine from HMDA which institution actually processed the application. This can create distortions, particularly if the institutions differ substantially in loan products and practices. If, for example, a commercial bank focusing on prime lending acquires an independent mortgage bank focused on

subprime lending in November, their combined filing would make it appear that all the loans were processed by the commercial bank and create a distorted picture of its ongoing business. Unfortunately, this is a problem without an easy solution. Without substantial external information, it would be impossible to disentangle the applications.

The HMDA lender file contains a “parent” field that can be used to identify affiliations and ownership patterns among files. However, this is also an imperfect indicator. Parents of depository institution filers are not shown so users need to append additional information to the file, such as structural data from the Federal Reserve’s National Information Center (NIC) website, to ascertain which depositories share ownership. Parent information for subsidiaries of depositories and bank holding companies is provided. However, often the parent listed is the direct parent, not the “high holder.” External information will generally be needed in these cases to fully identify affiliations.

Affiliations among independent mortgage banks are more difficult to identify. In many cases no parent information is provided even though the institution is indeed affiliated with other HMDA filers. For example, Long Beach Mortgage is a large HUD-filing institution affiliated with Washington Mutual Savings Association, which files with OTS. Yet information in the HMDA data shows no linkage. Another issue in identifying affiliations arises with HMDA filers, which are partially owned by another filing entity or are owned by a consortium. For years, HMDA parent rules required that an institution had to own 50% or more of a reporter to be listed as an affiliate. In 2005, this was changed with a new code that identifies partial ownership. However, such affiliations cannot be identified in prior years.<sup>8</sup>

### Geographic Issues

There are several data issues related to geography as reported under HMDA. Most notable is the focus on MSA lending in HMDA. Reporting is limited in both MSAs and rural areas, although the reporting is more complete and detailed for MSAs. As noted, small depository institutions need not report in either area. Moreover, any depository institution or mortgage company located exclusively in rural areas need not report. In addition, the geographic detail (census tract, state, and county) of loans outside of MSAs in which they have an office (including all rural areas) does not have to be reported by depositories under \$250 million in assets in 2004 and \$1 billion in assets in 2005 (however, most institutions, regardless of size choose to report the detailed geographic information). Mortgage companies, whether independent or affiliated with depositories, must report and fully identify the geography for all loans in MSAs in which they have five or more applications but need not provide detailed geography about other MSA lending or rural lending.

While HMDA coverage for MSAs is quite complete, these reporting exceptions lead to significant distortions in the coverage of rural areas in HMDA. For rural

areas located just outside MSAs and serviced by the lenders operating in and around those MSAs, HMDA coverage might be quite high; yet for more isolated rural areas coverage is likely less complete. For these reasons, rural areas are often dropped from analysis of HMDA, which examines the “market.” If the focus is on a given lender, however, and the lender provides full geographic detail on all its loans, then it is appropriate to include rural loans in such an analysis.

These considerations are particularly important in studies over different time periods. MSA boundaries were reasonably stable from 1996 to 2003. However, in 2004 new MSAs were designated and boundaries of existing areas were, in some cases, redrawn to reflect the population changes recorded in the 2000 Census. A total of 288 previously rural counties were added to HMDA MSA coverage as a consequence. On the other hand, 46 counties that had previously been part of an MSA were changed to rural designations. The number of lenders filing HMDA data increased by 9% from 2003 to 2004; in part, the increase reflects the larger number of counties in MSAs.

Finally, as a consequence of Office of Management and Budget (OMB) redefinitions, eleven larger metropolitan areas were subdivided and split into metropolitan divisions. These splits make it difficult to compare lending in these areas over time. They also make it particularly difficult to track lending to different income groups over time as most income classifications of borrowers or census tracts are done relative to the median income of the larger metropolitan area in which the individual or census tract is located. With a split, the denominator in such calculations changes greatly. Thus, it is important that any market analysis that spans time be restricted to counties in which there was consistent HMDA coverage.<sup>9</sup>

Other geographic issues in HMDA are less important. Lenders are allowed to suppress census tract numbers in small counties (those with populations under 30,000 individuals whether in or outside of MSAs), although most lenders do report such data. Similarly, some loans for manufactured houses show only the county or state of the property (since the lender may not have known precisely where the unit was to be located). Reporting of geography for pre-approval programs presents a more systematic issue. Under the reporting rules, lenders need not provide geographic information for applications for pre-approval that are denied or approved but not accepted. For 2004, to the extent that such information was provided, the level of detail was suppressed by FFIEC data processors. This means that the property location is unknown for pre-approval denials making it impossible to do a market-by-market analysis of pre-approval denial rates. Such analysis must be done at the aggregate national level for each institution reporting such applications.

### *Race and Ethnicity Definition Issues*

From 1990 to 2003 individuals and lenders were forced to choose among six possible racial or ethnic classifications—white, black, Hispanic, Asian or Pacific

Islander, American Indian and Alaska Native, or “other.”<sup>10</sup> In 2004, the reporting rules were changed significantly. Individuals were separately asked to report on their ethnicity (Hispanic or non-Hispanic) and race (white, black, Asian, American Indian and Alaska Native, Hawaiian or other Pacific Islander). HMDA follows OMB guidelines; individuals are allowed to choose more than one racial classification (all five could be checked).

The changes in reporting for race and ethnicity make it difficult to align the HMDA data for 2004 with those for earlier years. Most importantly, individuals who in 2003 were classified as Hispanic were not also classified by their race.<sup>11</sup> Consequently, a comparison of lending activity by race between 2004 and earlier years might lead some to conclude that lending to certain racial groups may have changed when, in fact, the only change was in the classification system.<sup>12</sup>

The new racial and ethnic groupings would present challenges even if there were no transition period issues. An issue arises in how to represent applications with multiple racial and ethnic classifications, either for one individual or between applicant and co-applicant. Many classification patterns are possible as shown in Exhibit 1. In practice, though, less than 0.1% of all individuals classify themselves as two or more minority races. The major duplications are individuals classifying themselves as white and a minority race (about 0.3% of applications) where one applicant is white and the other a minority (about 2.5% of applications) and where the individual classifies themselves as Hispanic with a minority racial code (about 1% of applications).

One possible solution, adopted by some, is to analyze race and ethnicity separately; that is, Hispanics versus non-Hispanics, and whites versus blacks and other races. The drawback to this approach is that, if there are three main groups (say non-Hispanic whites, Hispanic whites, and blacks) the two separate analyses distort the comparison. The ethnic test compares Hispanics versus a combination of white and black non-Hispanics and the racial test is a comparison of white Hispanics and non-Hispanics versus black Hispanics and non-Hispanics. Each of these comparisons combines two or more groups potentially watering down the distinction.

Another approach is to adopt a hierarchy. Of necessity, an ordering must be chosen, and it can be arbitrary. To maintain historic consistency in preparing summary tables for reporters and MSAs, the FFIEC has adopted a hierarchy where race “trumps” ethnicity. Individuals are first classified by race (mixed white and minority applicants are treated as minority and dual minority status is allowed) and ethnicity is treated as a subset of white. Thus, Hispanics can only be “white Hispanics.” A black Hispanic would be treated as black and so forth. The difficulty with this approach is that many Hispanics appear to regard Hispanic ethnicity as their race, and consequently fail to provide an indication of race. Moreover, when race is provided it may not be what is expected. The most prominent example is Mexican-Americans who descended in part from Mayans or Aztecs, and choose to report themselves as Native Americans. Because the Hispanic group is large

**Exhibit 1** | Classification Patterns

1st Lien Applications, Owner-Occupied, 2005 HMDA				Distribution of Co-applicant Characteristics (each row sums to 100%)											
Applicant Race/ Ethnicity	Number	Overall Percent	High Rate Incidence	Denial Rate	No Co-app. <sup>a</sup>	Same as App. <sup>b</sup>	Race/Ethnicity Different Than Applicant								
							Hispanic			Non-Hispanic			Missing Ethnicity		
							White	Minority	Missing	White	Minority	Missing	White	Minority	Missing
<b>Hispanic</b>															
Black	41,657	0.19	41.52	32.37	68.71	16.27	4.78	0.51	0.41	1.55	6.62	0.12	0.04	0.24	0.75
American Indian	113,139	0.51	25.51	32.76	57.29	35.52	1.27	0.33	0.39	2.59	1.86	0.05	0.05	0.10	0.53
Pacific Islander	34,108	0.15	30.09	31.29	61.69	27.93	1.60	0.69	0.38	3.66	2.54	0.09	0.27	0.22	0.93
Asian	13,368	0.06	28.91	26.99	62.78	20.11	3.65	1.03	0.64	3.46	7.29	0.06	0.13	0.34	0.52
2 or more Minorities	2,193	0.01	35.06	33.05	62.38	17.88	1.73	5.29	2.01	4.29	4.61	0.32	0.09	0.41	1.00
White	2,106,327	9.55	36.26	26.66	63.85	30.14	0.00	0.21	0.26	4.53	0.39	0.03	0.13	0.01	0.46
White & Minority	11,410	0.05	24.38	27.56	56.78	25.34	7.74	2.14	0.32	4.63	2.27	0.04	0.18	0.07	0.48
Missing Race	282,500	1.28	36.45	34.20	60.61	30.74	0.95	0.31	0.00	3.16	0.51	1.11	0.53	0.14	1.93
<b>Non-Hispanic</b>															
Black	2,006,142	9.10	46.35	37.47	72.62	24.21	0.21	0.17	0.05	1.48	0.36	0.14	0.02	0.18	0.54
American Indian	77,767	0.35	31.29	35.36	58.85	18.86	1.12	0.70	0.28	16.49	2.02	0.34	0.20	0.25	0.88
Pacific Islander	90,209	0.41	26.95	26.19	52.93	32.30	1.06	0.48	0.25	9.09	2.59	0.25	0.07	0.22	0.76
Asian	697,715	3.16	15.26	19.27	60.38	34.17	0.38	0.18	0.07	3.39	0.55	0.14	0.03	0.20	0.49
2 or more Minorities	12,534	0.06	24.07	31.63	61.94	17.66	0.65	0.73	0.24	5.86	11.85	0.14	0.02	0.25	0.66
White	12,018,974	54.50	18.93	22.30	52.23	45.17	0.73	0.05	0.09	0.00	0.77	0.15	0.28	0.01	0.51
White & Minority	41,490	0.19	23.94	32.51	59.23	8.22	1.26	0.45	0.23	21.89	7.74	0.16	0.14	0.12	0.57
Missing Race	371,222	1.68	32.91	35.94	62.42	30.90	0.10	0.02	0.53	2.58	0.57	0.00	0.14	0.05	2.70

**Exhibit 1** | (continued)  
Classification Patterns

1st Lien Applications, Owner-Occupied, 2005 HMDA			Distribution of Co-applicant Characteristics (each row sums to 100%)												
Applicant Race/ Ethnicity	Number	Overall Percent	High Rate Incidence	Denial Rate	No Co-app. <sup>a</sup>	Same as App. <sup>b</sup>	Race/Ethnicity Different Than Applicant								
							Hispanic			Non-Hispanic			Missing Ethnicity		
							White	Minority	Missing	White	Minority	Missing	White	Minority	Missing
Missing Ethnicity															
Black	90,873	0.41	50.09	40.35	71.64	19.67	0.05	0.08	0.21	0.21	1.51	0.09	0.98	0.35	5.21
American Indian	20,515	0.09	32.85	40.18	62.69	30.42	0.10	0.18	0.32	0.64	0.38	0.08	2.14	0.51	2.54
Pacific Islander	6,362	0.03	45.11	20.21	60.69	24.49	0.33	0.27	0.53	0.88	1.59	0.13	4.72	1.51	4.87
Asian	27,464	0.12	18.04	21.00	60.83	29.50	0.11	0.12	0.42	0.62	2.33	0.12	2.64	0.57	2.75
2 or more Minorities	669	0.00	27.68	33.88	65.47	10.91	0.00	0.75	0.90	0.45	2.69	0.00	2.54	9.27	7.03
White	432,786	1.96	23.40	27.62	54.58	35.64	0.31	0.02	0.35	2.49	0.10	0.10	0.00	0.73	5.68
White & Minority	1,844	0.01	29.08	35.65	60.52	7.43	0.92	0.54	0.81	4.66	1.19	0.27	16.05	5.31	2.28
Missing Race	3,552,082	16.11	30.78	42.63	58.56	40.21	0.05	0.01	0.06	0.41	0.08	0.17	0.23	0.21	0.00
<b>Total</b>	<b>22,053,350</b>	<b>100.00</b>	<b>24.45</b>	<b>27.39</b>	<b>39.53</b>	<b>57.03</b>	<b>0.50</b>	<b>0.09</b>	<b>0.11</b>	<b>1.04</b>	<b>0.62</b>	<b>0.15</b>	<b>0.23</b>	<b>0.09</b>	<b>0.61</b>

*Notes:*  
<sup>a</sup>No co-applicant  
<sup>b</sup>Same race/ethnicity as applicant.

relative to other American Indian groups, this can significantly distort data for the American Indian classification (54% of the American Indians in the 2005 HMDA classify themselves as Hispanic).

An alternative hierarchy, and one used by the Department of Justice for many years, is to have ethnicity “trump” race. The major issue here is for black Hispanics, primarily Cubans, Dominicans, and Puerto Ricans, who would be treated as Hispanics under this hierarchy. The HMDA denial rates and pricing information for black Hispanics more closely aligns with that of blacks than that of Hispanics (albeit only marginally). Thus, a variant on the Justice Department approach, and the one used by the authors in the 2005 and 2006 HMDA *Federal Reserve Bulletin* articles, is a hierarchy where black trumps Hispanic, but Hispanic trumps other races. Classification would be similar to the Justice Department approach except that black Hispanics are treated as blacks.

Further complications also arise when the applicant and co-applicant give different responses. The FFIEC hierarchy relies primarily on the applicant’s response. However, analysis of the patterns of different lenders reveals significant inconsistencies in the choice of applicant. For some lenders, the applicant is most often male for different sex couples, but for others the pattern appears to be random. This raises concerns about placing greater reliance on the applicant’s response than the co-applicant’s response. One possible solution to this problem is to incorporate both applicant and co-applicant into the hierarchy. To do this, though, it is necessary to create a hierarchy among all the race and ethnicity classifications. One reasonable ordering is: black, Hispanic, American Indian, Hawaiian and Pacific Islander, Asian, and finally white non-Hispanic. Thus, for example, if either the applicant or co-applicant checks black as one of their races, the application is treated as black, etc.

None of these alternative approaches is necessarily correct. One can circumvent the problem if minority/non-minority is the only comparison made (non-minorities are white non-Hispanics with no other designation, all others are minority). However, this distinction is not always appropriate, particularly because patterns for Asians are often far more like those for non-Hispanic whites than they are for other racial or ethnic groups.

### Missing Race or Ethnicity Data

Over the course of time, a growing share of reported applications and loans did not include race or ethnicity information. A principal contributing factor was the growing use of newer technologies to take applications, particularly for those taken by telephone and other forms of electronic media. From 1993 to 2002, the proportion of home loan applications of all types missing race or ethnicity information increased from 8% to 28%. The incidence of missing race or ethnicity varied some by the reported purpose of the loan. Such information was missing less often for home purchase loans and more often for loans for refinancings and

home improvements. Analysis of the patterns of missing racial information suggested that it was very noisy but not completely random. For example, racial data was disproportionately missing for denied applications in high minority census tracts.

To address this issue, reporting rules were changed for the 2003 data, requiring lenders to ask applicants in telephone applications for race or ethnicity information. Missing race or ethnicity fell in response to the new rules, declining from 28% of all home loan applications in 2002 to 17% in 2003 and 16% in 2005 (Exhibit 1).

Another problem in the more recent racial data is partially missing information. That is, applicants who provide ethnic but not racial information or racial but not ethnic information (about 5.6% of applications); or applicants who provide racial/ethnic information but have missing information for coapplicants (about 1% of applications). The most significant categories are Hispanics with missing race (1.3%), Non-Hispanics with missing race (1.7%), and applications with race but not ethnicity (2.6%). An imperfect but not unreasonable solution to this problem is to treat coapplicants with missing information as having the same information as the applicant, Hispanics with missing race as Hispanic whites, and applicants with missing ethnicity as non-Hispanic with races as specified. Applicants supplying no racial information and identifying themselves as non-Hispanic are treated the same as those providing no information on race or ethnicity.

### Missing Income Data

Regulation C generally requires lenders to report the income relied upon in making credit decisions. However, in certain circumstances income need not be reported. Income is not reported if the lender did not rely upon income in making the underwriting decision, the loan is for a multi-family property loan, the applicant is an employee of the lender, or the applicant or borrower is not a natural person (e.g., the borrower may be a trust or investment company). Also, lenders offer “stated income” loans and do not necessarily verify the reported income. Further, because lenders may only collect information about the income relied on to make a credit decision, the reported income may understate the financial circumstances of the applicant(s). Missing income is more likely for some types of loans than others, particularly home improvement lending where lenders may be particularly focused on collateral issues in underwriting. In total, for first lien home purchase loan applications, 5.6% were missing income information; for junior liens, 3.1% were missing income information.

### Product Definition

HMDA allows applications and loans to be classified by lien status, type of loan (conventional or government-backed), type of property (multi-family,

manufactured home, or other), and owner-occupancy status. It is also possible to use the dollar amount of the loan to identify those that are “size conformable;” that is, loans that meet the size criteria that makes them eligible for purchase by Fannie Mae or Freddie Mac. Each of these distinctions bears on credit evaluation and loan pricing and is a meaningful classification for many types of analyses.<sup>13</sup>

Beyond these distinctions, however, the ability to place applications in loan product groups bumps up against problems of consistency and incompleteness. For example, some analysts have treated non-owner occupied loans as “investor” loans.<sup>14</sup> Non-owner occupied loans have become an important element of the home lending market in recent years. Analysis of the annual HMDA data indicates that non-owner-occupied lending increased from about 5% of all home purchase loans in the mid-1990s to 17% in 2005. However, analysis of the geographic patterns of such loans shows that a significant share are in vacation and retirement areas, suggesting these are second homes. However, other than the imperfect inference from geography, there is no good way to separate the investor from second home loans.

Purchaser code identification also may raise issues. For example, one might believe that it is appropriate to use the purchaser code variable to identify “conformable loans” as those sold to Fannie Mae or Freddie Mac. However, just because a loan is not reported as sold to Fannie Mae or Freddie Mac does not mean that it is not conformable. Furthermore, the purchaser variable is available only for originated loans. Thus, this product distinction cannot be used for denial rate analysis.

Another distinction relates to loan purpose (home purchase, home improvement, or refinance). It is common to separate home purchase loans from the other two categories since underwriting and pricing may differ. The distinction between home improvement and refinancing is less clear. HMDA guidelines state that an application should be designated as “home improvement” if a purpose of the loan is, at least in part, for home improvement. Such loans cannot be home purchase loans since they would be reported in that category; thus, in all likelihood, most first lien home improvement loans are, in fact, refinance loans (a few may be new loans on an unencumbered home). Whether or not such loans are reported as for refinance or home improvement likely depends on lender policy rather than features of the loans. For example, lenders who do not routinely ask the purpose of refinance loans are unlikely to report many as home improvement loans. Moreover, it is not clear that the underwriting or pricing of home improvement loans differs from any other “cash out” refinance loan of the same lien status and term to maturity.<sup>15</sup> Thus, a case can be made for combining the home improvement and refinance categories of a given lien type.<sup>16</sup>

One class of loan that HMDA users should be aware of is “business-related” loans. Some small business owners may be required by lenders to pledge their personal assets, including their home, as collateral against a loan for their business. If such a loan is refinanced and meets other reporting requirements, then it

becomes eligible for HMDA reporting. Such loans are likely to go through a very different underwriting and pricing process than the typical mortgage loan. For example, the loan may not be subject to Regulation Z, the Truth in Lending Act, and thus not eligible for disclosure of pricing information in HMDA. Moreover the rate on the loan may depend on the nature of other account relationships. It is not always possible to identify such loans in HMDA. However, if an applicant or co-applicant is not a natural person, then HMDA requirements stipulate that the race, ethnicity, and gender codes for the applicant (or co-applicant) should be reported as not applicable or “N/A.” Business loans with personal guarantees would come under this rule. Data for 2005 indicate that about 264,000 loan applications out of some 30.2 million fit this definition of “business-related.”<sup>17</sup> Because such loans are underwritten quite differently than other loans, a good case can be made to treat these loans as a distinct category in analysis.

Finally, there are two relatively new product distinctions available in the expanded HMDA data that can potentially be used to classify loans or applications. As noted earlier, as of 2004, lenders are asked to report on applications for home purchase that go through a pre-approval program. It is possible to examine the pre-approval program activity directly, for example, by looking at pre-approval denial rates.<sup>18</sup> Another use of the variable, though, is as a product distinction. It might be the case, for example, that loans which go through a pre-approval process have different underwriting and pricing characteristics than other loans.

Another opportunity created by the new reporting rules is the potential to identify “piggyback” or “80-10-10” loans. These are loans where homebuyers (or refinancers) are simultaneously obtaining first and junior lien loans, generally, but not always, from the same lender. Such simultaneous borrowing has been a feature of the conventional mortgage marketplace for some time, but has grown in importance in recent years as lenders have marketed products intended to offer consumers an alternative to private mortgage insurance (PMI) or, in some cases, a line of credit that may be used for a variety of purposes. Typically, PMI is required on conventional loans with LTV ratios above 80%. Over the past few years, lenders have become more active in self-insuring by waiving PMI requirements if a borrower simultaneously takes out a first lien loan with an LTV ratio of 80% or more and a junior lien loan at a higher price to cover the remaining portion of the loan. The combined loans are often competitive on a price basis with a single loan involving PMI and may offer the borrower a tax advantage because the interest payments on the junior lien loan are generally tax-deductible, whereas the PMI premiums are not.

Piggyback loans are not identified as such in the HMDA data. However, the data provide a basis for identifying piggyback loans if one assumes that two conventional home purchase loans involving properties in the same census tract, from the same lender, and with identical time of application and closing, owner-occupancy status, borrower income, race or ethnicity, and sex involved the same borrower and the same home. Since 2004, the identification process has been improved by the addition of lien status, which earlier could only be approximated

by comparing the size of loans that were matched. For 2005, it is estimated that about 85% of the junior lien home purchase loans for owner-occupied properties can be matched to a first lien loan by this process.

It might appear that this matching process is possible only for those with access to the nonpublic date information in HMDA. However, examination of the data shows that piggyback loans are generally reported in sequence and that this sequencing is retained in the HMDA file released to the public. Calculations show that if the requirement that applications be “within 10 sequence numbers of each other” is substituted for date matching, the identification algorithm works almost as well.

Another product distinction that might be useful in analysis is to identify the channel by which a loan was originated, that is by an independent broker, loan correspondent or through a branch office of the lending institution. Unfortunately, direct information on channel is not available in the HMDA data. However, for large depository institutions, an indirect proxy can be created by matching the HMDA data with information on CRA assessment areas. The CRA data include a listing of the census tracts that comprise the CRA assessment areas of banking institutions. Generally, these are the areas in which the institutions maintain retail banking offices. Thus, loans originated in areas outside of the assessment are more likely to come from some channel other than retail branches (e.g., a broker, correspondent lender, or nonbanking affiliate). Since 2005, the assessment area data can be downloaded from the FFIEC website for CRA data.

Finally, beginning with the 2004 HMDA data, lenders must now disclose whether a loan is subject to the protections of the Home Ownership and Equity Protection Act of 1994 (HOEPA). Concerns about predatory lending led Congress to enact HOEPA, which amends the Truth in Lending Act and applies to closed-end home loans (excluding home purchase loans) bearing an APR or other dollar-amount fees above specified thresholds. Before 2004, little information was publicly available about the extent of such lending or the number or type of institutions involved in such activities. The expanded HMDA data show that many lenders grant loans covered by HOEPA (1,300 lenders in 2005) but just 10 institutions extended some 70% of all the reported loans.

Users of the data should be aware that HMDA does not capture all HOEPA-related lending. Some HOEPA loans are extended by institutions not covered by HMDA, and some HOEPA loans that are made by HMDA-covered institutions are not reported under the Federal Reserve Board’s Regulation C. In particular, if the proceeds of a home-secured loan are not used to refinance an existing home loan or to finance home improvement, then the loan may be covered by HOEPA but is not reportable under Regulation C. For example, if a homeowner takes out a HOEPA-covered loan to pay off outstanding credit card debt or some other type of consumer credit, and the loan does not involve the refinancing of an existing home loan or home improvement, then the loan is not covered by Regulation C and is thus not required to be part of an institution’s HMDA reporting.

---

## Background on Pricing Data in HMDA

Beginning with the 2004 HMDA data, lenders have been required to disclose limited information on loan pricing. Specifically, lenders are required to disclose the spread between the APR on the loan and the yield on Treasury securities of comparable maturity when the spread exceeds 3 percentage points for first lien loans or 5 percentage points for junior liens. Additionally, the new regulations require lenders to identify loans subject to the protections of HOEPA.

The availability of the new pricing information in HMDA increases the scope of analysis that can be undertaken either by researchers or in support of fair lending enforcement, and makes possible an assessment of pricing in the higher-priced segment of the home loan market. Important limitations on the usefulness of the data exist and several of these are discussed here.<sup>19</sup>

In selecting the threshold for reporting loans as higher-priced, the objective of the Federal Reserve Board was to require that pricing on most subprime loans *would* be reported and that most prime loans *would not* require disclosure of pricing information. Analysis at the time using the Annual Housing Survey (AHS) covering prime, near-prime, and subprime loans suggested that, based on historic experience, the chosen thresholds of 3 percentage points for first liens and 5 percentage points for junior liens would fall somewhere in the near-prime range and would require the reporting of about 10% of all loans backed by first liens and about 22% of all loans backed by junior liens. The evidence suggested that roughly 98% of first lien prime loans would have APRs below the reporting threshold, while 98% of first lien subprime loans would be reported as higher priced.

The 2004 and 2005 HMDA data, as well as industry data, suggest that both the near prime and subprime segments of the market grew rapidly from the time of the AHS survey that was used to help establish the pricing thresholds. Moreover, the mix of loans products in the subprime mortgage sector has changed over the past few years as adjustable-rate loans garnered a larger share of this market. As discussed below, the changing mix of fixed-rate and adjustable-rate loans affected the reported incidence of higher-priced lending. Overall, for the first two years of data collected, the incidences of higher-price lending for first lien loans was about 13% in 2004 and 24% in 2005. For junior lien loans, the incidences were about 30% in 2004 and 38% in 2005. All of these numbers are higher than the AHS projections.

For both lien types, the incidence of higher-priced lending for 2005 is substantially higher than in 2004, raising questions about which of the two, if either, is a typical year. As described below, some of the increase in higher-priced lending between the two years is a reflection of changes in the yield curve and not related to fundamental changes in consumer or credit behavior.

---

## Pricing Issues in the HMDA Data

Year-to-year changes in the number or proportion of loans with prices that exceed the thresholds for reporting price information under HMDA must be interpreted with great care. It is tempting to assume that a change in the incidence of higher-priced lending from one year to the next reflects changes in the volume of subprime lending activity. This simple interpretation ignores a number of factors that may influence the incidence of *reported* higher-priced lending.

Three factors may lead to a change in the reporting of higher-priced lending. The first is lenders' business practices, particularly lenders' willingness or ability to bear credit risk.

The second factor is consumers' borrowing practices or credit-risk profiles. The third factor is a change in the interest rate situation—specifically, the relationship between short- and long-term interest rates. Generally, interest rate changes can significantly affect whether loans are reported as higher priced but are less likely to affect the credit-risk component of loan pricing. The credit-risk component can be affected if interest rate movements influence the loan-product mix that borrowers use. In some years, for example, adjustable-rate loans, which tend to have higher default rates, may be relatively more attractive than fixed-rate loans.

---

### *The Interest Rate Situation and the Yield Curve*

The yield curve displays how the yield on financial instruments, such as U.S. Treasury securities, varies with maturity and, therefore, reflects the relationship between short- and long-term interest rates. The yield curve is typically upward sloping; that is, short-term rates are typically lower than long-term rates. Sometimes, however, the yield curve is relatively flat; that is, short-term rates are close to long-term rates. Occasionally, the yield curve inverts, and short-term rates rise above long-term rates.

Changes in the shape of the yield curve affect the reporting of higher-priced loans under HMDA. Because most mortgages prepay in a relatively short period (well before the stated term of the loan is reached), lenders use relatively short-term interest rates to set mortgage rates. But for most loans, Regulation C requires lenders to use longer-term rates to determine whether to report a loan as higher priced because the stated maturity of most loans, particularly first lien loans, exceeds twenty years. Thus, a change from one year to the next in the relationship between short- and long-term rates can cause a change in the proportion of loans that are reported as higher priced, all other things being equal.

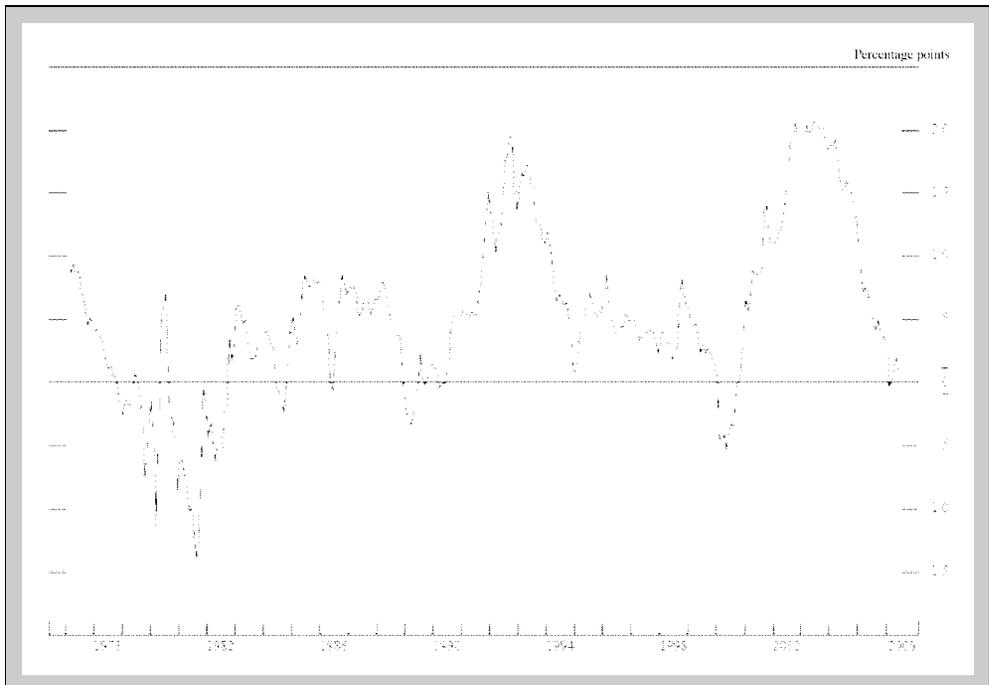
For example, if short-term rates rise relative to long-term rates, then the number and proportion of loans reported as higher priced will increase even if all other factors that may influence the number and proportion of higher-priced loans, such

as the business practices of lenders and the credit-risk profiles and borrowing practices of borrowers, remain constant. Conversely, if short-term rates fall relative to long-term rates, then the number and proportion of loans reported as higher priced will fall even if all other possibly influential factors remain constant.

### Changes in the Yield Curve from 2004 to 2005

A review of the rate spreads between five-year and thirty-year Treasury securities over the past two decades indicates that 2004 (and 2003) were somewhat unusual years by historical standards because the yield curve was particularly steep during this time because of low short-term rates (Exhibit 2). Consequently, the gap between longer- and shorter-term rates was particularly large. In mid-January 2004, for example, the yield on five-year Treasuries securities was 2.97%, and the yield on thirty-year Treasuries was 4.87%. Over the course of the year, the difference narrowed as shorter-term rates rose and longer-term rates fell slightly. By early January 2005, the yield on five-year Treasuries had risen to 3.71%, and the yield on thirty-year Treasuries had fallen to 4.72%. Shorter-term interest rates

**Exhibit 2** | Spread between Interest Rates on the 5-Year and 30-Year Treasury Bonds: 1977–2006



Notes: After March 2002, the spread is between the 5-year and the 20-year bonds. The source is the Federal Financial Examination Council's "FFIEC Rate Spread Calculator," found at: [www.ffiec.gov/ratespread/default.aspx](http://www.ffiec.gov/ratespread/default.aspx).

continued to rise through 2005, while longer-term rates were essentially unchanged. Thus, although at the beginning of 2004 short-term rates were well below long-term rates, by the end of 2005 short- and long-term rates were much closer.

Because of the changes in the relationship between short- and long-term interest rates, the gap between the effective interest rate (measured by the APR on the loan) on most mortgages and the HMDA threshold for reporting higher-priced loans narrowed markedly between 2004 and 2005. For example, for loans priced during the week of January 15, 2004, the average APR on conventional first lien fixed-rate thirty-year prime loans reported by Freddie Mac was 5.72.<sup>20</sup> As a result, a gap of 215 basis points separated the APR of the typical prime loan priced that week and the HMDA reporting threshold. By December 15, 2005, the gap between the calculated APR and the HMDA threshold had narrowed to 140 basis points. Although factors other than interest rate changes may also have influenced the proportion of loans reported as higher-priced under HMDA, this example clearly demonstrates that even if such factors (including business practices or consumer credit-risk profiles) had remained the same, the reported incidence of higher-priced lending would have increased in 2005.

To estimate the effect on loan pricing of the way the HMDA price-reporting threshold is determined, a modified set of the 2004 and 2005 HMDA data was created to identify those first lien loans that exceeded the pricing thresholds *solely* because of a change in the interest rate situation.<sup>21</sup> All reported higher-priced first lien loans were separated into two groups: (1) those that would have been reported under any interest rate situation that prevailed during the 2004–2005 period and (2) those that were reported only because of the interest rate situation that existed at the time the loan was made. In separating the higher-priced loans, several assumptions were made. First, all loans were assumed to be thirty-year fixed-rate. (Almost 80% of first lien prime mortgages have a fixed rate of interest, according to LoanPerformance, and most of these have a thirty-year term to maturity).<sup>22</sup> Second, the date the loan price was set (the lock date) was assumed to be midway between the action taken date and the application data.<sup>23</sup> Third, regardless of the timing of their loan, nonprime borrowers with a given risk profile were assumed to receive loan rates that were a constant markup over the rate on a “prime mortgage” as defined in the Freddie Mac interest rate survey. Under these assumptions, the “adjusted spread” for each loan in the HMDA data is calculated as the difference between the estimated prime fixed APR and the applicable HMDA threshold in effect on the date the loan was estimated to have locked.

A loan with an adjusted spread of 228 or more basis points above prime would have been reported as higher priced regardless of the date of origination during 2004–2005; that is, 228 basis points is the minimum spread for a loan to have been reported as higher priced during this period. Loans with adjusted spreads between 140 basis points and 228 basis points would have been reported as higher priced if originated on some days during the period but not on others. Loans with adjusted spreads below 140 basis points would not have been reported under any circumstances during this time frame.

Overall, the incidence of higher-priced lending for conventional home purchase loans on owner-occupied site-built homes was 11.5% in 2004 and 24.6% in 2005, an increase of 13.1 percentage points. The spread-adjusted estimates for the same period were 10.5% and 21.9% respectively, an increase of 11.4 percentage points. This comparison suggests that nearly 2 percentage points, or roughly 13%, of the total difference in reported higher-priced lending for this product can be attributed solely to the flattening of the yield curve. For refinancings for similar properties, about 2 percentage points of the 10.2% increase in higher-priced lending for refinance loans can be attributed solely to the yield curve.

Estimated mean APR spreads (the APR spread reported in HMDA for borrowers exceeding the price reporting threshold) are also lower after spread adjustment for the two conventional loan products. The mean APR spreads for conventional first lien home purchase loans, and for conventional first lien refinance loans, on owner-occupied site-built homes were both about 4.8 percentage points before spread adjustment and 3.8 percentage points and 4.0 percentage points respectively using the spread-adjusted 2005 data. Both comparisons show an increase in mean spreads of 70 or 80 basis points from 2004 to 2005.

The estimates derived here required knowledge of the relevant dates in order to calculate the spread adjustment. Unfortunately, the dates are not available to the public. Consequently, these adjustments cannot be easily made by others. The results suggest that mixing the 2004 and 2005 data on loan pricing is problematic. However, variations of the adjusted spread within the 2005 data are relatively small, suggesting the 2005 data are largely unaffected by yield curve effects.

### *Effects of Yield Curve Changes on the Reporting of Fixed- and Adjustable-Rate Loans*

Under the Federal Reserve Board's Regulation Z, in calculating the APR for adjustable-rate loans, lenders assume that the interest rate situation at the time of origination will continue for the term of the loan. When the yield curve is steep, it suggests that the market expects short-term interest rates to rise, yet the APR calculation for adjustable-rate loans assumes that interest rates will stay the same.<sup>24</sup> Because of this regulatory construct, when the yield curve is positively sloped, the APRs for adjustable-rate loans tend to be lower than those for fixed-rate loans of similar term and credit risk.

Thus, the flattening of the yield curve can have two effects. First, as noted earlier, it narrows the gap between the longer-term rates used for the HMDA reporting threshold and the shorter-term rates used for pricing loans. Second, the flattening narrows the APR gap between adjustable- and fixed-rate loans because, as short-term interest rates increase, it reduces the differences in APRs between fixed- and adjustable-rate loans.<sup>25</sup>

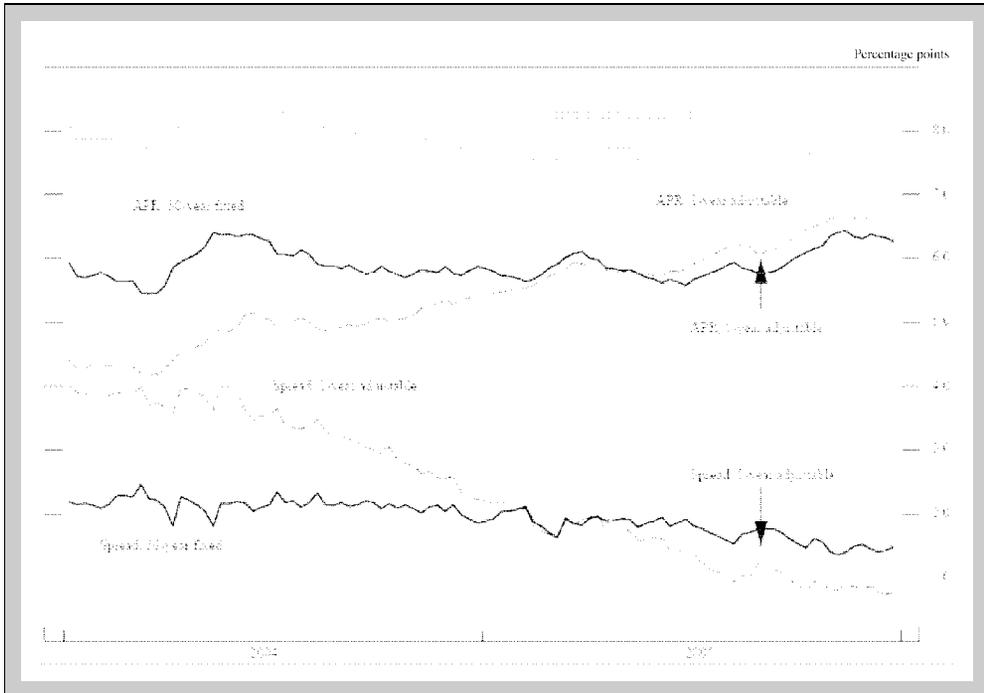
The likely result of the flattening of the yield curve from 2004 to 2005 was an increase in the proportion of adjustable-rate loans that exceeded the HMDA price-

reporting thresholds. The increase occurred because many relatively high-rate adjustable-rate loans that would not have been reported as higher priced in 2004 because of comparatively low APRs were reported that way in 2005.

Exhibit 3 shows the APRs of the prime fixed-rate thirty-year loans, the prime one-year adjustable-rate loans, and the prime five-year adjustable-rate loans reported in the Freddie Mac mortgage interest rate survey for 2004–2005.<sup>26</sup> The bottom three lines of the exhibit represent the differences (gaps) between the effective rates (APRs) reported by Freddie Mac and the HMDA reporting threshold. As noted earlier, the reporting gap between the typical prime thirty-year fixed-rate loan and the reporting threshold for 2004 narrowed from 215 basis points at the beginning of 2004 to 140 basis points at the end of 2005. For one-year adjustable-rate loans, the gap narrowed much more, from 404 basis points at the beginning of 2004 to only 75 basis points at the end of 2005.

Although the differences between the APRs on fixed- and adjustable-rate loans and the reporting threshold decreased for both types of loans, the decrease for

**Exhibit 3** | HMDA Price-Reporting Threshold, Interest Rates for Fixed- and Adjustable-Rate Loans, and Spreads between the Threshold and Such Rates: 2004–2005



Notes: For explanation for the HMDA price-reporting threshold, refer to text. Threshold and APRs are for the conventional first-lien 30-year prime loans. APRs are estimated from the Freddie Mac, Primary Mortgage Market Survey.

adjustable-rate loans was much larger. Thus, the gap between the APRs on fixed- and adjustable-rate loans, which was substantial at the beginning of 2004, had been virtually eliminated by the beginning of 2005. This finding suggests that, as an artifact of regulation, geographic areas with different percentages of fixed-rate versus adjustable-rate loans might have shown different incidences of higher-rate loans in 2004. That is, in 2004, areas with larger shares of adjustable-rate loans likely had fewer higher-priced loans than areas with larger shares of fixed-rate loans. This effect should have been much smaller in 2005 (and in the first half of 2006) because interest rates on adjustable- and fixed-rate loans were closer together.

Fully quantifying this effect would be difficult even if the HMDA data distinguished fixed- from adjustable-rate loans. As can be seen in Exhibit 3, applying the same method to one-year adjustable-rate loans as employed for fixed-rate loans would necessitate using an adjusted threshold of about 400 basis points above the APR on the Freddie Mac prime one-year adjustable-rate loan. This would potentially exclude a large share of the higher-priced adjustable-rate loans reported under HMDA and reflect only changes at the higher end of the subprime market.

To a limited extent, the differential effect of the flattening of the yield curve on the incidence of higher-priced lending among adjustable- and fixed-rate loans can be assessed using individual loan data. The Monthly Interest Rate Survey (MIRS) of the Federal Housing Finance Board is a monthly survey of major lenders that collects detailed information on each conventional, single-family non-farm loan used to purchase a home closed during the last five business days of each month.<sup>27</sup> The survey includes enough information to calculate an APR for each loan, to determine whether it is an adjustable- or fixed-rate loan, and, among the adjustable-rate loans, to identify the type of loan.

The focus of the survey is on conventional prime rate loans. For 2004, the data for one-year adjustable-rate loans included near-prime loans. But starting with the data for February 2005, most of the near-prime loans were excluded from the sample. Thus, the analysis of the one-year adjustable-rate loans is limited to those made before February 2005.

For the first few months of 2004, the percentage of one-year adjustable-rate loans included in the MIRS data that would have been reported as higher priced under HMDA was only about 1%. This percentage began to rise substantially in the middle of 2004, and by the end of the year the portion had risen to more than 18%. It appears that, at least within the MIRS data, the increase in the incidence of loans that would have been reported as higher priced under HMDA is driven almost entirely by the narrowing of the gap between the calculated APRs on adjustable- and fixed-rate loans. The distribution of rates of one-year adjustable-rate loans in the MIRS data *relative to the APR* of the Freddie Mac prime one-year adjustable-rate loans remained unchanged, that is, their movements mirrored each other.

The percentage of thirty-year fixed-rate loans in the MIRS data that would have been reported as higher priced under HMDA was little changed during the course of 2004 but rose, from about 1% to almost 4%, during 2005. The thirty-year fixed-rate loans were not pruned to exclude near-prime loans.

The patterns for the thirty-year fixed-rate and the one-year adjustable-rate loans in the MIRS data are consistent with what one would expect from the yield curve changes shown in Exhibit 3. Note that because the MIRS data do not include a full sampling of near-prime and subprime loans, the incidence of loans in the sample that would have been reported as higher priced under HMDA is not representative of all the loans included in the HMDA data. Nevertheless, the analysis here suggests that the flattening of the yield curve had a significantly larger effect on the reporting of adjustable-rate loans as higher priced than on the reporting of fixed-rate loans as higher priced. Therefore, the earlier estimate of the effect of the flattening of the yield curve is likely understated, perhaps substantially.

---

### Opportunities in Using HMDA Data

Based on the preceding discussion, one might conclude that the HMDA data are quite limited and fraught with technical difficulties. Despite the limitations, much can be done with the data to gain a better understanding of trends and practices in the home mortgage market. The potential uses of the data are too numerous to enumerate here; however, to illustrate the potential value of the information in this section, an analysis of racial differences in the incidence of higher-priced lending is presented along with a discussion of how the HMDA data can help account for a substantial portion of such differences. Clearly, HMDA data do not include all of the factors considered in evaluating and pricing credit. However, by controlling for differences along the dimensions that are available in the data, so as to make borrowers as similar as possible, one can account for much of the difference in the outcomes in the lending process across groups.

The HMDA data enable individuals of different races or ethnicities to be matched by loan type and purpose, type of property securing the loan, lien status, owner-occupancy status, property location (e.g., same MSA or even same census tract), the income relied on for underwriting, loan amount, and time of year when the loan was extended (for those with access to the date information), as well as whether the loan involved a co-applicant, the applicant and co-applicant's gender, and whether the loan involved a simultaneous second lien (a piggyback) or began in a pre-approval program.

Comparisons of pricing outcomes across groups can be conducted at the level of an individual institution, groups of institutions (e.g., manufactured-home lending specialists or specialists in higher-priced lending), geographic market, or populations as a whole. Further, a variety of statistical methodologies can be used to control for the effects of the credit-related or other factors in the HMDA data

noted above. In the analysis here uses a statistical technique commonly utilized in assessing differences in fair lending and employment discrimination investigations. Using 2005 HMDA data, a linear probability model is employed to predict the probability that a first lien, owner-occupied site built home purchase or refinance loan within the Fannie Mae and Freddie Mac “size” conformable limit will be higher-priced under HMDA *using only non-Hispanic white* borrowers.<sup>28</sup> A series of equations are estimated progressively adding explanatory variables. Coefficients from these equations are used to predict the likelihood that borrowers of other races will have higher-priced loans. The difference between the actual and predicted incidences of higher-priced lending for black borrowers is the unexplained difference across groups. The purpose here is to describe how much the unadjusted gross difference in higher-priced lending across the groups can be narrowed using the data items available in HMDA.

Group differences are evaluated at three levels successively controlling for more information. The first level is the gross difference controlling for nothing. The second accounts for product type (conventional or government-backed), piggyback and pre-approval status, market location (MSA or non-metropolitan section of the state), borrower income, loan amount, and applicant and co-applicant gender. The final comparison adds the specific lender.

Results of the analysis are shown in Exhibit 4. The analysis suggests that a significant portion of the differences in the incidence of high rate lending between

**Exhibit 4** | Explaining Racial Differences in the Incidence of High-Priced Lending, Site-Built, 1st Lien, Owner-Occupied, 2005

	Loans	Gross Incidence	Residuals from Whites Controlling for:		
			Nothing	Prod/MSA/Income	Everything
<b>Home Purchase</b>					
White	2,789,265	17.2	0.0	0.0	0.0
Black	312,451	54.7	37.5	29.8	10.0
Hispanic	464,634	46.1	28.9	17.0	4.7
American Indian	27,766	35.3	18.1	12.3	4.6
Pacific Islander	23,450	34.8	17.6	13.2	3.8
Asian	237,383	16.6	-0.6	-1.4	-0.6
<b>Refinance</b>					
White	3,496,425	21.0	0.0	0.0	0.0
Black	441,299	49.3	28.3	24.0	6.2
Hispanic	478,381	33.8	12.8	10.5	2.6
American Indian	37,213	28.9	7.9	11.1	3.1
Pacific Islander	31,453	28.4	7.4	11.2	3.3
Asian	165,011	19.3	-1.7	-2.1	1.4

racial and ethnic groups can be explained by the variables in HMDA. For example, the gross gap between blacks and white non-Hispanic home purchase borrowers of 37.5% is reduced to 10.0% when factors in HMDA are controlled for. For refinanced loans, the proportional reduction is even greater, from 28.3% to 6.2%. For refinance lending, the gap between non-Hispanic whites and Hispanics is almost eliminated when factors in HMDA are controlled for. Although controlling for borrower characteristics reduces the gap, it is clear that lender plays a significant role. Further, the relatively large reduction that occurs when lender is taken into account suggests that analyses of racial and ethnic differences in higher-priced lending should be done on an institution-specific basis. Nonetheless, it is clear that significant, unexplained gaps remain after all available HMDA factors are taken into account.

---

## Conclusion

In sum, the analysis presented here suggests that HMDA data can be useful in explaining differences in pricing patterns across groups. However, in conducting this exercise, a number of choices had to be made, such as assigning borrowers to a specific racial or ethnic group, how to treat loans that arose through a pre-approval program, determining the time period of the analysis, how to define a high rate loan, how to define product, which markets to use and how to define them, and which loans were “piggybacks.” As described earlier, many of these choices have elements of arbitrariness. None of these choices would be of sufficient force to invalidate the qualitative conclusions of the analysis. The exact numbers, however, are probably not robust and could well change if decisions were made differently. Ultimately all users of HMDA need to be aware of this and strike a proper balance between capitalizing on the rich source of information that HMDA is and properly informing their readers when their conclusions rest on arbitrary assumptions.

---

## Endnotes

- <sup>1</sup> See, for example, Goering and Wienk (1996).
- <sup>2</sup> Refer to Home Mortgage Disclosure Act (12 U.S.C. §§ 2801-11), Regulation C (12 C.F.R. pt. 203), and the staff commentary accompanying Regulation C (12 C.F.R. pt. 203, Supp. I). The final revisions to Regulation C were issued on February 15, 2002 and June 27, 2002.
- <sup>3</sup> Individual lenders covered by HMDA are required to make their data available to the public beginning on March 31 of the year after the calendar year for which the data apply. However, the data made available at that time have not been systematically checked by the supervisory agencies for errors or omissions, as have the HMDA data released by the FFIEC in September each year.
- <sup>4</sup> See Avery, Canner, and Cook (2005) and Avery, Brevoort, and Canner (2006).
- <sup>5</sup> Additional criteria for coverage are detailed in Regulation C.

- <sup>6</sup> A similar problem is created for the assessment of sales to the secondary market. A lender reports a loan as sold into the secondary market if it was sold in the same year that it was originated. However, if there is a lag between the origination and sale dates, this can mean that loans originated at the end of the year will appear to be less likely to be sold than those originated at the beginning of the year.
- <sup>7</sup> Only 22,000 of the 30,150,000 HMDA applications in 2005 were affected by the transition rule. For 2004, over 2,000,000 of the 26,100,000 loan applications were covered by the rule.
- <sup>8</sup> To deal with all of these issues, a file was created that reflects our best guess as to lender affiliations based on information in the Federal Reserve's NIC System and other sources. The lender file can be obtained by emailing Robert Avery at: [ravery@frb.gov](mailto:ravery@frb.gov). NIC information can be obtained at: [www.ffiec.gov/nicpubweb/nicweb/nichome.aspx](http://www.ffiec.gov/nicpubweb/nicweb/nichome.aspx).
- <sup>9</sup> Inter-temporal analyses also need to be careful using census tract data drawn from different years. Between the 2002 and 2003 HMDA reporting years, census tract definitions were shifted from the 1990 to 2000 Census. Boundaries were changed for about one-third of the tracts and even when boundaries did not change, such variables as relative median family income and racial composition changed. It is possible that a lender could look like they were significantly increasing their loans to high minority or lower-income areas from 2002 to 2003 simply due to tract definition changes even when there was no actual change in the lender's behavior.
- <sup>10</sup> In the event a borrower chooses not to provide race or ethnicity information, the lender is required to make a selection based on visual inspection. As noted below, lenders were not required to note race or ethnicity in mail or telephone applications prior to 2003.
- <sup>11</sup> This made it particularly difficult to analyze data covering Puerto Rico where all applicants were classified as Hispanic.
- <sup>12</sup> As discussed previously, the rule change actually took place for applications taken beginning January 1, 2004. Thus, all applications subject to the transitions in 2004 reflect racial information gathered under the old rules. Lenders were instructed to convert the pre-2004 answers to the 2004 reporting form by classifying all Hispanics as "Hispanic ethnicity, race N/A." Asian or Pacific Islanders were classified as "Asian race, ethnicity N/A." Whites, blacks, American Indians, and Alaska Natives were given the same racial classification under the new rules with ethnicity marked N/A. Those marked "Other" were treated as both race and ethnicity N/A. In addition, changes in the collection of race and ethnicity information in the 2000 decennial census affects classification of census tracts by minority population. This complicates comparisons across time, as prior to the 2003 HMDA data the census classifications from the 1990 census were available.
- <sup>13</sup> Lien status and manufactured home status have only been available since 2004. Prior to 2004, these variables have to be estimated. Typically, loans below a certain size cutoff (say \$50,000) have been assumed to be junior liens, particularly for home improvement and refinance loans. Since 1993, HUD has provided an annual list of subprime and manufactured home specialist lenders that can be used to identify, although imperfectly, such loans. Typically it has been assumed that all loans originated by the manufactured home loan specialist lenders are manufactured home loans and that no other lenders originate such loans. Analysis of the 2004 data, where individual loan status was known, however, shows this to be a dubious assumption. The fifteen manufactured home specialists on the 2003 HUD list who also filed HMDA data for 2004 accounted for only a little more than 15% of the reported manufactured home loans for 2004. The 2004 data showed an additional 35 lenders where more than 80% of their loans were

manufactured home loans but they were not on the 2003 HUD list, suggesting that the list is incomplete.

- <sup>14</sup> Lenders are not required to report owner-occupancy for loans outside MSAs or in MSAs where they did not have offices. However owner-occupancy is missing for only 30,000 applications in the 2005 HMDA data, indicating such information is typically reported even if not required.
- <sup>15</sup> Survey data show that about 45% of refinance loans are “cash out.” That is, the borrower increases their loan balance above that needed to pay off the previous loan to take out monies for some other purpose. The survey data also indicated that about one-half of all cash out loans involve some form of home improvement (Canner, Dynan, and Passmore, 2002).
- <sup>16</sup> As discussed earlier, the definitions of home improvement and refinance changed between the 2003 and 2004 reporting years. Reporting numbers suggest that on net, these changes may have shifted some loans from refinance to home improvement. The number of home improvement applications increased by almost 700,000 from 2003 to 2004 (a 45% increase), while the number of refinance applications fell by 8,500,000 (a 35% decrease). Of course, the vast majority of the decline in reported refinance activity undoubtedly owes to an increase in interest rates from 2003 to 2004 diminishing the potential gains from refinancing. Nonetheless, some of the decline may be due to the revisions in definitions.
- <sup>17</sup> In applying this rule, one should be aware that for 2004 reporting year loans covered by the transition rule, lenders were instructed to mark the ethnic code as N/A for all non-Hispanic responses and the racial code as N/A for all Hispanic and “Other race” responses. Thus, for transition period applications, only the N/A designation in the gender field should be used to identify business-related loans.
- <sup>18</sup> Lenders are not required to report requests for pre-approval that are approved but not acted upon by the borrower. Because of this omission, it may not be possible to fully reflect a lender’s approval process. Another problem is the absence of geographic information for all 2004 and most 2005 denied requests for pre-approval. This means that the pre-approval process cannot be examined on a market-by-market basis.
- <sup>19</sup> Prior to 2004, researchers interested in studying the higher-priced lending segment of the market would rely on a list published by the HUD to identify HMDA reporters that specialized in subprime lending. Users of the data often assumed that all of the conventional loans made by these institutions were subprime, and that all of the conventional loans made by other institutions were prime loans.
- <sup>20</sup> Data are from Freddie Mac’s Primary Mortgage Market Survey (PMMS). The effective rate (or APR) was calculated on the basis of interest rates and points reported in the survey for conventional first lien fixed-rate thirty-year prime loans. Since April 1971, Freddie Mac has surveyed lenders weekly to determine the average thirty-year fixed rate offered to prime consumers during the Tuesday of the surveyed week. Currently, 125 lenders are surveyed each week, and the mix of lender types—thrifts, commercial banks, and mortgage lending companies—is roughly proportional to the level of mortgage business that each type commands nationwide. Over time, the PMMS has expanded to include other types of loans. For more information, refer to: [www.freddiemac.com/pmms/pmms\\_archives.html](http://www.freddiemac.com/pmms/pmms_archives.html).
- <sup>21</sup> Because terms vary so much for junior lien loans, this exercise was conducted for first lien loans only.

- <sup>22</sup> Data from LoanPerformance suggest that about 90% of the first lien loans extended in 2004 and 2005 had a term of thirty years. For more information see: [www.loanperformance.com](http://www.loanperformance.com).
- <sup>23</sup> Within the HMDA data for 2004, the median time between the date of application and the date of loan origination for conventional first lien home purchase loans was thirty days; for 2005, the comparable figure was twenty-eight days. For refinancings, the median numbers of days for 2004 and 2005 were twenty-seven and twenty-six respectively. Less than 10% of home purchase loans had a difference in dates of application and origination of more than ninety days. For refinancings, less than 10% had differences in dates of application and origination of more than sixty days.
- <sup>24</sup> Under Regulation Z, borrowers are provided a variety of disclosures explaining the possibility of a rise in loan rates, the possible size of the increase, and the circumstances under which an increase might occur.
- <sup>25</sup> The flattening of the yield curve over 2004–2005 actually had a third effect: It also caused a general increase in the interest rates on adjustable-rate mortgages. This rise in real rates for adjustable-rate loans may have affected borrower behavior.
- <sup>26</sup> The Freddie Mac series for five-year adjustable rates did not begin until January 1, 2005. For 2004, estimates for the five-year Treasury securities are based on a statistical model using the one-year adjustable rates and thirty-year fixed rates reported in Freddie Mac's Primary Mortgage Market Survey and the one- and five-year rates for Treasury securities.
- <sup>27</sup> Information collected includes the contract interest rate, fees, loan terms (e.g., the LTV ratio and the term to maturity), property value, property type (newly constructed or previously occupied unit), loan type (fixed or adjustable rate), and type of lender (savings association, mortgage company, commercial bank, or savings bank). The data also include an estimated effective interest rate. For adjustable-rate loans, the survey includes information on the annual limit (the "cap") on how much the interest rate may increase, the margin, and the index used to set the contract interest rate. The survey excludes FHA-insured and VA-guaranteed loans, multifamily loans, and mobile home loans and is limited to home purchase loans. Refer to Federal Housing Finance Board: [www.fhfb.gov](http://www.fhfb.gov). The data in the survey reflect the shares of lending by lender size and lender type as reported in the HMDA data. Although the scope of the survey varies from month to month, it typically covers about 20,000 loans and about 100 lenders.
- <sup>28</sup> The sample excludes all loans with application dates prior to January 2004 and home improvement loans. The analysis is further restricted to loans in an MSA in the contiguous 48 states and the District of Columbia and loans with complete census tract information. The working data set includes about 9 million loans.

---

## References

- Avery, R.B., K.P. Brevoort, and G.B. Canner, Higher-Priced Home Lending and the 2005 HMDA Data, *Federal Reserve Bulletin*, 2006, 92, A123–A166.
- Avery, R.B., G.B. Canner, and R.E. Cook, New Information Reported under HMDA and Its Application in Fair Lending Enforcement, *Federal Reserve Bulletin*, 2005, 91, 344–394.
- Berkovec, Jim and Peter Zorn. 1996. How Complete is HMDA? HMDA Coverage of Freddie Mac Purchases. *Journal of Real Estate Research*, 11:1, 39–56.
- Canner, G., K. Dynan, and W. Passmore, Mortgage Refinancing in 2001 and Early 2002, *Federal Reserve Bulletin*, vol. 88, (December), 2002, 469–481.

Goering, J. and R. Wienk, eds., *Mortgage Lending, Racial Discrimination, and Federal Policy*, Washington, Urban Institute Press, 1996.

*The views expressed are those of the authors, and do not necessarily represent those of the Board of Governors of the Federal Reserve System or members of its staff. Sections V and VI are drawn substantially from the authors' earlier work published in the Summer 2006 issue of the Federal Reserve Bulletin.*

---

*Robert B. Avery, Board of Governors of the Federal Reserve System, Washington, D.C. 20551-0001 or [ravery@frb.gov](mailto:ravery@frb.gov).*

*Kenneth P. Brevoort, Board of Governors of the Federal Reserve System, Washington, D.C. 20551-0001.*

*Glenn B. Canner, Board of Governors of the Federal Reserve System, Washington, D.C. 20551-0001 or [Glenn.B.Canner@frb.gov](mailto:Glenn.B.Canner@frb.gov).*

