



Technical Report Series

National Survey of Mortgage Borrowers

Technical Report 15-02

August 17, 2015

This document was prepared by Robert B. Avery, Mary F. Bilinski, Brian K. Bucks, Tim Critchfield, Ian H. Keith, Ismail E. Mohamed, Forrest W. Pafenberg, Jay D. Schultz, and Claudia E. Wood. The analysis and conclusions are those of the authors and do not necessarily represent the views of the Consumer Financial Protection Bureau, the Federal Housing Finance Agency or the United States.

1.0 Introduction

The National Mortgage Database project is a multi-year project being jointly undertaken by the Federal Housing Finance Agency (FHFA) and the Consumer Financial Protection Bureau (CFPB). The project is designed to provide comprehensive information about the U.S. mortgage market based on a five percent sample of residential mortgages. It has two primary components: (1) the National Mortgage Database (NMDB) and (2) the quarterly National Survey of Mortgage Borrowers (NSMB).

The NMDB project will enable FHFA to meet the statutory requirements of section 1324(c) of the Federal Housing Enterprises Financial Safety and Soundness Act of 1992, as amended by the Housing and Economic Recovery Act of 2008, to conduct a monthly mortgage market survey. Specifically, FHFA must, through a survey of the mortgage market, collect data on the characteristics of individual mortgages, including those eligible for purchase by Fannie Mae and Freddie Mac and those that are not, and including subprime and nontraditional mortgages, and information on the creditworthiness of borrowers, including a determination of whether subprime and nontraditional borrowers would have qualified for prime lending.¹

For CFPB, the NMDB project will support policymaking and research efforts and help identify and understand emerging mortgage and housing market trends. The CFPB expects to use the NMDB, among other purposes, in support of the market monitoring called for by the Dodd-Frank Wall Street Reform and Consumer Protection Act, including understanding how mortgage debt affects consumers.

FHFA and CFPB considered existing databases but determined that none sufficiently support the above objectives.² The NMDB, when fully complete, will be a de-identified loan-level database of closed-end first-lien residential mortgages. It will: (1) be representative of the market as a whole; (2) contain comprehensive information on the terms and performance of mortgages, as well as characteristics of the associated borrowers and properties; (3) be continually updated; (4) have an historical component dating back before the financial crisis of 2008; and (5) provide a sampling frame for the NSMB.

The core data in the NMDB are drawn from a random 1-in-20 sample of all closed-end first-lien mortgage files outstanding at any time between January 1998 and June 2012 in the files of Experian, one of the three national credit repositories.³ The use of a sampling frame substantially reduces the privacy risk associated with any data collection. By contrast, a universal registry can present challenges for privacy since it is known that a particular loan must be in the dataset. However, for a 1-in-20 sample, the odds are 95 out of 100 that a particular loan is not in the

¹ FHFA interprets the NMDB project as a whole, including the NSMB, as the “survey” required by the Safety and Soundness Act. The statutory requirement is for a monthly survey. Other core inputs to the NMDB, such as a regular refresh of credit-bureau data, occur monthly, but not the NSMB.

² For a fuller description of the NMDB, including a discussion of existing sources and their limitations, see NMDB Technical Report 15-01.

³ Experian was chosen through a competitive procurement process to assist in creating the NMDB.

database. In addition, the sample used is large enough to support almost all types of statistically valid analyses but small enough to manage logistically, thus dramatically reducing both contract and personnel costs.

A random 1-in-20 sample of mortgages newly reported to Experian is added to the NMDB each quarter. Mortgages are followed in the NMDB database until they terminate through prepayment (including refinancing), foreclosure, or maturity. Information from credit repository files on each borrower associated with the mortgages in the NMDB sample is collected from at least one year prior to origination to one year after termination of the mortgage. The information on borrowers and loans available to the FHFA, CFPB, or any other authorized user of the NMDB data is de-identified and does not include any direct identifying information such as borrower name, address, or Social Security number.

The NSMB is a component of the NMDB project and is designed to provide policy makers, researchers and others with comprehensive de-identified data for analyzing housing and mortgage-related public policy and for improving lending practices and the mortgage process. The survey, conducted by mail, is designed to complement the NMDB by providing information, particularly related to mortgage shopping, that is not available in the database. The survey is completely voluntary and its target universe is newly originated closed-end first-lien residential mortgages and their associated borrowers. To achieve this objective, the NSMB draws its sample from mortgages that are part of the NMDB which draws its sample from the same target universe of new loans.

Beginning with loans originated in 2013, a simple random sample of about 6,000 loans per quarter is drawn from loans newly added to the NMDB for the NSMB. At present, this represents a sampling rate of 1-in-13 from the NMDB or 1-in-260 from the population given that the NMDB itself is a 1-in-20 sample of loans. Although information from other sources will ultimately be merged into the NMDB, the data from Experian are sufficient to select the NSMB sample.

This technical report provides background details on how the NSMB was developed. The second section presents a discussion of the development of the survey questionnaire, including the approval granted by the Office of Management and Budget (OMB) as required by the Paperwork Reduction Act. The third section discusses the survey sample frame and timeline, and the fourth section discusses the logistics of conducting the survey.

The fifth section presents an analysis of survey responses for the first four waves. The sixth section presents a discussion of how the useable population for analysis is derived. The seventh section describes the data cleaning, editing, and imputing processes used to refine the useable survey dataset. The eighth section presents a discussion of how sample non-response weights are computed. The ninth section of the document discusses sampling error of the survey.

There are two Appendices to this document: Appendix A presents the survey cover letters and the NSMB questionnaire; and Appendix B presents un-weighted frequency responses for all questions for the first three waves of the survey.

2.0 Development of the Survey

In reaction to the financial crisis of 2008, Freddie Mac developed a pilot version of what has become the NSMB. The pilot was administered as a mail survey to about 1,500 individuals drawn from data maintained by Experian, one of the three national credit repositories. The pilot used a sample frame similar to that currently used by the NSMB. The pilot survey response rate of 12 percent was much lower than hoped.

To improve the response rate, Freddie Mac retained the services of Don A. Dillman, of Washington State University, a leading expert in mail survey methods. Dr. Dillman focused on improving: (1) the contacting strategy; (2) the up-front cash incentives; (3) the communication strategy; and (4) the questionnaire format. His changes were incorporated into a second pilot survey in February 2011 with a sample of 1,000 new Freddie Mac loans split evenly between borrowers who had recently purchased a home and borrowers who had recently refinanced an existing mortgage. This second pilot survey resulted in a vastly improved response rate of 60 percent.

In the fall of 2012, Freddie Mac conducted a third pilot survey targeting a representative national sample of 5,000 new 2011 mortgage borrowers drawn from Experian files. The response rate for this survey was about 45 percent.

The improvements instituted in the later pilots confirmed the effectiveness of using credit repository records as the survey sampling frame as well as the effectiveness of the questionnaire and methodology.

The questionnaire for the NSMB draws heavily on the questionnaires piloted by Freddie Mac and leverages the input of an advisory group of industry experts from government, non-profits, advocates, trade groups, and academia that Freddie Mac convened when creating their questionnaires. This group played a significant role in ensuring that the NSMB provided information of ultimate interest to policy-makers, researchers, and data analysts.

The NSMB focuses on topics such as mortgage shopping behavior, mortgage closing experiences, and information that cannot be obtained from any other source: expectations regarding house price appreciation, critical household financial events, and whether “trigger” events, such as unemployment spells, large medical expenses, or divorce, have occurred. In general, borrowers are not asked to provide mortgage terms in the questionnaire, since these fields are available in the Experian data. However, the survey collects a limited amount of information on the mortgage to compare borrower’s views with those of credit and administrative records and to verify that the credit repository records and survey responses pertain to the same mortgage.

By interagency agreement between FHFA and CFPB, FHFA led the production of the NSMB.⁴ This included seeking public comments concerning information collection as required by the Paperwork Reduction Act. On April 25, 2013, FHFA published in the *Federal Register* a 60-day

⁴ An interagency agreement between FHFA and CFPB was signed on September 12, 2012 where the costs of the survey and the development of the NMDB are to be shared equally between the two agencies.

Notice of Submission of Information Collection for Approval from the OMB. No comments were received for this notice. Subsequently, on July 1, 2013, FHFA published a 30-day Notice of Submission of Information Collection for Approval from OMB indicating that FHFA had received no comments during the 60-day comment period.

Following these *Federal Register* notices, OMB reviewed the FHFA application and approved the request in December 2013, assigning the NSMB a control number of 2590-0012 with an expiration date of December 31, 2016. In April 2014 FHFA published a revised System of Records notification in the *Federal Register* extending the system of records entitled “National Mortgage Database Project” to cover the NSMB.

After obtaining OMB approval, FHFA modified an existing contract with Experian, which subcontracted the survey administration through a competitive process to Westat, a nationally-recognized survey vendor. Fair Credit Reporting Act (FCRA) rules dictate that the survey process, because it utilizes borrower names and addresses drawn from credit repository records, must be administered through Experian in order to maintain consumer privacy.⁵

The NMDB development staff consulted with Experian, Westat, and the Freddie Mac advisory group between December 2013 and February 2014 to finalize the survey questionnaire and supporting materials. The initial survey wave was mailed out in April 2014, with a new wave distributed each quarter since.

3.0 Detailed Survey Sample Frame and Timeline

Following the update of the NMDB at the end of each quarter, FHFA randomly selects 6,000 of the closed-end first-lien mortgage loans newly added to the NMDB for the NSMB.⁶ At present this represents about a 1-in-260 sampling rate from the population of such loans as a whole. Loans are selected at random from mortgages newly-reported to Experian, with the additional conditions that the mortgage be reported to Experian within a year of origination and that the borrowers have not been selected for an earlier NSMB survey.

After the sample is selected, Experian eliminates any potential respondents who have opted out of previous surveys or are deemed to not have legitimate addresses or names. Industry guidance (Metro 2[®] Industry Standards for Credit Reporting) requires that servicers must supply a billing address for each borrower on a trade line (including mortgages). Experian generally uses these borrower billing addresses as the survey mailing address. Sometimes, though, there are multiple addresses and borrowers associated with a survey sample loan. In these cases, Table 1 presents

⁵ The Fair Credit Reporting Act (FCRA), Public Law No. 91-508, was enacted in 1970, and substantially amended since, to promote accuracy, fairness, and the privacy of personal information assembled by credit reporting agencies (CRAs). The Act's primary protection requires that CRAs follow “reasonable procedures” to protect the confidentiality, accuracy, and relevance of credit information. To do so, the FCRA establishes a framework of requirements for credit report information that include rights of data quality (right to access and correct), data security, use limitations, requirements for data destruction, notice, user participation (consent), and accountability.

⁶ For a fuller description of how loans are selected for the NMDB, see NMDB Technical Report 15-01.

the rules for selecting the borrower(s) and address to which to mail the survey. The survey is sent to at most two borrowers who must share a common address.

Table 1
Rule for Best Address

Number of borrowers	Same or different address	Resulting survey recipient
1	n/a	One borrower with Experian’s associated best address
2	Same	Two borrower names with one common best address
2	Different	The one borrower and associated best address with the lowest number of open mortgages.
>2	Same	Two borrowers with one common best address that has the highest number of trade lines reported
>2	Different	The one borrower and associated best address with the lowest number of open mortgages

FHFA and CFPB never receive the names or addresses that are chosen for the survey. Only Experian and Westat as Experian’s subcontractor have access to this information.

4.0 Survey Logistics

The survey implementation strategy comprises four respondent contacts over a seven-week period (copies of the survey questionnaire and contact materials are provided in Appendix A):

Week 1	Printed questionnaire, cover letter, and cash incentive (entire survey sample population)
Week 2	1 st reminder letter (entire survey sample population)
Week 5	2 nd reminder letter, printed questionnaire, and additional cash incentive (sampled borrowers who have not responded by Week 4)
Week 7	3 rd reminder letter, which includes the due date for returning the questionnaire, to close the communication loop (sampled borrowers who have not responded by Week 6)

Participation in the survey is completely voluntary and respondents are assured of confidentiality in their responses. The first and the third contacts contain a printed survey questionnaire and a five dollar cash incentive, which the respondent is free to keep whether they return the questionnaire or not. The mailings and printed questionnaires detail how respondents can also complete the survey online in either English or Spanish (there is no printed Spanish questionnaire) using instructions and a unique “survey PIN number” provided in the questionnaire packet. About one quarter of survey responses are completed online.

Mail surveys are processed for four weeks after the third reminder letter, so the field period comprises 11 weeks in total. It takes between five and six weeks to draw the new NMDB sample, identify and combine duplicative records, draw the NSMB sample, process it at Experian, and print the survey materials. Thus, the survey cycle typically begins six weeks after the end of a quarter and extends about four weeks into the next quarter.

All returned questionnaires and any non-delivered mail are sent directly to Westat and not to FHFA, CFPB, or Experian. All survey responses received by Westat are purged of any information related to the name of the borrower, address of the borrower, or name of any financial institution. This is done to maintain the depersonalized confidential nature of the data and to ensure that the survey responses cannot be connected to a name or address.

During the first eight weeks of each cycle, Experian maintains a NSMB call center to address any questions by respondents. This call center also allows respondents to “opt out” of the survey and future surveys. Both FHFA and CFPB describe the survey on their websites so that respondents can independently validate the legitimacy of the survey.⁷ The agency officials signing the cover letter (Sandra Thompson at FHFA and David Silberman at CFPB) are identifiable on the websites as senior employees of the agencies.

Once the active phase of a survey cycle ends, it takes about 25 days for Westat to scan and edit returned questionnaires, combine them with on-line responses and create an electronic data file. This file is delivered to the NMDB development staff, through Experian. It takes a further eight weeks to complete additional cleaning and editing of survey responses, to create preliminary sample weights, and to assemble a preliminary user data file.

Since it takes between 90 and 150 days for the typical mortgage loan to be reported by the servicer to the credit repositories after origination, the first preliminary user data file will generally reflect mortgage originations of approximately one year earlier. Consider the fourth wave of 2014 as an example. The survey sample is drawn from the September 2014 archive and captures loans reported to Experian between June and September 2014, with most originated between March and June 2014. The fourth wave was put into the field in early November and closed at the beginning of February 2015. The electronic data file was delivered to the NMDB development staff in late February, and it took until the end of April 2015 to create a preliminary version of the survey data base.

The timeline just described applies to each quarterly wave data release. Because some loans can take longer than six months to be reported to the repositories, a usable data file fully representative of a calendar year will not generally be available until December of the following year.

5.0 Survey Response Analysis

In a typical cycle, the NSMB design calls for a sample of 6,000 cases each quarter as described in the previous section. However, in 2014, the first year of the survey, FHFA conducted modified

⁷ www.fhfa.gov/Homeownersbuyer/Pages/National-Survey-of-Mortgage-Borrowers.aspx and www.consumerfinance.gov/National-Survey-of-Mortgage-Borrowers

versions of the first three waves in April, June, and September. Wave 1 (April) included a sample of 15,000 mortgages. This was a catch up period to cover cases originated in 2013 and newly-reported to Experian in the archives for June, September and December 2013. For this first wave, 1.5 percent or 218 survey invitations were not delivered, resulting in a net delivered population of 14,782 (see Table 2). The survey was in the field for 11 weeks and yielded 5,793 completed surveys, with 173 borrowers opting out of the survey and the remaining 8,816 not returning a questionnaire. If the undelivered survey invitations are treated as ineligible, this represented a 39.2 percent response rate (5,793/14,782).

Wave 2 (June) included 3,000 surveys and was for mortgages that were originated in 2013 and newly-reported to Experian between January and March 2014. The postal non-delivery rate for this wave was somewhat lower than for the first wave at 1.2 percent. The questionnaire for this survey was the same as that used for the first wave except that, as described in the next section, a critical clarification was added to the initial survey filter question. The overall response rate for Wave 2 was 36.3 percent, resulting in 1,076 completed questionnaires. There were 31 borrowers who opted out of the survey.

For Wave 3 (September), Westat mailed out 6,000 surveys representing mortgages that were originated in 2013 and reported to Experian between March and June 2014 within a year of origination as well as any mortgages originated in 2014 and reported to Experian between January and June 2014. The postal non-delivery rate for this third wave was somewhat higher than Waves 1 and 2 at 1.8 percent, or 110 sample cases. The overall response rate for the third wave was 35.2 percent, resulting in 2,073 completed questionnaires. There were 42 borrowers who opted out of the survey.

A fourth wave was mailed in November 2014 and most closely represents the steady state for future surveys in that the sampling frame. It was comprised of any mortgage newly reported to Experian in the quarter just ended (July to September 2014) that was reported within a year of origination. It also represented the initial wave where Experian eliminated potential sample cases deemed to not have legitimate addresses or names prior to mailing. This resulted in a sample of 5,795 cases. Other than slight changes to two questions, the questionnaire was unchanged from prior waves. The response rate for this wave was similar to that of Waves 2 and 3. There was a fairly constant low percentage of the sample that was non-deliverable or elected to opt out of the survey. This confirms that Experian’s methodology for choosing the best mailing address has been working well.

**Table 2
Survey Return Analysis**

	Wave 1	Wave 2	Wave 3	Wave 4
Estimated Newly Reported Mortgage	6,963,150	888,420	1,685,760	1,527,736
Sample Weight Unadjusted for Sample Nonresponse	464.21	296.14	280.96	263.63
Sent	15,000	3,000	6,000	5,795

Postal Non-Delivery	218	37	110	86
Postal Non-Delivery - %	1.5 %	1.2 %	1.8 %	1.5 %
Net Delivered	14,782	2,963	5,890	5,709
Completed Surveys				
Mail	4,410	858	1,534	1,496
Online English	1,360	214	524	514
Online Spanish	23	4	15	10
Total Completed - #	5,793	1,076	2,073	2,020
Total Completed - %	39.2 %	36.3 %	35.2 %	35.4 %
Total Opt Out - #	173	31	42	54
Total Opt Out - %	1.2 %	1.0 %	0.7 %	0.9 %

6.0 Usable Population for Analysis

For each quarterly survey, all returned questionnaires and on-line responses were evaluated to determine the usable population for analysis. Table 3 below summarizes the results of this analysis. Based on this review, four criteria for rejecting a completed questionnaire for analysis were established.

The first criterion is a “no” response to the first question (Q1). Q1 is used as a screener question to confirm that the survey respondent took out a mortgage during the reporting period (which Experian records suggest that they did). In the first wave, a surprisingly high number of 764 respondents said that they had not taken out a mortgage. An analysis of the records suggests that some respondents who had refinanced their mortgage were not treating this as a new mortgage. Consequently in Wave 2, the wording of Q1 was changed to add the phrase “including any mortgage refinances.” With this change, the share of negative responses to Q1 decreased dramatically from 13 percent to 8 percent.

The next exclusion criterion was for respondents who broke off in the middle of the survey and only answered part of the questionnaire (breakoffs were defined as those that did not provide a response to almost all questions from question Q50 on). The third criterion for exclusion was for respondents who provided information on the wrong loan. The sampling frame was tied to a particular loan associated with the borrower. However, the questionnaire did not refer explicitly to that loan. Instead, respondents who had taken out multiple loans during the reference period were asked to report on the “most recent.” In some instances this was not the sample loan. This was a particular problem in Wave 1 which, as a “catch up” survey, had a relatively long reference period. Also, some respondents who had refinanced their mortgage reported on the original home purchase mortgage rather than the refinance. Finally, in a few instances it appears that the survey went to the wrong person, with answers bearing no resemblance to the sample loan features as

characterized by Experian records. In each of these circumstances the survey response was removed from the data set used for analysis.

The last category of unusable surveys comes from respondents whose sample loans were ultimately removed from the NMDB after the survey had been executed either because they were deemed to have duplicate trade lines and to not meet the criteria for remaining in the NMDB or where the sample loan was determined to be a second and not a first mortgage lien. In some instances the survey response itself led to the removal, as margin notes or comments indicated that the loan was a second lien.

Given this, the rate of usable responses in each wave is lower than the survey response rates reported earlier. Overall, 9,297 usable responses were obtained from 29,767 sample cases (for a rate of 31.3%) or from 29,344 survey invitations delivered (for a rate of 31.7%).

Table 3
NSMB Useable Population

Wave	Mailed	Not Returned	Answered No to Q1	Did Not Finish Survey	Wrong Loan	Duplicate or HELOC	Useable	2013 Loan	2014 Loan
1	15,000	9,207	764	61	218	35	4,715	4,715	0
2	3,000	1,924	88	4	40	7	937	937	0
3	6,000	3,919	117	17	67	13	1,867	486	1,381
4	5,795	3,775	176	21	1	64	1,778	14	1,764
Total	29,767	18,825	1,145	103	326	91	9,297	6,152	3,145

7.0 Cleaning, Editing, and Imputing Responses

The survey responses, once delivered to the NMDB development staff, were subjected to a thorough editing and cleaning process. The initial phase consisted of standard editing—correcting numbers reported in the wrong units, changing answers in responses based on margin notes and comments, assigning responses for questions with open-ended “other” responses, dealing with multiple responses to a question that calls for only one response and deciding how to handle situations where respondents followed the wrong skip pattern.

In some instances, examination of responses suggested questions that respondents may have frequently misunderstood or misinterpreted. Three questions were judged to be particularly problematic:

1. Question Q64 (how many separate units does your mortgage cover?): Inconsistencies between the self-reported loan amount and the amount reported in the credit repository

data suggested that the number of units that a mortgage covered in a property was sometimes answered incorrectly.

2. Q75 (owned other residential properties besides this one?): In many instances, credit repository data indicated that the borrower had previous mortgages contrary to the response to this question.
3. Q16 (a term of less than 30 years?): The term of the loan reported by the lender in many cases did not match responses to Q16.

Finally, there were also indications that respondents with sample loans on investment properties may have provided information on their primary residence property and neighborhood rather than that of their investment property. These problems are addressed in changes made to the questionnaire for Wave 7 based on the June 2015 archive. However, users should be aware of these interpretation inconsistencies when using data from the earlier waves.

One advantage that the NSMB has over other surveys is the availability of credit and administrative data, much of which appears to be quite reliable. These data can be used to assist in the editing and imputation process. Three primary sources of such data were available in processing the first four waves of the NSMB: (1) credit data from Experian on sample loans; (2) data collected by Experian from other data sources on the survey respondents, including loan servicers and data companies; and (3) information for loans that could be matched to Home Mortgage Disclosure Act (HMDA) files (only HMDA data through calendar year 2013 are available as of this writing).⁸ Ultimately, additional information from further administrative and property file matches will be available for this purpose but is not available at this time.

The credit and administrative data were used to determine which borrower in the Experian data corresponded to the respondent (and spouse/partner of the respondent) in the survey and to determine the loan they were reporting on. The data were also useful in determining if respondents correctly identified their loan as a home purchase loan or a refinance.

Tabulations of the raw un-weighted—but edited—responses to all the questions in the survey are presented in Appendix B. Data are presented only for usable observations in Waves 1, 2 and 3. Although Wave 4 has undergone substantial cleaning, information from HMDA matches are not yet available for 2014 originations in that wave (which is dominated by 2014 mortgages) to complete the process.

After editing and cleaning the survey response data, NMDB staff imputed missing responses using statistical models estimated based on credit and administrative data and answers to other questions in the survey. In order to preserve the original responses, the raw responses were

⁸ Merges between the NMDB or NSMB and HMDA rely on variables common to both datasets, including the original loan balance, the opening date of the mortgage and the general location of the property (census tract or state/county). Unfortunately, mortgage servicers report the billing address of the mortgage borrowers to Experian, but this is not necessarily the property address, particularly for mortgages on non-owner occupied properties. Additional address information maintained within Experian's databases is useful in supplementing the repository addresses, as is historical information on borrower location. Nevertheless, HMDA merges are less accurate than those employing directly identifying information such as name and Social Security number because the latter are less reliant on address.

retained (“Q” variables) with missing responses coded as such. A parallel set of variables (“X” variables) were constructed where all missing responses were imputed. Each instance in which an X variable differs from its comparable Q variable is recorded by a shadow variable (“J” variables) that indicate the method and reason whereby the change was made. Missing responses typically totaled about 3 to 5 percent for most questions and only in a few instances were more than 10 percent. The X variables were not created when a directly comparable credit or administrative variable was available for all respondents (*e.g.*, loan amount, loan payment, number of co-signers) as comparable credit or administrative variables could be used in lieu of survey responses in analysis.

Key demographic variables (age, gender, education, ethnicity, and income) were imputed first. For these variables, high quality administrative data were generally available and could be used directly to impute a value for the X variable. For example, lender reports provided high quality data on age and HMDA data, which were available for loans originated in 2013, provided high quality information on race, income, and gender.

For most variables, though, comparable relevant credit or administrative information was not available. Missing values for these variables were imputed statistically using an iterative process. Individual statistical models were developed for each question that used the key demographic variables as well as credit or administrative data such as loan amount and credit score as regressors in linear probability, logistic, or cell-based models (since almost all variables in the survey are categorical). In all instances the imputation incorporates a random component that reflects the accuracy of the imputation model. Variables were imputed in order, with higher-order variables that dictated a skip-pattern imputed first, before the variables conditioned on the pattern were imputed. Once the first round of imputations was completed, the process was repeated with expanded predictive linear or logistic models that incorporated some of the newly imputed variables as regressors for other variables. This iteration ensures that correlations among the imputed values will better reflect correlations among observations where responses were available.

8.0 Sample Non-Response Weights

There are several ways calculations based on the NSMB raw survey responses may not be representative of the population as a whole. First, as shown in the Table 2, the four survey waves did not have the same sampling rates. Second, only about one-third of the solicited borrowers returned a usable survey. Commonly, in survey sampling, some individuals chosen for the sample are unwilling or unable to participate in the survey. Non-response bias is the bias that results when respondents differ in meaningful ways from non-respondents. However, non-response is only a problem if the non-respondents are a non-random sample of the total sample. When non-response bias is present, rather than accept a poor match between the sample and the population, it is now common to use weights to bring the two more closely into line. This is known as “non-response weighting.” Such weights are generally calculated from statistical models.

Often, little is known about survey non-responders, thus the statistical models used to construct non-response weights are quite simplistic. Compared with many other surveys, however, the

NSMB has extensive credit and administrative data on both responding and non-responding borrowers that can be used to estimate non-response weights.

Sample non-response weights were estimated separately for each sample wave and within a wave for loans with a single borrower versus those with multiple borrowers with logistic models. The models estimate the probability of getting a usable response for each wave of the survey. The predictive equations had pseudo-R-square values ranging from .0379 to .0651. Key predictive variables included: loan amount, borrower age, the median income of borrowers census tract of the sample loan as captured in the 2013 HMDA data, whether or not the loan could be matched to HMDA (an indicator of investor status), and if so, whether it was a home purchase or refinance loan, whether a borrower kept a loan at the same time the sample loan was taken out (an indicator of multiple loans), and a measure of the number of days from loan origination to sending out the survey. The models also controlled for credit score, for geography using Census Divisions, and for demographic characteristics using Experian’s marketing-type variables on family composition, race, ethnicity, gender, and educational attainment.

The model’s predicted probabilities of response were grouped into quintiles. The average of the response rates from each of these five groups was used to calculate a response weight as the inverse of these five rates. Once within-wave sample non-response weights were estimated, they were multiplied by the wave sample weight to provide an overall weight.

Table 4 demonstrates the impact of differential sampling weights for the first three waves. Column one shows the distribution among various demographic and loan categories of the raw survey responses. Column two provides the distribution using estimated overall weights. Finally, column three shows the average overall weight for each category.

	Unweighted Percentage	Weighted Percentage	Average Weight
Loan Category			
Homeowner, Purchase, First-time Home Buyer	16.6	18.2	1382
Homeowner, Purchase, not First-time Home Buyer	20.8	18.8	1140
Homeowner, Refinance	53.2	54.4	1291
Investor, Purchase	4.6	4.0	1088
Investor, Refinance	4.8	4.6	1211
	100%	100%	
Loan Size			
\$50,000 or Less	4.0	3.7	1164
\$50,001 to \$150,000	40.2	39.9	1253
\$150,001 to \$300,000	38.3	38.5	1269
More than \$300,000	17.6	18.0	1287
	100%	100%	
Respondent Credit Score			

Less than 541	0.2	0.3	1844
541 – 680	15.9	19.3	1537
681 – 720	13.4	14.9	1407
More than 720	70.5	65.4	1170
	100%	100%	
Respondent Age			
Less than 35 years	16.2	19.9	1555
35 <= Age <= 50	34.3	37.3	1372
51 <= Age <= 65	35.5	31.8	1131
Older than 65	14.0	11.0	984
	100%	100%	
Respondent Race/Ethnicity			
White, non-Hispanic	80.1	77.8	1225
Other	19.9	22.2	1407
	100%	100%	
Respondent Education			
High School or less	12.3	12.4	1275
Some College	24.2	23.9	1247
College Degree	34.3	35.2	1296
Postgraduate	29.2	28.5	1228
	100%	100%	
Respondent(s) Income			
Less than \$50,000	17.8	17.7	1255
\$50,000 - \$99,999	38.6	38.5	1258
\$100,000 - \$174,999	28.0	28.3	1272
\$175,000 or More	15.6	15.6	1259
	100%	100%	
Household Type			
Single, no children	23.4	23.7	1273
Single with children	5.4	5.8	1357
Couple, no children	43.4	40.4	1175
Couple with children	27.8	30.2	1368
	100%	100%	
Property type			
Single-family detached house	83.4	83.6	1264
Townhouse, row house, or villa	6.0	6.0	1262
Mobile home or manufactured home	1.8	1.7	1225
2-unit, 3-unit, or 4-unit dwelling	2.3	2.2	1248
Condo, apartment house, or co-op	6.1	6.1	1242
Other	0.4	0.3	1184

	100%	100%	
Mortgage Term to Maturity			
Less than 15 years	5.6	5.1	1147
15 years	19.9	18.9	1200
Between 15 and 30 years	6.5	6.9	1326
30 years or more	68.0	69.1	1283
	100%	100%	

9.0 Sampling Error

Errors may be introduced into survey results at many stages. Sampling error—the variability expected in estimates based on a sample instead of a census -- is a particularly important source of error. For the NSMB two sources of such error are present -- the NMDB is itself a sample of loans in the Experian files and the NSMB is a sample from the NMDB.

Other errors occur because borrowers who respond to the survey are not random and those who chose not to respond to a particular question are also not random. Imputation and sample non-response weights correct for some of this error but not all. Other errors occur when respondents interpret a question differently from that intended by the survey or other respondents. As noted above, for some questions this problem was serious enough to call into question the use of the variable.

Analysis of these data with software that assumes the data are from a simple random sample will under-estimate the standard errors (statistical precision) of the estimates. Users are encouraged to use analytic procedures (so-called “survey” procedures in most major statistical analysis packages) that take into account the effect of the differential sampling and non-response adjustment weights on the estimates.

APPENDIX A

Questionnaire and Cover Letters

(In separate pdf attachment)

APPENDIX B

Survey Frequency Response Un-Weighted

(In separate pdf attachment)