Toward a More Useful HMDA

August 24, 2010
Ford Foundation
New York, NY

Working Session Proceedings

On August 24, 2010, in response to the Federal Reserve Board of Governors' announcement of a public comment period on Regulation C, which implements HMDA, the Ford Foundation hosted a researchers convening. This half-day working session brought together a mix of users of Home Mortgage Disclosure Act (HMDA) data, including researchers and representatives of the financial, mortgage and data management sectors. The proceedings of this meeting follow, and are respectfully submitted to the Federal Reserve Board under its call for public comment.

This document is intended to serve as a record of what was discussed, and not a consensus of opinion or a set of recommendations.

Introduction

In the context of these developments, the working session's agenda was designed to encourage the exchange of ideas in response to three basic, and related, questions:

1) How could HMDA better serve public purposes?
2) What are the costs and benefits of an expanded HMDA data set?
3) What lessons or models from other data sets could be applied to HMDA, particularly with regard to protecting privacy?

To address these questions, the day's agenda was organized into five parts:

I.	Opening Discussion: Potential Benefits of an Improved HMDA
II.	Recent Developments in HMDA and Other Mortgage Data Initiatives
III.	The Bank Industry Perspective
IV.	Privacy Concerns
V.	Concluding Comments

The structure of this document reflects the agenda.

This document is intended to serve as a record of what was discussed, and not a consensus of opinion or a set of recommendations.

1

I. Opening Discussion: Potential Benefits of an Improved HMDA

Participants shared their thoughts on what an expanded HMDA data set would look like and what additional purposes it would serve.

Critical to this exercise was participants' thinking about HMDA's potential: inclusive of, and more broadly, than its current statutory purposes. These statutory purposes are: 1) to help determine how lenders are meeting the housing needs of their communities; 2) to identify where additional public investment is necessary in order to attract private capital; and 3) to detect discrimination in the housing lending market.

Related or upgraded purposes identified included: to find out who is providing credit, where, and on what terms; to examine assumptions (e.g. about the relationship between prepayment penalties and interest rates); and to figure out how borrowers channel themselves or are sorted into different mortgage origination processes.

Some possible expanded applications identified included metrics on loan performance, geographic foreclosure patterns, and loan modifications. Discussants noted the important element of safely and soundness in Community Reinvestment Act (CRA) lending, and that HMDA can be integral in helping identify risks. First though, it is critical to know what risks the financial system needs protecting against: the subprime crisis revealed the dangers of relying on private industry for public policy.

This range of applications highlighted several key themes, identified below. Again, the points made below are a composite of all the comments made, representing differing (and often contrasting) opinions.

   i. **Improving HMDA's usefulness as a means of identifying risk and responding to crisis**
   Mortgage data reporting is relevant to the extent that risk in the mortgage market accumulates and threatens the stability of the financial system, as shown with the subprime crisis. For this reason, HMDA can and should be used as a proactive way of looking at the mortgage market to identify the industry sectors and/or actors that pose systemic risk and identifying where regulation is necessary. Related to this is the need for information that can be used to identify and respond quickly to crises.

      a. *Extending coverage to more types of loan and credit products:* Loan types and products named by group members that could be included in HMDA reporting, were home equity lines of credit and loans, unsecured manufactured home loans, reverse mortgages, and purchased loans. Auto loans will be a gateway to credit for many people and should maybe be tracked. There should be better data on

multi-family housing loans, as this is an indication of how an institution is meeting community needs. Reporting should cover what products are in use in mortgage market, but the tension is that new products and services are always being introduced; this would make it difficult for HMDA to keep up with the "next big thing".

b. *Performance and Outcomes:* In the absence of any national database on the status of individual properties, there is a clear need for more data on all activity related to default and foreclosure, including property status and loan modifications.

c. *Product Features and Better Pricing Data:* Related to the need for reporting on loan performance is the need for data fields that describe loan terms and pricing. As is, HMDA limits the ability to truly understand or describe adjustable rate mortgages (ARMs); information on fully adjusted rate, margins, indices, and reset frequencies should be included. Data on pricing and terms could also be used as a way to test assumptions about how certain terms relate to outcomes and disparities in performance across different products and services.

d. *Improved timeliness and/or frequency of reporting:* As it is, HMDA data is submitted on March 1 but is generally not released until September; in the 1990s, it was released in late July. Making it available earlier in the year would be a great step toward identifying and addressing systemic risk in a more timely and meaningful way. However, the Federal Reserve does do a lot of data correction, which is a benefit that might be forfeited by calling for quicker turn-around. Regulators do have access to the data in the quarter during in which it is reported, and it is available by request to the public within 30 days.

ii. **Allowing for traceability of loans and property status to understand outcomes for borrowers and properties**
Having a way to track loans throughout their lifecycle is critical for understanding outcomes, as is the status of and claims on properties to understanding what is happening on the ground in neighborhoods and cities. This has implications for how problems related to foreclosure are worked out—which is at the local level.

a. *Universal identifiers for properties:* A universal code (similar to a UPC) for properties could be linked to every entity that has a relationship to the property (for example the first several digits could identify the property, then, originator, lender, trust, etc.); this would be particularly useful if there are multiple mortgages on the property, and could also help track loan transitions (including

3

flipping). Another justification for this is that it would make it possible to identify the relationship between first and second liens; this would be valuable for investors, researchers, and policy makers. Currently, first lien holders are not required to notify second lien holders of the existence of the first lien. Without a geographic ID, there is no way of linking multiple liens to the same property. Note that privacy is a huge concern once you try to link loan-level data to granular geographic info. The keys to the system could be kept behind a government firewall.

b. *Universal identifiers for loans:* A universal loan identifier would allow for identification of which loan was going into foreclosure in cases where there are multiple mortgages. Furthermore, investors in securitized loans no longer want to see only group/aggregated data: loans must have unique IDs on the secondary market, so that their origination and terms can be scrutinized and tracked. Another advantage of having a universal ID for loans would be that it could be used to track loans in transition e.g refinances. (MISMO/MERS numbers were brought up. For example, HMDA could contain the MERS number, but not all loans have a MERS number, and MERS' role may change).

iii. **Identifying patterns that could strengthen HMDA's ability to prove discrimination**

Identifying possible discrimination is one of the original statutory purposes of HMDA. There are several ways in which the data could be expanded or disaggregated that would allow researchers to identify patterns in the provision of services as well as differentiation in the quality and riskiness of products.

a. *Disaggregation of race category:* The current race/ethnicity field in the HMDA dataset contains categories that are too broad to identify possible patterns of discrimination or product differentiation based on race—e.g. "Asian". Disaggregation can help identify groups that are more susceptible to high-risk lending. It is also increasingly important in areas with new immigrant populations.

b. *Additional socioeconomic data:* More detailed information on the socioeconomic status of borrowers could be useful to researchers in making links between certain characteristics and the products and services offered.

c. *Lender identification:* It can be difficult to identify who the lender is because of unclear codes and the fact that many subprime companies had more visible parent companies. Improved lender identification fields could help researchers see different patterns in loan and product dissemination between lenders, as well as allow for comparison across lenders.

iv. Make data easier to use and report through standardization

The idea of standardization was discussed in a number of ways: from the perspective of making the data easier to use and interpret for researchers, as well as streamlining the reporting requirements that HMDA-regulated institutions face from different federal agencies. In this sense, greater standardization could be a way of reducing redundancy in reporting requirements. However, that this would also add additional burdens to reporting institutions as they would need to construct and report data in a way that they currently do not for business purposes.

a. *Standardize how data fields are defined and/or calculated:* There is a great deal of variation in how lenders report data, including how they calculate certain key metrics and the document preparation systems that institutions use. This creates a big issue for system integration. The difficulty of accessing HUD1 data is another issue, largely due to variety of loan origination and document preparation systems; ideally, this would be integrated as part of any standardization process, but now it would require a tremendous amount of programming.

b. *Require institutions to provide data in a computer-ready format:* A modest change would be to require institutions to provide data by request in a computer-ready format after the 30 day period after it is submitted—this would allow users to have access to the data without having to wait for the Fed's version.

II. Recent Developments in HMDA and Other Mortgage Data Initiatives

After the initial discussion, conversation turned to several brief updates.

i. Statutory Changes to HMDA

Participants were updated on statutory changes that will be made to HMDA as a result of the Dodd-Frank Wall Street Reform and Consumer Protection Act (H.R. 4173); the summary is provided in Appendix B. In addition to the existing fields, the following fields will now be required, listed below by category:

- Borrower information: age; credit score.
- Property information: value.
- Loan terms: points and fees; spread price on all loans; prepay penalty term, introductory rate term and loan term (in months); and non-standard amortizing terms.
- Other information: broker/loan office license number; universal loan ID; parcel number.

In summary, the financial crisis has fundamentally altered the expectation of what is reported. State regulators have inserted themselves into data collection, and the Office of the Comptroller of the Currency and the Office of Thrift Supervision are collecting data on loan modification activity. These sets of data have hundreds of fields. Across sets, the data fields will be exhaustive, but their definition will not be uniform: the same field names will mean different things in different contexts, and not all specifications will make sense for HMDA.

Additionally, it is important to note that Congress started the Consumer Financial Protection Bureau to help achieve the goals of HMDA, and conduct policy debates on many of the issues surrounding its use. We must allow sufficient time for the CFPB to determine its structure and priorities, as well as complete the numerous studies related to expanding consumer protection, as mandated by Congress, before final judgments regarding the need for additional studies can be made.

ii.   Change to Secondary Market Reporting

Moreover, what investors want to see in private label securitization transactions has been fundamentally altered: loan-level data is required now in order for investors to be able to analyze risk. These transactions are covered by SEC disclosure laws and so privacy issues are paramount because of the loan-level data. Industry groups have worked with originators and investors to come up with a 16-digit unique identifier that captures underlying loan type, origination date and country of origin, in addition to randomized alphanumeric data.

One point worth noting is that Fannie Mae and Freddie Mac are not required to provide loan-level data because the government is guaranteeing these loans and therefore the associated risk is not relevant to investors.

Ongoing reporting of loan performance of originated loans is a separate issue that is not covered by SEC disclosure laws. The industry group is undertaking a year-long implementation process for reporting on servicing data. Servicers give monthly data to investors on whether or not payments are being made, if they are on time, and how much they are for, but loan modification is not clear. Investors will want to have granular loan-level data, moving forward.

A goal of the industry group is to ensure that changes to HMDA are the same as changes made to the SEC's reporting requirements, and that primary market data collection and secondary market data collection are aligned.

iii.   In the Works: National Mortgage Database

A National Mortgage Database is being developed to complement, not substitute for, HMDA data. Relying on a representative random sample approach, it is conceived as a way of creating a robust data set that could sound strong alarms to policymakers and politicians—in contrast to the isolated case studies and anecdotes of subprime mortgage distress that were not given much credence in the eight years leading up to the subprime crisis.

This effort would create a large body of publicly-accessible data and would not be focused on the same type of data or goals as HMDA, but rather would rely on data from credit repositories: these companies who have more data than anyone else, with most loan servicers reporting to them.

The data set will be a representative random sample, and because credit reports include data from the past seven years, the entire data set can be created as though it had been populated with data throughout the last seven years.

A second level of building the set would be to add additional data from the marketing side of the credit repositories. This would provide information on the mortgage characteristics, as well as demographic and income data on borrowers. A third level of building the set would be to merge it with HMDA data and other external sources, such as AVM models. A fourth level would be to add a survey, which will enable analysts to speak to the suitability of the mortgage product for the borrower.

The data set will be depersonalized; it will not provide a lot of geographic detail (maybe limited to state level) and will not provide information about the lender. This is in line with the goal of the database, which is to be able to provide answers to broad public policy issues.

One participant suggested that it would be useful to be able to differentiate between lenders—not for the purpose of identifying lenders associated with specific loans, but to know if loans are originated by different lenders.

More information on this project is included as Appendix B to this document.

III.   The Bank Industry Perspective

One of the goals of this convening was to ground the recommendations of HMDA data users in the reality of the challenges faced by banks and other HMDA-covered institutions in complying with HMDA reporting requirements. For this reason, the next session of the day solicited input from several representatives of the banking industry on the issues they face in reporting as well as some of the "wish list" items of researchers.

Several topics emerged; the conversation is synthesized according to these themes.

i. Reporting Capacity

Some data will be easy to report and not create any problems—particularly if investors are requiring it anyway. However, by definition, loan-level data is only requested for loans that are originated, which could create an issue of who is paying attention to the loan applications that are denied.

It is important to inject a dose of reality on the true return or benefit of the substantial increased burden on financial institutions, including costs to upgrade systems and technology. An alignment of definitions between what information is required for secondary markets and for HMDA would go a long way for banks in minimizing the incremental effort required with expanded requirements.

For large institutions, the concern is about how data is used, manipulated and interpreted with regards to making conclusions about discrimination. By way of example, one institution entered into a monetary settlement with state regulators as a result of an unintentional programming error, even though the institution promptly re-filed corrected HMDA data, conducted fair lending analysis on the re-filed data, and no fair lending issues were found.

There is also a difference in reporting capacities between big and small institutions. Of the 7500 institutions that report HMDA data, approximately 100 are responsible for 80 percent of loan originations. Nevertheless, what the Federal Reserve requires may ultimately be driven by what the small institutions can do—a lowest common denominator approach. Some small institutions are already struggling with HMDA requirements as they are.

This last point was something touched on by several participants: extensive reporting requirements may force lenders out of the lending business. The additional costs (technological and otherwise) that institutions will bear to comply with additional reporting requirements will inevitably be borne to some degree by the customer. In addition, added costs will be factored into an institution's business decisions regarding consumer lending programs and offerings. Therefore, it is especially important to weigh the substantial costs of increased reporting and disclosure requirements against any potential benefits.

Standardization

Speaking to the issue of integrating systems to ensure greater standardization, even though recordkeeping is electronic, at many institutions the systems are not as automated as they sound. At many institutions, key decisions, made by the people involved, are still recorded as if on index cards [comment fields]. Thus simple data analysis cannot really capture all of the process that is behind lending. Further, many loan originators are using multiple systems, many of which are old, and there remains a lot of variation in underwriting standards and calculations of key metrics e.g. income, loan-to-value ratios. For example, there are many different ways to

calculate debt-to-income ratio. In fact, even income determination varies; some use income to qualify, others may set a different standard for reporting/collecting income.

Institutions generally know the cost of complying with HMDA, but are foiled by the fact that different regulators ask for different data fields in other areas. The conversation turned to the issue of standardizing the set of variables that are reported across regulatory agencies.

However, even within a large bank, there are so many different products that coming up with a set of standardized variables would be very difficult; furthermore, regulators would never agree on a standard set of variables, as one participant suggested they should as a step towards streamlining reporting on the bank's end. The risk of standardizing definitions for data elements is that the "standard" definition may not match what is, in fact, used at the institution in its pricing decision, thereby leading to incomplete or inaccurate analysis. The OCC may come up with a standard set but as they drill down in their examinations, they ask for many more variables that banks must extract manually from loan files that are 30-60 pages long. Often times, the regulators pick obscure loan products to examine and use indicators that the banks themselves would never rely on to determine risk.

Substantial potential benefits would accrue from standardizing fair lending review methodologies. At present, regulators, academics, and advocates can use any one of a myriad of approaches and methodologies to evaluate performance. Standardization in this case will be helpful to the end-user of the analyses.

Could we find broad definitions for products that are not that dissimilar, for example? Or find a simple way to code them? Perhaps the list of variables reported could be standardized if the list of loan products could be standardized: loan products are different in name, but standards in how they are labeled could reduce the seeming differences (e.g. Adjustable=A, term goes on end of expression). The downside is that, as long as new products are being introduced, new data fields will always be up for discussion.

This led to a discussion of the need to add new data fields in a timely manner; as a case in point, pricing data was not added until 2004. This might mean talking about the process by which HMDA is updated and how frequently. For example, one participant noted that regulators could consider piloting fields, perhaps even with a limited number of reporters on a regular basis; this would make it so that opportunities to make changes to HMDA are not as seen as high stakes, or "all or nothing".

ii.   Accuracy and Precision

The related but not identical issues of accuracy and precision were touched on by several participants.

One question raised was that of what level of precision should guide regulators in setting HMDA requirements. One example used to demonstrate the point was income: even given the differences in how institutions calculate income, it has been a useful field for researchers. HMDA data users should want high-quality data and take advantage of opportunities to get good data and good measurements, but not necessarily be looking for it to be "precise to the atom."

A counterpoint made by another participant was the distinction between accuracy and precision. Describing a loan as an ARM is accurate, but the terms might not be described with enough precision to be useful.

Since credit scores are to be required as part of HMDA requirements, there were several comments on accuracy in credit score reporting. Someone pointed out that a lot of scores are different—which score should be used? Some lenders might report scores that they calculate themselves. The issue is further complicated when there are multiple borrowers.

Generally, there was a sense that accuracy was the greater value. Tolerances could be set for precision (though that might come at the cost of some ability to align HMDA standards with those of other regulators who might need different levels of precision for different data points).

IV.   Privacy Concerns

The next topic for discussion was privacy. Rather than focus on whether or not certain data fields should be reported because of their sensitivity the conversation focused on who should have access to data once it is reported, and on what terms. For the purposes of this conversation, participants were allowed to assume amendments to the major statutes that relate to financial privacy. Examples from other data sets were introduced for consideration of their applicability in safeguarding privacy, working under the assumption that, with an expanded HMDA set, adding more variables enables better matching with other data sets that can result in positive identification at the parcel level. Several people pointed to cases in which HMDA data have been matched to other data sets with up to 85% accuracy—highlighting the fact that as is, privacy is not protected.

Someone pointed out that more frequent reporting (as some would want to increase the timeliness of data) also increases privacy concerns, because it creates a smaller pool from which to match data from public records on origination dates.

Two basic possibilities for protection privacy were identified: restricting access; and restructuring data. Discussion points about each are summarized below.

i.   Restricting Access

One model of restricting access is to have gradations in access. For HMDA, this might look like a standard HMDA set that is publicly available, and an expanded set that is only accessible to approved researchers.

Many public agencies have protocol for granting access to sensitive information; these were discussed as possible models.

At the U.S. Census, a merged data set that includes economic data from the Internal Revenue Service and from employers, is only accessible to Census employees who have clearance; employees cannot even go to the floor where the dataset is located without clearance. Data is accessed in a separate room and output is reviewed to ensure non-disclosure. A representative of Census highlighted how important this level of restriction is, because there is a need to build public trust while also allowing people to use data to a good end.

Census also maintains a protocol for deputizing researchers to access this matched dataset. Data resides in Washington where the match takes place and is accessible to researchers who are rigorously vetted and then subject to all the same penalties as Census employees if there is a security breach. There are also very strong restrictions on what can be reported from the set. There are centers where the data can be accessed throughout the country, but the process of becoming deputized can take two to three years and costs money. In New York, a consortium of universities funds a Census research center. Someone pointed out that this model might work in places with a lot of institutions with financial and human resources, but could act as a barrier for smaller institutions as well as researchers outside of the major metropolitan areas where satellite access points are located. This would be a step backwards for HMDA, which has been a powerful tool because it is free and public; several participants made this point. Note that the census example was brought up as a principal to apply, and a variation on this approach might be workable.

Many public agencies such as departments of education and child welfare have a process by which researchers are vetted and once approved, can gain access to files with private information.

If the "trusted researcher model" for data access and evaluation is adopted, it must include clearly stated conditions that the methodologies used must be statistically valid, transparent, and fully disclosed.

Part of the issue of access is usability: HMDA data can be free and downloadable, but if it is not easy to understand and manipulate then it cannot serve its intended purpose or consumer interests.

This brought up the idea of user interfaces or interactive systems, where using a few commands or clicks, individuals could access aggregated data or data for a specific geographic area. For

example, such a system could be used if someone wanted to buy a house and was interested in seeing the other kinds of mortgages that had been made in their area. This would make the data set a public use dataset, and not restrict its use to policy makers and scholars. Related to this, someone mentioned the National Institute for Computer Assisted Reporting which trains journalists to use government databases so that they can be effective in using data to bring issues to the public's attention.

In certain cases, or for certain sensitive data, information could be reported aggregated at the tract level rather than loan level. This could be a role for the Consumer Financial Protection Bureau; the Bureau could aggregate data and make its analysis public to ensure that such analysis is conducted in a consistent and transparent manner across financial institutions.

A final point on restricting access is that by doing so, we restrict the possibility of HMDA (and the public, by extension) being the so-called third regulator, or the regulator of the regulators.

ii.   Restructuring Data

An alternative to restricting access is restructuring data.

One way of restructuring data is to alter it so that once released, it is not identifiable. Census has developed numerous techniques to that end.

There are a lot of ways that the data could be restructured so as to make it more useful to researchers: making loan variables less precise through binning, bottom-coding, top-coding, etc. can frustrate efforts to reengineer data so as to make it personally identifiable or able to be matched to other sets. The question for researchers then becomes what are the fields that are vulnerable to matching, what could they be merged with, and what level of imprecision can be tolerated?

Many participants felt that if researchers want HMDA to be available and useful moving forward, there needs to be a way to protect privacy; this is true not just for HMDA data, but for other datasets that the government is collecting.

One risk of restructuring data is that the ability to compare old and new data might be impaired. In response to this, perhaps an authorized researcher model (like with Census) would be appropriate: if blurred data prevents researchers from answering their research questions, they can apply for access.

Another way of protecting privacy would be for government agency to use a parcel number to match data. The match would happen in a black box, so that researchers could use the data as a more complete set but not be privy to the identifying information.

V. Concluding Remarks

The day's concluding discussions were opened with a synthesis of what had been discussed and what possible next steps are, as well as some basic points of agreement among the researchers in the room.

The starting point for the discussion was that from the researchers' perspective, more data is better. Based on this, the role of the researchers in the room was to advocate for more and/or better data with the caveat that it not expose anyone to undue risk. It would be useful to identify the tasks that HMDA should be doing—not just for public purposes but for research also. Researchers want to make sure that data that is extracted "makes the world better"—but they also just want to be able to better understand issues (in the name of academic pursuit). This needs to be balanced against the increased burden to financial institutions of reporting additional data, the need to keep customer information safe and the risk of misinterpreting or misusing the data.

There was recognition that if data are being collected by other stakeholders, it is easier to get them in HMDA. Thus, if there is interest among the group in building a platform for advocacy for a better HMDA set, it might be useful to do a stakeholder mapping exercise to see where stakeholders are already aligned in what they collect and/or want to collect. For example, now there is clearly a mutual interest on the part of securities investors and researchers in having loan-level data, whereas eight years ago securitizers said it was impossible to do loan-level analysis. If one stakeholder is interested in using loan level data to monitor systemic risk, another in using it to value his/her current investments and another in understanding pricing, what are the common data needs among them? These might be easy wins for additional fields, as opposed to some fields that consumer advocates might really want but that do not have a strong constituency otherwise.

Another stakeholder analysis would be with relation to standardization. Standardization is almost a non-negotiable issue if HMDA data are going to be at all useful. However, it becomes quickly complicated. There is a lot that could be standardized across stakeholders--fields, product types, documentation, reporting systems—but proposing standardization could create opposition from various interests: for example, the companies that make money from selling reporting systems. If standardization is going to be advocated, then there needs to be an analysis of where the points of resistance will be.

HMDA researchers have a joint responsibility of not proposing anything that they themselves could reverse engineer and in doing so, undermine public trust and jeopardize the possibility of enlisting even more people in the use of data to highlight important public policy and consumer protection issues.

The idea of figuring out what is non-negotiable is critical as a next step. For some people in the room, this was the current degree of geographic specificity (census tract), with the worry that

13

adding more variables would result in data only being available at a higher (and less useful) geographic level. Figuring out what is non-negotiable with HMDA also depends on what other data sets are available. Researchers are concerned about the mix of data they can get—not necessarily with all the data coming from one set.

Another issue that emerged as a common interest was that of enhancing HMDA's ability to anticipate risk, and the importance of reporting on new and more types of loan products in order to accomplish this. This tied in with the final point of the day, which is that participants generally felt that HMDA should be designed so as to be proactive, and in order for it to be so, there needs to be a regular schedule for revisiting the data set. As part of creating this schedule, participants felt that there needs to be an exercise to identify who the stakeholders are that will be brought to the table for each revision.

Appendix A

Partial Attendee List

Bob Aaronson, Consultant, Former JPMorgan and Chase

Dana Archer-Rosenthal, Independent Consultant

Howard Banker, Fair Mortgage Collaborative

John Carey, Citigroup

Marsha Courchane, Charles River Associates

Arthur T. Cresce Jr., U.S. Census Bureau

Tom Deutsch, American Securitization Forum

Lei Ding, Wayne State University

Ingrid Gould Ellen, Furman Center for Real Estate and Urban Policy

Keith Ernst, Center for Responsible Lending

Irene Fang, Citigroup

Wilhelmina Leigh, Joint Center for Political and Economic Studies

Patricia McCoy, Insurance Law Center, University of Connecticut School of Law

John Mollenkopf, City University of New York

Anthony Pennington-Cross, Marquette University

Kathy Pettit, Urban Institute

Janneke Ratcliffe, UNC Center for Community Capital

Christy Rogers, Kirwan Institute for the Study of Race and Ethnicity, The Ohio State University

Geoff Smith, Woodstock Institute

Jonathan Spader, Abt Associates

Michael Talieferro, ComplianceTech

Mary Weselcouch, Furman Center for Real Estate and Urban Policy

Mark Willis, Furman Center for Real Estate and Urban Policy

Michela M. Zonta, Virginia Commonwealth University

Peter Zorn, Freddie Mac

Appendix B

Home Mortgage Disclosure Act:

Current Fields and Additions from Recent Statutory Amendments

Data	Borrower	Property	Terms	Other
Existing HMDA	• Race (2 borrowers) • Ethnicity (2 borrowers) • Sex (2 borrowers) • Income ($k) • Purpose • Occupancy	• State • Metro area • County • Census tract • Type	• Lien position • Rate spread on higher-rate • HOEPA status • Loan amount ($k) • Application date (censored) • Action date (censored) • Type (conventional, government, insured by agency)	• Action • Reason for denial • Request for preapproval • Type of purchaser *(if sold same year)* • Lender name • Supervisory agency • Loan No.
Additions from HMDA Statutory Amendments	• Age • Credit score	• Value	• Points & fees • Spread price on all loans • Prepay penalty term (months) • Introductory rate term (months) • Non-standard amortizing terms • Loan term (months)	• Channel • Broker/Loan officer license number • Universal loan ID • Parcel number • Such other info as may be required

<u>Reporters</u>: Modestly sized and better depository lenders with a branch and at least one federally-related residential mortgage in an MSA in the preceding year; Non-depository for-profit lenders with modest and better residential mortgage activity in an MSA in the preceding year.

<u>Covered Loans</u>: Any loan for home purchase, home improvement (may or may not be secured), or refinancing of a dwelling-secured loan.

<u>Public Access</u>: Loan- and application-level data with three fields censored (two dates and loan number) and two fields kept rounded to thousands (income, loan amount).

Geography Detail: Census tract in counties with population of at least 30,000.

Sample of Recent and Proposed Publicly-Sponsored Mortgage-Related Loan-Level Data Collection Efforts

Sponsor – Effort & Status	Covered Reporters	Covered Loans	Public Access	Geography Detail	Brief Summary of Reported Information
SEC- ABS *Proposal*	Prospectus filers, servicers of loans underlying ABS	Loans serving as collateral in AB	EDGAR to provide machine readable data	Metro area	Hundreds of data points prior to offering for mortgages covering originator, channel, loan terms, post-closing pricing, borrower qualifications, modification activity, underwriting details including exceptions, valuation, property, servicing (e.g., fee), and, after offering, performance.
FHFA/GSEs - Uniform Mortgage Data Program (UMDP): *Full Implementation Pending*	Sellers of loans to GSEs or those who submit loans for securitization	All loans sold/securitized by GSEs and related loans (e.g., simultaneous second)	Unknown, but Freddie has disclosed some loan level data on pools	Street address	71-page data dictionary outlining variables that describe the secured property, valuation, borrower qualifications, loan terms, loan pricing, loan purpose, underwriting process, modification activity, escrow, securitization-related fields, mortgage insurance, performance, servicing arrangements, and various HMDA fields.
CSBS/AARMR—e.g., 04 NCAC 03M .0401: *Effective 1Q-2011*	State mortgage licensees including lenders and brokers	Residential mortgage loans	Not public	Census tract	Quarterly reporting on loan and lender identifying information; pricing information sufficient to calculate APR, schedule of payments, information disclosed on GFE, measurements from HUD-1, loan terms, information typically used in underwriting, and HMDA LAR.
OCC/OTS loan-level modification: *Ongoing*	National Bank and Thrift servicers	Presently, 34 million first-lien residential mortgages	Not public	Zip code	100 fields including measurements at origination and time of consideration for modification related to borrower qualifications, loan terms, underwriting, location, and performance.
Treasury-HAMP: *Ongoing*	HAMP participants	Loans considered for HAMP modification	Not public but some disclosure planned	Street address	Reporting on trial modification and permanent modification events (including current performance, pricing, and limited underwriting elements), as well as performance in trial and permanent modification status.