WRITTEN TESTIMONY OF PATRICIA A. MCCOY[*]

Hearing on the Home Mortgage Disclosure Act
Before the Board of Governors of the Federal Reserve System
September 24, 2010
Washington, D.C.

Thank you for inviting me to testify before the Board of Governors on the Home Mortgage Disclosure Act (HMDA) and needed changes. My remarks today address four topics. First, I will discuss the tradition of public access to HMDA data and the important purposes that public access serves. Next, I discuss obstacles that independent researchers faced during the housing bubble, due to HMDA's limitations and other lack of data, in analyzing the growth and risks of nonprime mortgage lending. The third part of my remarks calls for a comprehensive federal residential mortgage data set. Finally, I close by discussing several techniques for preserving public access to HMDA data while safeguarding borrower privacy.

I. HMDA's Long Tradition of Public Access to Data on Home Mortgages

Since its inception in 1975, the Home Mortgage Disclosure Act has made home mortgage data available to the public for free or at a modest price. This feature of HMDA reflects Congress' judgment that greater information would help enlist private citizens and local researchers and organizations to act as "citizen watchdogs" and thereby induce state and federal regulators to take stronger enforcement action.[1] It reflects the further judgment that making mortgage data public is important for the continued vitality of communities and the economy as a whole.

This "citizen watchdog" model of HMDA has had notable success. By shining a spotlight on troubling lending patterns, HMDA reporting and the resulting media coverage have heightened public awareness of the problems from disinvestment and lending discrimination. Public awareness, in turn, raised the reputational stakes of mortgage lenders and propelled many of them to improve their community reinvestment records and fair lending compliance programs. Meanwhile, HMDA data spurred regulators to increase Community Reinvestment Act (CRA) enforcement and revamp the design of fair lending examinations. Public debate about new, emerging home mortgage issues led to successive expansions in HMDA in 1989, 1992, 2002, and 2008 in order to allow HMDA to address those issues.[2] These changes broadened the concerns that HMDA serves to address, from redlining to discriminatory denials of credit,

---

[*] Connecticut Mutual Professor of Law and Director of the Insurance Law Center, University of Connecticut School of Law. Email: patricia.mccoy@law.uconn.edu.

[1] See generally Patricia A. McCoy, *The Home Mortgage Disclosure Act: A Synopsis and Recent Legislative History*, 29 JOURNAL OF REAL ESTATE RESEARCH 381 (2007), available at http://cbeweb-1.fullerton.edu/finance/journal/papers/abstract/forth/accepted/JRER_SI(0703S02R1)_5.htm.

[2] In 1989, in the Financial Institutions Reform, Recovery, and Enforcement Act (FIRREA), Congress expanded HMDA reporting to include these new data fields: mortgage applications, mortgage denials, race, ethnicity, gender, loan sales, and type of purchaser. In the Housing and Community Development Act of 1992, Congress enhanced disclosure of HMDA data to the public. A decade later, in 2002, the Federal Reserve Board amended HMDA's regulations (found in Regulation C) to require reporting of price spreads on higher-priced loans. In 2008, the Board again amended Regulation C to expand the reporting of price spreads to more loans.

discrimination in the pricing of credit, predatory lending, changes in mortgage industry structure, and concentrations of mortgage risk.

Over the years, a broad range of individuals, local groups, community institutions, and state and federal agencies have used HMDA data for a multitude of purposes. Local groups and the media use HMDA data in communities all over the country to monitor community reinvestment, fair lending compliance, and the growth in subprime lending. State and federal regulators rely on HMDA data to enforce CRA and to pinpoint originators with racial or ethnic lending disparities for further investigation. Financial institutions analyze their HMDA results to evaluate their lending records and make improvements. State banking regulators and state attorneys general routinely match HMDA data to other data sets to help them enforce state mortgage licensing and banking regulation laws.

Finally, independent researchers, including economists, sociologists, and legal scholars at universities, research institutes, and think tanks rely heavily on HMDA data to conduct increasingly sophisticated empirical research on a host of questions. These research topics have involved not only discrimination and reinvestment but other compelling issues involving mortgages including systemic risk and safety and soundness. However, as I discuss below, HMDA's relatively limited data fields, combined with the paucity of affordable, proprietary data that could help researchers supplement HMDA, hampered the ability of independent researchers to address these important questions on a timely and full basis.

The federal government also conducts substantial research using HMDA and other mortgage data. Internally, the federal government can supplement HMDA data through proprietary databases purchased from private vendors and information culled from other mandatory reporting and depository institution examinations. Not all of these sources have been aggregated (for instance, examination data) and other aggregate data are difficult to merge due to lack of standardization. In addition, as I will describe below, the government's use of proprietary mortgage databases has not been free from problems.

Given these challenges, the federal government's oversight efforts would be greatly aided by a comprehensive mortgage database that has the capacity to track each mortgage loan from loan application to maturity. If such a database existed, public access to that data would remain essential. The federal government does not have the budget or the staff to research all of the important questions on the residential mortgage market that need to be asked. Even if its resources were infinite, moreover, the federal government would still need outside independent researchers to act as its "eyes and ears," because they are better equipped to detect mortgage trends in local communities and have motivation as local residents to investigate them. These researchers, located in universities and other non-profit settings, also infuse the research enterprise with a diversity of views and methodologies that shed light on the full spectrum of public policy concerns better than the federal government can alone. Independent researchers often ask new and different questions and use different techniques to examine them. In turn, these issues often come to independent researchers' attention precisely because community groups identified the issues using HMDA data. Consequently, in HMDA reforms going forward, it is vital that public access to mortgage data be preserved as fully as possible, consistent with borrower privacy.

II. Restrictions on Data Access and Problems That Resulted During the Housing Bubble

Although HMDA has had demonstrated success in flagging mortgage market problems, it nevertheless had well-known limitations that hamstrung the ability of independent researchers to analyze the growing subprime threat. In all likelihood, the financial crisis would not have happened or been so severe if researchers had had access to better home mortgage data sooner that could have enabled them to document, analyze, and publicize the dangers of nonprime loans. During the housing bubble, no one outside of the lending industry – neither government researchers nor independent researchers at universities and non-profit organizations[3] – had sufficient mortgage data to fully document the mounting risk.

This problem was most acute for researchers at universities, think tanks, and other non-profit settings.[4] During the growth in the nonprime market, these researchers sought to analyze numerous pressing questions but were stymied because of lack of data access. Some of those questions included:

- What explained higher default rates in nonprime loans (the borrowers, the loan terms, or both)?
- How was the mix of different mortgage loan types changing and what were the implications for borrower welfare and systemic risk?
- Who were getting nonprime loans and could they manage the risk?
- What types of laws (if any) were successful in reducing default rates on mortgages?
- Did disparities in mortgage lending laws by jurisdiction and by type of lender accelerate a race to the bottom in lending standards?
- Were higher-quality borrowers steered into unnecessarily costly loans and what harm resulted?
- Were racial and ethnic disparities in the pricing of home mortgages explained by creditworthiness or rather the result of discrimination?

Independent researchers had trouble getting access to the data needed to study these questions for two major reasons. First, for all intents and purposes, HMDA was the only publicly available mortgage data set of national scope. HMDA had limited data fields, however, with insufficient data or none at all on borrower creditworthiness, loan terms, loan pricing, and loan performance. These limitations made it impossible for researchers outside of private industry and the federal government to tackle the questions just listed using HMDA alone.

Second, researchers found it next to impossible to get the private data they needed to supplement HMDA. Normally the only way researchers could tackle those questions was through access to rich proprietary data sets assembled by private companies such as Loan Performance or McDash. For researchers outside of the lending industry, however, those data were priced out of reach (costing $200,000 or more) or, in the case of Loan Performance data, were often not for sale at

[3] Loan-level data limitations in federal securities disclosures also severely hampered the ability of investors in private-label mortgage-backed securities, collateralized debt obligations, and related securities to perform meaningful due diligence on the securities they bought.

[4] For a general discussion, see Elizabeth Warren, *The Market for Data: The Changing Role of the Social Sciences in Changing the Law*, 1 WISCONSIN LAW REVIEW 1 (2002).

any price. As a result, in all but a handful of cases, independent researchers were not able to get access to the proprietary databases and had to put important research projects aside.

Even when independent researchers did gain access to a proprietary data set, sometimes the vendors obstructed their research. I know of one case where the vendor threatened to cancel a license – which the researchers had bought for a limited time at prohibitive expense with support from a foundation -- when it discovered what the researchers were asking. Similarly, just this past year, a vendor threatened to cancel its license to a federal agency working with university researchers when it learned that some of the research projects planned to match the proprietary data with other data sets. Some of the affected research had to be canceled or scaled back as a result.

Independent researchers sought a way around these problems by seeking mortgage data directly from major mortgage lenders, but those requests almost always hit a wall. Several years ago, for example, major banks rebuffed requests by researchers at one of the nation's top universities to provide loan-level data on risk levels in loans to racial and ethnic minorities. The researchers ended up altering their research design and were forced to rely on HMDA data instead.[5]

These obstacles to data access blocked significant research by independent researchers on urgent issues of systemic risk during the housing bubble. When such research was done, business interests often attacked it for lack of data variables and sufficient controls. Meanwhile, researchers affiliated with lenders were able to use proprietary databases to generate multivariate regression studies with results favorable to industry. Almost always, the authors of those studies refused to share the underlying data with the public, thwarting efforts to validate the results.

This unlevel access to data had devastating implications for the country. During the run-up to the financial crisis, independent researchers could not track mounting mortgage risks promptly and convincingly, due to lack of mortgage data. Had they had that information, the impending crisis might have been recognized sooner and stopped before it was too late. Instead, the risks mushroomed out of control, causing untold harm to millions of Americans and the U.S. economy.

Given the magnitude of that harm, we can no longer afford as a nation to allow private interests to withhold or control access to mortgage data that are vital to evaluating mortgage market trends that threaten the stability of the United States financial system. Opponents to broader public mortgage reporting often cite the added cost of higher reporting obligations. But what goes

---

[5] The federal government had greater success in gaining access to private mortgage data, but even then, sometimes its access was crimped. For instance, a 2005 study of 2004 HMDA data by Federal Reserve Board economists found significant disparities in the price of higher-cost home mortgages by race and ethnicity. To introduce controls for creditworthiness, the authors asked the Credit Research Center at Georgetown University (CRC) to merge HMDA with CRC data from eight unnamed subprime lenders and analyze the pricing disparities controlling for borrowers' credit scores. CRC performed the analysis and reported that controlling for creditworthiness eliminated disparities in some types of loan products. Robert B. Avery, Glenn B. Canner, and Robert E. Cook, *New Information Reported under HMDA and Its Application in Fair Lending Enforcement*, FEDERAL RESERVE BULLETIN 344, 385-387 (Summer 2005). Subsequently, an official at CRC told members of the Consumer Advisory Council that it did not allow such outside authors to analyze the CRC data directly or examine it for themselves.

unsaid is the high cost of *not* having public access to that data. Ordinary Americans are bearing that cost right now, in the form of TARP bailouts, lost wages, and lost jobs.

III. The Need for a Comprehensive Residential Mortgage Data Set

The Home Mortgage Disclosure Act has three essential purposes: to help evaluate reinvestment in local communities, to assist public officials when targeting public investment, and to help spot possible lending discrimination. These purposes continue to be important today.

At the same time, the need for public mortgage data has grown beyond the three purposes contemplated by HMDA. During the growth of nonprime lending, rapid changes in the mortgage market raised many new, important questions that HMDA was not suited to answer. Some of those questions, such as discrimination in loan prices, HMDA shed light on, although not very well due to its limited scope. Other questions, such as the effect of nontraditional loan terms on default risk or the success rate of loan modifications, HMDA could not address at all. Accordingly, in rethinking public reporting of mortgage data, it is vital to recognize that public mortgage reporting serves other compelling public purposes in addition to HMDA's stated goals. These purposes include safety and soundness, investor protection, consumer protection, and systemic risk.

Since 2008, the federal government has instigated other, scattered mortgage reporting programs as an *ad hoc*, partial response to the need for greater data. The Treasury Department collects data on loans considered for modification under the Home Affordable Modification Program (HAMP). The Department issues monthly reports on modification trends but does not make the data public. The Office of the Comptroller of the Currency and Office of Thrift Supervision also collect loan modification data, but do not release them to the public. The Securities and Exchange Commission has proposed expanding loan-level mortgage disclosures for asset-backed securities. Although these data would be available to the public on EDGAR, the SEC would not aggregate the data on an industry-wide basis because EDGAR reports that data only by individual issuer and deal.[6]

Meanwhile, the new Dodd-Frank Wall Street Reform and Consumer Protection Act, passed in July 2010, mandates more mortgage data collection. The new law expands HMDA's data fields in crucial ways. Separately, the law instructs the Department of Housing and Urban Development and the new Consumer Financial Protection Bureau (CFPB) to jointly establish a default and foreclosure database. Other aspects of the legislation will likely create a need for expanded data reporting as well. The CFPB will probably need additional data to assess emerging trends affecting consumer protection.[7] In addition, the Federal Reserve Board and the new Financial Stability Oversight Council, in their role as systemic risk regulators, will need expanded mortgage data to assess systemic risk.

---

[6] The Federal Housing Finance Agency, in tandem with Fannie Mae and Freddie Mac, are developing a separate uniform mortgage data reporting program for all loans sold or securitized by the two GSEs. Meanwhile, the Conference of State Banking Supervisors and the American Association of Residential Mortgage Regulators collect data on residential mortgage originators and loans through the National Mortgage Licensing System and Registry. That system recently launched a website where members of the public can research the enforcement history of mortgage originators. See http://www.nmlsconsumeraccess.org/.

[7] Under the Dodd-Frank legislation, the CFPB will take over responsibility for administering HMDA from the Board once the agency is created.

Having multiple federal mortgage reporting systems – one on loan originations, others on loan performance and modifications -- does not make sense. Currently, many research issues cannot be studied – not the least being systemic risk -- without merging the data sets. Doing so is difficult or impossible, however, because the data fields are not standardized and the data sets have radically different coverage. Meanwhile, the existence of these various databases puts duplicative reporting burdens on lenders. The public furthermore lacks access to the Treasury and OCC/OTS data sets.

Consequently, I urge the Board to consider creating an integrated mortgage database that serves other public purposes in addition to HMDA's, including safety and soundness, consumer protection, and systemic stability.[8] This comprehensive database should mandate reporting along at least five dimensions: (1) loan features, including loan terms and prices; (2) borrower characteristics; (3) geographic characteristics; (4) loan applications, originations, and denials; and (5) loan performance. It is furthermore essential that these data be available not just to the government, but to the public on terms affording adequate safeguards for borrower privacy.

IV. Balancing Public Access and Borrower Privacy

Public access to home mortgage data raises important issues of borrower privacy. HMDA itself does not publicly disclose individual borrowers by name or address. However, some of the loan level data disclosed in HMDA, when merged with other data sets or tracked through public registries of deeds that reveal borrowers' names, do raise legitimate privacy concerns. In some cases, it is possible to identify borrowers by searching registries of deeds using HMDA data identifying the census tract and amount of particular loans. As HMDA data expand to include more data about borrower characteristics, such as credit scores and debt-to-income ratios, privacy concerns will become more acute.

Barring collection of mortgage data of importance to public policy is not the right solution to privacy concerns. Neither is denying public access, which HMDA does not allow. Instead, the government should collect the full data elements it needs and then use privacy safeguards to permit as much public access as possible while preserving borrower anonymity.

To strike the right balance between access and privacy, the Board might consider a two-tier disclosure system that adjusts the level of public access according to the sensitivity of the data that is revealed. In the first tier of public access, members of the public would have full access to a scrubbed data set that adjusts certain data fields to protect the identity of the property and the borrower. As the tradeoff for privacy protections, this scrubbed data set would have limitations that would make it impossible to analyze certain research questions of importance to public policy. To permit that type of research under tighter privacy safeguards, the system's second tier would give access to the complete data set, but restrict that access to federal government researchers, state government analysts, and "trusted" independent researchers under strict safeguards such as the Census Bureau's research protocols.

---

8 Because of timing issues involved in federal securities law requirements for initial and periodic reporting at the issuer and deal level, it probably would be difficult to integrate securities reporting of mortgage data into a comprehensive federal database. It would still be desirable, however, to standardize data fields in securities and HMDA reporting, even if the two data sources could not be easily consolidated.

Tier One—Full Public Access to a Scrubbed Data Set: Under the first tier, the federal government would scrub the raw data that HMDA reporters provide to produce a loan-level data set with adequate privacy safeguards to make available to all members of the public. There are at least four techniques the government could use in scrubbing data to protect borrower privacy.

> (1) *Blurring*: This technique sorts individual data from sensitive data fields into "buckets" that make it difficult to trace the loan back to the property and the borrower. The key variables to blur would be loan amount, property value, and date of origination. HMDA already blurs the date of origination by only reporting the calendar year of the loan. As for public reporting of the loan amount, HMDA currently reports loans in $1000 buckets. These $1000 buckets may be too fine because they reportedly permit identification of borrowers from public records in some circumstances. To better preserve privacy, the Board might consider using bigger buckets, for instance grouping loans together in $5,000 or $10,000 increments. Before making any final decision, the Board should test buckets of various sizes to see how big a bucket needs to be to impede identification of the property or borrower through registries of deeds.
>
> In designing buckets, one issue needs to be kept in mind. Many census tracts have few loans and houses above a certain value or below a certain value. For example, the average West Hartford, Connecticut, census tract will have few first mortgages of less than $50,000 or more than $2 million. It is possible to blur the identity of those "outlier" loans and properties through techniques known as "top coding" and "bottom coding." Using bottom coding, the lowest loan bucket might be reserved for loan amounts of $1 to, let's say, $75,000. Using top coding, the highest loan bucket might be reserved for loan amounts of $1 million or higher. This makes it harder for users to identify the outliers by lumping them together with properties or loans in bigger or smaller amounts.
>
> (2) *Censoring Geography*: Currently, HMDA data are publicly disclosed at the census tract level. Sometimes, in census tracts with small numbers of loans, blurring will not be enough to mask the identity of properties and borrowers. In those cases, the government should also consider censoring geography on a selective basis. This technique would start by censoring public reporting of loans at the census tract level. If that was not enough to protect privacy, next the government would censor geography at the county level and if that did not suffice, next it would censor at the metropolitan statistical area level.
>
> It would be undesirable to censor census tract data across the board, because doing so greatly limits the ability of public researchers to study neighborhood conditions. Accordingly, the government should only censor geography selectively, on an "as needed" basis, when blurring is not enough to protect borrower privacy.
>
> The government could also produce companion data sets combining different levels of blurring and geographic censoring to permit closer study of different sets of variables. For example, one publicly released data set could blur borrower

characteristics more and provide full geographic information, to facilitate studies of geographic effects. A companion publicly released data set could provide greater detail on borrower characteristics while censoring geography, to permit study of borrower demographics.

(3) *Injecting Random Noise*: Another privacy safeguard, which the U.S. Census Bureau uses, consists of introducing random "noise" into publicly reported data by distorting the values reported for specific variables on a random basis.[9] This technique is more controversial than blurring or censoring geography, in part because econometricians and statisticians do not agree on the proper way to introduce random noise while avoiding bias and preserving the distribution curve and in part due to differences about how much privacy protection results.

(4) *Redacting Unique Geographic Identifiers*: In the interest of privacy, HMDA already redacts a handful of data fields on loan/application registers (LARs) from public view. These include the loan applicant's name, loan identification number, date of application, and date of loan decision. If the Board or the CFPB mandated reporting of parcel identification numbers or universal loan identification numbers, as the Dodd-Frank legislation allows them to do at their discretion, these two fields would also need to be redacted from public view due to their sensitivity.

Tier Two—Screened Public Access to the Full Data Set: Some research projects of vital public importance could not be performed with the scrubbed public data set due to its limitations. For example, independent researchers who wanted to examine racial disparities in high-cost lending at the census tract level, holding individual credit scores constant, could not do so using the scrubbed data set if census tract level data were suppressed. Nevertheless, the public has an interest in research projects like these on issues of public concern.

One way the government could facilitate that research while protecting borrower privacy would be to adopt a "trusted researcher" model, like that used by the U.S. Census Bureau.[10] Under such a model, full data sets would be made available to researchers working on approved research projects, using an application and clearance process similar to that used for accessing selected, confidential Census data sets at Census Research Data Centers around the country. Under this system, the Census Bureau makes census and other microdata available (including sometimes Internal Revenue Service data) to outside researchers in controlled, secure physical environments. Each outside researcher must first obtain "Special Sworn Status" after undergoing a security check and is subject to stiff legal penalties (including fines and imprisonment) for disclosure of confidential information. Once approved, researchers may only

---

[9] See, *e.g.,* John M. Abowd, Bryce E. Stephens & Lars Vilhuber, "Confidentiality Protection in the Census Bureau's Quarterly Workforce Indicators," U.S. Census Bureau Technical Paper No. TP-2006-02 (2006), http://lehd.did.census.gov/led/library/techpapers/tp-2006-02.pdf.

[10] In addition, federal government researchers and analysts at state regulatory agencies, state attorneys general offices, and other state law enforcement agencies should have access to the full data set under appropriate safeguards.

use the data inside a secure research facility and may not remove confidential data from the center or put it to any use that is not approved.[11]

Let me conclude by saying that this two-tier model is one proposal. It is also worth canvassing other possible techniques for protecting consumer financial privacy. The health privacy protections in the Health Insurance Portability and Accountability Act of 1996 (HIPAA), for example, mandate different privacy protocols for medical researchers that might be instructive.[12] Similarly, other agencies and organizations have developed their own, stringent privacy protocols.[13] Consequently, I urge the Board to survey a range of other privacy protocols to help design a mortgage reporting system that balances public access to sensitive data with individual privacy.

---

[11] For a more detailed description, see U.S. Census Bureau Center for Economic Studies, "Research Program: Overview," http://www.ces.census.gov/index.php/ces/researchprogram.

[12] See, *e.g.*, Gerald W. Woods, "Impact of the HIPAA Privacy Rule on Academic Research," commissioned by the American Council on Education (November 2002), http://www.acenet.edu/AM/Template.cfm?Section=Home&CONTENTID=10385&TEMPLATE=/CM/ContentDisplay.cfm.

[13] These agencies and organizations include the National Center for Health Statistics, the Department of Health and Human Services, the Social Security Administration, the Energy Information Administration, the Bureau of Labor Statistics, the National Agricultural Statistics Service, the Bureau of Justice Statistics, the National Center for Education Statistics, and the National Science Foundation. Agencies in many states and cities have also developed privacy protocols to permit independent researchers to use their data. See, *e.g.*, New York City Department of Education, "Doing Research in New York City Public Schools," http://schools.nyc.gov/Accountability/data/DataRequests.