Implications and Risks of
New HMDA Data Disclosure

By David Skanderson, Ph.D.

January 2018

A version of this paper appeared in ABA Bank Compliance, January/February 2018

Congratulations! If you are reading this article, you survived the "Big Bang" of Home Mortgage Disclosure Act data collection. Your bank is now smoothly and accurately collecting all the new HMDA data fields (as many as 110 possible data fields per application!), and you will be ready to start submitting all those data fields electronically to the Consumer Financial Protection Bureau by March 1, 2019. Right? Well, let's hope so.

Now that you have that job nicely squared away, it's time to start thinking about the implications of the huge expansion of publicly available data about your bank's mortgage lending that is coming in the future. The release of the new HMDA data could come as early as late-March 2019, if the CFPB holds to its ambitious plans. The expanded public data will have profound implications for mortgage lenders' regulatory, litigation, public policy and reputation risk exposure. The new public data will allow regulatory agencies to search more broadly for fair lending and other compliance risks, and to target institutions more precisely for examination and enforcement. It will also provide rich data mining opportunities for private litigants and advocacy groups, as well as for researchers and policy makers. Managing those risks is going to require some planning and analysis.

* * *

A Balancing Act

The CFPB described its approach to setting HMDA disclosure policy in terms of a "balancing test." The Bureau interpreted HMDA, as amended by the Dodd-Frank Act, to require that it balance HMDA's core disclosure purposes and consumer privacy in deciding which data fields to disclose and how to disclose them. The CFPB decided to issue Policy Guidance on the topic, which gives it flexibility in modifying its approach in the future if its view of that balance changes. As such, the Policy Guidance is both non-binding on the CFPB and is exempt from notice-and-comment rulemaking requirements under the Administrative Procedures Act. Nevertheless, the CFPB invited public comments for 60 days after publishing its proposal before issuing its final Policy Guidance – a window that is now closed. However, a recent statement by the CFPB indicates that the 2015 HMDA rulemaking may be reopened: On December 21, 2017, the CFPB issued a statement indicating that, among other things, it intends to engage in a rulemaking to reconsider various aspects of the 2015 HMDA rule, such as the institutional and

transactional coverage tests and the rule's discretionary data points (i.e., data points beyond those specified by the Dodd-Frank Act.) Therefore, to the extent that the data required to be collected changes, the data that would be reported publicly may also change.[1]

Proposal for Public Release of Loan-Level Data

What will actually be released to the public, and what risks will it pose to your bank? When the CFPB issued its final HMDA reporting rules on October 28, 2015, it deferred answering the question of how the expanded data would be disclosed. The much-anticipated answer arrived on September 25, 2017, with the publication of the Bureau's proposed Policy Guidance.

The short answer is that the great majority of the new data fields will be disclosed to the public in their raw, unmodified form. However, a handful of the most sensitive data fields will be excluded from public release, and others will be published in modified form, to reduce the risks to consumer privacy. The various free-form text entry fields also will be excluded from public release because of the practical difficulties of compiling, reviewing and publishing those fields.

The data fields proposed to be <u>excluded</u> from public reporting are as follows:

- Universal loan identifier
- Date of application
- Date of action taken
- Property street address and Zip code
- Credit score
- Automated underwriting system result
- Loan originator's NMLS number

Optional free-form text fields that may be used to report race, ethnicity, name and version of credit score model, reasons for denial, and automated underwriting system name.

The data fields to be <u>modified</u> for public reporting are as follows:

- **Loan amount:** midpoint of the $10,000 interval into which the actual value falls, with a separate indicator of whether the loan amount exceeds the applicable conforming loan limit.

---

[1] See: https://www.consumerfinance.gov/about-us/newsroom/cfpb-issues-public-statement-home-mortgage-disclosure-act-compliance/

- **Applicant age:** discrete ranges (less than 25, 25 to 34; 35 to 44; 45 to 54; 55 to 64; and 65 to 74) plus an indicator of whether the age is 62 or higher (which effectively splits the 55 to 64 bin into 55 to 61 and 62 to 64).

- **Debt-to-income ratio:** actual value if it is greater than or equal to 40% but less than 50%, and discrete intervals otherwise (less than 20%, 20% to less than 30%, 30% to less than 40%, and 50% to less than 60%).

- **Property value:** midpoint of the $10,000 interval into which the actual value falls.

Future disclosure of loan amount will actually be less granular than the approach of publishing loan amounts rounded to the nearest $1,000, as reported by the lender.

That leaves all of the following data fields to be reported in unmodified form:

- Legal entity identifier
- Loan type
- Loan purpose
- Preapproval request
- Construction method
- Occupancy type
- Action taken
- State, county census tract
- Ethnicity of applicant/co-applicant
- Race of applicant/co-applicant
- Sex of applicant/co-applicant
- Type of purchaser
- APR spread
- HOEPA status
- Lien status
- Name and version of credit scoring model
- Reasons for denial
- Total loan costs
- Total points and fees
- Origination charges
- Discount points

- Lender credits
- Interest rate
- Prepayment penalty term
- Combined loan-to-value ratio
- Loan term
- Introductory rate period
- Balloon payment
- Interest-only payments
- Negative amortization
- Other non-amortizing features
- Manufactured home secured property type
- Manufactured home land property interest
- Total property units
- Multifamily affordable units
- Application submission channel
- Whether loan is initially payable to your institution
- Automated underwriting system name
- Reverse mortgage
- Open-end line of credit
- Business or commercial purpose

The proposed Policy Guidance is silent on whether the Bureau would report that ethnicity or race were collected based on visual observation or surname, as opposed to self-reported by the applicant.

Fair Lending Monitoring: Now More than Ever

The annual public release of the HMDA data is typically accompanied by a statement like the following one from the FFIEC's press release on the 2016 HMDA data (September 28, 2017):

> "The current HMDA data alone cannot be used to determine whether a lender is complying with fair lending laws. The data do not include many potential determinants of loan application and pricing decisions, such as the applicant's credit history and debt-to-income ratio, the loan-to-value ratio, and other considerations. Therefore, when examiners conduct fair lending examinations, including ones involving loan pricing, they analyze additional information before reaching a determination about an institution's compliance with fair lending laws."

This statement will remain true, but to a far lesser extent. Regulatory agencies will have access to a much wider array of data than in the past and effectively have it at their fingertips. Regulators will not have to issue a pre-examination data request or Civil Investigative Demand to start screening financial institutions for risk indicators and building regression models to test for disparities.

The process of scoping and targeting fair lending regulatory examinations can be expected to change and accelerate dramatically as a result. Whereas agencies currently must rely on a limited set of data for initial scoping, and later obtain "HMDA Plus" data fields as part of a pre-examination information request, in the future they will have the rough equivalent of the current HMDA Plus data for the universe of HMDA reporters in a highly standardized format in March of each year. That will allow them to develop sophisticated data screening, data mining and statistical modeling routines that can be applied uniformly across the HMDA reporter universe to sift through and identify institutions with indications of elevated fair lending risk or other compliance risk. Their HMDA-based statistical models will still be subject to significant limitations because they will not include all underwriting and pricing factors, but they will be much more comprehensive than currently because the most important determinants of differences in credit outcomes across consumers will be available for modeling.

Opportunities for Regulators to Target More Precisely

To understand how the new HMDA data will allow more refined analysis by regulators, consider how HMDA-based pricing analysis will change. The old HMDA data provides very limited ability to perform analysis of pricing disparities. First, currently it is only possible to test whether there are prohibited basis disparities in the percentage of loans with reportable APR spreads and in the average sizes of those spreads. Second, there are only a handful of pricing-related factors that could be used in a regression model that attempts to explain disparities, and they do not include the most important loan-level pricing adjusters:

- Loan type

- Loan amount

- Property type

- Loan purpose

- Ratio of loan amount to income

- General occupancy status

- Lien status

In the new HMDA world, it will be possible for regulators to test for disparities in both average APR spreads (for all loans originated) and average interest rates after controlling for the following underlined additional variables:

- Specific loan purpose (including cash out)

- More specific property type

- CLTV

- Detailed occupancy status (differentiating investment properties and second homes)

- Credit score

- Loan term

- Number of property units

- Fixed or adjustable rate type

- Introductory ARM rate period

- Non-amortizing features

- Application channel

- Discount points

- Lender credits

Of course, other important pricing determinants will be omitted, such as the specific loan program features and mortgage insurance information, but the new HMDA fields include

those that typically are the most important determinants of mortgage pricing. Thus, the regulators' HMDA-based models will be much more precise in identifying potential fair lending disparities. Similar, but more limited, enhancements will be possible for underwriting models, with the addition of debt-to-income ratios, CLTV, detailed occupancy, detailed purpose and detailed property type being the most important additions relevant to explaining denial rate differences.

A more sophisticated regression model will allow regulators to identify and focus on those institutions with the largest unexplained pricing and underwriting disparities. In addition, it will become possible to analyze risk indicators based on specific pieces of the pricing puzzle, particularly total loan costs, points and fees, origination charges and lender credits, which may provide a window into the discretionary aspects of pricing.

Regulators will be able to do this analysis on a mass, market-wide basis using a set of computer programs that crunch through the full universe of data at once. Of course, this dark cloud also has a potential silver lining: lower-risk lenders can be more easily screened out, and might receive less regulatory attention.

Opportunities for Regulators to Look More Broadly

The new data will also allow regulatory agencies to search for a wider range of potential fair lending risk issues based on the HMDA data alone, such as

- Differences based on age,

- Differences based on specific race or ethnicity sub-categories,

- Denial disparities among marginal applicants,

- Cross-channel differences in pricing,

- Pricing and fee differences by and across loan originators,

- Broker pricing analysis based on aggregating a broker's loans across HMDA reporters,

- Redlining and reverse redlining analysis of reverse mortgages and home equity lines, and

- Product steering between fixed- and adjustable-rate mortgages, or between closed-end loans and home equity lines.

Regulators and others will also be able to search for indicators of potential predatory lending patterns based on loan product and borrower characteristics that some consider to be "higher risk," such as teaser rates, non-amortizing products, high-fee loans and high-debt-to-income borrowers. In the context of redlining analysis, it will be possible for lenders to refine peer comparisons based on the new channel information, by comparing a lender to other lenders that have the same types of distribution channels (retail, wholesale, or correspondent). Beyond that, regulators could use credit score, LTV, DTI and loan product

information (ARM versus fixed in addition to conventional versus government) to evaluate whether differences among lenders in minority area lending penetration may be attributable to differences in their product focus and credit policies.

Of course, members of the public will also have greater access to some of the key determinants of underwriting and pricing decisions, as well as much more complete data regarding loan characteristics, which can be expected to provide fertile ground for class action plaintiff attorneys. The new data-rich environment may allow litigators to present a more credible *prima facie* case to a court when disparities are found than is the case with the current HMDA data, because the disparities would be estimated after controlling for a broader range of pricing and underwriting factors. That, in turn, could make it easier for class action lawyers to survive a motion to dismiss and to get a class certified. The data may also provide opportunities for more narrowly targeted claims based on such things as disparities in points and fees or origination charges; product-specific claims relating to equity loans, reverse mortgages, ARM products, or "non-traditional" products; and age discrimination claims. Similarly, there will be increased reputation risk based on advocacy group studies of the data.

What's a Banker to Do?

The foregoing considerations should shape how financial institutions analyze their own data. Mortgage lenders will need to be more diligent than ever in evaluating themselves for fair lending and reputation risk. Continuing to perform regular monitoring of underwriting, pricing and redlining risk will remain as important as ever, but will not be enough. Lenders will need to get creative and think about the potential new risk indicators that could be mined from the new HMDA data fields and monitor for those additional risks. If there are prohibited basis disparities in terms of the new HMDA data fields, it will be important to evaluate the reasons for those disparities and whether they are explainable in non-discriminatory terms. This means performing more extensive analysis, understanding the processes and policies that underlie the data, and evaluating whether sufficient controls are in place to limit fair lending risk.

Here are some Key Risk Indicators to consider:
- Are there pricing or fee differences among loan originators or branches that create fair lending risk?
- Are there prohibited basis disparities in subcomponents of pricing, particularly those that may be subject to discretionary adjustments?
- Are discretionary pricing concessions, lender credits, and fees sufficiently controlled, justified, and documented to avoid fair lending risk?
- Are there disparities in broker compensation, at either the portfolio level or for individual brokers?

- Are there unexplained disparities based on age or for specific race and ethnicity subcategories in any of the risk indicators?

- Does the distribution of home equity or reverse mortgage lending based on neighborhood racial or ethnic demographics differ from the distribution of your forward mortgage lending or that of peer lenders?

- Are there sufficient controls regarding which products are offered to a borrower, and are there differences in incentive compensation across product classes that may create steering risk?

Time is of the Essence

Finally, the velocity of fair lending monitoring may need to increase. The CFPB has indicated that it plans to start releasing the data to the public earlier than has traditionally been the case – as soon as a month or less after the annual March 1 data submission deadline, instead of late September of each year. In addition, large institutions will be required to report their data on a quarterly basis starting in 2020, and the quarterly data submissions can be expected to be released publicly on a similarly timely basis. These regulatory schedule changes will put a greater emphasis on timely internal monitoring to stay a step ahead of the regulatory examiners and others who might use the data.

About the Author

David Skanderson is a vice president in the Washington, D.C. office of Charles River Associates, an economic consulting firm. He spends his workdays pouring through data and estimating complex statistical models to help lenders assess, monitor and manage their fair lending risk, including the evaluation of credit scoring systems. He also serves as an expert in mortgage litigation matters. Dr. Skanderson previously led departments responsible for fair lending analysis, HMDA compliance, and compliance loan review at Washington Mutual Bank. He can be reached at dskanderson@crai.com. The views and opinion expressed herein are those of the author and do not reflect or represent the views of Charles River Associates or any of the organizations with which the author is affiliated.