

Balancing Access and Privacy: The National Mortgage Database Project (NMDB)

Robert B Avery

Association of Public Data Users 2016 Annual Conference

September 13, 2016

*The views expressed are those of the author and do not necessarily represent those of the Consumer Financial Protection Bureau, the Federal Housing Finance Agency, or the United States. NMDB data calculations are preliminary. Please do not cite without permission.

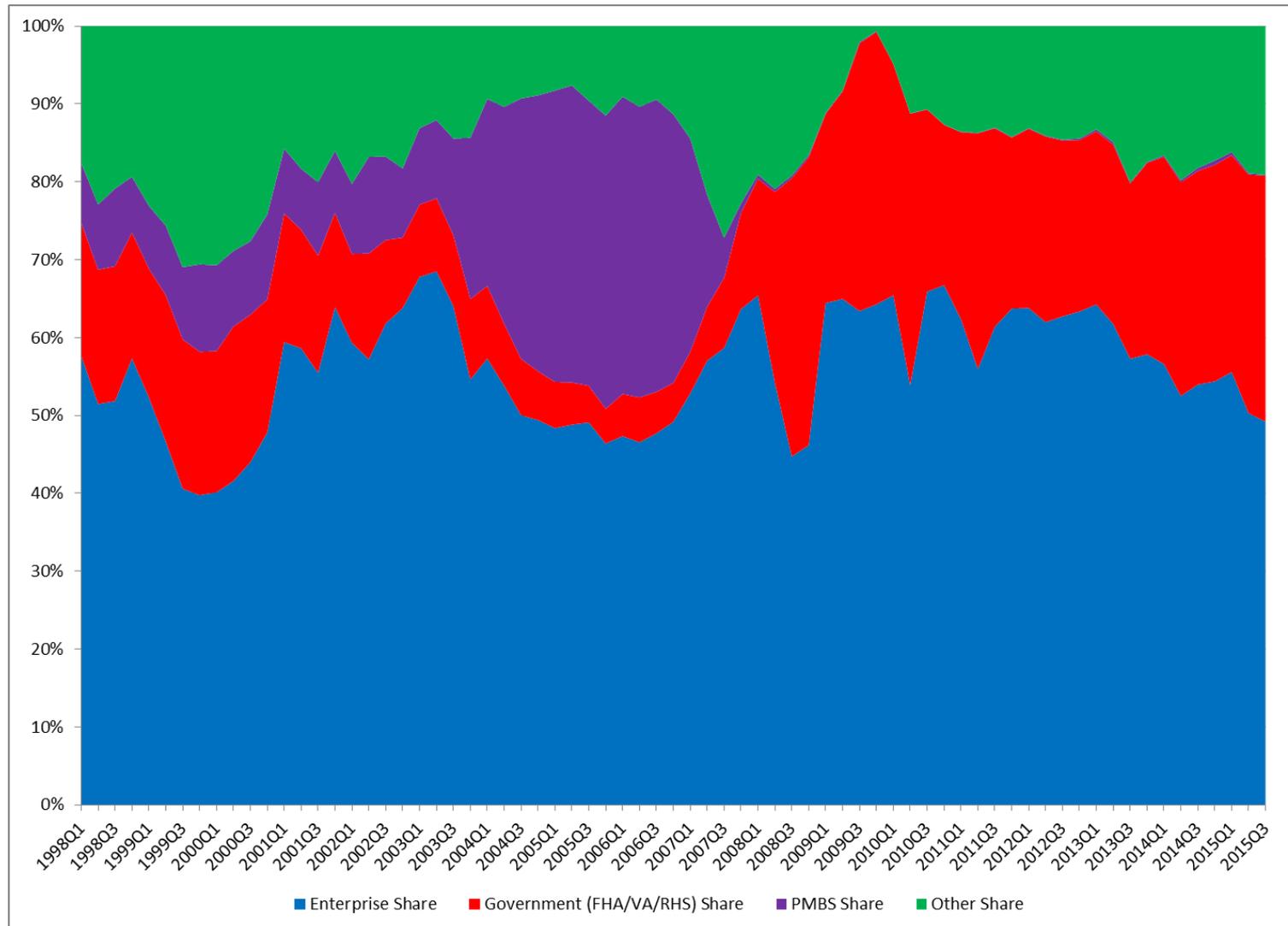
The Problem

- Under the 2008 HERA Act, FHFA is to collect data on “on the characteristics of individual mortgages, including... information on the creditworthiness of borrowers, including a determination of whether subprime and nontraditional borrowers would have qualified for prime lending.”
- Under Dodd-Frank, CFPB is “to monitor mortgage markets to gain an understanding how mortgage debt affects consumers and for retrospective rule review required by the statute.”
- The charges arose from the general agreement that the government (and private sector) lacked the information to see the financial crisis coming and to avert it.
- The Problem: how do you meet these directives?

The Problem

- Comprehensive datasets of the whole mortgage market (e.g. Home Mortgage Disclosure Act, Deed title records) focus on origination information and contain no data on subsequent performance.
- Servicer and administrative datasets contain performance information but are fragmented and disjoint (see the next slide).
 - Unlike credit cards mortgage servicers do little or no “account maintenance.” No updated information on the property backing the loan or other events going on in the credit life of borrowers.
 - Account maintenance is viewed as too expensive to justify given the low risk of failure in mortgage loans. Little effort made to “clean” data.
 - Hasn’t changed since the crisis.
- To monitor mortgage markets properly requires combining origination data with updated information not just on the mortgage loan performance but on the property value and other changes in the borrowers’ credit worthiness.
- So how to do it?

The Fragmented Mortgage Market



National Mortgage Database

The Solution

- Basic idea is to create a comprehensive database by combining information from different sources.
- The source for defining the universe of loans was selected to be mortgage loans reported to the credit bureaus. Why?
 - Mortgage coverage is almost universal. There is little cost to filing with a bureau for lenders and bureau data is generally cleaner than raw servicer files because of Fair Credit Reporting Act (FCRA).
 - The bureau frame is linked automatically to on-going performance data on the loan and on the borrowers' other credit obligations including home equity lines of credit (HELOC)s, second mortgage liens and credit cards.
 - The bureaus archive their data so it is possible to construct an historical database dating back at least 10 years.
 - The bureaus maintain very good information on borrower name, SSN, and mailing address which facilitates high quality matches to other data.
- But crucial data items are missing: No information on the property; limited data on borrower demographics; crucial mortgage characteristics including owner-occupancy, loan purpose and ARM status are missing. Type of mortgage is only imperfectly reported. Need to match other data sources.

- Select Experian as credit bureau partner. Define the universe of loans as closed-end 1st lien mortgages open at any time from January 1998 to the present.
- Instead of using all loans in the bureau, use a random 1-in-20 sample. This has the advantage of much less cleaning than a universe or registry, much better privacy protection, and almost the same statistical precision.
- Supplement the mortgage-based database with a separate overlapping database of all of the mortgage records of a 1-in-20 sample of borrowers with at least one active mortgage loan since 1998.
- Going forward, add a 1-in-20 representative sample of newly originated mortgages (and mortgagees entering the market for the first time) each quarter (currently through June 2016). Drop terminated mortgages from the sample.
- Jointly fund the project with the CFPB and FHFA and commit to a significant cleaning and processing effort with value added over the raw bureau data.
- Develop methods of obtaining variables not in the bureau files from other sources.

Adding Data from Loan Administrative Files

- Administrative loan file matching
 - » High quality matching using borrower information (e.g., borrower name, SSN, age, address) with 3rd party blind matching processes behind a firewall at Experian.
 - » Data are provided by Freddie Mac, Fannie Mae, FHA, VA, and RHS.
 - » Strict procedures are being followed to ensure that Fair Credit Reporting Act (FCRA) and Federal Privacy Act requirements are fully adhered to.
 - » All of the matching takes place at Experian. FHFA/CFPB never receive any sensitive borrower information (only de-identified data). Experian never receives linked mortgage information from data providers.
 - » New loans that are added to the database will be matched using the same procedures.
 - » It has taken over two years to work out agreements to do this.
- The challenges are the 20% of sample loans originated before 1998 (including 4 percent before 1990) and the estimated 9% that are non-owner-occupied. Experian maintains billing address for borrower whereas administrative records track property address. These can differ.

Adding Data from Property Administrative files

- » Public record property records and Multi-listing Realtor data for most US properties available from CoreLogic are being matched to NMDB loans using property address (available once an administrative file match has been made) or borrower billing address (from Experian files).
- » CoreLogic also maintains data on property and mortgage transactions which can be used to validate or improve the matches. Often transactions can be better matched to the billing address at Experian since property taxes are sometimes billed to the same place.
- » All matching takes place behind a firewall at Experian. This required CoreLogic to be willing to place their entire database behind the firewall and to regularly refresh it. Similar project with the US Census Bureau.
- » Our objective is to match loans to the property records twice—when the loan is originated and when it is terminated.
- » Problems with older loans (transaction files regularly collected only since the mid 2000's), “non-disclosure” states and counties where dollar amounts are suppressed, and rural areas with irregular collection.

Adding Data from other Sources

- Additional loan administrative file matching (also done behind a FCRA-compliant firewall at Experian).
 - » We estimate that about 14% of sample loans originated between 1998 and 2012 are in PLMBS loan-level data files obtained from CoreLogic and thus matchable to the NMDB using property address and/or account number.
 - » About 10% more loans are estimated to be available in the servicing database maintained by CoreLogic which will be merged using common matches to property address. Additional matching to Federal Home Loan Bank database.
 - » Since 2011, additional data from a national appraisal database covering all Fannie Mae/Freddie Mac loans is available and can be matched using property address behind the firewall. Starting in 2017 FHA loan appraisals can be added.
- Other data sources.
 - » Home Mortgage Disclosure Act (HMDA). We have obtained an 70% high quality match rate using fuzzy logic matching based on property census tract, loan origination date and amount. A HMDA match gives income and race.
 - » Household moving/address information on up to 25 last addresses and data on age, gender, and marital status are available from public records collected by Experian.

NMDB Data Progress Report: Where are we now?

- » The project was started in the fall of 2012 with an initial sample of 10.6 million loans active at any point from 1998 to 2012.
- » The loan database now contains of 11.4 million loans is current through June 2016. The database is updated quarterly and performance is updated monthly. Updating and cleaning is pretty mechanical at this point.
- » Fannie Mae/Freddie Mac administrative file matching is completed and quarterly updating has begun. FHA, VA, and RHS matching is taking place right now on the historic loan file.
- » Pilot testing of matching to CoreLogic property database has been completed. Production matching starting in September. Matches to HMDA back to 2004.
- » A preliminary version of a delinquency database was recently developed. The database consists of monthly loan performance, balance and payments, borrower age and credit scores, original loan balance and terms, and property location (county). This file will require testing and review.

NMSO and ASMB Consumer Surveys

- The NMDB Project also includes two regular consumer surveys of mortgagees sampled from the NMDB borrower database.

- National Survey of Mortgage Originations (NSMO)
 - » Starting in 2013 we have sent approximately 6,000 new mortgagees a mail survey each quarter. The respondents are drawn randomly from borrowers associated with mortgages newly reported to the NMDB database and tracked subsequently as part of the NMDB. This represents a sampling rate of approximately 1 out of 260 new mortgagees.
 - » The survey solicits information on mortgage shopping, consumer knowledge, expectations, demographics, life events and satisfaction. At least once a year a report will be prepared summarizing interesting results from the survey and disseminated publicly on the agencies' web sites.
 - » To date 71,500 surveys have been mailed out with a response rate of about 35 percent. Reports have been prepared on 2013 and 2014 lending. A cleaned dataset of the approximately 14,300 usable respondents from these initial waves is ready for research use. The dataset combines survey responses with credit data from the NMDB.

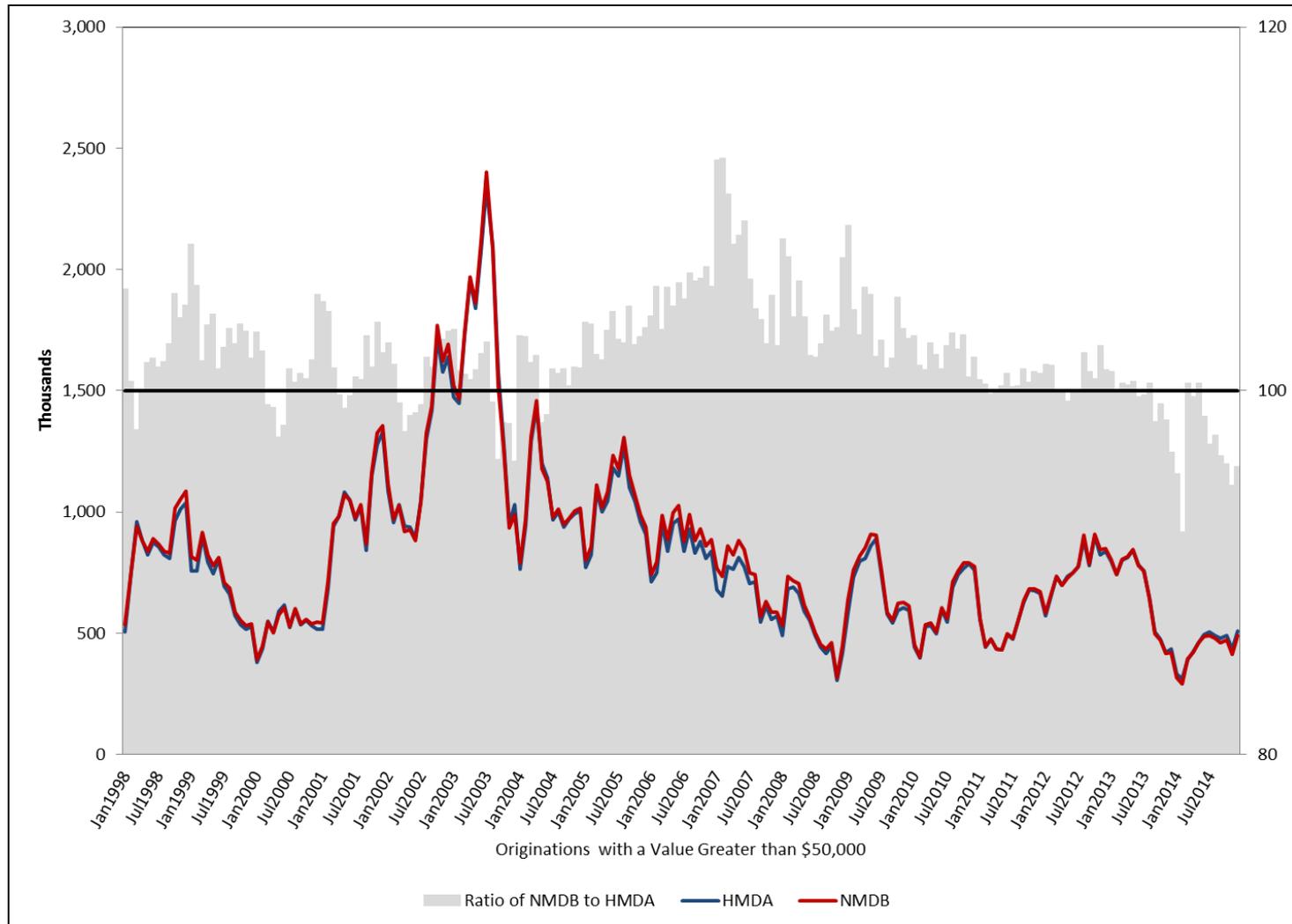
- American Survey of Mortgage Borrowers (ASMB)
 - » A second survey of 10,000 ongoing mortgagees who are disproportionately "in distress" was mailed out in early August. This survey will allow analysis of issues regarding servicer contact for borrowers who fall behind on their payments. The survey data will be ready for use next spring.

Who will Use the NMDB?

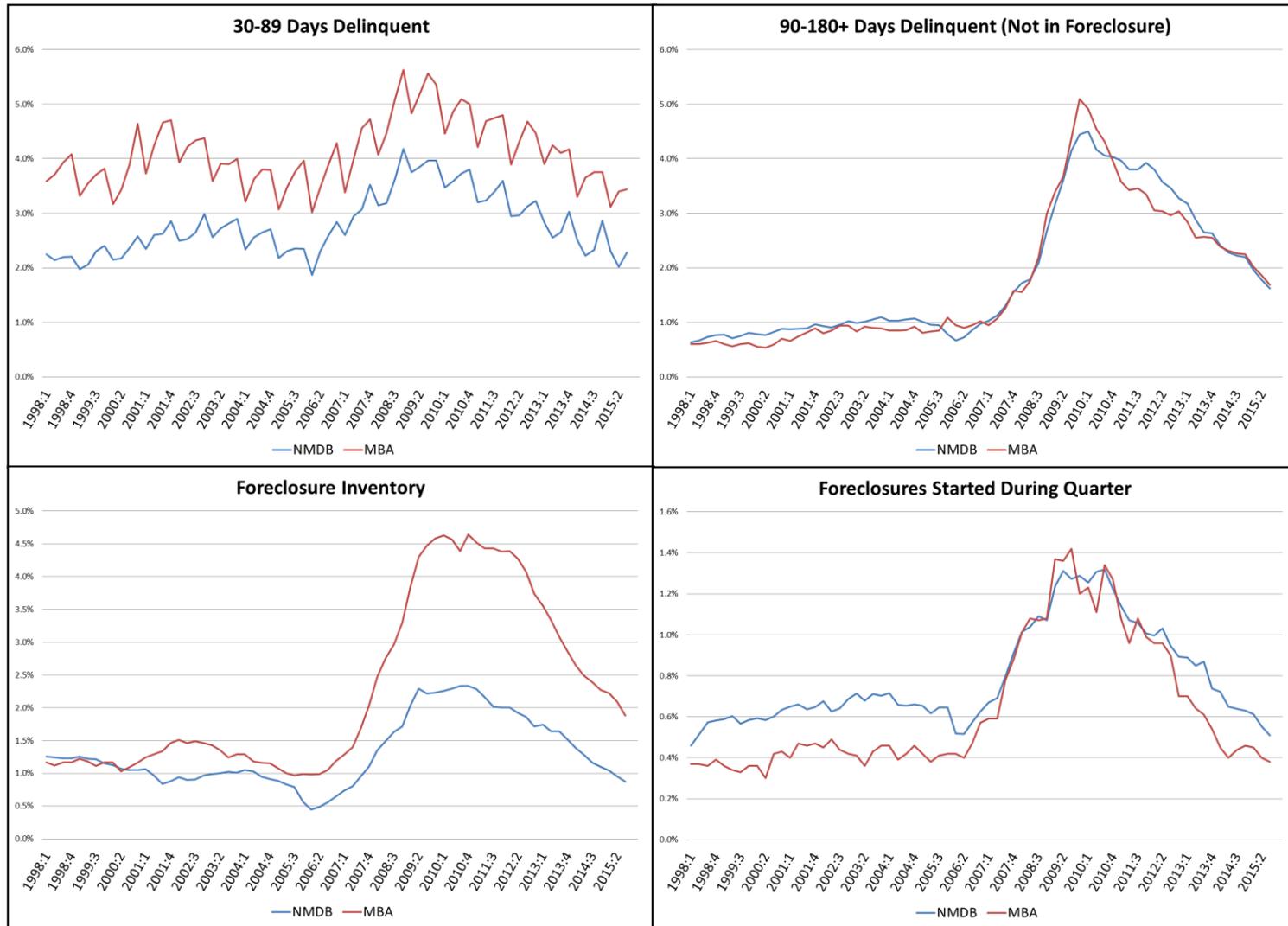
- NMDB Client base is Federal Employees.
 - » By contract with Experian the loan-level NMDB database can be made available to Federal Employees if they access it on a FHFA or CFPB server.
 - » Strict controls on removing data and all users must sign a terms of use agreement agreeing to not attempt to identify the borrower. All data delivered to FHFA/CFPB is de-identified and contains no sensitive variables such as name, address, date of birth or Social Security number. Users will not know servicer and, by contract, the dataset cannot be used for enforcement against servicers.
 - » Besides researchers, main clients are expected to be regulators in supervision responsible for tracking the mortgage market or large programs including FHA, VA, and RHS.

- However NMDB is a prototype of the kind of database that could be used to monitor a mortgage pool or portfolio.
 - » Dodd-Frank calls for mortgage pool trustees to provide information to investors which is “material.”
 - » There have been serious problems agreeing on appropriate disclosures because of concerns that they might violate the privacy act.
 - » NMDB-type database combining administrative data, updated information on borrowers credit circumstances and updated valuations of the property value in a de-identified database created behind a FCRA firewall might be the solution.

NMDB vs HMDA, by Month of Origination 1998-2014



Preliminary Evidence from the NMDB



Preliminary Evidence: Credit Scores of New Originations

