# The National Mortgage Database (NMDB)

Robert Avery
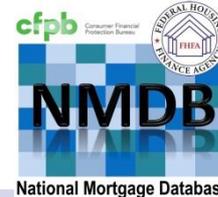
Federal Housing Finance Agency
Consumer Financial Protection Bureau

APDU Conference

September 1, 2015

# The Problem and the Challenge

- The Problem: Create a *nationally representative,* loan-level 1$^{st}$ lien mortgage database dating back before the financial crisis.
    - » Database must contain information on the mortgage characteristics (e.g. loan size, underwriting quality, price, investor vs owner-occupied, features etc.)
    - » Database must contain information mortgage performance (e.g. payments, delinquency, foreclosure, payoffs and termination).
    - » Database must contain information on the borrowers associated with the loan (e.g demographics, other credit obligations, other mortgages and shopping behavior).
    - » Database must contain information on the property collateralizing the mortgage (e.g. property value, other mortgages (2nd liens), location, neighborhood characteristics).
    - » Database should ideally involve no new data collection--make use of data that already exists.
    - » Database should be able to be constructed retroactively to at least early 2000's and be updated going forward with new mortgages added.
- The Challenge: No existing database meets all of these criteria or even comes close

# Why Can't Existing Databases be Used?

- HMDA:
  - Pretty close to a universe of all 1st lien mortgage originations.
  - However, has limited borrower, property, and loan information.
  - Critical flaw is that it has no performance data and lacks many underwriting variables (e.g., credit score, loan-to-value, ARM-type).

- LPS McDash and/or CoreLogic:
  - Servicing files from 25-30 large servicers versus 2,000 servicers in NMDB.
  - Not representative—no coverage of smaller servicers. Trade line not mortgage based.
  - No information on other obligations—particularly 2nd liens--previous or subsequent mortgages. Little information on borrowers

- NY Fed Equifax:
  - Similar source as NMDB, but unit of observation is borrowers, not loans. Trade line not mortgage based. Data is not cleaned.
  - Little or no supplementation with other data. Difficult to link files over time.

- GSE/FHA/VA Administrative data combined with PLMBS data
  - Dataset has no coverage of portfolio loans.
  - No information on other obligations—particularly 2nd liens--previous or subsequent mortgages. Limited information on borrowers and property characteristics.

# Why Can't Existing Databases be Used?



National Mortgage Database
Preliminary Numbers
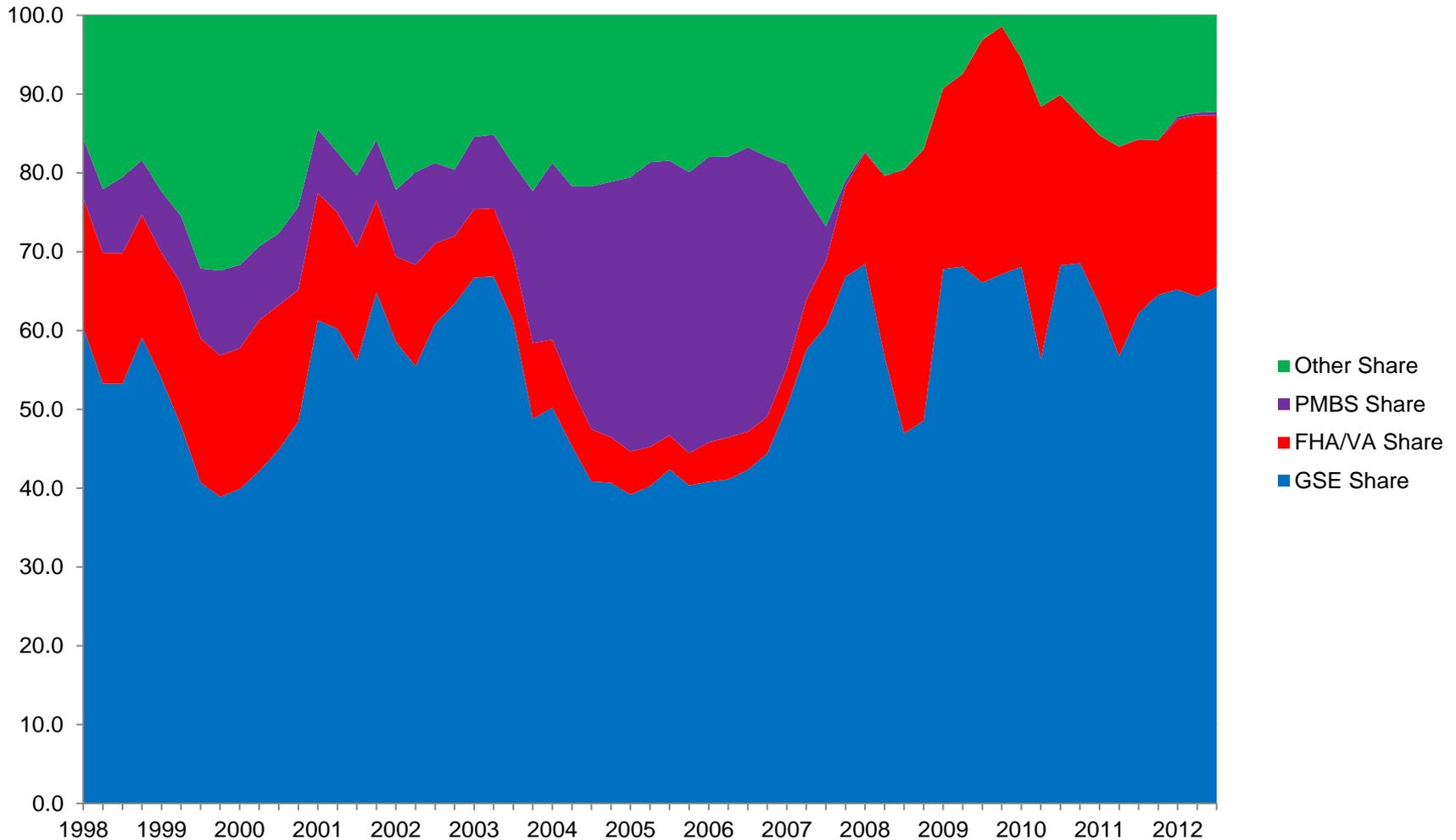
# Solution: National Mortgage Database
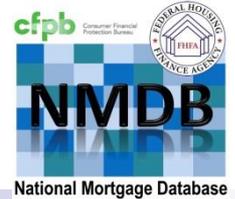
- A *nationally representative,* loan-level mortgage database jointly funded and managed by the FHFA and CFPB based on a prototype developed by Freddie Mac.
  - » 1st lien mortgages reported to the credit bureaus are used as both the sampling frame and the source of performance data.  Experian serves as the credit bureau partner.
  - » The primary database is a 1/20 sample.  This has the advantage of much less cleaning than a universe or registry, much better privacy protection, and almost the same statistical precision.
  - » Because the credit bureaus archive their data, the NMDB recovers data that would have been available had the project been started years ago.
  - » The initial 1/20 sample is representative of all mortgages open *at any time* from January 1998 to June 2012 and *any* borrower who had at least one mortgage during that period.
  - » Going forward, a 1/20 representative sample of newly originated mortgages is being added each quarter (currently through June 2015) and terminated mortgages exit the ample.
- Credit bureau data are comprehensive.  However, they are raw servicing data which require significant cleaning and **matching to additional data sources** to make them fully useful.
  - » Much of the cleaning is being done by FHFA staff.  Requires converting "trade line" data (record for one loan by one servicer) into "mortgages."  Approximately 20% have 2 or more trade lines.  Also have to deal with "person-merges" and integrating archived data over time.
  - » Working with the active cooperation of Experian staff.
- NMDB also has a survey component.  Starting with originations in 2013, each quarter a subset of 6,000 borrowers associated with loans newly added to the database has being sent a mail survey soliciting information on their mortgage shopping and origination experience.  Response rate 35%.

# How are Data Added?

- 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 were used to merge administrative data into the NMDB. 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. Fannie Mae/Freddie Mac completed, others in process.
  - » New loans that are added to the database will be matched using the same procedures.
- Property information (also done behind a FCRA-protected firewall at Experian)
  - » Public record property records and MLS datasets for most US properties are available from CoreLogic, is being matched to NMDB loans using property address (available once an administrative file match has been made) or borrower address (from credit bureau files). CoreLogic also maintains data on property and mortgage transactions which can be used to validate the matches. Matching will take at the completion of administrative file matching.
  - » Since 2011 additional data from a national appraisal database covering virtually all Fannie Mae/Freddie Mac loans is available and will be matched using similar methods. Starting in 2016 FHA loan appraisals will be added as well.
- Additional loan administrative file matching (also done behind a FCRA-protected firewall at Experian)
  - » We estimate that about 14% of sample loans originated between 1998 and 2012 are in private label MBS loan-level data files obtained from CoreLogic and thus matchable to the NMDB using property address and/or account number.
  - » About 10 percent more loans are estimated to be in available in the third party servicing database maintained by CoreLogic which will be merged using common matches to property address.
- Other data sources.
  - » Home Mortgage Disclosure Act (HMDA). We have obtained an 80% high quality match rate using 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 the credit bureau.
- The challenges will be the 20% of sample loans originated before 1998 (including 4 percent before 1990); the estimated 9% that are non-owner-occupied; and 10% that are under $50,000. Each of these loan types are more difficult to match to external records.

# Access and the Future
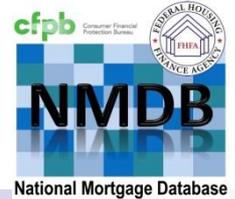
- NMDB is being set up as a public good. Project started in late 2012 and production database will take another 9 months to complete.

- The challenge is to (1) protect borrower/lender personally identifiable information and (2) provide users access to useful data. Key points of conflict on these objectives are items such as: property location, loan amount and timing of origination which are critical for mortgage analysis but potentially could be used to identify specific borrowers.

- Our solution:
    - NMDB loan-level data is physically housed *only* on a FHFA/CFPB server. No directly identifying borrower information and no geography below census tract is contained in the dataset. No datasets containing sensitive borrower information PII are allowed on the server.
    - Access to the loan-level data is potentially allowed for any federal government/Reserve Bank/Fannie Mae/Freddie Mac employee going through the following access process:
        » Must sign an agreement not to attempt to identify either the borrower or lender. Severe penalties for violations of agreement.
        » All work behind a firewall—data can't be removed. Users will be able to request that data be added to the servicer but follow strict protocols to do so.
        » Data cannot be used for "enforcement." Users will not have the identity of the servicer.
    - Access to aggregated NMDB data not representing privacy risk will be available to the general public through something like "American Fact Finder."