

Improving Research and Data For Policymaking

Robert B. Avery
Federal Reserve Board of Governors

December 3, 2010
The New Landscape for Consumer Credit and Payments
Federal Reserve Bank of Philadelphia

The opinions expressed are mine and do not necessarily reflect the opinions of the Board of Governors or the Staff of the Federal Reserve System.

Use of Institutional Loan Data



The Federal Reserve Board

- Financial Institutions maintain extensive electronic files of individual loan data---C&I, CRE, residential mortgages, consumer loans.
- 3 major credit bureaus contain virtual universe of every mortgage and consumer loan.
- Yet these data sources have been seriously underused in research, policy making and regulatory oversight.
- Why?

General Problems



- Data maintained for different purpose than research
 - Typically loan servicing files
- Designed for current servicing status of loan or portfolio
 - Difficult to link to application data or other information on borrower (e.g. loans, deposits).
 - Often missing data on disposition—loans just disappear from active files, e.g. transferred to collection/workout/REO
 - Can't tell why a loan became bad/good.
- Hard to follow “life” of individual loans.
 - Time series information only from repeated snapshots.
- Even within an institution, loans often maintained in different loan systems—hard to aggregate.

“Make Do” Fixes



- Approximate time series with snapshots at two different points in time
 - FICO and VantageScore model estimates from this type of data
 - “New” accounts or “newly bad” accounts are those appearing in 2nd point of time but not the 1st
 - Fraught with error as accounts are relabeled, transferred to other lenders/collection or reported with a lag (e.g. student loans).
- Expensive, manual construction of datasets
 - RMA/FICO data in mid 1990’s—17 banks with 300 loans each.
- Use of substitute variables because they are easier to measure—delinquency instead of default (PD not LGD)
- Extensive use of convenience samples—little concern about representativeness.

We Can Do Better



- Use of Samples not Universe
 - Makes cleaning the data more feasible
 - Less expensive to match to other data files
- Targeted sampling with an integration of electronic and manual data methods
 - FRB Loan underwriting study
 - Matched pairs of loans underwritten in two different years. Then used examiners to test subtle changes in underwriting.
 - Fair Lending analysis makes extensive use of this.
- Make better use of Credit Bureaus

Pros and Cons of CB Data



- Advantages

- Already cleaned and standardized. Relatively cheap.
- Can look at either consumer or account as the unit of analysis.
- Contains comprehensive information on all consumer obligations across different account types and lenders.
- Historical data goes back at least 7 years.

- Weaknesses

- Designed to measure credit risk not profit—little pricing data or data on account usage.
- Data *very* difficult to use—not user friendly. No manual.
- Significant duplication of accounts because of sale of servicing and transfers—needs cleaning. Need rules to handle joint accounts appearing in files of 2 different people
- Critical loan features (e.g. refi or home purchase, LTV, ARM, payment amount for revolvers) not reported.

Major New CB Data Initiatives



- Argus/TU database
 - Combines transaction data for credit cards with credit bureau data on other obligations
 - OCC version collected from national bank issuers
- Federal Reserve TU panel
 - Panel data of 300,000 representative individuals followed every 18 months since 2003. Demographic match to Social Security. Complete raw tradeline data. TU and VantageScores.
- FRB NY consumer credit panel database
 - 1 in 20 quarterly sample going back to 2000. Rollups except for mortgage tradelines. Quarterly credit score.

Major New CB Data Initiatives



- National Mortgage Database (NMDB)
 - Pilot project of Freddie Mac and the FRB initiated last Fall.
 - Designed to produce a Federally-funded, publicly-available representative sample of mortgages on an ongoing basis.
 - Working with a topflight team from Experian (CB) and an advisory group drawn from government, academe and non profits
 - Most of the pilot effort is being put into due diligence to validate the sampling frame and representativeness and accuracy of the data.
 - Scaled-down version of the database is already up and running. Should have a full proposal for production available in the Spring of 2011.

NMDB Overview



- Representative 1/100 sample of 1st lien closed-end mortgages outstanding on or after December 31 2003 selected from credit bureaus. Would select 1/100 new additions each month ongoing
- For **each** mortgage, collect monthly data from credit bureaus on the mortgage (terms and payment performance) and the mortgage co-signers (credit scores and other credit obligations such as 2nd liens and credit cards) over the life of the mortgage.
- Information is also collected from a **survey** of borrowers associated with each mortgage (borrower knowledge, shopping behavior, demographics, key life events, expectations).
- Supplemental match to (1) HMDA, (2) deed/title records, (3) MERS database on property location and GSE ownership, (4) AVM estimates of property value, (5) postal information on borrower location and moves, and (6) demographic information on borrowers

NMDB Pilot – Lessons learned



- Bureau Data needs cleaning
 - significant (10%) duplication of mortgages
 - 2nd liens need to be culled
 - Identifying home purchase vs. refi and owner-occupied versus non-owner occupied is not straightforward
 - Matched data (HMDA/Deed title records) only available for a portion of the sample—need to impute values for the remainder
 - Massive amount of data even with 1/100 sample—summary measures will need to be developed to make the data usable for researchers. Some data (e.g. geography) will be redacted
- Only 16% response rate on the Survey. Need to improve presentation to raise level. Content of surveys got a good response.

NMDB—More Lessons



- Property value, LTV, purchase value will probably have to come mainly from AVM models, as external match for mortgages too low
- It takes up to 6 months for servicers to report new mortgages—means unadjusted delinquency rates rise in a refi boom. Need to use termination data to estimate bias and to track market.
- Data on what happens to mortgages in foreclosure is incomplete. Information on moving likely to help recognize completed foreclosures. Loan modifications are hard to identify.
- Almost complete coverage of loans known to be owned by Freddie Mac. Some inconsistency on payments. Loan amounts and origination dates very accurate
- HMDA match higher than expected (60%). Provides only source on income and race. Confirms refi/home purchase, owner occupied.
- APR, origination points and fees, loan channel (broker) only available from survey—limited use

Coverage of CB dominates others



2008	Bureau	HMDA/Bur	HMDA	McDash
Jan	717,503	387,505	498,489	322,603
Feb	898,050	520,768	690,698	421,702
Mar	907,359	531,228	698,828	405,234
April	929,858	534,109	679,959	378,436
May	811,107	454,337	605,325	329,785
June	758,556	432,313	564,903	305,617
July	727,312	421,072	506,753	269,103
August	645,527	342,947	457,998	242,359
Sept	637,432	346,608	432,561	243,801
Oct	647,940	369,187	457,786	244,724
Nov	429,037	238,502	315,391	169,813
Dec	581,860	325,186	424,641	253,894
Total	8,691,540	4,903,761	6,333,332	3,587,071
		56.4	72.9	41.3

The Future?



- Credit Bureau and Argus-type data are critical for off-site regulatory safety and soundness and consumer compliance monitoring. No good substitutes.
- Available for a fraction of the cost of monies already spent on alternatives.
- Logistical and data cleaning problems are solvable but will require commitment of resources to make the data usable.