RMBS Data, Analytics Firms Enter Era of Innovation

by Larry Barnett

While RMBS markets have limped along over the past several years, the industry has seen dramatic innovation in the area of mortgage data and analytics. Due to improved data disclosure and the entry of new data distributors with enhanced data aggregation and cleansing techniques, investors, dealers, academics and others now have access to loan-level datasets and analytics tools that offer unprecedented clarity into private-label RMBS and improved risk management practices. That innovation is spreading to new areas, including agency RMBS and foreign ABS deals.

Key areas of innovation include access to premium loan-level data that is more timely and provides a significant new level of depth; the integration of loan data with third-party datasets; and the emergence of new tools that broaden access to sophisticated analytics.

The developments are primarily driven by an increased demand for loan-level data to enhance predictive analytics used to price securities and improve credit surveillance on the underlying assets backing securitized trusts.

Regulations and market demand have helped propel that innovation. Changes to the SEC's Regulation AB and other rules encouraged loan-level data providers, such as issuers and trustees, to offer investors improved access to loan-level data covering mortgage payment histories and attributes that describe loan originations. Creative vendors are now in a much better position to enhance the data sets by integrating loan information with complimentary data sources and using logic to remediate the data or fill in data gaps. A great example of this is the ability to generate information on loan modifications by comparing loan attributes from one period to the next.

Market participants and regulators alike are calling for better loan-level clarity. Data providers are delivering, as knowledgeable vendors innovate to add value to the atomic data being supplied.

Access to premium loan-level data

Loan-level datasets form the basis of any investor or broker's analytical framework. The quality of models they rely on to make buying and trading decisions is directly tied to the quality and depth of available loan-level data. During the height of the mortgage market in the mid-2000s, there were very few options for market participants who wanted comprehensive loan-level data. The options that did exist were expensive and required programming expertise to use, leaving many to rely solely on summary reports and rating agency reports.

Regulations including changes to the Security and Exchange Commission's Regulation AB allowed for much wider access to data, leading to significant innovation. The result has been more comprehensive and timely datasets that lead to improved analytic workflows and better information on non-scheduled cash flow predictions and bond valuations – which, in turn, lead to better buying and trading decisions. Small firms are enjoying the low cost of entry for premium loan-level data, and flexible tools eliminate the need for significant IT investments. Larger firms with in-house research teams are finding a new level of flexibility in data and customization that did not exist at the peak of the market and are benefitting from more comprehensive data.

Using new datasets, investors now can monitor past loan performance on a wider variety of attributes, including current views of loan attributes that have changed over time. Traders can more accurately and easily calculate and predict collateral and bond-level performance.

Importantly, the market is also benefitting from innovations in the reporting of payment data, a critical component to risk management and buying decisions. New algorithms allow investors to identify modifications even when they are not reported by issuers, servicers or trustees. These modifications traditionally have been difficult for investors to detect and often lead to deceiving results. Other advancements have corrected issues that have long plagued users of loan-level data, including timing issues around loss reporting and reasons for loan pay-offs, resulting in better data for modeling.

Third-party data integration

Partnerships with third-party data providers are producing richer data sets that tell investors much more about the underlying collateral of deals.

Companies such as Equifax can supply an anonymous view of updated credit scores, combined with updated views on a borrower's overall debt burden, to help analysts assess changes to a borrower's debt-to-income ratio, which provides up-to-date insights into the borrower's ability to repay a loan. Traditional analysis that compares attributes that describe the loan at origination, such as the original credit score, can therefore be replaced by producing a time series view of the borrower's credit profile. This enables the investor to more accurately determine a borrower's propensity and ability to repay mortgage debt.

Another important variable, used to determine both credit stability as well as likely loss given default (i.e. loss severity), is the loan to value ratio (LTV). Investors now understand the importance of the correlation between the amount of equity a borrower has in his home and future repayments. Loan-level data can now also be married to property and title databases that provide a time series analysis on the properties that collateralize the mortgage loans. This data can be used to accurately determine the value of the property and combined with the amount still owed on the property to calculate a current view of the borrower's LTV. Companies including DataQuick and Interthinx utilize sources such as county records to inform investors on attributes such as the current state of a property, additions made to the house and the values of recent appraisals to assist in this determination.

The market has just scratched the surface of what is possible in this arena, and more innovations in the integration with outside datasets are currently underway.

New analytics capabilities

In addition to providing deeper, more comprehensive datasets, data providers are packaging the data in new ways that allow for easier integration with analytic workflows.

Web-based software allows users with no programming experience to manipulate data in advanced ways, stratifying loans into nearly limitless categories. With a few clicks of the mouse, a portfolio manager, broker or researcher can now compare loans across different deals, vintages and a host of other attributes by hundreds of characteristics – something that was impossible to do just two years ago. For example, users of these new interfaces can generate views of historical performance by running queries on outstanding sub-prime deals in a given vintage year; the prepayment speeds of alt-A deals issued through individual shelves; and views of modification recidivism and the cumulative losses of all prime trusts in a given month. Formatting the output using graphical interfaces provides investors with easy-to-use comparisons that are essential in the decision-making process.

Analytic-ready loan-level data allows for much easier integration with downstream third-party vendors and in-house systems that analysts use to drive their default and prepayment models. The predictive vectors that document the unscheduled cash flows generated from the models can be easily integrated into third-party bond waterfall models to predict cash flows at the trust, bond and group levels that are then used to create bond valuations. Using this approach, deals can be compared using tools once limited to institutions with multi-million dollar analytics budgets.

Vendors such as Thetica offer infrastructure to integrate vector-producing models with cash flow waterfalls and comprehensive loan-level datasets. Third-party valuation software from providers such as 5 Bridges and Cooperstein Analytics utilize a self-contained set of applications to generate bond valuations for their clients.

The next frontier: Agency data, foreign markets

Innovation of data and analytics on the conforming side has been much slower than for private-sector RMBS, mostly because of the lack of availability of comprehensive loan-level data both historically and on newly issued deals backed by government agencies and government-sponsored enterprises (GSEs).

There is high demand for this innovation, as prepayment speeds have become at least as much a function of borrower credit profiles as they are a function of traditional prepayment incentives. Data and analytics providers are partnering to meet this demand.

Each agency has a different data disclosure strategy, and none currently discloses comprehensive loss information, which would provide a rich data source for the analysis of mortgage markets and provide insights into the relative health of the insuring institutions. Freddie Mac does provide a fairly comprehensive loan-level data set that can used to calculate delinquencies at the loan-level. The rich data set can be integrated into prepayment models allowing investors a newly available comparison between loans originated in the conforming and non-conforming markets. Fannie and Ginnie Mae do not currently provide loan-level data.

Black Box Logic is working with Five Bridges Advisors, LLC, to produce a standardized loan-level file covering all newly issued RMBS deals backed by Fannie Mae, Freddie Mac and Ginnie Mae, including the HUD loans. Data schematics were designed to represent loan values based on summary reports where no loan-level data was made available. The result is a monthly report updating 39 loan-level attributes for all loans backed by all three agencies covering nearly a half million new loans per month across thousands of new deals – a major breakthrough in the area of loan-level data for conforming markets.

Innovation is also happening in foreign markets, using breakthroughs in U.S. non-agency markets as a jumping-off point. BlackBox Logic is currently working with a partner to supply loan-level data for selected deals issued in the United Kingdom. Due to the success of the data prototype we are expanding coverage in that market and are studying ways of moving into other ABS deals throughout the rest of Europe.

Data and analytics companies will continue to produce new solutions that offer clarity to RMBS markets here and abroad, positioning the industry for better risk management and buying decisions on RMBS securities outstanding while preparing for the acute need for standard loan-level data when issuance volume returns.

*Larry Barnett is CEO of **BlackBox Logic**, a Denver-based provider of loan-level data for residential mortgage market participants. The company's core offering is BBx Data™, a database of more than 21 million non-agency residential loans covering more than 7,200 deals. Mr. Barnett can be reached at lbarnett@bbxlogic.com.*