3/17/2011

The Next Frontier of Loan-Level Data Advancements

Intro: The progress made in the private secondary markets when it comes to loan-level data and deal-performance predictors could be coming to the conforming and government-loan MBS markets. And it could come sooner than you think.

By Larry Barnett

While new residential mortgage-backed securities (RMBS) issues have remained at a virtual standstill following the 2008 credit crisis, companies have remained busy innovating in ways that critically matter to investors. This innovation has been directed at providing more complete, accurate and timely views of deals and the loans that underlie them, as well as offering advanced analytical tools to better predict deal performance.

So far, these advancements have largely focused on the market for non-agency conforming deals. Agency data are the new frontier.

Investors eager for better loan-level data on deals backed by government agencies and government-sponsored entities (GSEs) can expect to find an incredible amount of innovation in the next year, particularly as agencies follow Freddie Mac's lead in releasing comprehensive loan-level data.

There is high demand for this innovation. Loans insured by government agencies are under intense scrutiny, if for no other reason than loans backed by Fannie Mae, Freddie Mac and Ginnie Mae currently dominate the landscape for RMBS and likely will for the foreseeable future.

Why do investors and others care so much about loan-level data? For one, agency prepayment speeds are now significantly affected by loan credit characteristics and borrower behavior--two characteristics that are currently much easier to analyze in nonconforming deals because of access to comprehensive loan-level data. Investors, researchers and dealers in mortgage securities backed by conforming, conventional loans are calling for similar transparency for conforming loans.

To get an understanding of the types of developments to look for in the conforming markets, it is helpful to examine recent innovations in loan-level data for the nonconforming space.

Increased demand and new regulations have sparked innovation in the collection and delivery of loan-level data in nonconforming markets. Investors, academics, professional modelers and dealers seeking to reduce reliance on rating agencies, summary reports and other more simplistic indicators heavily used in the run-up to the credit crisis have demanded better data at the same time that regulations enabled wider access to mortgage data, including payment and loan setup data.

Enhancements to the Securities and Exchange Commission's (SEC's) Regulation AB, in addition to other changes, opened access for more loan-level data providers to offer new alternatives to traditional data providers. This general trend is expected to continue, as proposals in the Dodd-Frank Wall Street Reform and Consumer Protection Act push forward even more transparency in loan-level data. Vendors across the RMBS markets have embraced new data providers and helped develop some incredibly useful tools.

A key area of innovation in the nonconforming sector includes the accessibility of premium loan-level data integrated with third-party data sets, transforming traditionally static data into dynamic data that give the users time series views of key attributes. For example, users of some data sets can now access a time series of FICO scores for each borrower. Another examples is the ability to look at a time series of a comprehensive combined loan-to-value (CLTV), including all debt of the borrower (i.e. credit card, auto, additional liens, etc.)

These variables are then integrated into predictive modeling of unscheduled cash flows, bond-valuation routines and credit-surveillance applications that are more broadly available following innovations by companies whose services are largely available through an Internet connection.

Market factors such as price changes on real estate across the country, additional borrower leverage caused by second-lien loans taken out after the issuance of the first mortgage, and changes in borrowers' credit quality caused by changes in personal income and/or macro-economic factors can now be accounted for by integrating comprehensive nonconforming RMBS data sets with complementary data offered by property and title data companies and consumer credit-rating agencies.

This has led to the availability of enhanced versions of two of the key analytic attributes used in mortgage credit analysis: loan-to-value (LTV) ratios, which compare current property value with the current amount owed by the borrower; and borrower credit scores, which help determine a borrower's propensity and ability to repay a loan.

The marriage of comprehensive nonconforming loan and deal data with property and title databases supplied by companies such as Data Quick Information Systems, Dan Diego; Interthinx, Agoura Hills, California; FNC Inc., Oxford, Mississippi; and others provides users up-to-date perspectives on important information. Such information includes the quality of a home, additional liens taken out on the property since mortgage origination, and additional attributes recorded with county recorders. These data sets lead to current and accurate time-series views of LTV data tied to each individual property backing each loan in a trust.

The integration with consumer credit-rating agencies such as Equifax Inc., Atlanta, provide users of nonconforming RMBS data with views of the changes to a borrower's economic circumstances that can lead to changes in the borrower's propensity to default, such as an unexpected partial or complete loss of family income. These factors can be determined based on anonymous views of the borrower's credit scores supplied by the consumer credit-rating agencies. Anonymous, in this instance, refers to the fact that the loan-level data and the consumer credit data were matched without releasing any personal information. This allows the user of the data to know how a particular consumer's credit situation tied to a particular loan is changing over time without having any information that allows the user of the data to identify the consumer. For example, if a house was originally reported as owner-occupied, the consumer credit rating agency can determine if the owner remained in the home by verifying the payee address. The credit bureau would not provide the data user any personally identifiable information**.**

Other attributes can help users compare the data supplied by the bond trustee or mortgage issuer with data supplied to the credit-rating agency that provide insight to possible fraud at mortgage origination**.** For example, if the FICO score reported by a credit agency differed significantly from what is reported in loan documents, this can be a red flag that one of the numbers was reported incorrectly. If the occupancy type in the trustee data indicated owner-occupied and the rating agency indicated investor property, that could also be a red flag for possible fraud.

More accurate, timely and comprehensive loan-level data describing nonconforming RMBS deals has increased transparency into deals, bonds and the underlying loans, and enabled investors and researchers to more accurately model cash flows and therefore value individual securities and their underlying assets.

One area where loan-level data provides significant benefit is in the area of delinquencies and loan modifications. Typically, when a borrower receives a loan modification, the borrower is immediately reported as current on its monthly payments, even though many payments may have been simply forgiven. Research has shown that the majority of borrowers in this category will eventually re-default, usually within twelve months of the modification. Because of this industry convention, high rates of loan modifications can make reported delinquency rates appear as if they are improving when they are not. When analyzing more granular data, it becomes clear in many cases that this is just a temporary phenomenon. More than 20 percent of loans in the subprime market fall into this category, where they are considered current on payments solely due to a loan modification. On a deal-by-deal basis, many deals have close to half of their loans in this category. If relying only on reported aggregate delinquency rates, this phenomenon remains completely hidden.

Less innovation on conforming loan side

Innovation on the conforming loan side has been slower to come by, mostly because of the lack of availability of loan-level data on newly issued deals backed by government agencies and GSEs.

Each agency has a different data-disclosure strategy (see sidebar), and none reports comprehensive loss information. Loss information would provide a rich new data source for the analysis of mortgage markets and provide insights into the relative health of the GSE's and other institutions that provide public funds to cover non-payments by borrowers due to mortgage defaults. For example, HUD provides mortgage set-up data and payment histories on mortgages insured by the FHA and the VA but don't yet provide loss severities on properties that are covered by their mortgage insurance programs.

Loan-level data currently are available on loans backed by Freddie Mac and the Department of Housing and Urban Development (HUD), which supplies data on loans insured by the Federal Housing Administration (FHA) and guaranteed by the Department of Veterans Affairs (VA). Data released by HUD describes loans backing Ginnie Mae pools but that data cannot be linked to RMBS deals, making it of limited value to trading analysis.

It appears likely, based on conversations with industry experts, that Fannie Mae will provide loan-level data in the near future In the meantime, the market is able to utilize the 18 summary files provided by the agency to generate a statistical representation of loans in RMBS backed by Fannie Mae to model future pool behaviors.

Unfortunately there is little standardization of the data that are made available, so there is much work to be done by the data providers to normalize the myriad of formats that describe conforming pools and the underlying collateral.

The industry is innovating to overcome these significant obstacles. BlackBox Logic LLC, recently teamed with mortgage adviser and analytics firm Five Bridges Advisors LLC, based in Bethesda, Md., 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.

Because of the limitations of data, BlackBox developed data schematics to represent loan values based on summary reports where no loan-level data were made available. For Fannie Mae loans, this was accomplished through 17 summary (quartile) files that describe various views of the agency new pool issuances. For Ginnie Mae loans, data were based on HUD loan files that cannot be cross-referenced back to new pools but can be used for predictive modeling of Ginnie Mae new issuances in general.

The result is a monthly report updating 39 loan-level attributes for all loans backed by all three agencies, covering nearly one-half million new loans per month across thousands of new deals.

In partnership with 5 Bridges, we have been supplying this data to a government agency for the past two quarters. 5 Bridges has been writing a comprehensive report describing new mortgage originations, and that monthly report is expected to be made publicly available in the future. 5 Bridges is also using the data to calibrate internal mortgage valuation models to become the first to bridge the analysis gap between conforming and non-conforming mortgages within the same dataset. This should lead to more accurate predictive models, which allow better valuations on securities backed by both conforming and non-conforming collateral.

With innovations like these, investors and others seeking better transparency into conforming deals are starting to get access to the data they need to better analyze newly issued deals. This is happening despite the lack of comprehensive loan-level data provided directly from the agencies. An easier fix, of course, would be for the agencies to release this data in a standardized format in much the same way they currently release new pool information through the Securities Industry Automation Corporation (SIAC), Brooklyn, New York, and their own websites.

When I served at Fannie Mae, my understanding was that the agency did not release loan-level data because of concerns about violating the privacy of borrowers and because of the possible effect on TBA (to be announced) trading markets. I spent approximately 12 years at the agency starting in late 1993. The majority of my time at the agency, I ran the securities trading operation, a department somewhat analogous to the bond trustees covering non-conforming space.

The private market, however, has shown that data can be released on an anonymous basis, protecting borrower privacy while providing investors information necessary to properly evaluate their investments. Freddie Mac's issuance of loan-level data demonstrates that releasing such information does not materially affect the spreads at which TBA products trade. **After conversations with many industry participants, I believe other government agencies will follow Freddie Mac's lead in releasing comprehensive loan-level data. The is a necessary step forward if for no other reason than to standardize the loan level information coming out of the three government agencies.**

Regardless of these considerations, the cat is essentially out of the bag. The industry is sophisticated enough to overcome many of the remaining barriers to obtaining loan-level data on collateral backing agency new issuances.

Over the next year, the market can expect more advancements mirroring the private market when it comes to obtaining loan-level data. These include the possible marriage of loan-level data to third-party databases containing property and credit-score information. The innovations will continue to improve the market's ability to analyze deal performance in ways that were unavailable or overly cumbersome leading into the crisis. MB

Bio: Larry Barnett is chief executive officer of BlackBox Logic LLC, a Denver-based provider of loan-level data for residential mortgage market participants. He previously served as vice president for secondary mortgage trading operations at Fannie Mae. He can be reached at lbarnett@bbxlogic.com.

SIDEBAR:

Agency Data Availability on New Pools

-Freddie Mac provides loan-level data on new issues on a daily basis, and has done so since 2006. The data are rich in information that describes loan attributes at origination but provides only loan balances in the ongoing payment files. The user can back into delinquency information but cannot calculate loss severities. Freddie Mac also provides extensive pool-level summary files that can be used for higher-level analysis and are complementary to the loan-level data currently being supplied by the agency.

-Fannie Mae provides extensive pool-level data on new issues on a daily basis. **We believe that Fannie Mae is now actively considering ways to provide loan-level data for new issuance in much the same way that Freddie does. In the meantime it is possible to derive a statistical representation of the loans within Fannie MBS which we believe is useful in modeling future performance of individual Fannie issued mortgage backed securities.**

-Ginnie Mae provides pool-level data on new issues in a variety of ways. The summary data are rich but not as extensive as pool-level summary data provided by Fannie Mae and Freddie Mac.

-The Department of Housing and Urban Development (HUD) provides extensive loan-level data on loans backed by the Federal Housing Administration (FHA) and Department of Veterans Affairs (VA). The data are comprehensive and more than adequate to build loan-level data sets for FHA/VA-insured collateral. It is not yet possible to cross-reference the HUD data to Ginnie Mae new security issuances.