Community Investments Vol. 15, Issue 1 Credit Scoring and Model Development and Maintenance

March 2003

Credit scoring is an underwriting tool used to evaluate the creditworthiness of prospective borrowers. Utilized for several decades in granting certain forms of consumer credit, scoring has come into common use in the mortgage lending industry only within the last 10 years. Scoring brings a high level of efficiency to the underwriting process, but it has also raised concerns about fair lending with regard to historically underserved populations.

To explore the potential impact of credit scoring on mortgage applicants, the Federal Reserve System's Mortgage Credit Partnership Credit Scoring Committee is producing a five-part series of articles. This is the second. An important goal of the series is to provide the industry and concerned groups and individuals with the opportunity to comment on issues surrounding credit scoring.

The first article provided a context for the issues to be discussed in the series and gives further background information on the Mortgage Credit Projects.

Each representative for this article received a request to comment on the following text:

Lending institutions face various pressures in the course of their credit operations. They must consistently achieve and increase profitability, comply with a complex regulatory framework, and contend with new sources of competition. An institution's loan underwriting policy, and, in particular, its credit-scoring model, reflect the institution's appetite for risk, targets for profitability, and role in serving the credit needs of its market.

Credit-scoring models have predictive power; they give lenders the ability to expeditiously assess the likelihood of borrower default. There is general agreement that to retain their predictive power, models must be maintained and adjusted to reflect changes in loan performance and in market demands and demographics. In addition, observers argue that absent proper maintenance, a lender risks using a model with diminished predictive capability, which may produce an unjustifiable disparate impact on prohibited basis groups.

From your perspective and experience, what can lenders do to ensure that the credit-scoring models they develop or purchase will accurately predict the performance of their applicant base? What steps might lenders take to effectively update and maintain their models? Finally, what methods should lenders employ to monitor the performance of their credit-scored loans, particularly with respect to the fairness and accuracy of their models?

This article incorporates statements requested from representatives of three organizations, selected because of their interest in and differing perspectives on credit scoring and fair lending.

James Wheaton
Neighborhood Housing Services of Chicago

Mr. Wheaton has worked for and with nonprofit community development organizations since the mid 1970s. He now serves as the associate director

of Neighborhood Housing Services of Chicago, Inc. (NHS), a position he has held since 1993. Mr. Wheaton's responsibilities include administration of NHS's home-improvement and purchase/rehab lending programs, as well as new program and product development. NHS of Chicago was established in 1975 as a nonprofit corporation that partners with financial institutions, community residents, city government, and Chicago businesses. NHS of Chicago has citywide lending programs as well as targeted neighborhood programs operating in 11 of Chicago's neighborhoods. NHS also recently created a program for victims of predatory lending. NHS of Chicago originates 500 loans annually, totaling $15 million.

Thomas P. Fitzgibbon, Jr.
Manufacturers Bank

Mr. Fitzgibbon is a senior vice president and chief retail banking officer for Manufacturers Bank, and is the president of Manufacturers Community Development Corporation. Mr. Fitzgibbon is a 30-year veteran of the banking industry, having served as a principal banking officer in lending and retail banking operations for institutions in Washington, DC and Minnesota prior to moving to Chicago in 1990. He has served on the Steering Committee of the Mortgage Credit Access Partnership and the Small Enterprise Capital Access Partnership for the Federal Reserve Bank of Chicago since 1995, and currently he is on the boards of directors for Bethany Hospital, DevCorp North, NHS of Chicago, the Northwest Housing Partnership and Regional Redevelopment Corp., and the Woodstock Institute. Manufacturers Bank, a $1.4 billion community bank with 13 offices, is ranked as the one-hundredth leading small-business lender in the nation (American Banker) and the third leading small-business lender in low- and moderate-income markets in Cook County, IL. Manufacturers Community Development Corporation is a six-year-old subsidiary of the bank, managing more than $40 million in direct-equity investments and loans in real-estate and small-business ventures.

Alex Stricker
Fannie Mae

Dr. Stricker is an economist for credit policy at Fannie Mae. He has worked on development of Fannie Mae's automated underwriting models for the past two years, with emphasis on fair-lending implications. Prior to joining Fannie Mae, he pursued doctoral studies at Syracuse University specializing in urban economics and housing discrimination. Fannie Mae is a stockholder-owned corporation chartered by the Congress to create a continuous flow of funds to mortgage lenders in support of homeownership and rental housing. It serves as a secondary market for mortgage loans by purchasing mortgages from lenders across the country, aggregating groups of loans into mortgage-backed securities, and selling the securities to investors.

Response of James Wheaton
Neighborhood Housing Services (NHS) of Chicago

Along with the pressures to increase profitability, comply with complex regulatory requirements, and contend with new and ever more aggressive sources of competition, mortgage lenders, like other businesspeople, must also manage rapid change in technology. In the lending arena, this change is evident in the approval of loans through automated underwriting, made possible in part by the use of credit scoring. The past few years have seen a dramatic increase in the use of credit scoring in mortgage lending, yet there is substantial anecdotal evidence that credit scoring may not be a particularly responsive tool for the low- to moderate-income borrower.

Credit-scoring proponents point to the speed, accuracy, and fair treatment it brings to the lending process, but credit-scoring models require regular maintenance, testing, and updating to reflect changing market conditions, without which both lender and borrower will suffer. Nonetheless, it appears

that some lending institutions rely on scoring models with limited predictive power, and they miss significant business opportunities as a result.

NHS of Chicago's direct lending is targeted to low- to moderate-income (LMI) neighborhoods and borrowers. Many of these communities did not, until fairly recently, have a neighborhood banking or lending branch. The primary providers of credit to many residents were financial entities that were aggressive in pursuing LMI borrowers; today, many of them would be characterized as subprime lenders. Because credit-scoring models factor in the types of credit used by a borrower in the past (and subprime credit has a negative impact on the score), many borrowers from these neighborhoods may be adversely affected when dealing with a conventional lender who relies on credit scores. Further, my own observation of credit scores of first-time buyers and LMI homeowners is that negative factors have an immediate effect on scores, while positive factors influence the score much more gradually.

Supporters of credit scoring also maintain that its use frees the lender to more closely examine the marginal borrower and spend the time and effort necessary to close the loan. At NHS, though, we have seen too many situations where credit scoring has actually been used to limit access to first-tier credit. In the Spring 2000 issue of the Federal Reserve Bank of Boston's Communities & Banking, Calvin Bradford argues that the use of credit scoring does not always result in more underwriting time being spent on applicants with marginal credit but may actually serve as a tool to identify candidates for higher-cost loans. Absent proper maintenance of a scoring model and its underlying assumptions, and without diligence to ensure its fair application across all applicants, credit scoring could further widen the gap between low- and high-income borrowers.

I believe that scoring models' predictive power is worse for low-income borrowers than it is for the average mortgage applicant. NHS understands

and appreciates that the acquisition of a home and the opportunity to thereby build both financial and social wealth is a powerful incentive. I do not believe that any credit-scoring model factors in the emotional impact of potential homebuyers when they are the first members of their families for generations to own a home or buy a home in the newly revitalized neighborhood in which they grew up. Human judgment is still essential in weighing these factors. And as Peter McCorkell of Fair, Isaac & Company, Inc. states in the article mentioned above, the scoring models most often used in mortgage lending were not specifically designed to assess mortgage risk.

Lending institutions that use credit scoring to identify customers who would benefit from a second look, prepurchase, or credit counseling are to be applauded. With government-sponsored enterprises such as Fannie Mae and Freddie Mac currently offering products with more flexible terms for the credit-challenged borrower (such as Fannie Mae's Timely Payments Rewards product), lenders can offer conventional pricing more readily than before.

Credit scoring proponents further maintain that a primary benefit of scoring is that it increases people's access to credit. I take this to mean that its primary goal is to provide credit that is reasonably priced and without excessive fees or burdensome loan terms. To reach this goal, all parties with a vested interest in the activities of lenders using credit-scoring technology need to ensure that the credit-scoring tool is working as effectively and fairly as possible. While a scoring system may be developed on the basis of statistics, the developers' role cannot be ignored. Just as lending institutions and secondary-market investors are held to a standard of fairness, scoring-system developers should share in the obligation to ensure that their models do not unfairly exclude borrowers.

It has been our recent experience that lending institutions most sensitive to the needs of LMI borrowers are increasingly those institutions that rely less

on credit scoring and more on individual assessment of the borrower. Community lenders (such as NHS) that are focused on LMI neighborhoods have an understanding of the local environment and neighborhood dynamics, and they provide competitively priced mortgages to LMI borrowers in considerable volume. For national lenders, this kind of hands-on approach is not feasible. An underwriter in St. Louis cannot be expected to know and understand the characteristics of a buyer and a property on the West Side of Chicago; there needs to be some adjustment to the automated system that might wrongfully deny that buyer access to credit.

If credit scoring is going to be a factor in credit decisions for the foreseeable future, models that more adequately assess mortgage risk need to be developed and put into general use. Scoring system developers need to develop methodologies that are more responsive to a borrower's positive credit behavior and that incorporate some of the more subjective, but very relevant, data that often factor into a human being's decision about someone's creditworthiness.

Underwriting and Training Policies with Respect to Credit Scoring
Lending institutions clearly need to do a better job of training their personnel about the purpose and limitations of credit scores. I do not suggest that underwriters be divested of the capacity to override a credit-scored decision. However, excessive overrides raise serious concerns about disparate treatment of borrowers. Access to credit for a borrower who is qualified by a credit score (even marginally) should not be denied because of the underwriter's or loan officer's personal assessment of the borrower's gender, ethnicity, lifestyle, personality, temperament, family connections, and the like. Human nature being what it is, a lending policy allowing for "high-side" overrides-in which an applicant's score suggests they deserve a loan yet they are denied it-opens the door to potential misuse, and I do not believe a responsible lending institution would either tolerate such decisions or accept such liability.

Second review of all adverse actions should be standard operating procedure for lending institutions, both to ensure fair and equal access to credit and to ensure that acceptable business opportunities are not missed. For lenders that offer subprime products, I would suggest that their second review be conducted in the context of trying to qualify their customers for a conventional product. Lending staff involved in second reviews should have special training in the use of credit scores, including some education about how scores are developed, what a score is designed to predict, and what factors in a borrower's credit history will affect the score (either positively or negatively). The scoring-system developers are key in this process, and an acceptable middle ground must be struck between protecting their proprietary systems and educating lenders on the use and limitations of credit scoring.

In summary, access to credit continues to be a critical need in many LMI communities. The recent increase in the homeownership rate in this country indicates that there is a large population striving to be homeowners and making some progress to achieve that goal. To the extent that credit-scoring technology has made this possible, that is very positive. However, lenders, especially those who have developed their own credit-scoring model on the basis of their own experience and portfolios, must maintain and upgrade the credit-scoring model in the same way that they maintain other systems. Maintenance and regular upgrades of credit-scoring models to reflect market conditions should be part of the business plan and evaluated on a regular basis. Such evaluation should include an analysis of the performance of credit-scored loans versus those that were overridden, and especially an analysis of the performance of those credit-scored loans that were identified as marginal. Just as no institution would attempt to run its business with outdated hardware, it should not be using an outdated scoring model to direct credit decisions.

Response of Thomas P. Fitzgibbon, Jr.
Manufacturers' Bank

What can lenders do to ensure that the credit-scoring models they develop or purchase will accurately predict the performance of their applicant base?

For the successful use of predictive scoring models in the credit decision-making process, the models must be based on similar products, environments, and populations. In addition, the attributes and application of the criteria parameters in the models must be refreshed routinely to ensure that the applications produce results consistent with the expectations when the models were developed or purchased.

Model use is a two-step process. First, the lender must select the right model for the loan product. Second, the lender must consistently refine the model, which requires dedicating resources long after original development. This refinement requirement can be easy to ignore, especially in the early stages of a product rollout when there is little product performance to point to as indicators of performance shortfalls. However, this initial stage is the time when even more due diligence needs to be devoted to fine tune the model and avoid unintended results. Higher than anticipated pull-through rates or adverse action rates are early indicators that the model has serious flaws requiring immediate attention.

Most purchased credit-scoring models have solid data to support their predictability. In addition, the best model vendors require lenders to supply the results of their experience so the vendor can improve and enhance its own data for future models. This feedback improves the quality of the predictive factors and model fairness. Consistent feedback is part of the model-refreshing process; however, modification of the model criteria by the lender can degrade the model's results.

Lenders who develop their own models often need to compensate for their small population performance base by comparing experience for an extended time, and even more care should be given to reviewing results during the initial product rollout. Comparing customer performance results, as well as application approval and pull-through rates, will yield richer data. These data will help the user identify fairness issues (adverse impact), adverse selection (capturing undesired applications), and low pull-through (closing) rates that could indicate a competitive disadvantage of the product.

Senior management and boards of directors should be wary of "proxy-like" models, either in-house or purchased from a vendor, that were developed for a loan product or population somewhat similar to another lender's product or population. Because such similarities can be hard to define, this practice can have disastrous results in both fairness to applicants and the bottom line. Management should perform adequate due diligence on the criteria and, if not convinced, employ outside resources to provide evaluation and recommendations related to the model.

What steps might lenders take to effectively update and maintain their models?

As I stated previously, most model vendors insist that lenders provide specific information related to model performance, including applications received, approval rates, pull-through rates, and servicing results. These data will also provide the lender with information that can be employed to change the criteria of the lender's model, product price, collateral value (if included in the model), population attributes, brokers or mortgage bankers who bring applications to the lender, and other levers, in order to achieve the desired results.

Most lenders employ models to develop results based on return on assets (ROA) objectives, understanding there will be losses in any model that is

employed. Loan pricing should reflect performance expectations and results. Therefore, consistent review of pricing (rate, fees, and so on) will be necessary to achieve the ROA and to ensure that the pricing reflects the risks associated with the population and security characteristics, thus ensuring fairness to all populations.

Lenders who develop their own models need to take steps to consistently review adverse actions: comparing protected-class applicants to the applicant pool, reviewing approval and pull-through rates related to the expectations, and comparing the servicing results to the ROA projections. Deviations from model projections should guide the lender to change the model, including credit score (FICO, Delphi, and the like), loan-to-value categories, applicant attributes, and vendors (if used).

In the initial stages of the product rollout, the lender needs to review early performance indicators that do not meet the expectations of the design phase. Even small indicators of performance shortfalls, such as low application rates from prohibited basis groups, higher-than-expected adverse action rates (especially where protected-class populations are concerned), or lower-than-expected pull-through rates, are indications that the model may have flaws that need to be addressed.

What methods should lenders employ to monitor the performance of their credit-scored loans, particularly with respect to the fairness and accuracy of their models?
The methods lenders should employ include the following:

- Due diligence review of all adverse actions to ensure that the model is applied correctly,
- Comparative analysis of adverse actions to evaluate model results on protected-class applicants,

- Comparison of computer records (data input) with application sampling to ensure quality control,
- Review of any subjective decision-making performed on scored applications that changes the model decision or modifies the pricing or product parameters, and
- Review of closed-loan packages (quality control) to ensure that the loan parameters approved are the same as the parameters in the closed loan.

Consistency and diligence are imperative in developing and using credit-scoring models. Early indications of performance that are different than predicted allow action to be taken early in the process to change the model parameters and modify elements that caused the deviations. Vendors and lenders need to stay alert to changes and intervene quickly.

Response of Alex Stricker
Fannie Mae

Automated technologies in credit-granting institutions have expanded dramatically in the past 10 years and credit-scoring applications are now common. These applications aid significantly in the effort to streamline origination processes and cut costs while delivering consistent and objective decisions about an applicant's creditworthiness. Scoring models relate an applicant's past credit performance and current financial characteristics to future debt repayment. They are often characterized as generic or custom. Generic scores are created to be predictive of delinquency for generic consumer debt, using large amounts of credit data. Custom scores are designed to be predictive of repayment performance for specific types of credit or perhaps for a specific lender's customer base. With custom scores, additional non-credit-report information may be used in the modeling effort. Regardless of who builds a scoring model, there are common considerations in the development process and maintenance of the model.

Follow a Clear and Explainable Development Process

Scoring-model development occurs with the coordination of market analysts, credit-risk managers, statisticians, database administrators, and computer programmers. Each part of the process must be carefully planned to ensure development and implementation of a successful model.

Objective

The first step in the technical development of a scoring model is to determine what measure of performance to model. Models may predict the probability of default (nonperforming loans that terminate and do not prepay in full), the probability of becoming delinquent, the financial losses an institution expects for each loan, or some combination of delinquency, default, and losses. A lender that uses another company's underwriting system to make loans to hold in its portfolio should be aware of the implications of the scoring model objectives for lending patterns. For example, models designed to predict serious mortgage delinquency tend to place more importance on past-credit-history variables than models designed to predict default. By contrast, mortgage default models give more weight to loan-to-value ratios.

Data Collection and Sample Design

The data available for use in statistical modeling are the single most important technical element of model development. Lender data retention is crucial for model construction and testing. Typically, the more information available, the more precise the results can be. Lenders developing their own system are best served by data that come not only from their existing customer base but also from other segments of the market that represent potential applicants. The selection of risk factors included in a scoring model is determined in part by their availability to the modeler. Therefore, it is vital to capture and retain as much origination and subsequent performance information as possible.

After a sample has been constructed, the scoring limitations created by the available data sample need to be identified. For example, at this time, Fannie Mae's Desktop Underwriter does not process 95 percent loan-to-value ratio refinance loans with a cash-out component on non-owner-occupied, three- to four-unit housing. Our experience with this product is currently too limited to model, but as we learn more and acquire more data, the risk of this product may become better understood and be modeled appropriately.

Statistical Tools

Most scoring applications predict the likelihood of an event. Many statistical tools are available. For example, default probabilities can be estimated by means of logistic regression. The logistic procedure, well known and understood by economists, is fast and straightforward to implement. The specific tool chosen depends on the goal of the scoring model and any deficiencies in the development sample. In the case of sample deficiency, data-augmentation methods are available to improve estimation on thin samples, as are procedures to account for potential biases stemming from missing information. The result of a scoring model is the generation of a scorecard. Thus, the scorecard's combination of points may be influenced by the statistical tools and methods employed in the model.

Validation and Testing

A variety of statistical tests are available to aid in the validation of a model. No single test provides a complete answer. Fannie Mae has estimated hundreds of models, with all potential variables, divided and clustered, to yield the statistically strongest model. The typical measures of qualitative-dependent-variable modeling are used, such as gini coefficients, K-S statistics, and concordance. The overall idea is that the model must do the very best job of separating high-risk and low-risk loans. Since many model variations may be tested using several criteria, it is important to have rules for what constitutes a more predictive model. Equally important is how well

the model predicts for subgroups of the intended population. For example, does a model designed to predict delinquency for borrowers of all income levels produce an appropriate ordering of risk when it is applied only to low-income borrowers? The answer depends in part on how diverse the development data are with respect to income. Testing a model's differential validity is necessary before implementing it in production.

Cutoffs and Overrides

During model development, attention should be given to determining how much risk to tolerate. The model itself may predict how likely default is for a particular loan. However, consideration must be given to how much collective credit risk the company is willing to take. This is determined by market analysis of likely application volumes, the length of time loans are expected to stay in the book of business, capital requirements, and pricing and revenue targets. A periodic review of these targets is necessary to ensure that the approved mix of business continues to meet revenue objectives.

Limits within the scoring engine can be reached if the scoring model tries to evaluate values for certain risk factors that are improbable in the scorecard application. At Fannie Mae, our system filters out for manual review all applicants with total debt-to-income ratios greater than 65 percent. The Desktop Underwriter program refers the application to the underwriter to determine whether the data were entered incorrectly or if the relatively high debt-to-income ratio is manageable for the applicant.

Monitor Application Decisions

Is the production-decision process working in a way similar to the process tested? Generic creditworthiness scores might be used only in part to make a decision, so it is important to keep track of how these scores relate to the final decision. Custom systems may be used to support a comprehensive evaluation of applications and to monitor who is being approved or denied at

the recommendation of the automated-scoring system. At Fannie Mae, we have monthly reports on applications through our Desktop Underwriter system. We examine the system's recommendations across various financial and demographic characteristics. When changes or irregularities are observed, more detailed examination follows. Such monitoring is vital to remedy problems or irregularities.

Monitor Performance

Regardless of what the system is designed to predict, performance can be tracked from one month after origination. The most important report will show how loan performance varies by the scoring system's recommendation. Are the approved loans performing differently than the loans made with an automated recommendation for further review? If generic scores were used in the decision to make the loan, are higher-scored loans performing better than lower-scored loans? Other analysis should focus more narrowly on loans scoring near the cutoff to be sure that those marginal loans are performing as expected. A complete examination will involve tracking performance for numerous loan subsets across product, financial, demographic, and geographic segments of the market. The particular array of reports depends on the financial institution's lending goals and regulatory requirements. Simple reporting, done regularly and completely, will alert management, marketing personnel, and model developers to potential problems and areas to investigate further.

Model Evolution

Expect to update your model. Experience will improve the effectiveness of a scoring system. As such, the development process must be flexible to allow for changes suggested through the learning. At Fannie Mae we are continuously investigating and developing new models. Every new model we generate is an evolution of the model it replaces. Approximately annually, the Desktop Underwriter scorecard is re-estimated to utilize additional performance data that come with the passage of time and variation in the

economy. There is no secret formula for success. Able statistical analysis is necessary to generate a system. Its success requires the coordination of market analysis, data retention and reporting, and skilled risk managers.