

reasonable assurances of their rating systems' accuracy. The ongoing process to confirm and ensure rating system accuracy consists of:

- The evaluation of developmental evidence,
- Ongoing monitoring of system implementation and reasonableness (verification and benchmarking), and
- Back-testing (comparing actual to predicted outcomes).

IRB institutions are expected to employ all of the components of this process. However, the data to perform comprehensive back-testing will not be available in the early stages of implementing an IRB rating system. Therefore, banks will have to rely more heavily on developmental evidence, quality control tests, and benchmarking to assure themselves and other interested parties that their rating systems are likely to be accurate. Since the time delay before rating systems can be back-tested is likely to be an important issue—because of the rarity of defaults in most years and the bunching of defaults in a few years—the other parts of the validation process will assume greater importance. If rating processes are developed in a learning environment in which banks attempt to change and improve ratings, back testing may be delayed even further. Validation in its early stages will depend on bank management's exercising informed judgment about the likelihood of the rating system working—not simply on empirical tests.

#### Ratings System Developmental Evidence

The first source of support for the validity of a bank's rating system is developmental evidence. Evaluating developmental evidence involves making a reasonable assessment of the quality of the rating system by analyzing its design and construction. Developmental evidence is intended to answer the question, Could the rating system be expected to work reasonably if it is implemented as designed? That evidence will have to be revisited whenever the bank makes a change to its rating system. If a bank adopts a rating system and does not make changes, this step will not have to be revisited. However, since rating systems are likely to change over time as the bank learns about the effectiveness of the system and incorporates the results of those analyses, the evaluation of developmental evidence is likely to be an ongoing part of the process. The particular steps taken in evaluating developmental evidence will depend on the type of rating system.

Generally, the evaluation of developmental evidence will include a body of expert opinion. For example, developmental evidence in support of a statistical rating model must include information on the logic that supports the model and an analysis of the statistical model-building techniques. In contrast, developmental evidence in support of a constrained-judgment system that features guidance values of financial ratios might include a description of the logic and evidence relating the values of the ratios to past default and loss outcomes.

Regardless of the type of rating system, the developmental evidence will be more persuasive when it includes empirical evidence on how well the ratings might have worked in the past. This evidence should be available for a statistical model since such models are chosen to maximize the fit to outcomes in the development sample. In addition, statistical models should be supported by evidence that they work well outside the development sample. Use of "holdout" sample evidence is a good model-building practice to ensure that the model is not merely a statistical quirk of the particular data set used to build the model.

Empirical developmental evidence of rating effectiveness will be more difficult to produce for a judgmental rating system. Such evidence would require asking raters how they would have rated past credits for which they did not know the outcomes. Those retrospective ratings could then be compared to the outcomes to determine whether the ratings were correct on average. Conducting such tests, however, will be difficult because historical data sets may not include all of the information that an individual would have actually used in making a judgment about a rating.

The sufficiency of the developmental evidence will itself be a matter of informed expert opinion. Even if the rating system is model-based, an evaluation of developmental evidence will entail judging the merits of the model-building technique. Although no bright line tests are feasible because expert judgment is essential to the evaluation of rating system development, experts will be able to draw conclusions about whether a well-implemented system would be likely to perform satisfactorily.

#### Ratings System Ongoing Validation

The second source of analytical support for the validity of a bank rating system is the ongoing analysis intended to confirm that the rating system is

being implemented and continues to perform as intended. Such analysis involves process verification and benchmarking.

#### Process Verification

Verification activities address the question, Are the ratings being assigned as intended? Specific verification activities will depend on the rating approach. If a model is used for rating, verification analysis begins by confirming that the computer code used to deploy the model is correct. The computer code can be verified in a number of established ways. For example, a qualified expert can duplicate the code or check the code line by line. Process verification for a model will also include confirmation that the correct data are being used in the model.

For expert-judgment and constrained-judgment systems, verification requires other individual reviewers to evaluate whether the rater followed rating policy. The primary requirements for verification of ratings assigned by individuals are:

- A transparent rating process,
- A database with information used by the rater, and
- Documentation of how the decisions were made.

The specific steps will depend on how much the process incorporates specific guidelines and how much the exercise of judgment is allowed. As the dependence on specific guidelines increases, other individuals can more easily confirm that guidelines were followed by reference to sufficient documentation. As the dependence on judgment rises, the ratings review function will have to be staffed increasingly by experts with appropriate skills and knowledge about the rating policies of the bank.

Ratings process verification also includes process override monitoring. If individuals have the ability to override either models or policies in a constrained-judgment system, the bank should have both a policy stating the tolerance for overrides and a monitoring system for identifying the occurrence of overrides. A reporting system capturing data on reasons for overrides will facilitate learning about whether overrides improve accuracy.

#### Benchmarking

S. Banks must benchmark their internal ratings against internal, market and other third-party ratings.

Benchmarking is the set of activities that uses alternative tools to draw inferences about the correctness of ratings before outcomes are actually