Implementation note

Title	Validating Risk Rating Systems at IRB Institutions
Category	Capital Adequacy Requirements
Date	January 31, 2006
Sector	Bank Holding Companies Banks Trust and Loan Companies
No	A-1

Table of Contents

I. Introduction

II. Background

III. Principles

- 1. Purpose
- 2. Responsibility
- 3. Integrity
- 4. Documentation
- 5. Timing
- 6. Scope
- 7. Response
- 8. Perspective

IV. Conclusion

Appendix I: Use of Scoring Models for which Institutions have Incomplete Information

Appendix II: History of Major Changes

Appendix III: Procedures Generally Required in a Validation

Appendix IV: External Data and Retail Validation

Footnotes

> This implementation note is valid until the end of the reporting period on January 31, 2027. A newer version is available.

I. Introduction

The term "rating system" comprises all of the methods, processes, controls, data collection and IT systems that support the assessment of credit risk, the assignment of risk ratings, and the quantification of default and loss estimates.

This Implementation Note elaborates on Section 5.8.8 of Chapter 5 of OSFI's *Capital Adequacy Requirements* (CAR) Guideline A-1. An institution's[1] degree of adherence to these principles, both initially and on an ongoing basis, will be a key consideration in OSFI's decision whether to approve the use of the internal ratings-based (IRB) methodology to establish minimum regulatory capital under CAR. The principles apply to all rating systems under the IRB method.

II. Background

Institutions use various rating methodologies and credit risk modelling approaches to differentiate credit quality, and to quantify default likelihood and loss severity. However, a rating system that has not been validated is not suitable for IRB standards. Under CAR, ratings will drive minimum capital requirements for credit risk for institutions that are qualified to use the IRB method. Institutions will need to demonstrate the validity of rating systems as one of the minimum standards they must meet in order to obtain OSFI's approval to use the IRB method. Institutions' adherence to the broad principles outlined in this implementation note will be an important consideration in OSFI's initial approval of institutions for IRB and ongoing use of the IRB approach.

Institutions may look to CAR for specific standards applicable to IRB. However, these standards are subject to interpretation, and institution implementation is subject to OSFI approval. This Implementation Note sets out the principles that OSFI expects institutions to apply to validation, including discussion and general examples. They are provided with the understanding that the application of these principles will be tempered with good judgment. This does not negate the principles, but may limit their application to avoid undue costs or perverse results. Institutions may encounter situations in which the suggested procedures have negligible impact or do not help validation. In such cases, the institutions may consider other procedures. Documentation is essential for process review, validation, other aspects of good governance, and future risk quantification, but only to levels of detail that could plausibly be useful. Lists of what "might" be done are not exhaustive and are not meant to discourage institutions from proposing better approaches to validation.

III. Principles

Institutions will use different methods to validate their rating systems according to their history and current portfolio. To do this, all institutions need to establish an effective validation framework that observes principles of purpose, responsibility, independence, documentation, continuity, scope, response, and perspective. OSFI's supervisory processes to approve and monitor the ongoing use of the IRB method for the calculation of regulatory capital under CAR will include a review of adherence to the principles outlined below.

1. Purpose

Validation confirms that rating systems:

- *Identify factors to help discriminate risk;*
- *Appropriately quantify measures of risk;*
- *Produce measures of risk that have a response to macroeconomic conditions consistent with an institution's intentions, and that meet the standards of CAR for the calculation of IRB capital.*

Institutions should have robust systems[2] to validate the consistency and accuracy of rating systems, including rating assignment processes and the quantification of all relevant risk parameters. Validation should confirm that assigned risk ratings and risk measures[3] react to changes in the credit environment in a manner consistent with a ratings

philosophy formally adopted by the institution[4]. Consequently, an institution's expectation of the performance of its rating systems should be consistent with its ratings philosophy.

2. Responsibility

Institutions validate the performance of their rating systems.

Institutions should designate specific groups to be responsible for the design and performance of the validation process, including the outputs. As rating systems are integral to the management of credit risk, economic capital and other vital matters, CAR specifically requires that an institution's Board of Directors (or a designated committee thereof) and Senior Management understand the operation of the rating system and have a detailed comprehension of its associated management reports. This understanding should include the validation process. Under CAR, Senior Management is also required to ensure that the rating system continues to operate properly. This would include verification that validations are timely and effective, and that the rating system is suitably adjusted to the findings of validation studies. (See Appendix I on the use of scoring models for which institutions have incomplete information.)

3. Integrity

The validation process should be independent of the design, operation and consequences of the rating system.

The goal of the validation process is to deliver an effective challenge to the design and operation of the rating system. IRB institutions should therefore demonstrate that the validation process for ratings systems is independent from the personnel and management functions responsible for originating exposures. Those who validate should have the knowledge, resources, accountability and independence to effectively challenge risk rating design, operation and risk quantification.

Overall responsibility for independent review of an institution's validation processes lies with Internal Audit, which provides a link to the Board of Directors. While internal auditors may be able to review processes and controls related to validation, they may lack the technical expertise to review highly quantitative elements of validation. In such cases, the review of validation processes and outcomes should be conducted by other groups within the institution's organization that are independent of those groups responsible for designing, operating and validating

institution rating systems.

4. Documentation

Institutions should document their validation of rating systems to ensure that parties reviewing the material can understand the objectives of the rating systems, the scope and methodology of validation, and the conclusions drawn from validation activities.

In order to approve the use of parameters drawn from a rating system to drive regulatory capital under the IRB method, OSFI and the institution need clear and comprehensive documentation in order to understand the design of the rating system and the validation of the system. Part of the documentation will be a record of major changes to the risk rating system, as illustrated in Appendix II.

5. Timing

Institutions should establish regular processes to validate their rating systems, but validation should also respond to special events or circumstances.

As noted above in Principle 3: Integrity, a process is required to show that rating systems and the risk parameters they generate remain valid, and policy should establish a schedule for formal reviews of validation, which should be performed at least once a year. More frequent reviews may be required depending on emerging results, availability of data, changes to validation procedures, and plausible impact on the institution. Institution policy should establish a minimum frequency for the comparison of experience to expectations.

A material change in products, or their distribution, should prompt special analysis to ensure that performance remains adequate. A major change in the rating system itself should also prompt special analysis to ensure that performance remains adequate.

6. Scope

Institutions should consider all data and issues that may be material and relevant to the validation of their rating systems.

Institutions may be unable to provide conclusive proof that their rating systems are valid by applying statistical tests, owing to data scarcity and the shortcomings of the tests themselves. Nonetheless, institutions should use whatever statistical tools can assess the likelihood of emerging results, supposing various hypotheses, to inform assessments of the performance of systems and the accuracy of estimates. They should also examine related data from internal and external sources to establish a context for assumptions, calculations and results.

Generally institutions will arrive at a decision to revise their rating systems after reviewing them from many angles and seeing too many results that are unlikely under the assumed model. Institutions may also decide to revise their rating systems after concluding that this would improve their ability to discriminate risk. No combination of tests will prove conclusively that a rating system is valid, but institutions may construct a mosaic of evidence that provides reasonable confidence to the institution's Senior Management and regulators.

Institutions should examine a variety of issues, including:

- the relevance, completeness, consistency and adequacy of inputs;
- the assumptions embedded in the rating systems;
- the ability of the rating system to predict future outcomes for the business to which it is applied over a range of conditions;
- the consistency between the theoretical models and implemented applications; and
- the appropriate and intended use of the rating system.

To address these issues, institutions will generally need to perform many procedures. A discussion of some possible validation procedures is included in Appendix III. Institutions should consider the application of these procedures to their own portfolios. In some cases, institutions will need to use other techniques. More elaboration on retail validation is included in Appendix IV.

7. Response

Institutions should adjust their ratings systems to take account of reasonable conclusions drawn from validation activities. In particular, they should identify and respond to deviations of experience from expectations that call into question the validity of their rating systems.

Institutions should develop and follow a formal policy to compare realized rates with estimated PDs (LGDs, EADs or other measures) for each obligor grade. They should demonstrate that the realized default rates are within the expected range for the relevant grade, taking into consideration current conditions and the sensitivity to current conditions consistent with the embedded rating philosophy. Comparisons should also be performed for aggregations of grades. Institutions should prepare, in advance, criteria to identify outcomes that may be inconsistent with the rating model or the estimates used in risk management. Appropriate adjustments should be made when these results occur. Institutions should compare experience against expectations according to an established schedule. Outputs from a validation, including recommendations from the validation function of the institution, should play an important role in the use and development of the rating system.

8. Perspective

Institutions should validate the overall performance, as well as the details, of their rating systems.

As noted above in Principle 6: Scope, validation assessments are required for all material and relevant rating system elements. However, estimates of details are never exact, and cumulative errors across a number of components may seriously flaw aggregate results. Consequently, institutions should validate at different levels of granularity, as well as validating the overall performance of each rating system, to confirm that aggregate results are reasonable.

IV. Conclusion

The validation of a rating system requires a continuing commitment of resources. The use of these resources will be most effective if the process is carefully planned, with due attention to ratings philosophy, governance and data integrity along with more technical issues of statistical inference. Once the validation is completed and documented, the outputs of the rating system will obtain credibility and applicability and acceptance in the institution's risk management systems.

Appendix I: Use of Scoring Models for which Institutions have Incomplete Information

Institutions are required to segment risks into homogeneous pools for the calculation of PD. For this segmentation, some institutions would like to use credit scores[5] developed by external vendors to distinguish high risks from low risks. The developer may use data from other institutions. To protect the confidence of contributing institutions and their customers, developers may not share the full development dataset. Users of the scores may look at summary statistics or extracts from the development file, but the cost of doing this is material. Whether or not the institution can see full details of the dataset, the developer may consider the logic underlying a score as valuable intellectual property, and will not share the details with institutions.

Paragraphs 418 - 420 of CAR require documentation of design, rating criteria, and inputs to a system. Compliance with these requirements may be difficult when data collection is in the hands of a third party and details of the model generating scores are considered proprietary information. However, institutions may not ignore these requirements. Paragraph 421 states that the fact that a model uses proprietary technology from a third party vendor does not exempt an institution from standards for rating systems or documentation.

The use of a credit score for retail segmentation is similar to the use of expert judgement in corporate underwriting, mapping to external rating systems, and using external benchmarks.

These are expressly permitted or required by CAR, even though it is unlikely that institutions will be able to document all the processes behind an expert's judgment, the decisions of an external rating system, or the development of an external benchmark. Similarly, institutions may use credit scores to segment retail risks into homogeneous pools to develop IRB parameters, even without seeing all the development data, and without knowing the precise details of the scoring formula.

A model may use an input if it works reliably under all anticipated conditions. Although comforting, knowing the details of how an input was produced is neither sufficient nor absolutely necessary. For example, an input to structural models of credit risk is stock price, determined by thousands of individual investor decisions that may never be known, much less understood.

The use of credit scores to segment retail risks into homogeneous pools for the estimation of IRB PD depends on the empirical observation that the scores and PD are highly correlated. Scores are developed to predict the risk of default. Given the similarity of the definitions, one would not expect them to be independent. However, institutions should confirm the high degree of correlation.

Institutions are unlikely to want to use credits scores in the management of their retail accounts without having confidence in the integrity of the scores' development and their continuing accuracy in predicting the odds of an account going bad. It is in the interest of developers to provide assurance to institutions.

In summary, institutions may use credit scores to segment risks for IRB estimation without having complete information about the underlying data or model. However, the institution should obtain information and perform analysis to ensure that the scores are relevant to the risks and are properly used. Normally this would include:

1. **From the developer:**
    1. An exposition of the general methodology for developing scores, e.g., a specific type of neural network, logistic or probit.
    2. An understanding of the data available for modeling.
    3. A statement of the purpose of the model, its intended output, and the conditions under which it is expected to work.
    4. Historical performance of scoring models that the developer has built using this methodology.
    5. A statistical profile of the development population.
    6. An explanation of the developer's process to monitor and change the model when necessary.
    7. Contractual undertakings to report the performance of the model and the statistical characteristics of the population against which its performance is measured.
    8. Contractual undertakings to report any changes to the model.
2. **To assure the validity of the behaviour score as a relevant basis for segmentation, the institution should:**
    1. Review industry and academic literature to understand the strengths and weaknesses of the methodology used to develop the score.

2. Review the statistical profile of the development population.
3. Regularly recalibrate the score to the institution's own customers, and confirm that the score accurately predicts the odds of going "bad" (or whatever event the score is designed to predict).
4. Periodically calculate the correlation of the score to the probability of default as defined in CAR.
5. Track the relation of the score to CAR definition of PD, through time.
6. Test the power of the score to discriminate the risk of default.

Appendix II: History of Major Changes

"An institution should document a history of major changes in the risk rating process, and such documentation should support identification of changes made to the risk rating process subsequent to the last supervisory review "(CAR, paragraph 418). Further, under Principle 5: Timing of this Implementation Note, and under the principles included in OSFI's Implementation Note, Risk Quantification at IRB Institutions, institutions should track events and conditions that are likely to affect risk characteristics of their portfolios.

Institutions are expected to use this history as a tool to perform the following:

- identify the need to change rating systems for adjusting estimates;
- decide whether data remain relevant for estimating future outcomes for various exposures;
- adjust parameters as the characteristics of the exposures to which they are applied change; and
- interpret comparisons of observed outcomes against predictions.

Institutions should use their best judgment in deciding what changes, events and conditions should be tracked; however, the following data could be useful for tracking purposes:

- Date of change
- Portfolio affected
- Size of portfolio affected
- Expected effect on PD, LGD, EAD
- Type of change or event
- Institution induced

    - Distribution method
- Adjudication
    - New rating criteria
    - New cut off score
    - New behaviour score
- Management of accounts
    - Reporting
    - Covenant policy
    - Collateral requirements
- Environmental
    - New competing products or method of distribution
    - Changes in employment, housing prices, etc.

Appendix III: Procedures Generally Required in a Validation

This Appendix includes a list of potential validation procedures. Institutions should consider the application (and relevance) of this list to their own portfolios and may need to add or amend procedures according to their internal validation requirements. Some of the under-noted activities may also be performed in the independent review of rating performance as described in Paragraph 443 of CAR. This list is provisional and may not be adequate for all institutions. It is the institution's responsibility to validate, and this may require other procedures.

Replication:

Verification that the rating assignment and risk quantification processes can be replicated following documented procedures and policies.

A review of the logic and conceptual soundness of the rating system:

This should include a review of the implied 'rating philosophy'.

An audit of the information technology providing inputs to the system:

See OSFI's Implementation Note, Data Maintenance at IRB Institutions.

Accuracy testing:

The validation should assess the discriminative power of the rating system and the reasonableness of the estimates of PDs and LGDs using prevailing tests. Institutions should also assess whether their ratings philosophies have been successfully and consistently implemented. For this, an analysis of regularly updated rating transition matrices may be of assistance.

Sensitivity testing:

A validation should analyse the sensitivity of model outputs to model assumptions and to model inputs.

Scenario testing:

A validation should identify possible events or future changes in economic conditions and assess the effect of these scenarios on rating assignment and risk quantification.

Back testing:

A validation should regularly compare model outputs against subsequent real world events and the rating system's actual, realised performance.

An inventory and analysis of the use of the rating system:

Please see OSFI's Implementation Note, The Use of Ratings and Estimates of Default and Loss at IRB Institutions.

A review of comparable external data:

The relevance of external data used and its consistency with internal data should be investigated and fully documented. Often, this will require a comparison of the definitions of default and loss. Institutions should attempt to reconcile internal and relevant external estimates of risk parameters covering comparable risks. In some circumstances, a formal benchmarking to external public rating systems will help confirm internal ratings and PDs. If internal data is limited, institutions should consider using estimates that incorporate some external results.

Special attention to overrides and other exceptions:

Institutions should develop and implement a policy regarding how overrides and other exceptional business are fed back into the ongoing validation framework. All exceptions to the standard model or processing should be identified, documented and reported to those responsible for the design and performance of validations.

Appendix IV: External Data and Retail Validation

CAR calls on institutions' validation procedures to incorporate all relevant, material and available data, information and methods. These principles of validation call on institutions to use data from internal and external sources. Appendix III: Procedures Generally Required in a Validation, calls on institutions to review external data.

Institutions recognize the need to refer to external data in the quantification and validation of ratings and estimates for corporate portfolios, because, on their own, corporate portfolios generally have too few exposures and losses for credible estimates of PD, LGD and EAD. The need to look at external data is not as obvious for the validation of retail portfolios, which usually generate ample data.

Although the sampling error in their internal data will be small, retail institutions may need to look outside their institution (for example, consider macroeconomic data) in their validation of IRB estimates. A review of events and results outside the institution may be useful for:

1. **Establishing the position of the dataset used to estimate parameters in the economic cycle.**

   The retail market and the management of retail accounts evolve rapidly. It is therefore difficult for an institution to distinguish the fluctuation of loss events that arise from changing market conditions or account management from the effects of the economic cycle. A review of the experience of other providers of retail credit may inform the institution's assessment of the relative impact of management and economic factors and the calibration of estimates to achieve a long-term average. Although the external data may not be directly comparable to the product in question, the changes from year to year may inform the institution about macroeconomic events that do not depend on product. Securitisations may provide information about credit card experience. Performance metrics from The Bank of Canada, Statistics Canada, and the Canadian Bankers Association also provide useful information about retail credit performance. Although this data will include the exposures of other institutions with different marketing strategies, the observed fluctuations in aggregate results will likely reflect very general drivers.

2. **Interpreting discrepancies between an institution's long-term averages and results observed in particular years**

   Observed losses that come close to expected losses calculated from IRB estimates do not confirm the accuracy of the IRB estimates *as long-term averages* if all other credit institutions report lighter losses than usual. The results from peers may suggest that the current conditions are favourable, and that long-term losses should be well above current results. Similarly, losses that exceed what is predicted by IRB estimates do not show that the IRB estimates are inadequate if all credit institutions report unusually heavy losses.

3. **Anticipating the effects of changes in the marketplace**

   Changes in the marketplace will affect the amount and quality of business acquired by individual institutions. Institutions may see signs of these changes from a review of internal evidence, but they will have better knowledge of these changes from a survey of industry practices, tabulations of market share, and other external data. Changes in the marketplace may suggest adjustments to estimates based on aging data.

1 Banks and bank holding companies to which the *Bank Act* applies and federally regulated trust or loan companies to which the *Trust and Loan Companies Act* applies are collectively referred to as "institutions".

2 'System' is defined as the combination of people, processes and technology.

3 Risk measures refer to probability of default ("PD"), loss given default ("LGD") and exposure at default ("EAD").

4 For example, a PD conditional on macroeconomic conditions should rise as business conditions deteriorate, while an unconditional PD should remain reasonably stable.

5 Credit scores are often called behaviour or acquisition scores depending on the time of their use and the information available to develop them.