Advancing Credit Risk Management through Internal Rating Systems

August 2005

Bank of Japan

For any information, please contact:

Risk Assessment Section

Financial Systems and Bank Examination Department

Mr. Oyama	+81-3-3277-3078	tsuyoshi.ooyama@boj.or.jp
Mr. Yoneyama	+81-3-3277-2039	masao.yoneyama@boj.or.jp

Table of Contents

I. Introduction

This report updates "Development of Credit Risk Management Based on Internal Rating System" released by the Bank of Japan in October 2001. Since then, financial institutions have been rapidly improving risk management techniques in order to accommodate dramatic changes in their credit risk profiles. In addition, the Basel II Framework[1] was published last year and financial institutions' preparation for adopting the framework has been in progress in tandem with the authorities' domestic rule making. Considering these recent developments, this paper tries to present sound practices of credit risk management through internal rating systems, and also provide some important risk management issues to be further discussed and studied. It is our intention to use the topics in this paper to start in-depth discussions of risk management with financial institutions at the time of our on-site examinations and off-site monitoring and thereby encourage their advancement of credit risk management.

The contents of this report are as follows. In Chapter II, we discuss an outline of internal rating systems, which are a basic tool for enhancing credit risk management. The following chapters draw on sound practices of risk management through internal rating systems, focusing on the architecture of internal rating (Chapter III), rating process (Chapter IV), rating models (Chapter V), estimation of risk components (Chapter VI), uses of internal rating systems (Chapter VII), and validation of internal rating systems (Chapter VIII). In the last chapter, we also discuss the quantification of credit risk. In addition, there are many textboxes throughout the report, designed to elaborate issues which are secondary to the main text but still very important, such as our answers to FAQs from financial institutions and some new and thus not yet established ideas on advancing risk assessment techniques.

To facilitate readers' understanding of the entire picture of credit risk management, Appendix 1 contains two charts, one illustrating the administrative structure surrounding credit risk management, and the other providing an image of advancing credit risk management.

[1] See "International Convergence of Capital Measurement and Capital Standards: A Revised Framework" (June 2004), issued by the Basel Committee on Banking Supervision.

II. Outline of Internal Rating Systems

A. Definition of an Internal Rating System

An internal rating system helps financial institutions manage and control credit risks they incur through lending and other operations by grouping and managing the creditworthiness of borrowers and the quality of credit transactions.

For a long time, many financial institutions managed credit risks by monitoring only the creditworthiness of each borrower. The process for making lending decisions was quite basic, and often only yes/no decisions were made. Even if borrowers went bankrupt, however, losses were often sufficiently covered by real estate collateral.

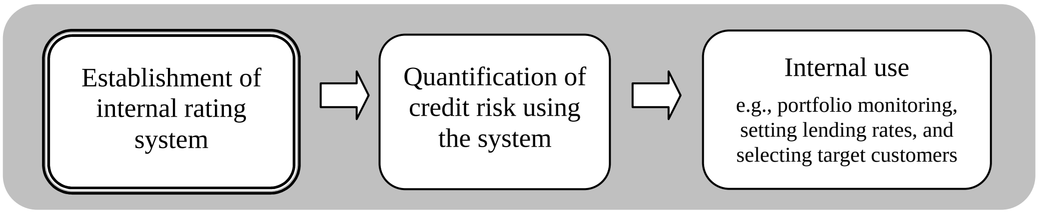
After the bursting of the economic bubble, credit costs increased significantly at financial institutions due to the rising number of bankrupt borrowers and the falling value of collateral. Credit risk control therefore became an important management requiring reconsideration. In addition, changes in the macroeconomic environment forced financial institutions to seek a more reliable credit risk management framework. Hence, the method of statistically managing and analyzing credit risk through an internal rating system became widely used. More financial institutions, including small banks, started adopting internal rating systems to categorize borrowers especially after 1998, when the self-assessment guideline was introduced by the Financial Services Agency (FSA).

B. Benefits of Using an Internal Rating System

What are the benefits for financial institutions of introducing internal rating systems? First, such a system enables banks to efficiently make lending decisions and manage loans with less administrative work. Second, it makes it easier to grasp the creditworthiness of borrowers and the quality of credit transactions using a single yardstick. Moreover, the assessment of the credit quality of the entire portfolio becomes possible by monitoring changes in the amount of credit exposure and number of borrowers in each rating grade. It also becomes possible to quantify credit risk through estimation of the possibility of default by rating grade. These analyses with consistency, comprehensiveness, and objectivity should serve as the foundation of sound bank management.

Improvement of credit risk management is not the only benefits of introducing an internal rating system. The system also gives financial institutions more ideas and a more solid foundation for bank management strategies, such as setting lending rates according to the borrowers' ratings or expanding the target customer base that is the most profitable from the point of view of risk and return (Chart 1).

Chart 1: Advancing Credit Risk Management



The internal rating system is the prerequisite for advanced credit risk management, and each financial institution is expected to develop its own internal rating system. Every institution faces a different business environment, so each system should have its own design. For example, a more simple framework might be suitable for small institutions (Box 1). There is no single answer for the framework of internal rating systems, such as the number of rating grades, a definition of each rating grade, and the method of rating assignments. Financial institutions need to introduce their own system depending on the characteristics of their loan portfolios, their operations, the objectives of the rating system, and other factors. Obviously, the institutions need to make necessary adjustments flexibly due to changes in the business environment.

The following are four important points for consideration in establishing an internal rating system.

- Architecture of the internal rating system
  What are the key factors in building the structure of the rating system?
- Rating process and rating models
  How should information on individual borrowers and loan contracts be linked to the ratings?
- Estimation of risk components and quantification of credit risk
  How should components and sizes of risk be calculated using rating information?
- Active use and validation of the system
  How should the system continue to operate smoothly and be improved continually?

Box 1: Does the Size of Financial Institutions Matter in Discussing the Need to Establish Internal Rating Systems?

It is sometimes argued that small financial institutions with a relatively small number of borrowers do not need internal rating systems. This is based on the perception, for example, that a statistical approach is not necessarily efficient in managing credit risk of loan portfolios with a limited number of borrowers and that the establishment and management of a system is far too expensive.

However, even small institutions can have a considerable number of borrowers, whose credit quality may vary significantly. Thus, it is useful for banks of any size to have a system that categorizes loans into rating grades, and assesses objectively the relative creditworthiness of borrowers. This also enables financial institutions to conduct more efficient risk management through close monitoring of high-risk borrowers, for example, by taking preemptive actions against further downgrades of already low-rated borrowers or recently downgraded ones. Furthermore, financial institutions can gain a better understanding of the quality of their entire loan portfolios in a comprehensive way by monitoring developments in internal ratings, including the degree of concentration on certain borrower ratings and the number of upgrades/downgrades.[2]

Some might argue that it is possible for lending officers and bank management to grasp their bank's loan portfolios and be aware of the risk involved in low-rated borrowers without relying on an internal rating system. Nevertheless, we can make the case for internal rating systems as they avoid a monopoly of important credit risk information by a limited number of staff, thereby contributing to consistent and transparent management. Financial institutions should also be aware that such a system helps increase the accountability to their stakeholders and the financial authorities.

We certainly understand that the establishment and management of internal rating systems could be expensive for some institutions. Thus, each institution needs to determine the cost relative to the return and come up with its own design. Some banks obviously could start with a simpler approach than described in this paper and gradually move to a more advanced one when they judge it could repay their investment.

[2] Internal rating systems with a higher level of accuracy enable financial institutions to estimate risk parameters that characterize loan portfolios in a quantitative and thus more objective way.

III. Architecture of Internal Rating Systems

A. Borrower and Facility Ratings

Internal rating systems are used to assign grades either to individual borrowers (borrower ratings) based on creditworthiness, or individual loan transactions (facility ratings) based on riskiness. Borrower ratings focus on the credit risk of borrowers, in other words, whether borrowers will default or not (the possibility of default) (Chart 2). This is the most standard type of rating system and is used by many financial institutions. Meanwhile, facility ratings focus on risk exposures of each transaction. In assigning grades, facility ratings take into account the collateral or guarantee pledged to loans, and the maturity, in addition to the creditworthiness of borrowers. With this rating system, different grades can be assigned to loan contracts with one borrower depending on the degree of risk exposure for each transaction.

Chart 2: An Example of Borrower Ratings

Borrower rating	Level	Borrower classification
1	Excellent	Normal
2	Prime	
3	Good	
4	Above standard	
5	Standard	
6	Below standard	
7	Needs attention (1)	Needs attention
8	Needs attention (2)	
9	In danger of bankruptcy	In danger of bankruptcy
10	Bankrupt	De facto bankrupt & bankrupt

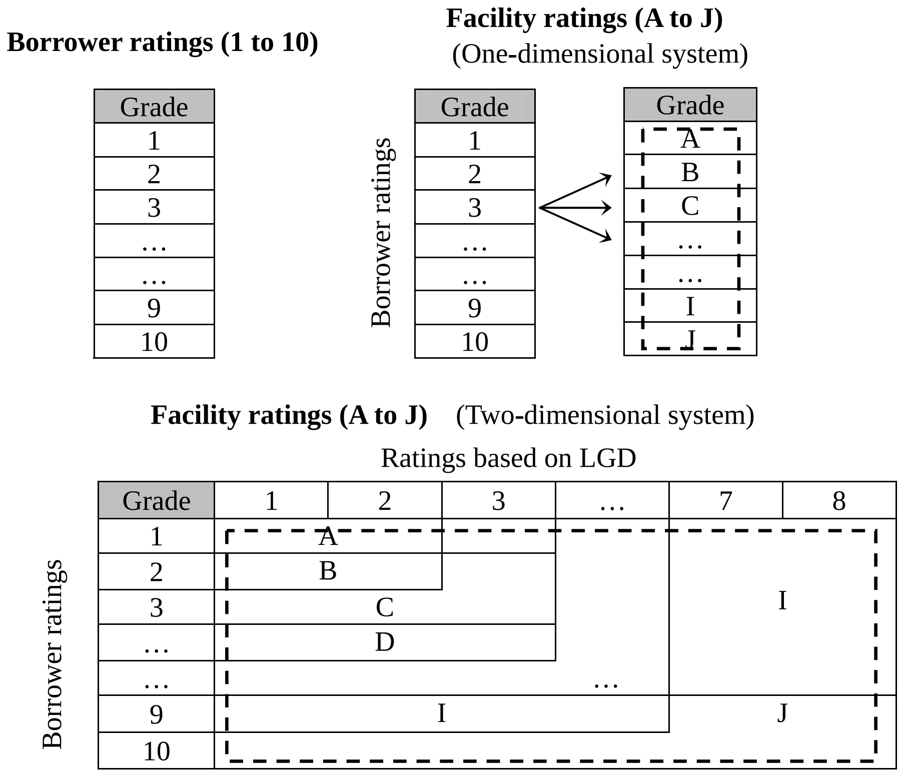
When judging a grade, probability of default (PD) is a criterion for borrower ratings. Expected loss (EL) rate, that is, PD multiplied by loss given default (LGD), is used for facility ratings to measure risks associated with individual transactions.[3]

There are two types of facility ratings, which are more elaborate than borrower ratings, namely, one-dimensional and two-dimensional systems (Chart 3). A one-dimensional system bases facility ratings on borrower ratings and makes upward or

[3] Estimation method of risk components, such as PD and LGD, is discussed in Chapter VI.

downward adjustments to the grades as necessary to reflect the characteristics of the loan transaction concerned (such as the secured level of loans in terms of the ratio of collateral or guarantee pledged to loans). A two-dimensional system combines borrower ratings with evaluation of the features of individual loan transactions independent of borrowers (e.g., ratings based on LGD).

Chart 3: Borrower and Facility Ratings



Financial institutions in Japan widely use borrower ratings mainly for loans extended to firms (corporate loans), while facility ratings are only used for specialized loans that require management on a per loan basis. The latter include loans extended for project finance, real estate finance, and structured finance.

One of the factors hindering the spread of facility ratings for corporate loans in Japan is its unique pledge system in lending, such as “pooled” collateral and revolving guarantee where each collateral or guarantee is not linked to each transaction but to each borrower. This makes it difficult to assess the exposure of one transaction individually and separately, which is essential to facility ratings.

B. Scope of Ratings

In principle, ratings should be assigned to all borrowers and transactions. A rating system embodies a financial institution's stance toward evaluation of credit risk. The system should, therefore, be applied to all credit risk with consistency. This will in turn promote accuracy and efficiency in the ratings. Depending on the size and nature of transactions, however, it may be too costly to assign grades to all transactions. Such cases need to be treated as exceptions, but distinct criteria should be established for such treatment. Nevertheless, all major borrowers and transactions should be subject to ratings.

Unlike the case of corporate loans, however, small-sized loans including personal loans for individuals and uncollateralized business loans for small firms could be better fit into credit risk management based on segmentation by product type (e.g., residential mortgage and card loans) and characteristics of borrowers (e.g., occupation and age) (Chart 4).[4] In managing such retail loan portfolios, statistical similarities by transaction type and borrower characteristics for each portfolio, for example average PD for a specific segment, rather than differences among borrowers, are more useful in achieving more accuracy and efficiency in credit risk management. To accurately grasp risks using statistical data, risk profiles of transactions in each group should be sufficiently homogenous. Such homogeneity may be assured by, for example, evaluating the stability of default rates and other data on risk factors for each group.

Chart 4: An Example of Segmentation of Retail Loan Portfolios

<table>
<tr><th>Product type</th><th>Type of borrower characteristics</th><th>Delinquency</th></tr>
<tr><td rowspan="6">Residential mortgage loans</td><td rowspan="3">A</td><td>No delinquency</td></tr>
<tr><td>Past due 1 month</td></tr>
<tr><td>------</td></tr>
<tr><td rowspan="3">B</td><td>No delinquency</td></tr>
<tr><td>Past due 1 month</td></tr>
<tr><td>------</td></tr>
<tr><td rowspan="3">Credit card loans</td><td rowspan="3">A</td><td>No delinquency</td></tr>
<tr><td>Past due 1 month</td></tr>
<tr><td>------</td></tr>
<tr><td>------</td><td>------</td><td>------</td></tr>
</table>

[4] Segmentation of retail loan portfolios may also be based on the starting date of credit extension, number of years elapsed, and the delinquency.

C. Rating Grades

Having an appropriate number of grades and their definitions also affects the adequacy of credit risk management (Box 2). The number of grades should ensure that borrowers and transactions with a similar level of risk are grouped together.[5] Obviously, financial institutions dealing with many borrowers with varying degrees of creditworthiness should have a relatively large number of grades. If, despite this, borrowers concentrate in a particular grade, it may be a sign that credit risk is not being evaluated efficiently and accurately. The definition of each grade should be reviewed in such cases.

Too many grades for a small number of borrowers may also be inappropriate, because accuracy of risk estimation may decline if the number of borrowers in each grade is too small. In some cases, external sample data may be utilized and adjacent grades may be combined as long as the homogeneity of credit risk in each grade is maintained.

Even when a portfolio has no specific concentration in one grade at the time the rating system is established, factors such as changes in cyclical business conditions can cause a shift in grades to one direction, which then causes a concentration. In such cases, the review of rating definitions and the number of grades should be considered, although the stability and sustainability of rating systems are also important.

Box 2: Optimal Structure of Ratings

Although there are no specific criteria in deciding the number of grades and their definitions, the range of each borrower's probability of default and the number of borrowers in each grade are two important factors for borrower ratings.

These two are usually in a trade-off relation to each other. The more the number of grades, the smaller the range of default probability, and the more homogenous each grade. However, in this situation the smaller the number of borrowers in each grade, the larger the estimation error of default probability.

It is therefore desirable to establish the number of grades and their definitions in a way that minimizes the sum of the variance of default probability and the

[5] Banks which adopt the internal ratings-based (IRB) approach in the Basel II Framework are required to have seven or more non-default grades and one or more default grades.

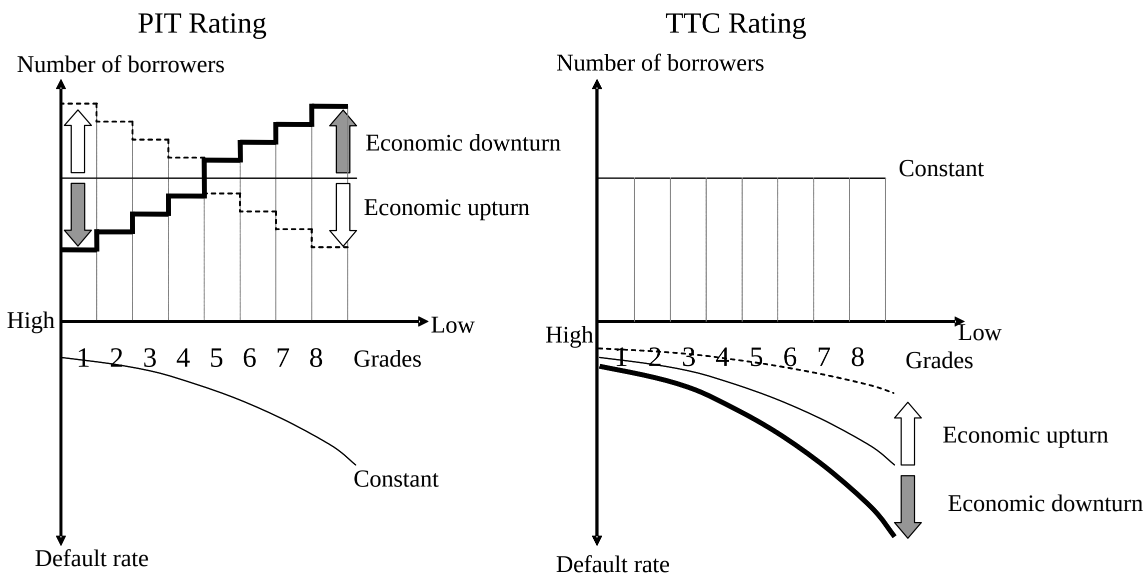
estimation error of default probability. Consequently, the rating structure using these factors should be considered optimal if the sum of disparities between estimated figures of default probability per grade and each borrower's default probability is kept minimal.

D. Rating Assignment Horizon—Relationship with the Business Cycle

The time horizon of assessing the creditworthiness of borrowers in assigning ratings is also important. Two different approaches may be taken in considering the effect of the business cycle in assigning ratings. One is a point-in-time system (PIT rating). In PIT rating, risks are evaluated based on the current condition of a firm regardless of the phase of the business cycle at the time of evaluation. The other is a through-the-cycle system (TTC rating). In TTC rating, risks are taken into account on the assumption that a firm is experiencing the bottom of the business cycle and is under stress.

Chart 5 illustrates changes in grades based on PIT and TTC ratings in relation to the business cycle. In PIT rating, grades fluctuate reflecting the business cycle and thus ratings tend to be upgraded at economic upturns and downgraded at economic downturns. In addition, ex-post default rates per grade are stable regardless of the business cycle. On the other hand, in TTC rating, grades of firms remain the same through the business cycle, and ex-post default rates within the same grade fluctuate reflecting the business cycle. In this way, the difference between PIT and TTC ratings shows up in this behavior of rating changes and default rates per grade.

Chart 5: Point-In-Time (PIT) and Through-The-Cycle (TTC) Ratings



Rating agencies usually assign grades from a long-term perspective and their choices, therefore, are considered to approximate those based on TTC rating. Meanwhile, few financial institutions seem to clearly make a choice between PIT and TTC ratings. They seem to evaluate the creditworthiness of borrowers over some period, for example, three to five years, indicating that their choice is somewhere between the above two types of ratings.

Choice between PIT and TTC ratings or a mixture of the two depends on the length of time financial institutions are exposed to credit risk. If the majority of a bank's loans have a long time before maturity, it is desirable to assign a grade considering creditworthiness over the whole period. However, it is difficult in practice to assess the future change in a borrower's conditions over the long term including the business cycle as in the case of TTC rating. One alternative might be to adopt an approach in which ratings are based on recent conditions and assess the degree of credit risk under the economic downturn by use of a stress test. This approach is based on PIT rating but takes account of TTC components in application.

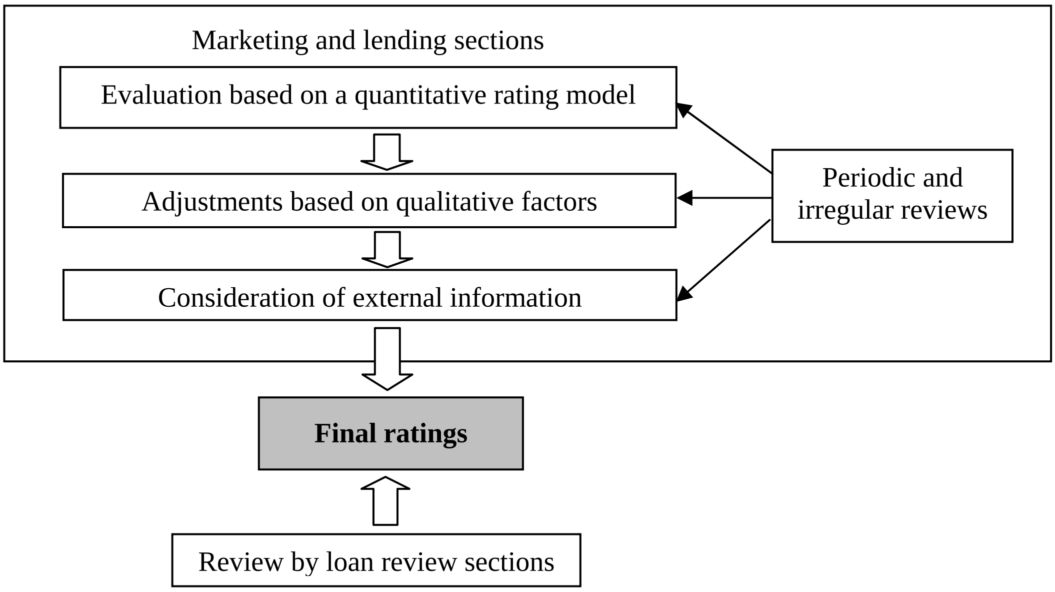
Though we cannot judge a priori which rating method is better than the other for certain banks, it is still very important for financial institutions to understand whether their own internal rating systems are more PIT-oriented, TTC-oriented, or follow a mixed approach, in other words, how their systems are affected by the business cycle. This is because the assessment of an internal rating system using actual default rates per grade depends on whether PIT or TTC rating is used. With PIT rating, default rates per grade are expected to be stable and hence their stability is confirmed, whereas with TTC rating, they will fluctuate over the business cycle.

IV. Rating Process

A. Assigning and Reviewing Ratings

In the process of internal rating, marketing and/or lending sections assign grades in accordance with established policy and procedures. The process generally includes the following steps: (1) evaluation based on a quantitative rating model; (2) adjustments based on qualitative factors, such as industry trends; and (3) consideration of external information,[6] such as ratings by outside agencies and stock prices (Chart 6).[7] Grades assigned through the process are reviewed by loan review sections.

Chart 6: Example of a Rating Process



Rating grades should always appropriately reflect risks associated with borrowers.[8] Reviews, therefore, are necessary to reflect changes in the creditworthiness of the borrowers. These include periodic reviews of the timing of disclosure of financial statements as well as irregular reviews carried out when there are significant changes in the creditworthiness of borrowers, such as default of a large trade counterparty.

---

6 It is desirable to use external information in verifying assignment of grades even if it is not used to determine initial ratings.

7 Internal rating systems are not only a tool for advancing credit risk management but also bases for making accurate "self-assessment of credit risk" and adequate write-offs and loan-loss provisioning. Thus, it is necessary to maintain consistency between borrower categorizations and self-assessment of credit risk.

8 If assignments are based on PIT rating that reflects changes due to the business cycle, it would be

The process and architecture of the internal rating system should be (1) documented, for example, in the form of internal rules or manuals; (2) approved by the management; and (3) made widely known to related sections. Important information for designing the system should also be made available in written form.

B. Quantitative and Qualitative Evaluation

For borrower ratings, grades for each borrower are usually decided based on evaluation using quantitative information, such as financial indicators regarding the borrower, and qualitative information, such as industry trends, and extension of financial support from parent companies. A quantitative rating model is often used for quantitative evaluation of individual borrowers. Financial indicators that have a close statistical relationship with defaults are used in this model (Chart 7). Furthermore, in many cases, it is better to use substantive financial data reflecting borrowers' financial conditions that are not necessarily captured by accounting data, such as nonperforming assets and unrealized losses.[9]

If quantitative financial data are insufficient to accurately measure the creditworthiness of borrowers, qualitative analysis should be used to make necessary adjustments (Chart 8). Specifically, (1) qualitative factors may be expressed in terms of scores that are either added to or subtracted from scores reflecting quantitative evaluation, or (2) grades based on quantitative evaluation may be upgraded or downgraded to reflect qualitative factors.

Chart 7: Examples of Quantitative Factors Determining Borrower Ratings

Type of factor	Examples
Size of operation	Amount of capital and net assets
Safety	Current ratio, capital adequacy ratio, and current account balance ratio
Profitability	Return on assets, operating profits, years required to pay back interest-bearing liabilities, and interest coverage ratio
Others	Rate of growth in sales and profits

necessary to revise grades relatively frequently.

[9] Substantive financial data provide more reliable information than accounting data. However, they have disadvantages in terms of low objectivity and high cost of data collection. In the choice of financial data, these pros and cons should be taken into account.

Chart 8: Examples of Qualitative Factors Determining Borrower Ratings

Type of factor	Examples
Industry	Growth potential, size of market fluctuations, and entry barriers
Firm	Ownership relations with parent companies or affiliate firms, management's ability, and existence of an external audit system

For facility ratings, quantitative and qualitative information on each transaction is necessary in addition to that used in borrower ratings. For corporate loan exposures, it includes information on type of collateral, guarantees, seniority, and maturity. This information is highly related to LGD for the transaction, whereas data used to assign borrower ratings reflect PD.

Loans for real estate finance and project finance are, unlike corporate loans, mostly managed solely by facility ratings. Quantitative and qualitative factors as listed in Chart 9 are often used to assess the quality of a project.

Chart 9: Examples of Factors Determining Facility Ratings

Type of factor	Real estate finance	Project finance
Quantitative	Credit extension period, loan to value (LTV), and debt service coverage ratio (DSCR)[10]	Credit extension period and DSCR
Qualitative	Characteristics of real estate, e.g., locations and other conditions, adequacy of cash flow schedule, and risks attached to sponsors of the project	Risk attached to the project, e.g., risks attached to sponsors and operators of the project, the risk of being unable to complete the project, and transfer risk

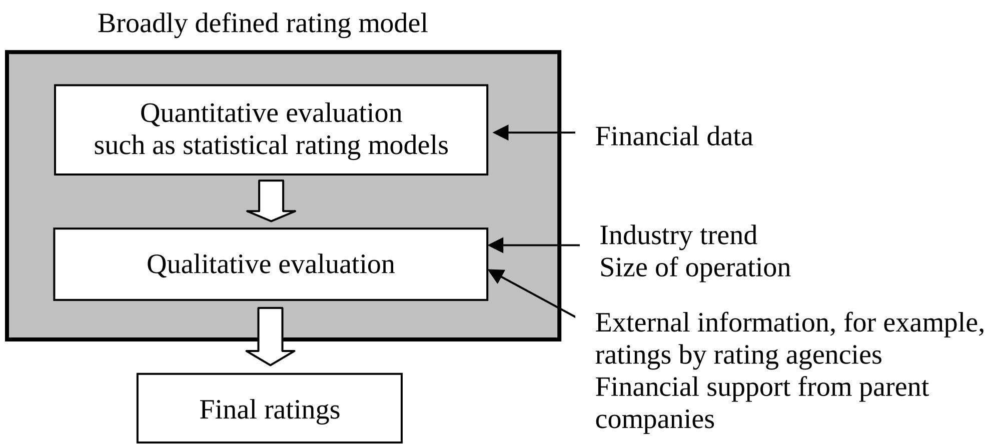
Detailed evaluation criteria are necessary for qualitative evaluation. It is often difficult to secure objective and consistent judgments in qualitative evaluation. Criteria, therefore, should be as specific as possible and documented in detail to avoid subjective judgments. In addition, it is necessary to ensure common understanding among credit rating staff of the criteria of qualitative evaluation by, for example, holding internal training seminars.

[10] LTV is calculated by dividing outstanding credit by appraised value of real estate. DSCR is calculated by dividing annual cash flow generated from the assets concerned by the amount of repayment of principal and interest.

V. Rating Models

A rating model can be broadly defined as a systematic process of quantitative and qualitative evaluation in rating assignment (Chart 10), and is the centerpiece of internal rating systems. These models use financial and other information on the creditworthiness of firms to determine ratings objectively. Rating models are expected to promote (1) efficiency in the rating process in marketing and lending sections, and (2) stability and objectivity in credit risk assessment within a financial institution by decreasing discrepancies in evaluation made by rating staff. This article uses this broad definition for rating models to cover the whole process, not just statistical rating models to which the term sometimes refers.

Chart 10: Outline of a Rating Model



A. Outline of Rating Models

There are many types of models including the ones that use information on the financial conditions of firms, that is, statistical rating models,[11] and scoring models. The use of financial data also varies. In some cases, data may be input directly into calculation formulas. In other cases, each item of financial data is analyzed and transformed into credit scores for some categories and then all the data are input into calculation formulas.

Financial institutions may use any of these models. To pursue accuracy, they can use different models for different industries, size of operation, and loan categories.

---

[11] Statistical rating models include discriminant analysis models, logistic regression models, and neutral network models.

Others may use only one model to maintain consistency. In the former case, the amount of estimation sample data available for each model usually decreases, and as a result, the robustness of each model might decrease. On the other hand, in the latter case, there is concern that the accuracy of the model declines since a single model is not able to fully capture the characteristics of different industries and firm size. Each financial institution should choose a model that is most appropriate to it, in consideration of the advantages and disadvantages relative to the risk profile of its loan portfolio.

B. Validation of Rating Models

Rating models are critical tools of internal rating systems and efforts should, therefore, be made to improve their accuracy. Specifically, models should be validated both when they are being developed (ex ante validation) and when they have been in operation (ex post validation) (Chart 11).

Specifically, models are mainly validated in the following ways. At the development stage, the logic behind the model is checked and the adequacy of data used for designing the model is confirmed. The adequacy of financial indicators used as input (explanatory) variables should be checked for the quantitative rating model. Ex post validation includes examination of the performance of the model based on the default conditions of borrower firms in each grade. For example, data on defaulted firms and non-defaulted firms are collected to check the accuracy of rating using scoring models.

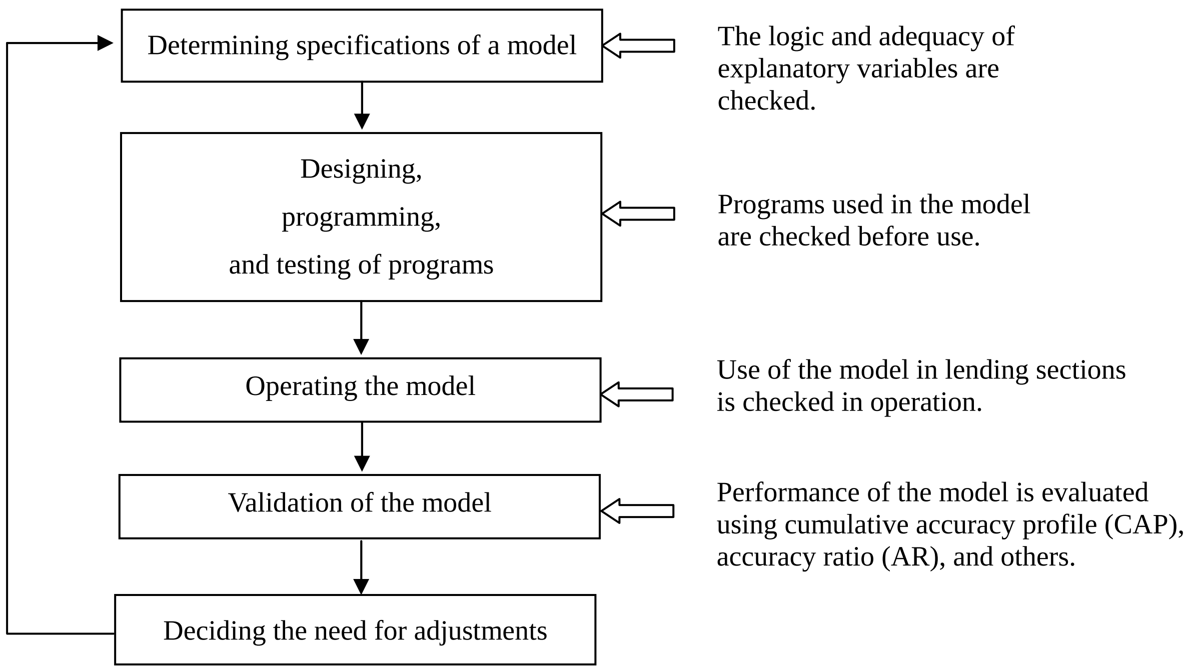
Various approaches should be taken in validating rating models. In credit risk analysis, statistical validation often faces some difficulties. This is due to insufficient accumulation of data on defaults for statistically examining a model. This case might be supplemented by qualitative analysis and warrant examination on a continuous basis.[12]

Sections below present examples of ex ante and ex post model validation methods.

---

[12] Some rating models use more than one quantitative financial model and some use subsystems that make qualitative adjustments. The function of each subsystem should be validated in such cases.

Chart 11: Building and Validating a Rating Model



1. Ex ante validation at the model development stage

a. Validation of the logic behind the model

Ex ante validation includes verifying the adequacy of the logic supporting the model and the development of the model to see whether it properly reflects this logic. When a third party such as loan review sections conducts the validation, it should check whether computer programming codes are appropriately developed based on specifications on the logic. It is also necessary to conduct checks by inputting test data into the model.

b. Validation of input data

The adequacy of financial indicators and other input data used in the rating model is also checked. Specifically, financial indicators should be reviewed based on the conventional wisdom and those clearly unrelated to the creditworthiness of firms should be eliminated. In the case of logistic regression and other regression models, qualitative checks are made on explanatory variables by confirming that plus and minus signs are correct and quantitative checks are made by using statistical values such as t-values, F-values, and Akaike's information criterion (AIC).[13] It is also important to check the problem of multicollinearity, which refers to a decrease in the model stability due to a choice of correlated explanatory variables.

Outliers (extremely different values from the overall average) should also be eliminated from sample financial data that are used to develop the model. This is because outliers may weaken the accuracy of the model.

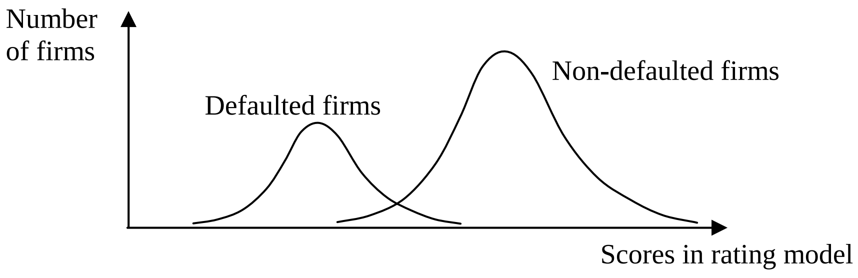
2. Ex post validation using default samples

a. Backtesting

Backtesting is a comparison of estimates and actual outcomes in order to measure the accuracy of the model. Models with low accuracy evaluate defaulted firms as highly creditworthy, and vice versa.

One way to check the model performance is to sort all borrowers in the sample with the smallest to largest scores calculated using a rating model and to analyze the distribution of defaulted and non-defaulted borrowers (Chart 12). If the model predicts defaults with high accuracy, firms with low scores are more likely to default and the overlapping area in Chart 12 decreases. It is important to confirm such a tendency on a graph and to calculate statistical values on the basis of this distribution to objectively measure the accuracy of the model. See Appendix 2 for examples of the accuracy ratio (AR) and other generally used quantitative indicators.

Chart 12: Distribution of Scores of Defaulted and Non-Defaulted Firms

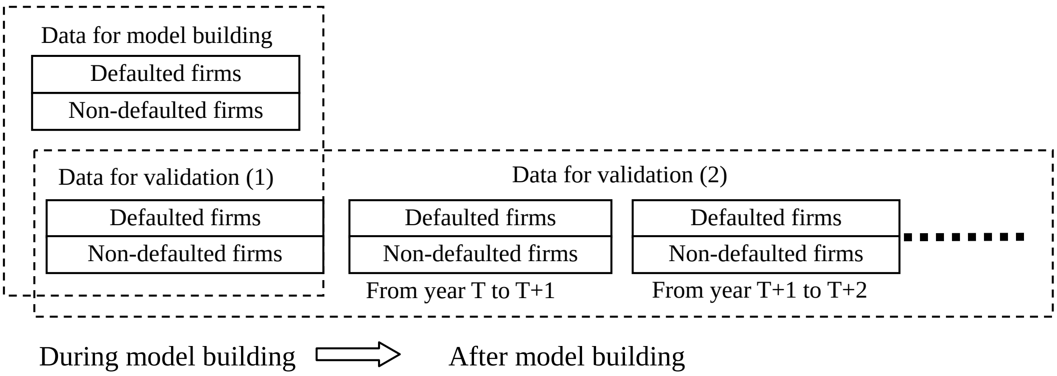


[13] t-values and F-values are used in a regression model to validate the statistical significance of explanatory variables. AIC is an indicator evaluating the efficiency of use of information, in other words, validity of the combination of various explanatory variables.

Box 3: Caveats concerning Use of Quantitative Indicators for Model Validation

In using accuracy ratio (AR) and other quantitative indicators for model validation, it is important to note that the evaluation results based on the data for model building and the ones based on the data for validation could be quite different. Specifically, a model that shows high accuracy when tested with sample data at the development stage of the model may not always show the same results when tested with other sample data. It is, therefore, necessary to use a different set of samples for ex ante and ex post validation of the model using statistical values.

Also, data for validation used to analyze the stability of AR should include historical data covering different periods. A model may require fine-tuning if it performs badly with one set of data even if it performs well with another covering a different period.



3. Ex post validation using data on rating migration and default rates

In addition to default samples used above, data on rating migration (changes in grades) and default rates may be used for ex post validation of the accuracy of a rating model. For example, the accuracy of the model should be doubted if many borrowers estimated to have low PD are downgraded considerably in the following fiscal term. Methods for validating a model focusing on irregular changes in creditworthiness include monitoring of rating migration and order of default rates by rating (Chart 13).

A rating migration matrix shows changes in grades in a financial institution's portfolio over a certain period. Using this matrix, firms with large changes in grades (outliers) are specified. Migration of ratings is then analyzed in detail to find factors behind the migration. These procedures help evaluate the adequacy of quantitative models and qualitative adjustments, in other words, the overall rating model.

In addition, the matrix generally shows that the migration rate becomes small as the distance lengthens between the rating at the beginning and the rating at the end of the year. Based on the outcome of the actual rating transition, it is necessary to confirm this tendency and also to look into what causes the irregular order, if any. Another validation method is to check the consistency between actual default frequencies and rating scales. In other words, default rates should not be larger for the upper grades than for the lower ones.

It should be noted that inaccuracy of the rating model is not the only cause of inconsistency. Inconsistency in the rating scale may occur when default rates in a certain grade become unstable because the number of borrowers within the same grade is too few or a strong default correlation exists among the borrowers.

Chart 13: Example of Validation Using Data on Rating Migration and Default Rates

Check the order of migration rate from rating 1 to other ratings

Check the order of migration rate to rating 7 from other ratings

Check the order of default rates

Rating at the end of the year

Rating at the beginning of the year	1	2	3	4	5	6	7	8	Default
1	83.1	12.8	2.1	0.3	0.4	0.3	0.5	0.5	0.0
2	4.7	75.4	15.1	3.3	0.7	0.3	0.2	0.2	0.1
3	0.2	11.9	66.5	13.9	4.2	1.5	1.0	0.8	0.0
4	0.0	1.4	13.3	63.1	13.1	4.4	2.5	1.9	0.3
5	0.0	0.4	4.4	24.5	44.0	15.7	6.0	4.5	0.5
6	0.0	0.1	1.5	7.5	20.4	43.9	16.0	9.5	1.1
7	0.0	0.0	0.5	2.8	6.8	18.9	47.8	20.0	3.2
8	0.0	0.0	0.4	1.6	2.1	2.6	3.8	74.7	14.8

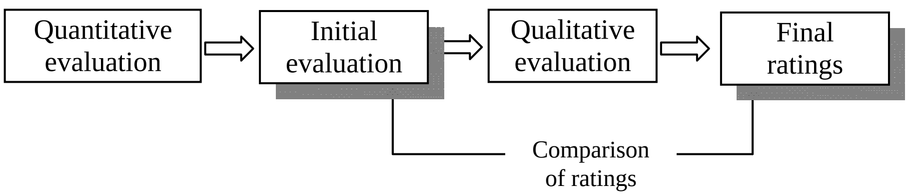
4. Performances of quantitative and qualitative evaluation

For a rating model that combines quantitative evaluation and qualitative adjustments, it is also important to analyze the process followed by rating staff in assigning final ratings. For example, the adequacy of a quantitative rating model may be questioned if final ratings tend to differ greatly from the result of initial evaluation using the model and the weight of qualitative adjustments becomes larger as a result (Chart 14). These model validation and application review processes are especially important for financial institutions that employ quantitative models as a main tool of internal rating systems.

Chart 14: Comparison of Initial Quantitative Evaluation and Final Ratings

%	Final ratings				
Initial evaluation	Down by two or more grades	Down by one grade	Same as initial evaluation	Up by one grade	Up by two or more grades
1	0	15	85	---	---
2	1	9	88	2	---
3	3	10	82	4	1
4	5	11	78	5	1
5	9	22	62	6	1
6	10	27	60	3	0

Initial evaluations are very different from final ratings under 5 and 6



It is also necessary to examine whether the final ratings adjusted by qualitative evaluation are appropriate ex post (Chart 15). One of the methods is to check whether actual default rate for each group differs, for borrower groups which have the same rating in the initial evaluation but a different one in the final rating stage. In addition, it is necessary to check the default rate for borrower groups which have a different initial rating but the same final rating.

Chart 15: Validation of Final Ratings Adjusted by Qualitative Evaluation

Ratings before migration			After migration	
Initial evaluation	Final ratings	Number of borrowers	Number of defaults	Default rate (%)
6	5 or more	156	0	0.0
	6 (no change)	223	3	1.3
	7 or less	147	8	5.4
	Total	526	11	2.1

Check the order of default rates

Ratings before migration			After migration	
Initial evaluation	Final ratings	Number of borrowers	Number of defaults	Default rate (%)
5 or more	6	112	2	1.8
6		223	3	1.3
7 or less		117	7	6.0
	Total	452	12	2.7

Check how far the default rates are from average rate

Possibility of excess upgrade of "7 or less" under the initial evaluation

Box 4: Caveats concerning Use of External Data and Models

An external database may be used in the internal rating system. It is used in various ways. For example, financial institutions with limited accumulation of data on firms' financial conditions and defaults may use an external database to supplement their data. In such cases, the adequacy of external data to be added to the user's portfolio should be confirmed. Specifically, characteristics of firms in the sample data that greatly influence default frequencies, such as the firm size, industry, and region, should be similar to the user's portfolio.

Even external rating models may be used in some cases. For example, external models designed by rating agencies that have a large set of historical data may be used to assign grades to large firms with good standing for which only a few default samples exist. Financial institutions, in doing so, should have a sufficient understanding of the rating philosophy underlying the model, for example, PIT or TTC ratings, and the design of the model used. It is also necessary that they validate the models themselves rather than depend only on reports made by the model vendor. The same applies to financial institutions that develop rating models by outsourcing the procedures to an outside agency.

C. Adjusting Rating Models

The final step is to decide the need for adjusting the rating model given the results of model validation. Specifically, it is important to judge which part of the models, for example, the quantitative rating model or the logic behind the qualitative evaluation, should be adjusted and the degree of adjustment to be made. Frequent adjustments are costly and may break the continuity of the rating method. Adjustment of the model should be carefully decided in view of these aspects. A realistic approach is to make drastic changes to the model only if its performance has clearly declined and to make small changes in other cases.

When major changes are made to the logic behind the model or its parameters, for example, financial indicators, or when the model is replaced by one built by an outside third party, careful comparison of the results of the old and new models is necessary.

VI. Estimation of Risk Components

As already mentioned, one of the important points in the validation of a rating model is to examine whether the model appropriately reflects risk components in ratings, for example, the PD for each borrower. In view of the validation of credit risk quantities for loan instruments, the LGD and the exposure at default (EAD) are also key factors to be evaluated. Thus, these three factors, called "risk components," play a crucial role in the assessment of the credit risk quantities for individual loans.

In this chapter, we describe the basic procedures for estimating and validating the three risk components.[14]

A. The Basic Process of Estimating Risk Components

The basic process for estimating and validating PD, LGD and EAD is summarized below (Chart 16).

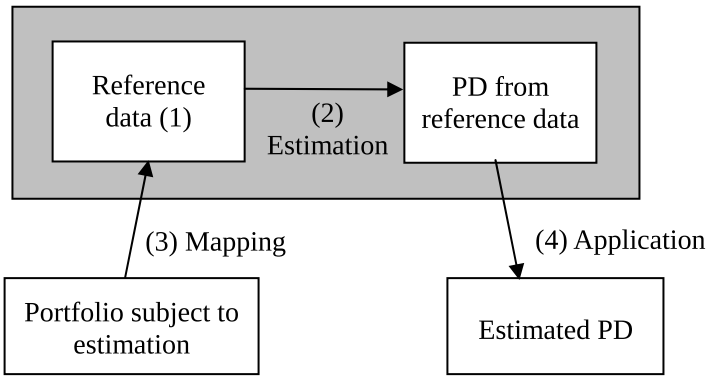
(1) Collect reference data.

(2) Estimate risk components with reference data.

(3) Map out the correspondence with the reference data and current portfolio data.

(4) Apply and adjust the estimated value of risk components in (2) according to the results of (3).

Chart 16: Estimation Process of Risk Component (PD)



In using the reference data not originating from the loan portfolio concerned, steps 3 and 4 above need to be implemented. We show the specific procedures to

[14] The three risk components are essentially random variables. Thus, "estimating component values" in this paper means calculating expected values of random variables.

estimate PD in the case of using internal data and external data as below.

Example 1: PD estimation using internal data

(1) Obtain reference historical data, for example, covering a period of five years.

(2) Calculate a simple annual average for PD using the above historical data.

(3) Assume that the internal rating structure of the current portfolio is equivalent to the past one.

(4) Apply PD estimated in (2) above to the current portfolio.

Example 2: PD estimation using external data

(1) Obtain data provided by a rating agency as reference data.

(2) Find out the long-term average of PD for each grade released by a rating agency.

(3) Map out internal and external rating scales.

(4) Apply PD in (2) above by making necessary adjustments according to the results of (3), for example, to reflect the differences in the definition of a default.

The following sections give a detailed explanation of the estimation process for each risk component.

1. Probability of default (PD)

PD is the likelihood of a borrower defaulting within a certain period in the future. In the estimation of PD, the definition of a default needs to be clarified by making clear what specific state of a borrower should be considered default. A common practice is to define the downgrading of a borrower to a certain rating grade or below as a state of default.

The choice of a rating grade for the default threshold depends on the policy of business and risk management of financial institutions. Several candidates for a criterion are considered, for example, legal bankruptcy or deterioration of loans to the following categories: loans to borrowers who became "bankrupt," borrowers who are "de facto bankrupt," , borrowers "in danger of bankruptcy," and borrowers that need "special attention."

One common idea is that default is a situation in which the recovery of a loan is no longer affected by the creditworthiness of borrowers, but instead by factors independent of borrowers' creditworthiness, such as collateral. This approach seems dominant among Japanese financial institutions. Accordingly, many institutions regard borrowers "in danger of bankruptcy" or lower under the credit risk self-assessment framework as being in default.

Another example for a definition is that default is a situation in which financial institutions have an incentive to remove loan assets from their balance sheets or take drastic measures to revitalize borrowers after reevaluating the loans to reflect the market price. In this case, default is considered to be borrowers that need "special attention" or those in lower categories under the self-assessment framework.[15]

There are various methods for estimation of PD, from the calculation of long-term averages based on internal data, the mapping to external ratings, and the model approach to estimate PD directly. Most of the methods are subject to inaccuracy arising from the paucity of default data. Therefore, the accumulation of sufficient default data is critical to PD estimation. When samples of internal data are too small for statistical inference, external data from a shared database and others can be used as supplementary data. However, the use of external data requires careful attention to the similarity in the composition of firm data, such as industrial sectors, size of business, and geographical areas, between internal and external databases. Equally, the difference in the definition of default between the internal and external data should be carefully examined. If a large discrepancy of this kind exists between databases, a necessary modification should be implemented in order to make estimates of PD reasonably conservative.

2. Loss given default (LGD)

LGD refers to the ratio of expected loss relative to credit exposures at the time of default. It may be defined in terms of recovery rate by denoting "1 – recovery rate." Here, losses should be measured in economic terms. LGD estimation, therefore, includes all the costs for the process of collection including payments to loan servicers.

[15] Definition of default according to the Basel II Framework is as follows: (1) when an obligor is unlikely to pay its credit obligations, for example, distressed restructuring, or (2) when an obligor is past due more than 90 days on a credit obligation. This definition may be considered as being close to the definition of loans to borrowers that need "special attention," which is a borrower categorization used in Japan.

As in the case of PD, long-term historical data on defaulted assets is essential. Such data should include peaks and troughs of the business cycle in consideration of the fact that economic conditions may affect the amount of collection by changing the value of collateral and other factors. In building a reference database for estimation, other information potentially affecting LGD also needs to be gathered. This includes provision of collateral, collateral type, collateral coverage, and borrower characteristics,[16] such as industry, geography, and creditworthiness (see Appendix 3 for an example of LGD estimation).

In Japan, real estate collateral continues to be a major form of collateral. It is, therefore, very important to monitor trends and distributions in real estate values and to find out how long it takes to sell the real estate collateral in the case of default. In the past, Japanese financial institutions often experienced additional losses than expected in the final disposal of nonperforming loans. This may be attributed to optimistic estimates concerning (1) developments in the value of real estate collateral, and (2) time required for the final disposal of loans. These additional losses should be included in LGD estimation. If the value of real estate collateral strongly reflects the creditworthiness of the borrowers and may drop at times of default, estimation of LGD should take such possibilities into consideration.

3. Exposure at default (EAD)

EAD is the amount outstanding of credit at the time of default. For on-balance assets such as loans and bond securities, EAD equals the amount of principal or the book value. Estimation, however, is needed for off-balance assets such as commitment lines because EAD equals the current outstanding plus an estimate of additional drawings up to the time of default.

It should be noted that the correlation between the amount of the off-balance assets and the creditworthiness of a borrower is one factor in EAD estimation. This is because the exposures tend to increase rapidly as the default of a borrower becomes more likely. Financial institutions should make efforts to build a sufficient database on these transactions, data for which are generally rare. In the meantime, financial institutions are expected to make conservative estimates.

[16] Borrower information, such as industry, is basically linked to PD but may also affect LGD in some cases. For example, the amount of collection at the time of default tends to be relatively small for credit extended to borrowers belonging to an industry, which provides collateral with relatively low liquidity.

B. Validation of Risk Components

Risk components are estimated based on historical data, and thus validation of their adequacy should be made on a continuous basis. Validation approaches may vary and the following are some examples of PD validation.

1. Order of estimated PD for each grade

Consistency between the rating scale and PD estimated for each grade should be confirmed. Specifically, financial institutions should confirm that PD for a lower grade does not exceed that for a higher grade.

2. Backtesting

Backtesting is a common approach of PD validation, which represents a comparison between the estimated PD and the actual default rate for each rating grade. Unlike the case of market risk where there are daily data for backtesting, the frequency of historical data is normally yearly for corporate defaults. The lack of long time-series data may pose a serious problem for the reliability of backtesting based on statistical inference.

3. Stability of PD in time-series perspective

As mentioned earlier, PIT rating assumes that the actual default rate for each rating grade is stable over time, not subject to change with the business cycle. The use of PIT rating, therefore, should accompany the assessment for stability of PD over the business cycle. On the other hand, TTC rating assumes that the actual default rate for each rating grade fluctuates yearly, reflecting the influences of the change with the business cycle. Accordingly, under the TTC ratings, the estimated PD should include some margin for its volatility arising from the business cycle. In this respect, it is important to ensure that the magnitude of the fluctuation of the actual default rate is within this margin. If the actual default rate exhibits variation that exceeds the acceptable margin, then investigation should determine whether it is caused by (1) the unexpected change with the business cycle, (2) the occurrence of anomaly, or (3) the shortcomings of the rating design or the estimation method. The case of (3) requires making modifications to the rating model or estimation approach itself.

Box 5: Statistical Approaches for Backtesting

Binomial tests and normal tests are among the exemplary statistical methods for PD validation. The binomial test is an approach to test whether the observed variation of PD for each rating grade is within a certain bound of error. For example, given the setting that the number of borrowers is 1,000 and PD for each borrower is 1 percent, the average number of defaults is 10 (1,000 $\times$ 0.01). In reality, however, the realized number of defaults might be greater or less than 10, depending on the sample size. The binomial test helps to examine whether the observed number of defaults could occur with some statistical significance, given a certain confidence level.

Likewise, the normal test provides a way to check whether the calculated mean of PD is valid given the observed variation of actual PD over time. For example, even if average PD is known to be 1 percent, observed PD in the past five years might be 1.2 percent, 1.8 percent, 0.8 percent, 0.7 percent, and 1.1 percent successively. The normal test helps to assess the validity of the calculated average of PD, given the variation of historical data over time.

It should be stressed, however, that statistical approaches of this kind rely on the quality and availability of sample data. Moreover, many of the statistical tests for validation are based on the assumptions that defaults are independent events, while in reality the correlation of the default events among borrowers is highly likely. Unless these limits on the assumptions of the methods employed are taken into account, the results obtained by statistical tests may lead to the wrong assessment for the estimates of risk components.

Box 6: How to Distinguish Characteristics of Ratings Based on Ex Post Rating Results—PIT or TTC?

As explained above, the difference between PIT and TTC ratings is whether changes in the economic environment surrounding borrowers are absorbed by rating migration as in PIT, or are incorporated in changes in the actual default rate of each grade as in TTC. Therefore, it is possible to analyze characteristics of ratings afterward by decomposing total qualitative changes in loan portfolios over the business cycle into two parts, the PIT part which can be explained by rating migration and the TTC part which can be explained by changes in the actual default rate.

Through the Bank’s examinations carried out for some financial institutions equipped with established internal rating systems, some correlations were observed between the business cycle and the default rate for each rating grade. This finding implies that the financial institutions’ internal rating systems have, more or less, some sort of TTC characteristics.

The relative size of PIT part or TTC part based on the above definition varies depending on each financial institution. In addition, some institutions have different sensitivities to the business cycle depending on the ratings. For example, the actual default rate has a high sensitivity to the business cycle in higher ratings (closer to TTC), while it has a relatively low sensitivity in lower ratings (closer to PIT). Therefore, it is advisable that institutions consider these factors in verifying the validity of PD estimates.

VII. Uses of Internal Rating Systems

Internal rating systems provide financial institutions with the foundation for advancing credit risk management and also for efficient lending operations and strategies. There is no point in introducing such systems if financial institutions cannot make full use of their rating systems in practice for the internal management of risk and business activities. Hence, institutions should recognize the importance of introducing the system and maximizing its use.

For example, marketing and lending sections are expected to use rating systems as a tool of risk identification and pricing. In addition, using output information[17] of internal rating systems, such as loans outstanding for each grade, migration matrix, and PD by grade, middle offices can monitor the quality of loan portfolios, establish a pricing guideline by grade, quantify credit risk, and estimate credit cost.

The following provides more detailed examples observed in recent bank examinations of the Bank of Japan of how internal rating information and PD estimated for each grade are used.

1. Uses of internal rating systems

a. Loan origination

- Setting upper credit limits based on rating grades: For example, institutions can extend a smaller amount of loans to low-graded borrowers and thereby avoid the risk of credit concentration in them.
- Setting authority ranks for loan approval by rating grade: For example, loan officers at bank branches can make loan decisions for only a limited amount of loans to low-graded borrowers.
- Simplifying the loan review process for higher-graded borrowers: Risk-based allocation of risk management resources can improve efficiency of the overall loan review process.

[17] Output information of internal rating systems is expected to be defined and used consistently within a financial institution. If an institution makes certain adjustments to output data, it needs to ensure the appropriateness of the adjustment in advance. For example, if an institution uses different PDs for pricing and for risk quantification, it should have a reasonable justification for doing so.

b. Monitoring

- Monitoring individual borrowers based on rating grades: For example, even among "normal" borrowers, institutions can monitor more carefully downgraded or relatively low-graded borrowers. In addition, institutions might be more deeply involved in those borrowers' management at an early stage of their financial troubles so as to prevent them from being further downgraded.

- Monitoring the overall loan portfolio: For example, institutions can spot impaired assets in the loan portfolio by monitoring the migration matrix on ratings and changes in loans outstanding by rating for each industry and geographical area.

2. Uses of PD for each rating grade

- Quantification of credit risk and allocation of capital: Institutions can use PD for each grade as input data for calculating credit risk (for quantification of credit risk, see Chapter IX). In addition, they can allocate economic capital to each section of the institution based on the calculated risk amount.

- Pricing of loan rates reflecting credit risk: Institutions usually set reference interest rates for each loan by adding the credit spread (credit cost rate and capital cost rate) to the funding cost, expense ratio, and target return ratio. They can estimate credit cost rates by using PD for each grade. In a case where they quantify the credit risk amount using PD for each grade and allocate corresponding economic capital, they also use this PD for each grade to estimate the capital cost rate to be considered for pricing. Widespread use of loan pricing based on accurately estimated credit risk is vital in preventing an unexpected increase in unrealized losses that could eventually materialize as large losses in the stage of loan disposal.

- Evaluating the economic value of loans:[18] Institutions can evaluate the economic value of loan assets using the discounted cash flow (DCF) method, which calculates the discounted value of future cash flow after considering future credit cost. This process requires the estimation of credit cost and hence PD for each grade.

[18] For details, see "Evaluating the Economic Value of Loans and Implications: Toward Transformation of the Business Model for Banks and Nonbank Firms" in the August 2003 issue of the *Bank of Japan Quarterly Bulletin,* and the Bank's web site (http://www.boj.or.jp/en/ronbun/03/ron0304a.htm).

VIII. Validation of Internal Rating Systems

In the continuously changing environment surrounding financial institutions, it is necessary to keep validating the appropriateness and relevance of internal rating systems in order to maintain their effectiveness. In chapters V and VI, we discussed the validation of rating models and estimation of risk factors mainly from a technical aspect. In this chapter, we look at the validation of internal rating systems comprehensively (charts 17 and 18).

The following are important points in validating an internal rating system.

1. Active involvement of the management

The management should have a full understanding of the need for an internal rating system and its validation, and actively take part, for example, in establishing a validation framework and system.

2. Validation by middle offices and loan review sections

The credit risk management sections (middle offices) need to plan the overall design of an internal rating system, establish necessary policies and procedures for the system, and conduct validation. Also necessary are third-party validation by the loan review section and reporting of the results to the management.

3. Clarification of the validation framework

The policy and procedures for validation need to be clarified and disseminated to all related staff.

4. Continued validation

Once an internal rating system is introduced, the system should be validated on a continuous basis and validation results should be used to make necessary adjustments. The frequency of validation depends on factors subject to validation. Factors that have significant influence on the internal rating system as a whole need to be validated frequently.

5. Validation of the architecture and usage

Architecture and models of an internal rating system as well as the usage in marketing and loan review sections are subject to validation.

6. Continuous review of the validation method

There is no single absolute method of validation. Financial institutions, therefore, need to review and make necessary changes to methods they use. Statistical validation needs to be revised as new data are accumulated, since the accuracy of statistical validation depends greatly on data volume.

7. Preparation of data needed for validation

Sufficient amount of data are necessary for validation (Box 7). They include financial data used in rating models and rating migration data that reflect quantitative and qualitative evaluation. These data may be used to analyze the performance of quantitative models and qualitative evaluation. Data should continue to be collected to enable retroactive assignment of grades even when changes are made to the rating models.

8. Documentation

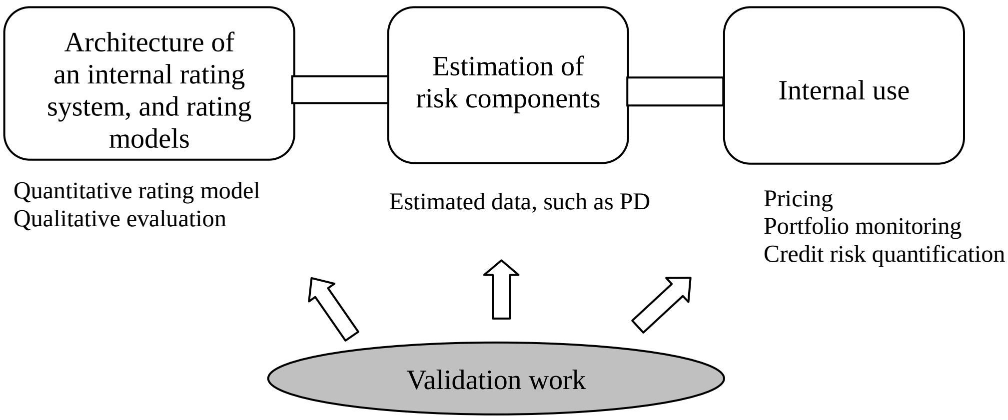
Methods and results of validation need to be documented.

Chart 17: Validation Framework

Purpose (Why)	1. To secure the adequacy of the internal rating system. 2. To improve the accuracy of credit risk evaluation.
Scope (What)	1. The framework of the internal rating system 2. Borrower ratings and facility (LGD) ratings 3. Rating models, such as quantitative rating 4. Data accumulation 5. Management of the internal rating system and rating process (whether marketing and lending sections are assigning grades accurately) 6. Outputs of the internal rating system, such as rating migration matrix, PD, and LGD 7. Use of output data (portfolio monitoring, loan rate pricing, and reports to the management)
Structure (Who)	1. Credit risk management sections (marketing and lending sections) 2. Loan review sections 3. The management

Method (How)	1. On-site and off-site validation 2. Quantitative (statistical) and qualitative validation 3. Checking of various documents 4. Verification of the logic behind the model (programs) 5. Checking of the usage of sample data 6. Benchmarking, for example, comparison with external data sources 7. Backtesting, for example, comparison of estimated to actual figures
Timing (When)	Periodical basis (frequency depends on the purpose of the validation and significance of the element subject to validation).

Chart 18: Scope of Validation

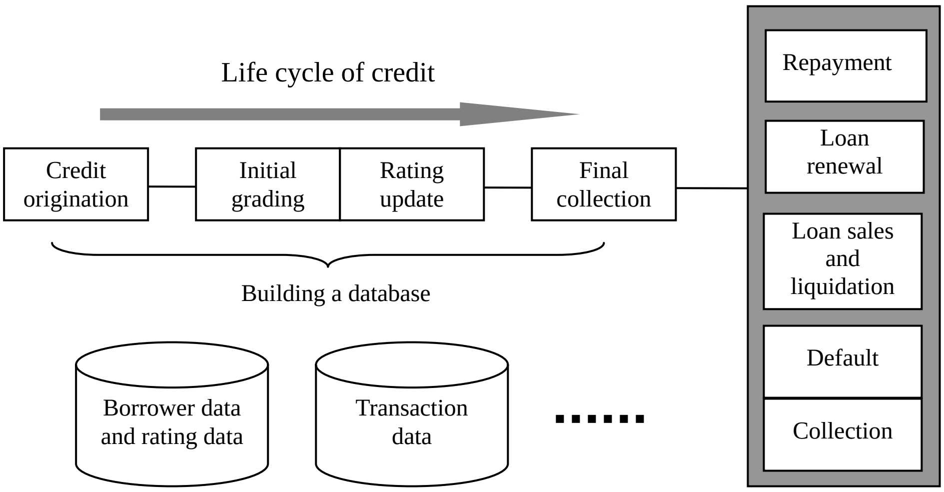


Box 7: Data Maintenance

Accumulation of data used for internal ratings is essential in increasing the accuracy of internal rating models and estimation of risk parameters. A detailed and highly reliable database should be built, in particular, to realize appropriate and efficient validation by a third party.

For example, accumulated data on borrowers' rating records and the results of qualitative and quantitative evaluation may be used to validate the performance of rating models and PD estimation. In addition, financial and other borrower-specific data used for rating will be useful when making drastic changes to the internal rating system (especially when changing the rating model) and when building a new system given, for example, mergers among financial institutions.

The keys for gathering borrower and transaction information are to track the life cycle of credit, for example, from origination of a loan to collection of its claim, and gather long-term information on initial grades, rating records, and final collection. In addition, the database should be arranged in a way that enables various analyses in an efficient manner.



The following are examples of data items for collection.

Box 7 (Cont.): Data Maintenance

1. Borrower Data

Borrower ratings; initial ratings and rating updates; changes in evaluation from initial to final ratings, for example, first and second evaluations; quantitative information (financial data); qualitative information, such as the industry and characteristics of borrower firms; and external information, such as external ratings.

2. Transaction Data

Facility ratings; facility amounts; cash flow information; amount of commitment line; purpose of credit extension; facility type; senior/subordinate structure; collateral type, for example, cash, government [bonds/securities], and real estate; collateral value; timing of appraising collateral; coverage ratio of collateral; and the secured level of collateral and guarantees.

3. Default Data

Final disposition process; timing of default; default conditions, for example, legal bankruptcy, loans in arrears, and relaxing of conditions; exposures at default; amount and timing of collection; collection method, such as collateral and guarantees; collection cost; and discount rate for obtaining the present values of collections at the time of default.

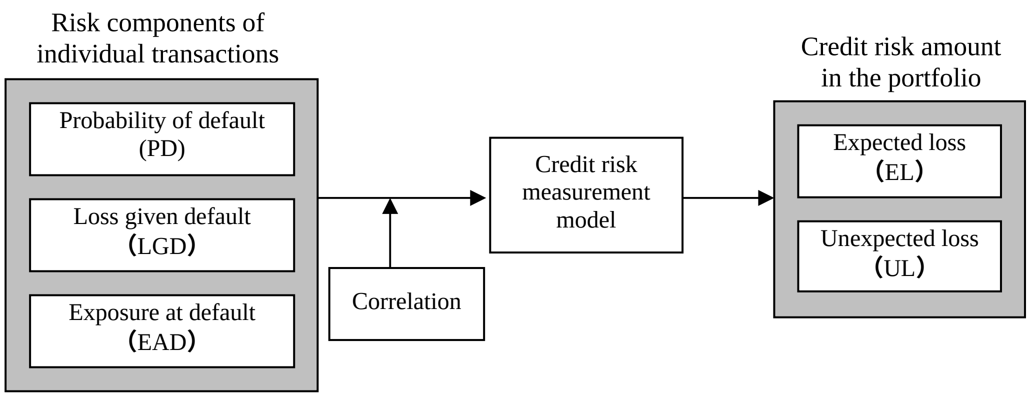
IX. Quantification of Credit Risk

We have introduced the various aspects of internal rating systems as the foundation of credit risk management at financial institutions. In this last chapter, we discuss the quantification of credit risk, which becomes possible by the development of internal rating systems.

A. Expected Loss (EL) and Unexpected Loss (UL)

It becomes possible to quantify the amount of credit risk in the entire loan portfolio once we estimate all risk components mentioned above (Chart 19).[19] Credit risk amount is usually explained by expected loss (EL) and unexpected loss (UL). EL is the average amount of loss forecast for a certain period of time,[20] for example, one year. UL is defined as the maximum loss with a certain probability[21] minus EL. For example, 99 percent or 99.9 percent is used as this probability, and this figure is called a confidence level (Chart 20).

Chart 19: Quantification of Credit Risk

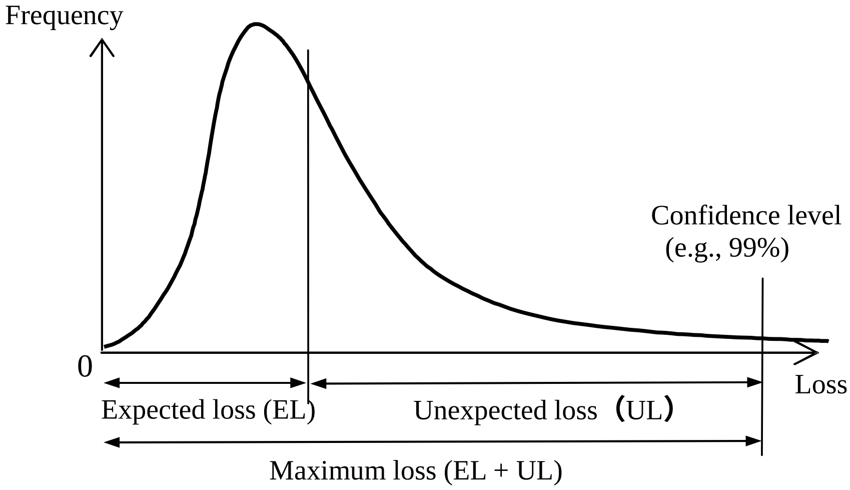


[19] There are two ways of recognizing credit risk. One is a default mode, which recognizes losses as realized only when borrowers default. The other is a mark-to-market mode, under which losses are recognized when the value of loan assets decreases due to deterioration in the creditworthiness of borrowers because of downgrading.

[20] The period during which loan assets are held (the period for which risks are evaluated) is generally one year in view of the twelve-month business cycle or the period equivalent to the remaining maturity on a contract basis. In some cases, the remaining maturity may actually be longer than that on a contract basis. The criteria for determining the remaining maturity for such cases are an important point for discussion.

[21] The maximum loss with a certain probability, for example, 99 percent, is equivalent to the loss amount that could be exceeded by actual losses with a probability of 1 percent (= 100 percent – 99 percent).

Chart 20: Loss Distribution of Credit Portfolio



The maximum loss (EL + UL) with a 99 percent confidence level for a year of lending tenor implies that a larger loss than this figure would affect banks only once in 100 years.

EL is expected to be covered by loan-loss provisioning as a necessary cost for risk taking while UL is expected to be covered by capital as a potential loss associated with risk taking.

The size of UL depends on the credit portfolio. For example, UL will be larger relative to EL, in other words, the right-side tail in Chart 20 will be extended further, when the weight of loans to the same borrowers is high or when there is a strong default correlation among the borrowers (proxy variables include concentration of industry, geographical areas, and affiliated firms that increases the possibility of a chain bankruptcy reaction). In contrast, UL tends to be small if small-size loans are extended to a large number of borrowers or if default correlation is small.[22]

---

[22] Let us take an example of a hypothetical portfolio with a loan amount of 100 billion yen, 2 percent of yearly PD for each borrower, and 100 percent LGD across the board.

Case 1: A portfolio consists of a sole borrower, A. A's PD is 2 percent. This implies the loans could go into default twice in the next 100 years, and the loss for each year is 100 billion yen (the loss for the rest is zero since A does not default for 98 years). Thus, the maximum loss (EL + UL) with a 99 percent confidence level (i.e., the second-worst-case scenario in the next 100 years) is 100 billion yen.

Case 2: A portfolio consists of two borrowers, A and B, and the lending amount for each is 50 billion yen. If there is no default correlation between A and B, the second biggest loss in the next 100 years is 50 billion yen when either A or B goes into default. The probability of both going into default with a loss of 100 billion yen is less than 1 percent (2 percent × 2 percent

B. Risk Calculation

EL is calculated for each borrower using the following formula.[23] EL of the overall credit portfolio is the sum of EL of each borrower.

$$EL = EAD \times PD \times LGD$$

UL calculation requires not only the three risk components (PD, LGD, and EAD) used for calculating EL but also the shape of probability distribution of each component and default correlation among the borrowers. In this sense, the calculation is more difficult than that for EL from a technical aspect. Therefore, in many cases the Monte Carlo simulation method is used to calculate UL.

Box 8: Risk Weight Function of the IRB Approach in the Basel II Framework

Under the internal ratings-based (IRB) approach in Pillar 1 of the Basel II Framework, the credit risk amount is calculated by the equations (risk weight function) defined by the framework. Only PD, LGD, EAD, and other risk components are needed for the calculation provided that there are certain assumptions regarding credit concentration and default correlation.

Pillar 2 of the framework requires assessment on risk concentration. This means that the above-mentioned formula may not be sufficient to measure the overall risk faced by each financial institution. Financial institutions, therefore, should be sufficiently aware of their risk profiles and estimate the risk attached, especially UL.

Risk weight functions are theoretically based on a one-factor Merton model. This model describes the changes in the corporate value of borrowers as the weighted sum of one macro factor common to all borrowers (systematic factor) and a specific factor of borrowers (idiosyncratic factor). An assumption of the model is that changes in these factors cause the changes in the corporate value, and default happens when the corporate value falls below a certain level. For example, if the business cycle is

= 0.04 percent).

In this way, the diversification of a portfolio reduces the maximum loss with a 99 percent confidence level. However, the maximum loss in Case 2 becomes 100 billion yen, the same as in Case 1, if the default correlation between A and B is one. This is because B always goes into default when A does (and vice versa). Thus, the higher default correlation causes a larger maximum loss.

[23] The assumption is that the three risk factors are all independent probability variables.

considered a systematic factor, the default of a borrower can be explained by the business cycle together with the conditions specific to the borrower. If the business cycle has a strong impact by influencing a relatively large number of borrowers “at the same time”, the UL of each borrower and the entire loan portfolio increases. Similar factor models are used by most financial institutions which calculate EL and UL.

C. Stress Testing

Uncertainties always remain in calculating risk using a credit risk measurement model. This is because risk calculation is an estimation that requires certain prerequisites. It is also because stability is not fully secured for the model and data on risk components. Stress tests and other measures are necessary against this uncertainty. A stress test is a way of evaluating the sufficiency of portfolio management policy and financial strength against a certain assumed but probable stress situation, such as a sudden change in the market condition. Examples of stress scenarios include a decline in GDP growth, a drop in stock and real estate prices, deterioration in the business condition of large borrowers, and large fluctuations in risk components. It is important that a scenario be constructed in view of the characteristics of the credit portfolios, for example, industry and size of operation of borrowers.

The results of stress tests need to be actively used by the management, risk management sections, and marketing section in discussing portfolio management for the near future. Middle offices are required to prepare a stress test that is persuasive to a third party. Only the results of such stress tests will be taken seriously by the management as information significant in determining an appropriate level of capital buffer and necessary countermeasures.

Box 9: Stress Test for Special On-Balance Assets

As an example of a stress test, we look at a special kind of on-balance asset for which EAD could increase in the future.

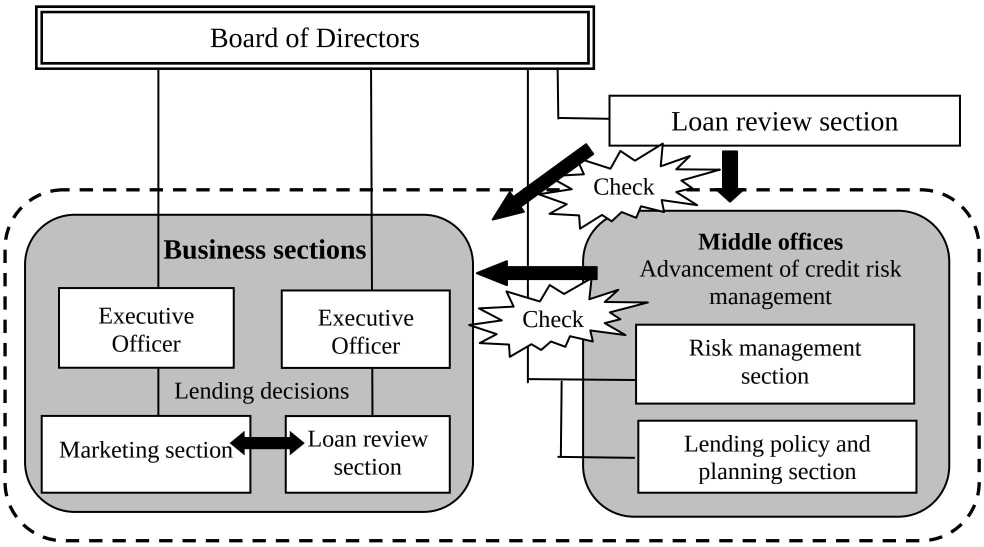
As mentioned earlier, a notional or face value is directly used as EAD in normal on-balance assets such as normal loans, since face value remains intact during a tenor of transaction. However, in some cases, financial institutions could experience the increase in the lending amount at the time of the borrower's default due to special relationships between institutions and borrowers, though the original exposure amounts are not necessarily contingent on the borrowers' conditions.

When a financial institution, A, is a main bank of a large borrower, B, and B's financial status deteriorates, it is quite possible that A will be forced to take over lending of other banks because (1) A may be most affected by the default of B, and (2) A may feel a moral responsibility to support B as a main bank to avoid the materialization of reputational risk. As a result, a shift of loans from non-main banks to main banks occurs. In reality, at many financial institutions, the lending amount to a large borrower with strong ties but deteriorating financial status tends to increase until the borrower defaults or is revitalized after debt restructuring.

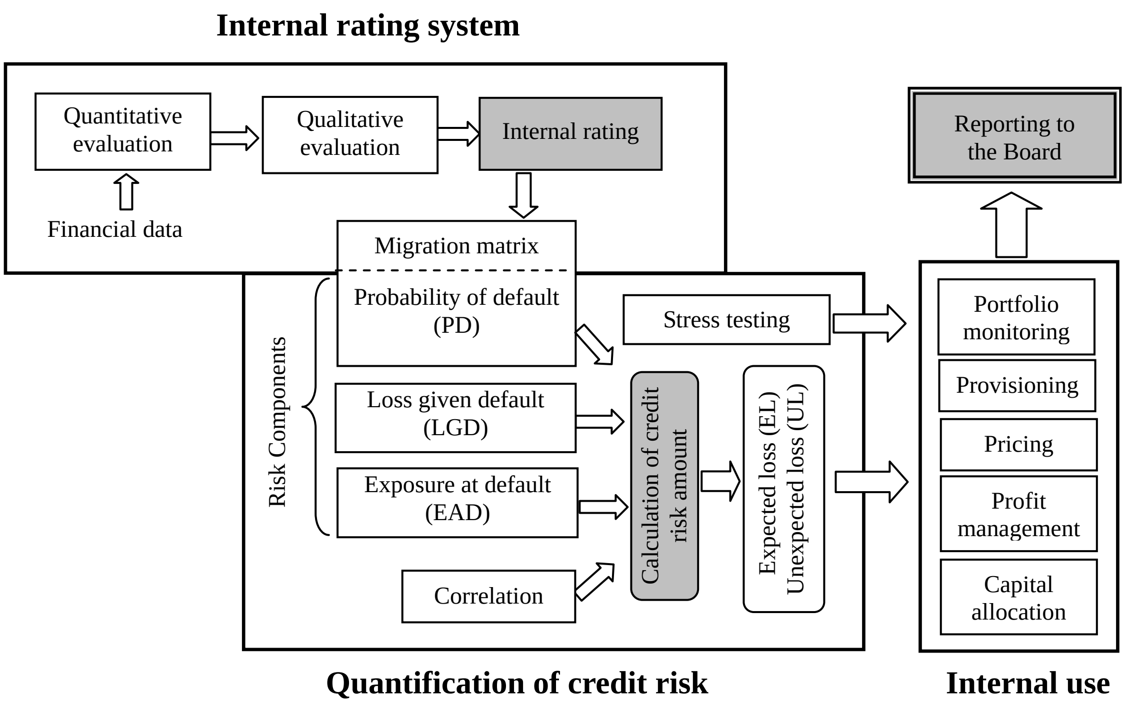
It is not easy to identify a possible borrower that could cause such a shift due to its strong relationship with the bank or to estimate the increase in lending due to such a shift. However, the past experience of Japanese financial institutions showed that the risk of such a shift, in other words, of being a main bank, is quite significant. Therefore, it is necessary to consider such risks in their risk management by evaluating them by means of stress tests.

Appendix 1: Promoting Credit Risk Management

The Present Structure of Credit Risk Management: An Example



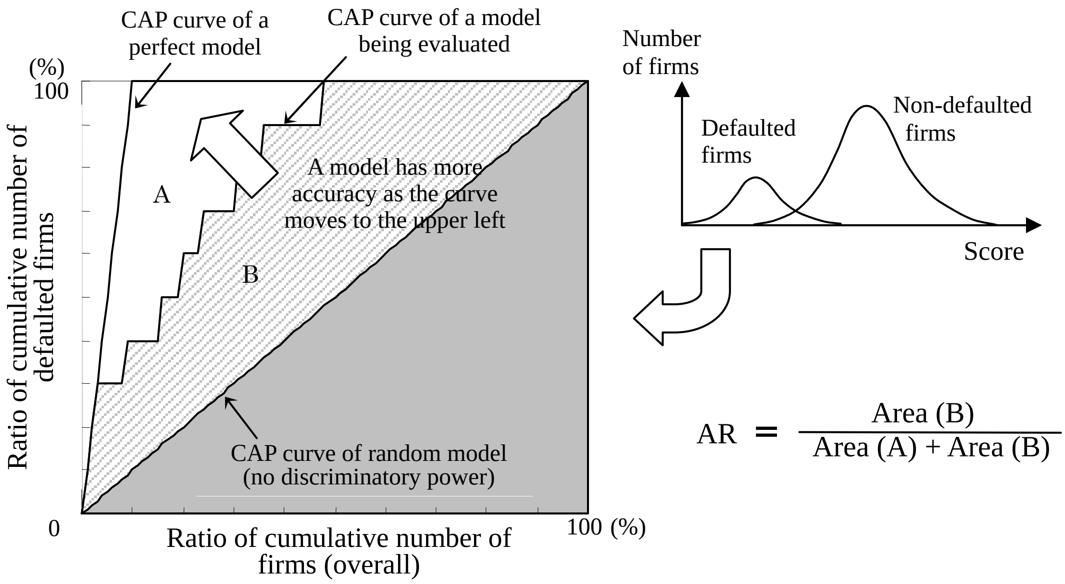
Advancement of Credit Risk Management



Appendix 2: An Example of a Statistical Test for Validation

A. Cumulative Accuracy Profile (CAP) and Accuracy Ratio (AR)

One method of evaluating the accuracy of a rating model is to draw a cumulative accuracy profile (CAP) curve (also called the Gini curve, power curve, or Lorenz curve) and to calculate the accuracy ratio (AR), a statistical value that quantifies the performance of a model. In this method, accuracy of the model in detecting default is evaluated by arranging borrowers in an order by their creditworthiness that is expressed in scores and comparing the number of defaulted borrowers.



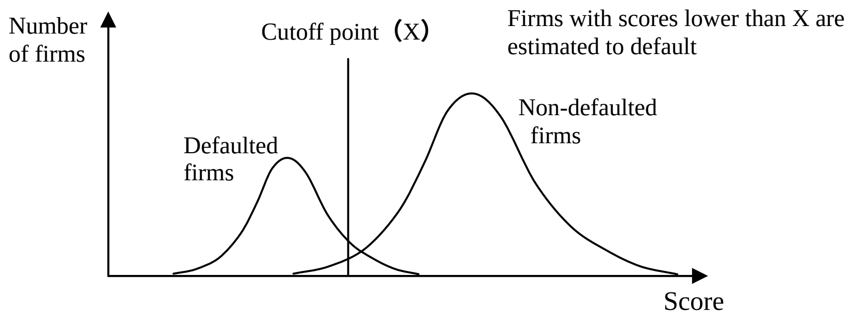
$$AR = \frac{\text{Area (B)}}{\text{Area (A)} + \text{Area (B)}}$$

Suppose the number of borrowers used for the estimation is N and the number of firms that defaulted is n. On its horizontal axis, the CAP curve plots the ratio of the number of the firms with the worst x creditworthiness given by the model to the total number of firms used for the estimation ("x/N"). On its vertical axis, the CAP curve plots the ratio of defaulted firms (n_x) that are low-graded by the model to the total number of defaulted firms ("n_x/n").

The CAP curve shows the accuracy of the model. For example, a highly accurate model will show a curve in the upper left corner. On the other hand, an inaccurate model will show a straight line with an angle of 45 degrees. AR calculation is based on the area between the 45-degree line and the CAP curve in the above chart. This means, therefore, that a rating model with larger AR is more accurate.

B. Receiver Operating Characteristics (ROC) and Area under Curve (AUC)

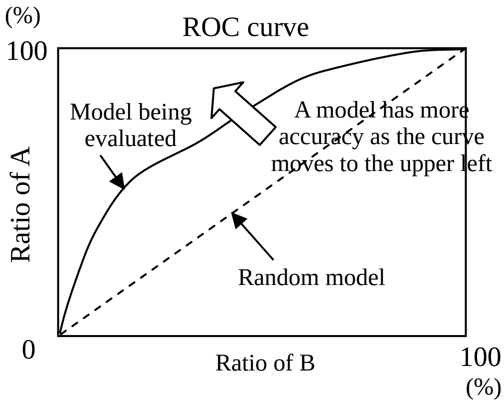
Some models may assume that firms with scores lower and higher than X (the cutoff point) are those estimated as defaulted and non-defaulted firms, respectively. In this case, the relationship between estimated and actual numbers of defaulted (or non-defaulted) firms is helpful in analyzing the performance of models.



As shown in the chart below, there are four combination patterns of estimates and actual outcomes. The ratio of zones A and D increases when the accuracy of a model is high. The ratio of zone A (=A/(A+C)), which shows correct estimates, and the ratio of zone B, which shows a false estimate (=B/(B+D)), are plotted to show an ROC curve by moving the position of a diverging point X from the lower to higher scores (from the left to the right side of the graph).

The ROC curve will be on the upper left side of the graph for a model with high accuracy and will overlap the 45-degree line for a random model. AUC is the area below the ROC curve.

Estimation		Actual outcome: Default	Actual outcome: Non-default
	Default	A (Correct)	B (False)
	Non-default	C (False)	D (Correct)



Appendix 3: An Example of LGD Estimation

The following is an example of LGD estimation in a case in which reference data are internal historical data, that is, data collected during the past seven years. Data should be arranged to estimate LGD with the aim of finding the relationship between the type and conditions of loan assets and the degree of economic loss at the time of default.

$$\text{LGD} = 1 - \text{recovery rate}$$
$$= 1 - (\text{amount collected} - \text{cost of collection}) / \text{exposure at default}$$

1. Collecting Data on LGD (Building a Reference Database)

Some examples of factors affecting LGD are below. An important point here is to collect data on defaulted assets from various perspectives.

- Provision of collateral
- Collateral type, for example, financial assets, real estate, other physical collateral, and guarantees
- Ratio of assets covered by collateral
- Total amount of credit extended
- Borrower information, for example, industry, geography, and creditworthiness[24]

2. Grouping Defaulted Assets

Main categorization factors are decided and defaulted assets are grouped accordingly. Average LGD is then estimated for each group (the calculation method is described in items 3 and 4) and this is treated as estimated LGD for each category. The following is an example of categorization of LGD estimation by coverage ratio of collateral and type of collateral.

		Type of collateral	
		Real estate	Financial asset
Collateral coverage ratio	High (More than x%)	LGD = A%	LGD = B%
	Middle (y% - x%)	LGD = C%	LGD = D%
	Low (Less than y%)	LGD＝E%	LGD＝F%

[24] Borrower information is basically linked to PD but should also be taken into consideration to estimate LGD if it is expected to affect LGD.

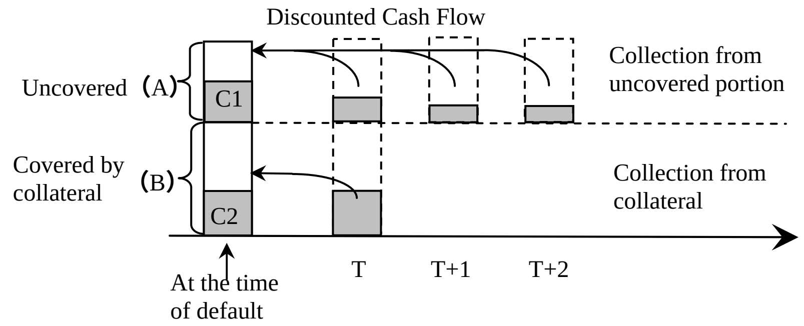
Another method of LGD estimation is to build the LGD rating framework made by combining the above factors and then estimate average LGD. In addition, there is a method to directly calculate LGD for each factor using a regression model based on the multiple explanatory variables, for example, coverage ratio of collateral, and type of collateral.

Consideration should be given to the number of categorizations because data on defaulted assets are limited. Too many categorizations will decrease the number of samples in each group. This will lessen the chance of obtaining statistically meaningful estimates of LGD.

3. Measuring Economic Loss

Definition of a default should be equivalent to that used for estimating PD. In addition, LGD should be based on the economic losses rather than the accounting losses. Therefore, not only the amount of collection from collateral and guarantees but also the cost of collection should be taken into account in measuring LGD.[25] Each collection amount after default should be evaluated at the time of default using a discount factor reflecting uncertainty of recovery.

Chart: LGD measurement of each defaulted loan



C1 = amount collected from A, the portion uncovered by collateral

C2 = amount collected from B, the portion covered by collateral

LGD = 1 - (C1 + C2) / (A + B)

---

[25] The following are some points to be carefully considered in estimating the amount and cost of collection.
(1) How to deal with the collections materializing many years after the default. One example might be to cutoff the collections materialized after a certain threshold year.
(2) How to widely cover the collection cost.
(3) How to determine the discount factors for calculating the amount of collection to reflect the time value from default to collection.

4. Long-Term Average of LGD

Economic loss or LGD is calculated for all default assets in the past. LGD is affected by economic conditions. Data should, therefore, cover a long time frame in order to estimate LGD reflecting economic downturns as well as average LGD.