Outline of CRD Models

I. Model 3 for Corporations

II. Model 4 for Sole Proprietors

— BS Model —

March 2006

CRD Association,
a Limited Liability Intermediate Corporation

This document outlines CRD's models that credit guarantee corporations use in determining their guarantee charge rates, with a summary of the features of CRD-developed default probability estimate models, model structures and their performance.

Contents

I. Model 3 for Corporations

1. Features of Model 3

Model 3 is a model for estimating default probability, (PD) developed by improving Model 2. The main features of Model 3 are described in Table 1.

Table 1. Main features of Model 3

	Main features	Specific contents
1	Able to estimate multiyear (two-year, three-year cumulative) default probability, going beyond one year.	In order to meet membership requests, this model can estimate two-year, three-year cumulative default probability.
2	Able to estimate default probability by intermediary business classification.	To meet requests for estimating default probability for smaller businesses, this model is designed to enable the estimation of a default probability (PD) according to an intermediary business classification.
3	Detects with a high degree of accuracy companies likely to default.	The huge database accumulated at CRD over many years enables the construction of high-performance models.
4	Has an easy-to-understand model structure.	This model is designed to ensure easy understanding of its structure and the rate of contribution of financial indicators.

2. Structure of Model 3

2.1. Data used for the construction of Model 3

To construct Model 3, we used the data accumulated at CRD as shown in Table 2. With this large data set, we can achieve a high degree of discriminating power as a whole.

Table 2. Data used for the construction of Model 3

		Volume of data in use
	Number of cases used to construct Model 3	About 2.93 million
	Number of default cases	About 40,000

2.2. Steps in estimating default probability

In calculating default probability, we followed the three steps as described below (Chart 1).

Step 1: Categorizing businesses according to a primary business classification model.

Target companies are categorized into two groups based on their creditworthiness — "good" and "ordinary."

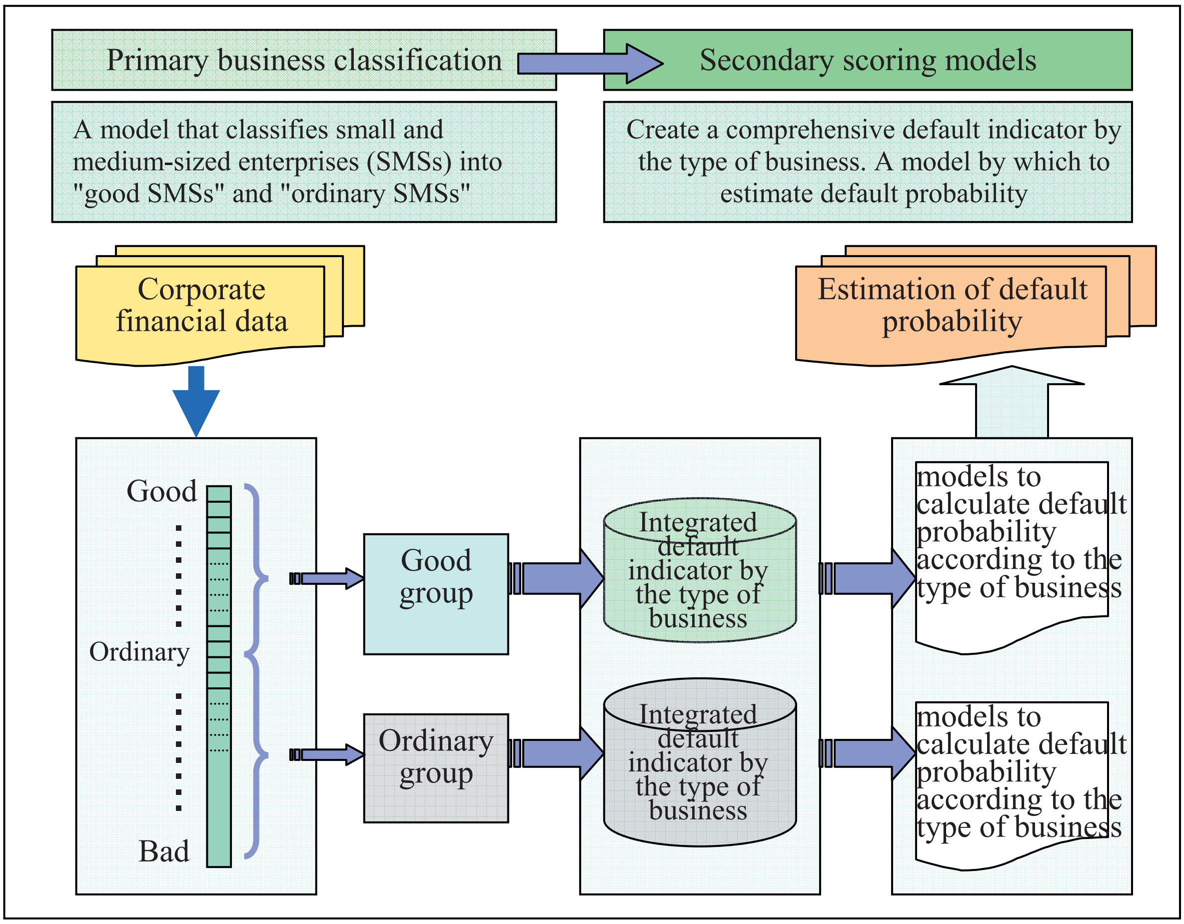
Step 2: Creating a "integrated default indicator"

A variable named "integrated default indicator" for the explanation of corporate default probability is created by combining varieties of financial indicators according to the classification in each business sector.

Step 3: Calculating default probability using secondary scoring models.

Using the default indicator, we calculate the default probability for a target company in the model for its business industry.

Chart 1. Flowchart for the calculation of default probability



2.3. Primary business classification model

We distinguish good companies from ordinary companies from the viewpoint of foreseeable bankruptcies. We can obtain a more accurate estimate of these two types if we use different financial indicators for each type.

Therefore, a primary business classification model has been introduced with the aim of roughly classifying companies into a "good" or an "ordinary" group, based on their current financial position. The classification of companies into two types — good or ordinary — enables the estimation of default probability more accurately in the secondary scoring model.

2.4. Integrated default indicators

An integrated default indicator is a variable to indicate corporate default likelihood by combining a number of corporate financial indicators.

For processing missing values, conventional wisdom had to exclude financial items with many missing values or to replace them by some values. These processes decrease the number of indicators available, thereby often posing impediments in constructing models.

Model 3 introduces an integrated default indicator to avoid these problems — a variable that indicates default likelihood. The model calculates a comprehensive default indicator by using 22 to 28 financial indicators for both good and ordinary companies in each business sector, and this indicator has specific features as mentioned below.

(1) An integrated default indicator synthesizes various financial indicators to assess a target company without dependence on specific financial indicators.

Usually financial items with many missing values were not available in model building. Because of these shortcomings, at most 10 financial indicators were used as appropriate indicators for credit assessments. However, this single indicator can accommodate the maximum number of financial indicators available for each company, without being affected by missing values.

(2) This indicator is less subject to outlying or missing values.

Every financial indicator is discretized in advance: the range of values is divided into a few groups, with a new particular value assigned to each group. This value, relevant to each group of financial indicators in respective business sectors, is named default index.

Each financial indicator of a debtor is translated into the correspondent default index, respectively. And these default indexes are averaged by the number of indicators available for a debtor. That is why this average is called an integrated default index.

Therefore, this single indicator makes the resulting credit score less vulnerable to the changes in financial indicators due to the descretization, or to extreme values due to the averaging procedure. And this integrated indicator can avoid the difference in the number of financial indexes available among debtors because of missing values, due to the averaging.

2.5. Classification of businesses used for the Model

After classifying all the debtors in the "ordinary" or "good" group, Model 3 applies different sector-specific models to a debtor in each of these groups.

Table 3. Classification of businesses for Model 3

	Type of model business	Remarks (large grouping)
1	Manufacturing	Manufacturing
2	Construction	Construction
3	Real estate	Real estate
4	Wholesale	Wholesale
5	Retail	Retail
6	Services	Restaurants and services
7	Others	Agriculture, forestry, and fisheries; mining; electricity, gas, heat supply and water utilities; transportation and telecommunications; finance and insurance; civil service; and nonclassifiable business

2.6. Calculation of default probability by a secondary scoring model

The secondary scoring models calculate default probability for each debtor. The default probability is calculated by using an integrated default indicator as the single explanatory variable in one of 12 secondary models applicable.

With x (comprehensive default indicator) as a single explanatory variable, default probability (PD) is calculated by using a logistic function as indicated below (logit model). Parameters such as α, β and γ vary in every secondary model so that estimation can reflect actual default rates accurately in each sector with each business model.

$$PD = \frac{1}{1+\exp(-Z)} \qquad Z = \alpha + \beta x^{\gamma}$$

2.7. Calculation of multiyear cumulative default probability

Model 3 calculates the two-year cumulative default probability (PD) — the probability of default occurring within two years. It calculates the three-year cumulative PD, too.

Multiyear PD can be used to estimate loss amount from the default of a specific long-term loan, for example.

2.8. Score calculation by Model 3

Model 3 gives each debtor's PD reflecting its financial conditions. However, some users want to use 0-to-100 grading, rather than a percentage. Model 3 gives 0-to-100 scores by translating PD according to a certain standard. PDs or the grading scores are comparable across the sectors: users do not have to adjust the level of two debtors' PDs or scores from different sectors to make the comparison.

3. Verification for Model 3

3.1. Performance of the model

When all the debtors' PDs are ranked, an excellent model can catch all the defaulters with higher PDs than all the other non-defaulters'.

Accuracy is usually evaluated based on an index called AR (Accuracy Ratio). The larger the value, the higher is the accuracy of the model to detect would-be defaulters.

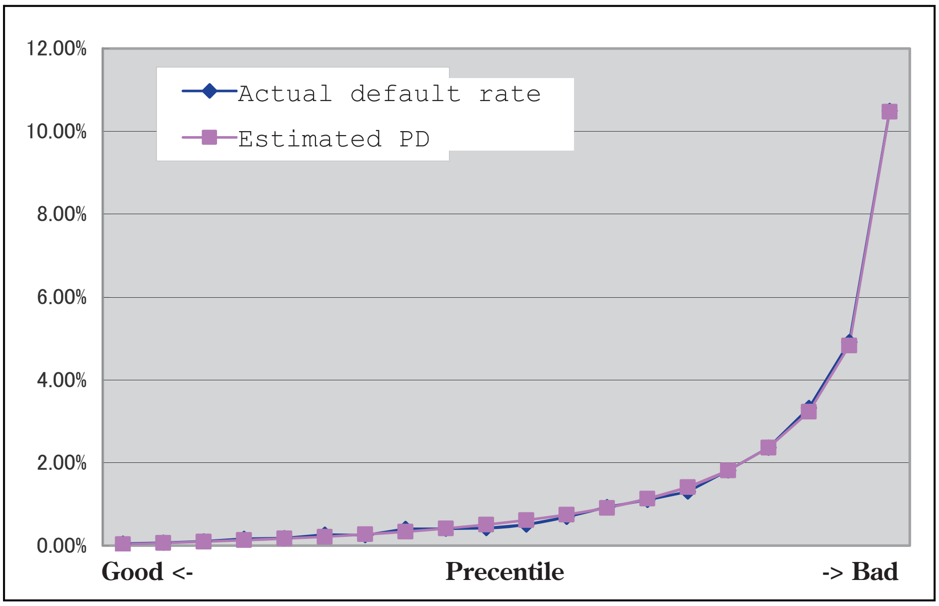
The result shows that Model 3 has good AR statistics and outperforms the other preceding CRD models.

3.2. Fit between an estimated PD and an actual default rate

It is a problem in practice if the actual default rate does not follow the estimated rate, regardless of AR value. Calculated AR value might be high, but when a calculated PD does not fit reality, this poses a problem in business practice.

Chart 2 below compares the average estimated PD from Model 3 for each segment with its actual default rate. These samples are evenly divided into 20 ranked groups in terms of their PD. The chart shows a good fit between the estimated PD and the actual default rate.

Chart 2. Fit between estimated PD and actual default rate



II. Model 4 for Sole Proprietors (BS Model)

1. Features of Model 4 (BS Model)

Model 4 was newly designed, built on the preceding CRD model for sole proprietors.

Model 4 improves upon the weakness of the old model that consisted in large part of BS and PL data. The new model adds qualitative data and the type of business as an explanatory variable to allow for the correlation between the type of business and actual defaults.

(1) Use of qualitative data

CRD, in collaboration with its members, has been accumulating data on sole proprietors in recent years. In the construction of Model 4, CRD utilized qualitative data to the maximum.

(2) Use of a detailed business sector classification.

Model 4 uses the detailed business sector classification to allow for the correlation between the sector and defaulters.

The major points of Model 4 (BS Model) are summarized in Table 4.

Table 4. Major points of Model 4 (BS Model)

Major points of Model 4 (BS Model)		Relevant descriptions
(1)	Use of qualitative data	2.4.2. Creation of explanatory attribute variables (qualitative information) 2.4.3. Reflection of qualitative data 3.1. Performance of the model
(2)	Use of a detailed classification business sector	2.4.2. Creation of explanatory attribute variables (qualitative information) (described in the following chapter)

In the next chapter, we outline the BS Model used to determine the credit guarantee charge rates by credit guarantee corporations.

2. Structure of Model 4 (BS Model)

2.1. Data used for the construction of Model 4

In the construction of Model 4, we extracted data, as shown in Table 5, from the CRD database. Such an abundance of data ensures a high degree of overall accuracy.

Table 5. Data used for the construction of Model 4

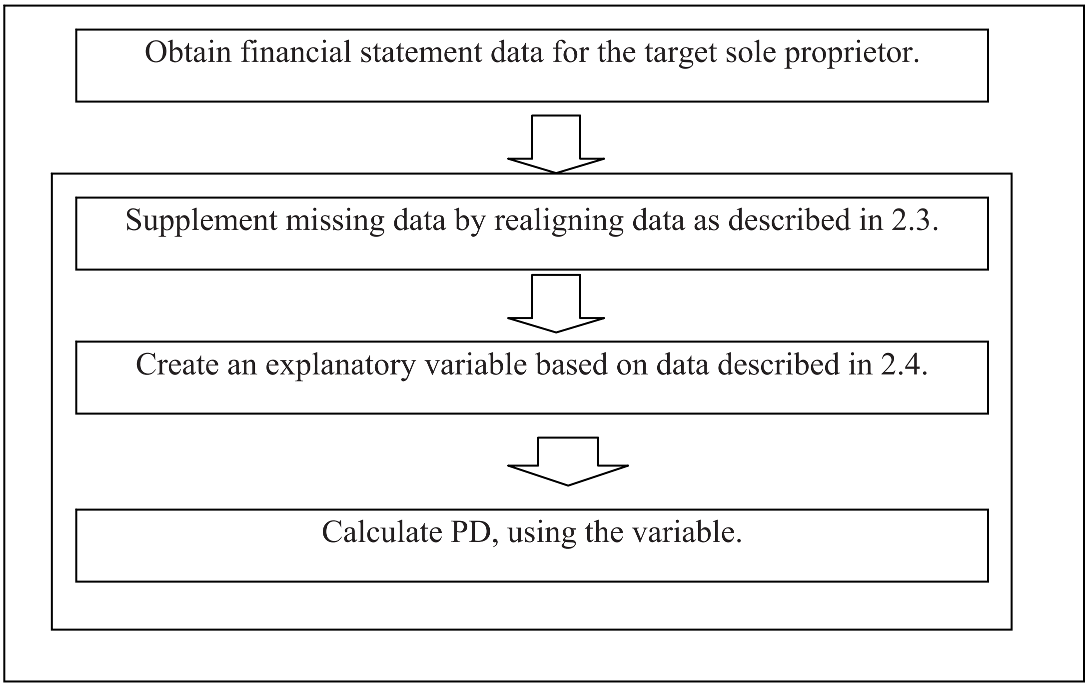
		Volume of data in use
	Number of cases used to construct Model 4	About 1.2 million
	Number of default cases	About 20,000

2.2. Steps to calculate default probability

Model 4 has been designed for estimating the default probability within one year of a sole proprietor. Estimation uses a logistic regression (logit model) — a common method in creating credit risk estimation models.

Steps to calculate PD are illustrated in Chart 3 below.

Chart 3. Flow of calculating default probability



2.3. Complementation of sole proprietors data

Many sole proprietors do not file balance sheets. Therefore, unlike the case with corporations, it is difficult to construct a highly accurate model without BS information. One of the critical issues involves the way to fill in the lack of information of sole proprietors.

(1) Because debt on the balance sheet and interest expenses on the income statement indicate debtors' ability to repay debt, the two accounts are essential as explanatory variables in estimating default probability. Therefore, Model 4 complements one of the two, if it is missing, with an estimated value from the other.

(2) Likewise, operating profit, if missing, is estimated from information in other lines of the income statement.

Operating profit = gross profit on sales – total expenses.

Data complementation like this applies to 20 items in the balance sheets and income statements.

2.4. Creation of explanatory variables

2.4.1. Creation of explanatory financial variables

Model 4 selects a set of useful financial indicators, as explanatory variables, from 272 indicators.

2.4.2. Creation of explanatory attributes qualitative variables

In addition to the financial data, non-financial data make up the explanatory variables.

Analysis of 27 qualitative indicators identified some of those relevant to defaults or those relevant to sales and profit. These indicators, given a particular score according to the information, are synthesized into an integrated qualitative score, an explanatory variable in the model.

2.4.3. Addition of qualitative data

Model 4 produces two different results, depending on whether to allow qualitative data or not. Model 4 with qualitative information is a "Comprehensive Model" and the model without it is a "Financial Model."

2.5. Calculation of scores with Model 4

Model 4 produces an estimated PD for a sole proprietor. Some users want to use, in addition to PD, 0-to-100 scores for easy comparison of creditworthiness across debtors. Model 4 also gives such scores by converting PD according to a certain standard.

3. Verification for Model 4 (BS Model)

3.1. Performance of the model

As was the case with Model 3, we checked Model 4's ability to detect foreseeable defaults, using AR statistics.

Results of AR show that Model 4 outperforms the preceding models when actual BS data are available. And Model 4 with qualitative data outperforms the financial model.

3.2. Fit between an estimated PD and an actual default rate

Chart 4 compares the averaged estimated PD from Model 4 for each segment with its actual default rate. These samples were evenly divided into 10 ranked groups. This chart indicates that the estimated default rate closely fit the actual default rate.

Chart 4. Fit between estimated PD and actual default rate