

Expert Judgment Based Scoring Model

Narumon Saardchom

(NIDA Business School, National Institute of Development Administration, Bangkok, Thailand)

Abstract: The literature of scoring model seems to favor the model built based on statistical method. However, every bank has its internal rating system, which is not entirely based on statistical methods. This paper provides an alternative scoring model based on expert judgment when the bank cannot build a statistical model either because it does not have high quality data or it decides to rely on the expert judgment for certain types of loans. Existing validation techniques are also mainly statistical for measuring the discriminatory power and calibration. When the model is not built based only on data, the regulatory validation of the internal scoring model can be problematic. The validation method for expert judgment based model is introduced in this paper.

Key words: analytical method; validation; risk analysis

JEL codes: C65, C52, G11

1. Introduction

Recent instabilities in financial sector lead to the improvement of a credit scoring system. The main principle of a credit scoring system is assigning to each borrower a score in order to separate the bad loan and the good loan. As a result, a scoring system is a classification tool providing indications of the borrower's probability of default in the future.

The bank also has primary responsibility to validate its developed credit scoring model under the internal rating based approach of the BASEL framework. Fundamentally, the objective of this validation requirement is to confirm the predictive ability of the bank's credit scoring model and the uses of credit scores in its credit approval processes.

When the bank can develop statistical credit scoring model using its high quality data, the data is divided into development and validation samples. There are several statistical methods for building and validating statistical credit scoring models. A variety of statistical methods for building credit scoring models include linear regression models, logit models, probit models, and neural networks. The validation methods are classified into three main dimensions of validation purposes; its ability to separate between good loans and bad loans (Kolmogorov-Smirnov (KS) statistics), its ability to provide accurate credit ranking (C statistics or ranking statistics), and its ability to predict the number of the bad loans in each score range (Hosmer-Lemeshow (HL) statistics).

The banking industry tends to favor statistical credit scoring models because when a credit scoring model is established upon statistical models and not on opinions, it offers an objective way to measure and manage risk. The validation requirement by BASEL framework also favors the statistical credit scoring model because it can be easily validated. Third, the statistical models used by credit scores can be improved over time as additional data are collected. There are several statistical methods for building and estimating scoring models, including linear

Narumon Saardchom, Ph.D, Associate Professor, Finance Department, NIDA Business School, National Institute of Development Administration; research areas: banking, risk modeling. E-mail: narumon@nida.ac.th.

regression models, logit models, probit models, and neural networks. The most popular one is the logit model which assumes that the probability of default is logistically distributed. However, when a bank does not have high quality data or its data does not contain enough number of bad loans to develop statistical model, its credit approval process usually relies on expert judgments. Even if the banks have high quality data, no bank relies only on statistical models for its credit approval decisions because some types of loans cannot rely only on a statistical modeling procedure. Unlike the more automated and more frequent credit granting decisions for credit card, medium and long term bank loans require a much deeper and more time-consuming analysis, in which qualitative expert judgments concerning performance attractiveness of the application play a very important role.

A credit scoring model based on this expert judgment can be developed using a technique like analytical hierarchy process (AHP). For this type of model, the validation requirement can be problematic. Although there are consistency ratio suggested by Saaty (1980) and geometric consistency index suggested by Crawford and William (1985) under the AHP method, they are only used to check if the experts are consistent in their priority settings for pairwise comparison matrices. In other words, they can only check the consistency of information given by the experts at the time the model is built, but neither consistency ratio nor geometric consistency index can guarantee that these experts will actually make their decisions the exact same way they set the priorities. As a result, the validation of the AHP scoring model is more complicated than those of statistical models.

This paper proposes a detail step of how a credit scoring model can be developed based on expert judgments. The AHP technique is introduced as one alternative along with the validation method needed to complied with BASEL IRB framwork. Other existing validation methods can later be used when the bank has enough number of bad loans in its data. The paper is structured as follow: Section 2 provides a comparison of credit scoring modeling choices. Section 3 describes how expert judgments can be organized into information hierarchy to build a credit scoring model. Section 4 discusses consistency test for expert judgment based scoring model. Section 5 explains how the expert judgment can be validated. Finally, section 6 discusses the conclusions.

2. Model Selection

Credit scoring model is a key component of the automated loan approval system. There are several choices of credit scoring models, which can be divided into two main categories: parametric and non-parametric models. Parametric models include linear probability model, logit or probit model, discrimination analysis-based model, and neural networks while non-parametric models include mathematical programming, classification trees, nearest neighbor model, analytic hierarchy process (AHP), and expert systems.

There are several articles in the literature that try to compare these model choices. Srinivasan and Kim (1987) and Henley and Hand (1997) conclude that classification tree is the best method for credit scoring model while Boyle et al. (1992) and Yobas, Crook, and Ross (1997) conclude that linear regression is the best. However, Desai et al. (1997) argue that logistic regression is the best. Lovie and Lovie (1986) suggest that a large number of scorecards would be almost as good as each other as far as classification is concerned. This means that there can be significant changes in the weights around the optimal scorecard with little effect on its classification performance, and it perhaps explains the relative similarity of the classification methods. This relative stability of classification accuracy to choice of method used has prompted experts to wonder if a scoring system is also relatively stable to choice of customer sample on which the system is built. The systems are very sensitive to differences in the population that make up the scorecard. The regression approaches, both linear and logistic, have

all the underpinning of statistical theory. Thus, one can perform statistical tests to see whether the scores of an attribute are significant, and hence whether that attribute should really be in the scorecard. This allows one to drop unimportant characteristics and arrive at lean, mean, and robust scorecards.

Linear programming deals very easily with constraints that lenders might impose on the scorecard. For example, they may want to bias the product toward younger customers and so require the score for being under 25 to be greater than the score for being over 65. Regression approaches find it almost impossible to incorporate such requirements. One of the other advantages of linear programming is that it is easy to score problems with hundreds of thousands of characteristics, and so splitting characteristics into many binary attributes causes no computational difficulties, whereas statistical analysis of such data sets with large number of variables can cause computational problems.

The methods that form groups like classification trees and neural networks have the advantage that they automatically deal with interactions between characteristics, whereas for linear methods, these interactions have to be identified beforehand and appropriate complex characteristics defined.

2.1 Logit Model

Given all these modeling choices, the logit model is the most popular in banking industry because it is easy to developed, validated, calibrated, and interpreted. Logit model is simply a statistical model to predict qualitative outcomes: default and no-default. When the default events occur, y is equal to 1. Let Π be the probability that the event y occur (y = 1), then the odd ratio can be define as $\frac{\Pi}{1-\Pi}$. Then, logit model can be written as follow:

$$\ln\left(\frac{\Pi}{1-\Pi}\right) = \beta_o + \beta_i x_i + \varepsilon$$

Thus, $\hat{\Pi} = \frac{1}{1+e^{-\hat{\beta}_o - \hat{\beta}_i x_i}}$, which makes $0 \le \Pi \le 1$

 β can be solved by maximum likelihood estimators, or solving $\frac{\partial \ln L}{\partial \beta_i} = 0$, with the likelihood function L

$$L = p(y_1, ..., y_n) = \prod_{i=1}^n p_i(y_i)$$
$$= \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1 - y_i}$$
$$\ln L = \ln \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1 - y_i}$$
$$= \ln \prod_{i=1}^n \left(\frac{\pi_i}{1 - \pi_i}\right)^{y_i} (1 - \pi_i)$$
$$= \sum_{i=1}^n y_i \ln \left(\frac{\pi_i}{1 - \pi_i}\right) + \sum_{i=1}^n \ln(1 - \pi_i)$$
$$= \sum_{i=1}^n y_i \left(\beta_0 + \beta_i x_i\right) + \sum_{i=1}^n \ln(1 + e^{\beta_0 + \beta_i x_i})^{-1}$$
$$= \sum_{i=1}^n y_i (\beta_0 + \beta_i x_i) - \sum_{i=1}^n \ln(1 + e^{\beta_0 + \beta_i x_i})$$

given as:

2.2 AHP Model

Among all of available modeling choices, only the AHP and expert system models do not require historical data of applicants' characteristics in the model developing process. Therefore, both methods can be adopted by banks that do not have enough historical data to develop a credit scoring model. An expert system is particularly applicable when the decision maker makes decisions that are multiple and sequential or parallel and where the problem is ill-defined because of the multiplicity of decisions that can be made. Relatively few examples of an expert system used for credit scoring have been published, and because the details of such system are usually proprietary, none that have been published give exact details.

A modeling procedure based on a group analytic hierarchy process (AHP) is well suited to address these issues because it allows the consideration of such qualitative expert judgments and makes the complex decision making possible by allowing for qualitative measures to derive the scale of priorities. Therefore, the AHP model is able to combine expert judgments concerning the performance attractiveness of the application and convert such combined judgment into credit scores.

The analytic hierarchy process (AHP) was introduced by Thomas L. Saaty (1980). It is a structured process for organizing and analyzing complex decisions. The AHP model is based on the principle that when we make a decision on a given matter, we consider a lot of information and factors, which can be represented as an information hierarchy. The most important step in the AHP is arranging a problem in a hierarchical structure. The decision makers have to decompose their decision problem into a hierarchy of more easily comprehended sub-problems, each of which can then be independently analyzed. Therefore, building this model requires the involvement of experts who define the mapping most suited to the problem. We are interested in how elements at the lowest level affect the top-level factor. Since this impact varies across factors, we need to define their weight or their priority, which are derived by a pairwise comparison of these elements. The decision makers can use their judgments to compare the elements' relative importance. The key element of the AHP is that human judgments, not only the underlying information, can be used to perform the evaluations.

The AHP has been widely developed and become one of the most commonly applied multicriteria decision-making techniques. Its applications include personal or business decisions, public policy decisions, planning economic policies, determining consumer preference, estimating the economy's impact on sales, selecting portfolio, finding conflict resolution, benefit/cost decisions, resource allocation, military decision, and many more. Carlos A. Bana e Costa, Luis Antunes Barroso, and Joao Oliveira Soares (2002) developed a qualitative credit scoring model for business loans based on the similar concepts of the AHP.

The AHP framework also provides a measure for consistency of the decision maker when eliciting the judgment. The group version of AHP is applied when the analysis requires an aggregation of individual qualitative judgments (AIJ) and the aggregation of individual priorities (AIP).

3. Expert Judgments

The bank can construct the information hierarchy of its credit approval process based on the experts. At each level of the hierarchy, the relationships between the elements are established by comparing the elements in pairs. A pairwise comparison can be done by forming a matrix to set priorities. The comparison starts from the top of the hierarchy to select the criterion, and then each pair of elements in the level below is compared. These judgments will then be transformed to the scale of 1 to 9 that represent the relative importance of one element over the other

with respect to the property. The scale of 1 to 9 has the following qualitative meaning.

1: Equal importance of both elements.

3: Weak importance of one element over the other.

5: Much more important of one element over the other.

7: Very much more important of one element over the other.

9: Absolute importance of one element over the other.

2. 4, 6, 8: Intermediate values between two adjacent judgments.

If item i has one of the preceding numbers assigned to it when compared with item j, then j has the reciprocal value when compared with i. We should always compare the first element of a pair (the element in the left-hand column of the matrix) with the second (the element in the row on top).

Overall priorities can be made through synthesizing or pooling together the judgment made in the pairwise comparisons. That is, the weighting and adding are needed to come up with a single number to indicate the priority of each element. With this establishment, we can represent the relative impact of the elements of a given level on each element of the next higher level. These pairwise comparisons are repeated for all elements in each level. We can get the result of a vector of priority, of a relative importance, or of the elements with respect to each property. Then, we need to weigh each vector by the priority of its property to derive the net priority weights for the bottom level. If the number of elements to be ranked is n, then the number of judgment needed is (n*n - n)/2.

The calculation of such priorities (weights) can be done by two methods: the eigenvector method (EVM) and the row geometric mean method (RGMM)

4. Consistency

Perfect consistency is difficult to achieve for the decisions based on human judgments. Human judgments typically change according to circumstances, new experiences, the season, or the time of the day.

4.1 Consistency Ratio

The prioritization procedure was introduced by Saaty (1977, 1980)

Let $C_1, C_2, ..., C_n$ be n elements to be compared, a_{ij} be the relative weight (or priority) of C_i with respect to C_j , $A = (a_{ij})$ be an nxn square matrix, in which $a_{ij} = 1/a_{ji}$ for $i \neq j$, and $a_{ii} = 1$ for all I. A is consistent if $a_{ik} = a_{ij} a_{jk}$

Let ω be an eigenvector (nx1) and λ_{max} be an eigenvalue.

$$A\omega = \lambda_{\max}\omega$$

The measurement of the overall consistency under the AHP is called a consistency ratio.

The inconsistency is captured by a single number, $\lambda_{max} - n$, which reflects the deviations of all a_{ij} from the estimated ratio of priorities ω_i / ω_j . If $\lambda_{max} = n$, A is a consistent matrix. If $\lambda_{max} > n$, a consistency index (C.I.) can be calculated as

$$CI = \frac{\lambda_{max} - n}{n - 1}$$

The larger CI means greater inconsistency. Saaty (1980) proposed the use of a normalized measure, the Consistency Ratio (CR), to provide the inconsistency measure that is independent of the order of the matrix, n.

Consistency Ratio (CR) =
$$\frac{CI}{RI(n)}$$

The random CI, called RI(n), is the expected value over a large number of positive reciprocal matrices of

order n, whose entries are randomly chosen in the set of values $\{1/9, \dots, 1, \dots, 9\}$

$$RI(n) = E[CI(n)]$$

The consistency ratio gives a measure of where the judgments in pairwise comparison matrix lie between totally consistent and totally random. When CR =1, then CI = E[CI(n)] = RI(n), and the judgments are totally random, meaning low precision. High values of CR reflect even more inconsistency and thus we are interested in values of CR as low as possible. Saaty (1980) proposed a rule of thumb for the CR which is a 10% threshold. To improve the consistency when CR is greater than 10%, the most inconsistency judgments are modified and a new ω is derived.

4.2 Geometric Consistency Index

There are several other prioritization procedures, among which is the Row Geometric Mean Method (RGMM). The use of RGMM has significantly increased due to its psychological and mathematical properties. When the prioritization procedure is not the EVM, the aforementioned CI is no longer appropriate, and new consistency measures are required. Priorities under RGMM are given by

$$\omega_i = \left(\prod_{j=1}^n a_{ij}\right)^{1/n}$$

Crawford and Williams (1985) suggest that the estimator of the variance of the perturbations can be used as a measure of the consistency, where the lower the value, the better the consistency of the judgments.

$$s^{2} = S / df = \frac{2 \sum_{i < j} \left(\log a_{ij} - \log \omega_{i} / \omega_{j} \right)^{2}}{(n-1)(n-2)}$$

The degrees of freedom are the differences between the judgments included n(n-1)/2 and the estimated parameters (n-1)

$$n(n-1)/2 - (n-1) = (n-1)(n/2 - 1) = (n-1)(n-2)/2$$

The smaller the s², the shorter the distance between the judgments a_{ij} and the ratios ω_i/ω_j and the smaller the variance of perturbation will be and the better will be the fit between the judgments and the priority vector ω .

A measure of consistency proposed by Crawford and William (1985) are given by

$$\text{GCI} = \frac{2}{\left(n-1\right)\left(n-2\right)} \sum_{i < j} \log^2 e_{ij}$$

where the error terms are $e_{ij} = a_{ij}\omega_j / \omega_1$. GCI can be seen as an average of the squared difference between the log of the errors and the log of unity.

$$GCI = \frac{2}{(n-1)(n-2)} \sum_{i < j} (\log e_{ij} - \log 1)^2$$

Aguaron and Moreno-Jimenez (2003) follow the proposal of Crawford and Williams without entering into the analysis of the validity of the CR as a consistency measure in AHP. They proposed the threshold called Geometric Consistency Index (GCI). GCI can be normalized in a way analogous to that carried out with Saaty's consistency ratio by dividing the value that measures the log quadratic distance between the errors e_{ij} and unity (s²) by its expected value.

The expected value of s^2 is a constant: if the judgments of a pairwise comparison matrix follow independent, reciprocal and identical distributions, the mean of the GCI is given by

 $E[GCI] = Var(log a_{ii})$

2n

Theoretical relation between the GCI and the CR is given by

$$GCI = \frac{2n}{n-2}CI + o(\varepsilon^{3})$$
where $\varepsilon = \max_{ij} \left\{ \log e_{ij} \right\}$ and $e_{ij} = a_{ij}\omega_{j} / \omega_{i}$. Recall that $CR = CI / RI(n)$; thus,

$$GCI = \frac{2n}{n-2}CR * RI(n) + o(\varepsilon^{3})$$

$$= k(n)CR + o(\varepsilon^{3})$$
where $k(n) = \frac{2n}{n-2}RI(n)$

aat

This k(n) shows the relationship between CR and GCI, where RI(n) = E[CI(n)]. The k(n) results of simulation of 100,000 matrices for each order (n) are shown in Table 1.

	Tuble 1 R(II) Simulation Results	
N	RI(n)	k(n)
3	0.525	3.147
4	0.882	3.526
5	1.115	3.717
6	1.252	3.755
7	1.341	3.755
8	1.404	3.744
9	1.452	3.733
10	1.484	3.709
11	1.513	3.698
12	1.535	3.685
13	1.555	3.674
14	1.570	3.663
15	1.583	3.646
16	1.595	3.646

Table 1 k(n) Simulation Results

If the judgment matrices are close to consistency (small errors), then the two measures, CR and GCI are proportional. In general, the behavior of the two measures is similar for the different values of n. For low values of Saaty's CR, the relationship between CR and GCI is linear. This relationship is particularly significant when CR is below 0.1. As the range increases, the slopes estimated through regression decrease due to the small concavity of the relation. With this relationship, Aguaron and Jimenez (2003) proposed the thresholds, shown in Table 2, for GCI corresponding with Saaty's CR:

		Table 2 CR and G	FCI	
CR	0.01	0.05	0.1	0.15
GCI (n=3)	0.0314	0.1573	0.3147	0.4720
GCI (n=4)	0.0352	0.1763	0.3526	0.5289
GCI (n>4)	~0.037	~0.185	~0.37	~0.555

The interpretation of the GCI used for RGMM is analogous to the CR used with the EVM proposed by Saaty. When the values of the GCI is greater than the corresponding threshold, the most inconsistent judgment (that with larger e_{ij}) has to be modified in the sense of approximating a_{ij} to ω_i / ω_j . Recall that $e_{ij} = a_{ij} \omega_j / \omega_i$ which is the error obtained when the ratio ω_i / ω_j is approximated by a_{ij} .

An aggregation of individual decisions can be done either by aggregation of Individual Judgment (AIJ) or aggregation of individual priorities (AIP)

An aggregation procedure of either AIJ or AIP can be carried out by weighted geometric mean method (WGMM). Saaty (1980) and Aczel and Saaty (1983) argue that the WGMM is the only separable synthesizing function that satisfies the unanimity, the homogeneity, and the reciprocal properties.

Let $A^{[k]} = \left(a_{ij}^{[k]}\right)$ be the judgment matrix provided by the kth decision maker when comparing n elements (*i*, j = 1, ..., n) with $\omega^{[k]} = \left(\omega_1^{[k]}, \omega_2^{[k]}, \mathsf{K}, \omega_n^{[k]}\right)$ being its priority vector where $\omega_i^{[k]} > 0$ and $\sum_{i=1}^n \omega_i^{[k]} = 1$, and β_k

being the weight of the kth decision maker, k = 1, 2, ..., m in the group where $\beta_k > 0$ and $\sum_{k=1}^{m} \beta_k = 1$. If all decision makers are equally weighted, then $\beta_k = \frac{1}{m}$ for all k.

 $p_k = \frac{1}{m}$

Group judgment matrix can be represented by

$$A^{G} = \left(a_{ij}^{G}\right) \text{with } a_{ij}^{G} = \prod_{k=1}^{m} \left(a_{ij}^{\left[k\right]}\right)^{k}$$

Group priority vector can be shown as

$$\omega^{G} = \left(\omega_{ij}^{G}\right) \text{with } \omega_{ij}^{G} = \prod_{k=1}^{m} \left(\omega_{ij}^{\lfloor k \rfloor}\right)^{\beta_{k}}$$

From the individual judgment matrices, $A^{[k]} = (a_{ij}^{[k]})$, k =1, 2, ..., m, an aggregating individual judgment (AIJ) can be done by obtaining $A^G = (a_{ij}^G)$ with $a_{ij}^G = \prod_{k=1}^m (a_{ij}^{[k]})^{g_k}$ from the individual judgment matrices $A^{[k]} = (a_{ij}^{[k]})$ by using WGMM. Then, using the RGMM to derive the group priority vector, ω^G . On the other hand, an aggregating Individual Priority (AIP) can be achieve by obtaining the individual priority vectors, $\omega^{[k]} = (\omega_1^{[k]}, \omega_2^{[k]}, \mathsf{K}, \omega_n^{[k]})$ for each decision maker using the RGMM, then deriving the group priorities $\omega^G = (\omega_{ij}^G)$ with $\omega_{ij}^G = \prod_{k=1}^m (\omega_{ij}^{[k]})^{g_k}$ using the WGMM.

Barzilai and Golany (1994) prove that using the RGMM, AIJ and AIP provide the same priorities, which can be easily proved as shown below.

$$\omega_i^G \left(AIJ \right) = \left(\prod_{j=1}^n a_{ij}^G \right)^{1/n} = \left(\prod_{j=1}^n \prod_{k=1}^m \left(a_{ij}^{\lceil k \rceil} \right)^{\beta_k} \right)^{1/n} = \prod_{i=1}^m \left(\left[\prod_{j=1}^n \left(a_{ij}^{\lceil k \rceil} \right)^{\beta_i} \right]^{1/n} \right)^{\beta_i}$$
$$= \prod_{i=1}^m \left(\omega_i^{\lceil k \rceil} \right)^{\beta_k}$$
$$= \omega_i^G \left(AIP \right)$$

However, this result is not true for EVM. The two approaches present the same order of complexity for their

respective algorithms. However, it is usual in practice to start with checking whether each individual judgment is consistent. This means that the $A^{[k]}$, $\omega^{[k]}$, and $GCI^{[k]}$ are known in advance, so it might be simpler and more efficient to work with the AIP approach which requires only o(mn) operations than with the AIJ approaches which requires o(mn²) operations.

Again, if the individual judgments are of acceptable inconsistency, the group judgments are also of acceptable inconsistency.

Xu (2000) suggests that the EVM should be used as prioritization procedure and the WGMM as the aggregation procedure. If the decision makers have an acceptable inconsistency when eliciting the judgments, then so has the group. Escobar, Aguaron, and Jimenez (2004) prove that when using the RGMM as prioritization procedure and the WGMM as the aggregation procedure, the inconsistency of the group is smaller than the largest individual inconsistency. In other words, the group inconsistency is at least as good as the worst individual inconsistency for both aggregation approaches (AIJ and AIP).

At the lowest level of the information hierarchy, the experts define the value where each component is considered bad and neutral so that we can arbitrarily assign the score of 0 and 100 for each component respectively. Then, the piecewise linear interpolation is applied to transform these weights into credit scores. The credit scores of each criterion can be calculated by summing the multiplication of the value scores of each component.

5. Validation

The AHP credit scoring model that passes both consistency ratios (CR) or geometric consistency index (GCI) tests is not always the model that can provide the credit scores that are consistent with the experts' actual decision. These two tests only confirm that the experts are consistent in their priority setting at the time the model is built. The model that can be used in the automated loan approval process should also be consistent with the experts' actual decisions. This paper suggests the application of the KS statistics to perform such a test.

If the model does not present anything, the KS statistics is equal to zero. This can be shown in Table 3. The model provides the same probability of acceptance or rejection in each score range. In other words, this credit scoring model is totally not consistent with the experts' actual decisions, and that the model must be rebuilt.

Table 5 Model with No Predictive Power $(KS = 0)$							
Score	# accept	# reject	% accept	% reject	CDF accept	CDF reject	Difference
0–99	10	10	9.10%	9.10%	9.10%	9.10%	0%
100–199	10	10	9.10%	9.10%	18.20%	18.20%	0%
200–299	10	10	9.10%	9.10%	27.30%	27.30%	0%
300–399	10	10	9.10%	9.10%	36.40%	36.40%	0%
400–499	10	10	9.10%	9.10%	45.50%	45.50%	0%
500–599	10	10	9.10%	9.10%	54.50%	54.50%	0%
600–699	10	10	9.10%	9.10%	63.60%	63.60%	0%
700–799	10	10	9.10%	9.10%	72.70%	72.70%	0%
800-899	10	10	9.10%	9.10%	81.80%	81.80%	0%
900–999	10	10	9.10%	9.10%	90.90%	90.90%	0%
1000	10	10	9.10%	9.10%	100%	100%	0%
Total	110	110	100%	100%	Max D	ifference	0%

Table 3Model with No Predictive Power (KS = 0)

The objective of an actual decision validation is to confirm that the AHP model, which is derived from the experts' decisions, should be at least consistent with their actual decisions. We would like to have the model that has the ability to separate between the likely bad loans and the likely good loans. Figure 1 shows the example of the model's ability to separate between loans with accept and reject decisions. The actual rejection decisions by the experts should be more pronounced among the likely bad loans with low scores while the actual accept decisions by the experts should be more likely to have high scores from the model. When we apply the KS statistics to this actual decision validation, the model with larger KS values will be the model that has more predictive power for actual decisions, providing more confidence for banks to implement it in the automated loan approval process.



Figure 1 Actual Decision Validation

Table 4 shows an example of the KS statistics calculation for the model that has more predictive power for actual decisions.

Table 4 Model with Fredictive Power ($KS = 75\%$)								
Score	# accept	# reject	% accept	% reject	CDF accept	CDF reject	Difference	
0-99	0	31	0%	31%	0%	31%	31%	
100-199	1	25	1%	25%	1%	56%	51%	
200-299	2	17	2%	17%	3%	73%	71%	
300-399	5	10	5%	10%	8%	83%	75%	
400-499	6	6	6%	6%	14%	89%	75%	
500-599	6	5	6%	5%	20%	94%	74%	
600-699	7	3	7%	3%	27%	97%	70%	
700-799	8	2	8%	2%	35%	99%	64%	
800-899	15	1	15%	1%	50%	100%	50%	
900-999	20	0	20%	0%	70%	100%	30%	
1000	30	0	30%	0%	100%	100%	0%	
Total	100	100	100%	100%	Max Differen	ce	75%	

When the bank does not have enough number of bad loans in its data to develop statistical credit scoring model, the AHP credit scoring model can be an alternative model that tries to replicate the decision process of the

experts in lending business. The AHP model should be implemented as one of the decision criteria in automated loan approval process. The bank can decide to use the model to automate the loan approval decision for loans with certain characteristics; for example, the application with loan amount less than a certain threshold (5 million baht) or the application with loan amount larger than a certain loan to collateral value (90%).

6. Conclusions

When the bank has high quality data, the bank can apply a variety of statistical model choices. However, when the bank does not have such high quality data, the bank does not have that luxury of statistical modeling choices. This is true for the banks that do not have good database system, or even if they do, the number of bad loans in the data may be too small to be used to develop any statistical credit scoring model. This is also true for certain types of loan that require a longer time horizon for the loans to become default.

In reality, there is no single bank that has its scoring/rating system built entirely on statistical models. Some types of loans require a much deeper and more time-consuming analysis, in which qualitative expert judgments concerning performance attractiveness of the application play a very important role. This paper provide a detail step of how those expert judgment can be aggregated and used to develop a scoring/rating model for the bank.

Since the validation of credit scoring model is the primary responsibility of the bank to confirm that the developed model is applicable and robust. While the validation should encompass both quantitative and qualitative elements, the validation methods available in the literature heavily based on statistical methods, which makes it almost impossible for the expert judgment based scoring model to be validated.

This paper introduces an actual decision validation based on the KS statistics to confirm that the expert judgment based model is consistent with the actual loan approval decisions by the experts.

References:

- Aczel J and Saaty T. L. (1983). "Procedures for synthesizing ratio judgments", *Journal of Mathematical Psychology*, Vol. 27, No. 1, pp. 93–102.
- Barzilai J. and Golany B. (1994). "AHP rank reversal, normalization and aggregation rules", *Information System and Operational Research*, Vol. 32, No. 2, pp. 57–64.
- Boyle M., Crook J. N., Hamilton R. and Thomas L. C. (1992). "Methods for credit scoring applied to slow payers", *in: Credit Scoring and Credit Control*, Oxford: Clarendon, pp. 75–90.
- Costa B., Carlos A., Barroso L. and Soares J. O. (2002). "Qualitative modelling of credit scoring: A case study in banking", *European Research Studies Journal*, Vol. 5, No. 1–2, pp. 37–51.
- Crawford G. and William C. (1985). "A note on the analysis of subjective judgment matrices", *Journal of Mathematical Psychology*, Vol. 29, pp. 387–405.
- Desai V. S., Convay D. G., Crook J. N. and Overstreet G. A. (1997). "Credit scoring models in the credit union environment using neural networks and genetic algorithms", *IMA Journal of Mathematics Applied in Business and Industry*, Vol. 8, pp. 323–346.

Dyer J. S. (1990). "Remarks on the analytic hierarchy process", Management Science, Vol. 36, pp. 249–258.

- Escobar M. T., Aguarón J. and Jiménez J. M. M. (2004). "A note on AHP group consistency for the row geometric mean priorization procedure", *European Journal of Operational Research*, Vol. 153, No. 2, pp. 318–322.
- Glennon D. C. (1998). "Issues in model design and validation", in: E. Mays (ed.), *Credit Risk Modeling Design and Application*, Glenlake Publishing, pp. 207–221.
- Henley W. E. and Hand D. J. (1997). "Statistical classification method in consumer credit scoring: A review", *Journal of the Royal Statistical Society*, Vol. 160, No. 3, pp. 523–541.
- Lovie A. D. and Lovie P. (1986). "The flat maximum effect and linear scoring model for prediction", *Journal of Forecasting*, Vol. 5, pp. 159–186.
- Saaty T. L. (1980). The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation, McGraw-Hill.

Saaty T. L. and Vergas L. (1984). "Inconsistency and rank preservation", *Journal of Mathematical Psychology*, Vol. 28, pp. 205–214. Saaty T. L. (1990). "Eigenvector and logarithmic least square", *European Journal of Operational Research*, Vol. 48, pp. 156–160.

Srinivasan V. and Kim Y. H. (1987). "Credit granting: A comparative analysis of classification procedures", *Journal of Finance*, Vol. 42, No. 3, pp. 665–681.

- Takeda E. (1993). "A note on consistent adjustment of pairwise comparison judgments", *Mathematical and Computer Modelling*, Vol. 17, pp. 29–35.
- Takeda E. and Yu P. L. (1995). "Assessing priority weights from subsets of pairwise comparisons in multiple-criteria optimization problems", *European Journal of Operational Research*, Vol. 86, pp. 315–331.
- Yoba M. B., Crook J. M. and Ross P. (1997). "Credit scoring using neural and evolutionary techniques", *Working Paper 97/2*, Credit Research Center, University of Edinburgh.
- Zahedi F. (1986). "The analytic hierarchy process—A survey of the method and its applications", *Interfaces*, Vol. 16, No. 4, pp. 96–108.
- Xu Z. (2000). "On consistency of the weighted geometric mean complex judgment matrix in AHP", *European Journal of Operational Research*, Vol. 126, No. 3, pp. 683–687.