THE JOURNAL OF FINANCE • VOL. XLII, NO. 3 • JULY 1987

Credit Granting: A Comparative Analysis of Classification Procedures

VENKAT SRINIVASAN and YONG H. KIM*

ABSTRACT

Financial classification issues, and particularly the financial distress problem, continue to be subject to vigorous investigation. The corporate credit granting process has not received as much attention in the literature. This paper examines the relative effectiveness of parametric, nonparametric and judgemental classification procedures on a sample of corporate credit data. The judgemental model is based on the Analytic Hierarchy Process. Evidence indicates that (nonparametric) recursive partitioning methods provide greater information than simultaneous partitioning procedures. The judgemental model is found to perform as well as statistical models. A complementary relationship is proposed between the statistical and the judgemental models as an effective paradigm for granting credit.

THE CREDIT GRANTING process involves a tradeoff between the perceived default risk of the credit applicant and potential returns from granting requested credit. The main objective in credit granting is to determine the optimal amount of credit to grant. The amount of credit requested along with other financial and nonfinancial factors influences the assessment of default risk. While default risk assessment can be appropriately modeled using classificatory models, integration of such risk assessment with potential returns can be accomplished using a dynamic expected value framework.[1]

Academic research on credit granting can be grouped into two basic categories:[2] (i) attempts to apply classification procedures to customer attribute data and develop classification models to assign group membership in the future; and (ii) attempts to explicitly recognize the need to integrate such risk assessment with potential return as well as allow for differing degrees of credit investigation depending on the costs of such investigation. To our knowledge, this study represents the first attempt at examining the relative performance of classification procedures for corporate credit granting. Relative performance is measured in terms of how well the models replicate expert judgement. We examine four statistical models: multiple discriminant analysis (MDA), logistic regression

* Northeastern University and University of Cincinnati, respectively. We are grateful to Robert Eisenbeis and Sangit Chatterjee for their insightful comments on an earlier draft. We also thank an anonymous Fortune 500 corporation as well as the Credit Research Foundation for their support. The usual disclaimer applies.

[1] This paper is primarily concerned with the credit granting decision in an industrial (nonfinancial) setting. A more detailed version of this paper is available from the authors. Further, for a review of the various motives for corporations to extend credit, refer Emery [10].

[2] We do not make any attempts to present a comprehensive review of the literature. Interested readers are referred to an excellent review by Altman et al. [1].

(logit), goal programming (GP) and recursive partitioning algorithm (RPA), and a judgemental model based on the Analytic Hierarchy Process (AHP). The data employed for the study has been provided by an anonymous Fortune 500 corporation.

The remainder of this paper is organized as follows. Section I presents the credit granting problem formally. We present the four classificatory models and the judgemental model in Section II. The data and methodology are summarized in Section III. Section IV contains a discussion of the results as they pertain to classificatory ability and the determination of credit limits. The paper concludes with a summary.

I. The Corporate Credit Granting Problem

The prescriptions of normative theory in credit granting (e.g., Altman et al. [1]) can be summarized as follows: (i) credit granting is a multiperiod problem, implying that granting credit may not only enable the firm to make the current sale but also to sell in the future to the same customer; (ii) the extent of credit investigation must be determined by the tradeoff between the incremental costs of investigation and the amount of credit involved; (iii) estimate the present value of benefits and losses from granting credit for all the periods in the planning horizon, given the customer's action in the immediately prior period; (iv) integrate probabilities of collection and default with benefits and losses each period to compute the net present value from granting credit; and (v) grant credit, if computed net present value is positive and reject full credit, otherwise.[3]

The problem of setting credit limits can be formulated as a dynamic program (Bierman and Hausman [2]; Srinivasan and Kim [24]). The dynamic program considers the influence of past collection experience on default probabilities in the future. It is, however, very difficult (if not impossible) to estimate future benefits and losses from granting credit other than at a qualitative level. Elimination of future benefits and losses from the dynamic program reduces it to a single-period formulation which is, of course, computationally more tractable. The recurrence relationship in the single-period formulation can be stated as:

$$f_i(P_{1i}) = \text{Max}[(P_{1i}k_1) + (P_{2i}k_2), 0] \tag{1}$$

where

$f_i(P_{1i})$ = maximum discounted expected payoffs from state i to ∞, given the current state variable (P_{1i}) and that optimal decisions are made from now on (the return function);

P_{1i} = probability of payment in period i, given customer action in period $i-1$;

P_{2i} = probability of default in period i, given customer action in period $i-1$;

k_1 = present value of immediate benefits in the event of collection; i.e., discounted profits from sale or s a^{-c1} − v a^{-c2};

[3] Alternatively, the maximum credit limit to be granted will be determined by equating NPV to zero.

k_2 = present value of loss in the event of default or $-\text{v}\ a^{-c2}$;
$c1$ = average selling cycle;
$c2$ = average payables cycle;
$a = 1/(1 + d)$ = appropriate discount factor;
d = appropriate before-tax discount rate per period;
s = sales; and
v = variable costs.

In the context of a single period, expression (1) is simply the E(NPV) from granting credit in the first period. Assuming for convenience that variable costs (v) can be expressed as a proportion of sales (s), formulation (1) can be restated as follows to determine the maximum credit limit:

$$P_{1i}k_1 - P_{2i}CL_{\max}(\text{v/s}) = 0 \tag{2}$$

where, $CL_{\max}$ is the maximum credit line that should be granted to the customer.

The above description implies that the credit granting process consists of two stages: (i) estimation of default probabilities; and (ii) integration of default probabilities in (1) or (2) to estimate credit limits. Two basic approaches have evolved to facilitate the assessment of default risk: statistical and judgemental systems. All credit analysis, whether based on judgemental or statistical systems, operate on similar principles. Further, in estimating default probabilities, the suggested multistage investigation process essentially determines the feasible set of measurements, X, that can be used to estimate the probabilities, P_{1i} and P_{2i}.

II. Classificatory Models for Credit Evaluation

A. Statistical Models

The two parametric methodologies we employ, MDA and logit, have been extensively investigated in prior classification studies. Interested readers are referred to Altman et al. (1981) for a detailed discussion. It will suffice to mention here that the default probabilities for each observation using MDA and logit will be the posterior probabilities of group membership generated from the respective MDA and logit models. These probabilities can then be substituted in (1) or (2) to determine credit limits.

Both MDA and logit necessitate restrictive assumptions of distributional form, being parametric in nature. In particular, the problems in using MDA for financial classification are well documented (see, e.g., Altman et al. [1]). Several nonparametric classification alternatives have been suggested in the literature. Among these are the rank transformed discriminant analysis (Conover and Iman [6]), the recursive partitioning algorithm (RPA) (Breiman et al. [2]), the Quinlan algorithm (Quinlan [17]), and goal programming (GP) (Freed and Glover [12]). We focus on the GP and the RPA in this study.

MDA, logit and GP can be characterized as simultaneous partitioning procedures since they consider all the potential discriminatory variables simultaneously in the classification model. However, models like the RPA can be charac-

terized as sequential partitioning procedures that are combinatorial in nature.[4] There is a fundamental reason why repeated partitioning procedures like the RPA may yield greater information than simultaneous partitioning procedures.[5] Simultaneous partitioning procedures implicitly assume convexity of group spaces in the p-variate measurement space. This may be unrealistic in the context of many financial classification issues, particularly in multigroup classification problems where there is likely to be considerable non-convexity in group spaces. Further, it is feasible that the underlying decision process may be a conditional and sequential one instead of being a simultaneous process. The data may also have a bimodal or multimodal distribution. In such cases, greater information and hence better classification accuracy, may be obtained by examining characteristics in an incremental manner using procedures like the RPA.

1. Mathematical Programming (MP)

Recently, the interest in MP-based alternatives to parametric classification procedures seems to have been revived by Freed and Glover [12]. Formally, in a MP context, the discriminant problem can be stated as one of finding a linear transformation X and boundary value b to categorize group k, given points A_i and groups G_k. Thus, for a 2-group problem, the objective is to find X such that:

$$A_iX \leq b \qquad A_i \in G_1$$

$$A_iX > b \qquad A_i \in G_2$$

where b is any positive constant. Since, however, the A_i's may be distributed in a manner that may not permit complete discrimination, we need to relax the constraints to ensure feasible solutions. The slack (surplus) variable for the purpose, a, can take two forms: (i) it can be constrained at the group level, a_k; or (ii) it can be constrained at the observation level, a_i. Constraining the a's at the group level yields LPI:

$$\text{Min } \textstyle\sum_{k=1}^{2} a_k$$

such that

$$A_iX \leq b + a_k \qquad i = (1, \ldots, Q), \qquad i \in G_1, \qquad k = (1, 2).$$

$$A_iX > b - a_k \qquad i = (Q + 1, \ldots, R), \qquad i \in G_2, \qquad k = (1, 2).$$

where the X's are unrestricted in sign. Constraining the a's at the observation level, on the other hand, yields LPII:

$$\text{Min } \textstyle\sum_{k=1}^{2} a_k$$

such that

$$A_iX \leq b + a_i \qquad i = (1, \ldots, Q), \qquad i \in G_1, \qquad k = (1, 2).$$

$$A_iX > b - a_i \qquad i = (Q + 1, \ldots, R), \qquad i \in G_2, \qquad k = (1, 2).$$

[4] A number of sequential partitioning algorithms have evolved in the pattern recognition literature. For a review of some of these procedures, refer Diettrich and Michalski [7].

[5] For details, see Chatterjee and Srinivasan [5].

where the X's, as before, are unrestricted.[6] Typically, the b values have been set at 1 or 10. Estimation of the probability of group membership in the GP formulations can be approximated by the relative frequencies of the groups in each of the two halfplanes. These probabilities can, as before, be substituted in formulation (1) or (2) to obtain credit limits.

2. *Recursive Partitioning*

RPA yields a binary classification tree, similar to a spanning tree, that enables the assignment of objects into one or k known groups. Binary classification trees are constructed by repeated splits of X, where X is the measurement space containing all relevant measurement vectors. The construction of the binary classification tree revolves around the following three elements:

1. the selection of splitting rules;
2. the decision when to declare a node as terminal; and
3. the assignment of each terminal node to a group.

The performance of the model critically depends on the first two elements. It turns out that the third element is trivially resolved, given solutions to the first two elements. The fundamental notion behind each split of a subset is one obtaining descendant subsets that are 'purer' than the data in the present subset.

The best splitting rule for the given sample is defined as the one which maximizes the decrease in the sum of the impurities of the two resulting subsamples compared with the impurity of the parent sample. In order to find the best splitting rule, the algorithm first searches for the best splitting point for each explanatory variable and then the best of these splits is selected. The splitting process terminates when further splitting does not lead to any decrease in the impurity of the current tree. The current tree then is referred to as T_{fin}.

The next step in the process is to search for a tree that is less complex than T_{fin} but has a smaller cross-validated classification error rate. This step is motivated by the observation that T_{fin} usually overfit the data and are complex. The final optimal tree is selected based on a criterion of minimizing cross-validated resubstitution risk. The cross-validation risk of tree T is computed using a V-fold cross-validation process, where V is the number of folds. All cases in the sample are randomly divided into V groups of approximately equal sizes. Observations in $V - 1$ groups are used to construct the tree corresponding to the penalty chosen from the range of values of penalty parameters for which tree T was optimal, based on all of the observations. The observations in the group left out are classified by the newly constructed tree. The procedure is repeated V times, each time with a different group being left out. The resubstitution risks from all cross-validation tests for a particular candidate T are averaged to obtain cross-validated risk for T.

[6] In some exceptional cases, LPI and LPII can be shown to be degenerate (Markowski and Markowski [15]). However, simple normalizations can be undertaken to remedy the problem (see, Freed and Glover [11]). In the data under study, we did not have to resort to any normalizations.

B. A Judgemental Model Based on the AHP[7]

AHP is a multiattribute modeling (MAD) approach that has substantial intuitive appeal and is theoretically rigorous. Like other MAD approaches, AHP also attempts to resolve conflicts and analyze judgements through a process of determining the relative importance of a set of activities, players or criteria.[8] The AHP starts by decomposing the principal problem into a hierarchy. Each level consists of a set of elements and each element, in turn, is broken into sub-elements for the next level of the hierarchy. The final level consists of the specific courses of actions that are being evaluated for adoption. Within each hierarchical level, priorities are established for each of the elements using a measurement methodology. This methodological process constitutes the core of the AHP and determines the scope of data collection and analysis.

The basic assumption underlying the measurement methodology in AHP is that relative dominance can be measured by pairwise comparisons. Assume that we wish to conduct pairwise comparisons of a set of n attributes and establish their relative weights. If we denote the attributes by $O_1, O_2, \ldots, O_n$ and their weights by $w_1, w_2, \ldots, w_n$, the pairwise comparison matrix $\mathbf{O}$ may be expressed as the reciprocal matrix shown below:

$$\mathbf{O} = \begin{array}{c} \\ O_1 \\ O_2 \\ \vdots \\ O_n \end{array} \begin{array}{c} \begin{array}{cccc} O_1 & O_2 & \cdots & O_n \end{array} \\ \begin{bmatrix} w_1/w_1 & w_1/w_2 & \cdots & w_1/w_n \\ w_2/w_1 & w_2/w_2 & \cdots & w_2/w_n \\ \vdots & \vdots & & \vdots \\ w_n/w_1 & w_n/w_2 & \cdots & w_n/w_n \end{bmatrix} \end{array}$$

Note that if we post-multiply $\mathbf{O}$ by the column vector $\underline{w}$, we obtain the vector $n\underline{w}$. To recover the weights, $\underline{w}$, we can solve the system $(\mathbf{O} - n\mathbf{I}) = 0$, which must have a non-trivial solution, since $\mathbf{O}$ has unit rank and all the eigenvalues of $\mathbf{O}$, $(L = 1, 2, \ldots, n)$, are zero except one $(L_{\max})$. Further, the trace $(\mathbf{O})$ = sum of the eigenvalues = sum of the diagonal elements of $\mathbf{O} = n$.

The matrix $\mathbf{O}$ satisfies the 'cardinal' consistency property $o_{ij}o_{jk} = o_{ik}$. Thus, given any row of A, we can determine the rest of the entries from this relation. However, suppose we have estimates of the ratios in the matrix. In this case, the cardinal consistency relation above need not hold, nor need an 'ordinal' transitivity relation of the form: $O_i > O_j$, $O_j > O_k$ imply $O_i > O_k$ hold (where the O_i are the rows of $\mathbf{O}$). It can be shown that in a positive reciprocal matrix, small perturbations in the coefficients imply small perturbations in the eigenvalues and hence the eigenvector is insensitive to small changes in judgement [18, p. 192].

[7] This model is more fully described in Srinivasan and Kim [23]. Further, note that even though we have included the description of the AHP-based model under classification methods, it is strictly not a classification model. The model is, however, directly usable to determine credit limits.

[8] AHP has been applied to many varied problems including planning for a national waterway (Saaty and Vargas [20]) and marketing decisions (Wind and Saaty [27]). Srinivasan and Kim [22] illustrate the applicability of AHP to many financial decisions. For an extensive listing of applications of AHP, see Zahedi [28]. For theoretical details, refer Saaty [18]. For an axiomatic treatment of the process, refer Saaty [19].

In general, the weights are estimated from data or from experienced decision makers or a group of experts. A 9-point measurement scale is used, where 1 stands for equal importance of element i over element j and 9 stands for absolute importance of element i over element j. The principal eigenvector of this matrix is then derived and weighted by the priority of the element in the higher level with respect to which the evaluation has been made. Continuing this process of eigenvector extraction and prioritization by weighting leads to a unidimensional priority scale for the elements in each of the hierarchical levels.

The next step in the process is to obtain a measure of the consistency of the judgemental data. It can be shown that $L_{\max} \geq n$ for all possible states and that $(L_{\max} - n)/(n - 1)$ serves as an index measure of consistency of the comparison data (Saaty [18]). The index (C.I.) indicates the departure from consistency of the comparison ratios, w_i/w_j, and the ratios are deemed consistent if $L_{\max} = n$. The consistency index is compared to a level of consistency that can be obtained merely by chance. Saaty and Mariano [21] have established for different order random entry reciprocal matrices, an average consistency index which ranges from 0 for 1 to 2 element matrices to 0.9 for 4 element matrices and to 1.49 for 10 element matrices. In general, a consistency ratio of 10% or less is considered very good. If consistency is poor, additional data or another round of comparisons may be required.

In the context of the credit granting problem, the final level in the hierarchy will consist of two elements whose relative weights can be interpreted as the subjective probabilities of granting and rejecting full requested credit. These probabilities can again be substituted in (1) or (2) to derive the optimal credit limits.

III. Data and Methodology

A. Data

The actual credit granting process in the participating corporation (hereafter PC) is reported in detail in Srinivasan and Kim [23]. We only describe the relevant facts here. The process can be defined as a multistage process where the stages have been established as a function of the size of the credit request. There are four stages in the process:

Stage I: Less than \$5,000
Stage II: \$5,001–\$20,000
Stage III: \$20,001–\$50,000
Stage IV: Greater than \$50,000

The extent of investigation varies across the stages with stage I undergoing very little or no investigation and stage IV undergoing extensive investigation. Even extensive investigation in many cases is limited by the amount of financial information that can be obtained. Many customers are privately held corporations and the firm's main source for financial information is the Dun and Bradstreet Corporation (DBC). DBC, however, only provides balance sheet information, net profits and sales for the year. Income statement details are not provided.

Complete financial evaluation is done (with available information) in addition to nonfinancial evaluation (e.g., trade and bank references, pay record, customer background) only in the case of stage IV customers. Analysis of stage IV customers is supervised by senior level credit management staff and all credit decisions relating to these customers are made by such staff. For purposes of this study, an appropriate senior level credit manager was chosen as the 'expert'.[9] If a stage IV customer is perceived to be 'poor' risk, the customer is classified as a 'high risk' (HR) and one of several actions is taken:

a. Future credit shipments are stopped effectively eliminating the customer's credit line;
b. Future credit shipments are only undertaken if proper collateral is furnished;
c. All future shipments are stopped. Appropriate legal action is initiated. This is often the course of action for customers who file for bankruptcy.

For the purpose of this study, the procedure yields two groups of interest: (i) customers where the firm has taken restrictive action on their credit lines (note that these customers need not have filed for bankruptcy); and (ii) customers where the firm has not taken restrictive action on credit lines (a very small portion of this group are customers who at some prior period belonged to group (i)).

Data were collected on the two groups of interest as of July 1986.[10] Several conditions were imposed on the data:

1. Only credit lines in excess of $50,000 and not previously analyzed in Srinivasan and Kim [23] were considered;
2. Complete financial information (DBC) was available for up to within 7 months of July 1986;
3. Restrictive action on credit lines was significant.[11]

The above process yielded a total of 215 customer files out of which 39 customers were HR customers and the remaining 176 customers were non-HR customers.

[9] Note that the study assesses the ability of statistical and judgemental models to replicate expert judgement. In this respect, it is similar to prior studies that have attempted to predict investment quality ratings (e.g., Srinivasan et al., [25]) and also to the loan classification study by Marais et al. [14]) that attempts to replicate loan officer judgements. Our objective could be argued to raise a potential concern of determining whether an error is due to the models or an error on the part of the credit manager. However, we feel that the objective is justified under the circumstances prevalent in the PC. The PC is convinced that its existing system is best suited for its business conditions. It has experienced very negligible losses. The interest in this study stems from the desire on the part of the PC to identify and capture the underlying expert judgement process so that the entire process can be replicated using artificial intelligence. Therefore, the appropriate objective for this study is to attempt to replicate the expert's judgement. Further, note that the classification of interest in the study is defined by restrictive credit action initiated by the expert.

[10] Theoretically, it can be argued that the study ignores the population of customers who were denied credit in the first place. However, this argument is not valid in the PC's case as far as stage IV is concerned. In the PC's case, customers normally start in stage I and gradually move through the upper stages over time. The PC very rarely has received first-time credit requests that fall into stage IV. Further, the PC has never denied initial credit to the few first-time stage IV customers.

[11] There were a few cases (=5) where the firm continued to sell on credit because of marketing needs even though the customers were classified as HR.

For each customer, available information was used to compute the following ratios:

a. current ratio (CR);
b. quick ratio (QR);
c. net worth to total debt (NWTD);
d. logarithm of total assets (LOGTA);
e. net income to sales (NIS);
f. net income to total assets (NITA)

In addition to the above financial variables, two nonfinancial variables were also included: (i) pay record which is indicative of the firm's past collection experience; and (ii) customer background proxied by the number of years the customer has been in business. Pay record was treated as a categorical variable with three possible values (Mehta [16]):

Good: when payments have been consistently made within the prescribed period of one month;
Fair: when the formal credit period has been frequently exceeded, but there has been no need for stronger than formal payment reminders; or the customer has given a satisfactory explanation for delinquency;
Poor: when stronger and personal reminders have been necessary in the past, and credit has been at times exceeded beyond a three-month period without reasonable or satisfactory explanations.

Similarly, customer background was also treated as a categorical variable with three values: (i) less than 2 years; (ii) 2 to 5 years; and (iii) greater than 5 years.

Exploratory data analysis was conducted by reviewing the sample data behavior of each variable. Obviously, the two categorical variables are nonnormal. All the six financial variables also violated univariate normality. The financial ratios could be transformed to approximate univariate normality. The data even after transformations is not likely to be multivariate normal due to the categorical variables. Moreover, transformations may change the relationships among the variables and may also affect the relative position of observations. Further, exploratory MDA models based on transformed data did not yield significantly better results. The remainder of the analysis was, therefore, done only based on the untransformed data.

B. Methodology

The real test of the classification models in the context of this study is their ability to replicate the expert's judgements. Marais et al. [14] identify three types of overfitting bias that can result if expected misclassification losses are estimated from the same sample used for model estimation: (i) overfitting in the choice of explanatory variables; (ii) sensitivity to a particular loss function; and (iii) statistical overfitting bias in the computation of the expected loss rate.

The first type of bias is clearly not relevant for this study as we do not use any data reduction procedure. Further, loss functions are typically very hard to estimate as also admitted by Marais et al. [14]. A variety of methods have been proposed to deal with the statistical overfitting bias. In most prior classification

studies, some form of holdout samples has been used. Alternatively, some form of statistical resampling schemes may be used, e.g., cross-validation, bootstrapping. We use the bootstrap for the MDA, logit and GP. Cross-validation is used for RPA.

Bootstrapping is a computer intensive technique that is especially appropriate when the data violate the distributional assumptions of the statistical procedure employed. Efron [8] finds various bootstrap estimators to dominate cross-validation or jackknifing, particularly in small samples. Chatterjee and Chatterjee [4] suggest a superior procedure where the model developed from the bootstrap sample is used to classify observations not included in the bootstrap sample. The bootstrap procedure adopted in this study can be described in terms of the following three steps:

1. From the sample being studied, S, with a total of N_s observations, draw an independent sample of size n_s each observation being drawn without replacement with a probability of $1/(N_s - K_i)$, where K_i is the total number of observations already selected as of the ith selection, $i = \{1, 2, \ldots, n_s\}$. The sample thus drawn constitutes the bootstrap sample and is used for estimating the discriminant function.
2. The resubstitution loss rate of the discriminant function is assessed by applying it to classify all the observations not included in the bootstrap sample.
3. The above procedure is repeated 25 times. Each replication provides an estimate of the misclassification probability for both the bootstrap and validation samples. The average of all the replications is taken as the bootstrap and validation estimates. The standard deviation of the estimates provides an estimate of the standard error of misclassification probabilities.

Further, for the purposes of this study, n_s was determined by drawing from a uniform distribution such that 50% of the sample in each of the two groups constituted the bootstrap sample.

IV. Results

A. Classification

The cross validated results of a linear MDA model using proportional prior probabilities are presented in Exhibit 1. On the average, the linear MDA correctly classifies 88.89% of all the customers in the bootstrap sample. The average classification accuracy drops to 85.05% in the case of the holdout sample. There is some variation in the classification accuracy within the two groups: 90.9% of the non-HR customers were classified correctly and only 80% of the HR customers were correctly classified. Cross validated classification accuracy is lower but not appreciably. The overall resubstitution loss rate is about 3.84% but is more severe in the HR cases.

The homogeneity of dispersion matrices was tested using a likelihood ratio test [16] which is approximately chi-square distributed. The results indicated that all the 25 bootstrap samples did not meet the homogeneity assumption. A quadratic

Exhibit 1
Summary of Classification Results (% Correctly Classified)

	LMDA	QMDA	Logit	LP I	LP II	RPA 1	RPA 2
i. Bootstrap Sample							
Non HR	90.90	92.05	93.18	89.77	90.90	95.45	94.32
	(1.66)	(1.29)	(1.73)	(1.40)	(1.51)		(0.99)
HR	80.00	85.00	90.00	80.00	80.00	89.74	90.00
	(2.48)	(2.32)	(4.21)	(2.83)	(2.45)		(0.00)
Total	88.89	90.74	92.59	87.96	88.39	94.44	93.52
	(1.66)	(1.37)	(2.11)	(1.59)	(1.67)		(0.80)
ii. Validation Sample							
Non HR	87.50	88.64	89.77	86.36	88.64	93.18	93.18
	(1.01)	(1.52)	(1.22)	(1.46)	(1.27)		(0.34)
HR	73.68	78.95	78.95	73.68	78.95	89.74	89.74
	(2.53)	(2.46)	(2.36)	(3.36)	(2.58)		(2.48)
Total	85.05	86.92	87.85	84.11	86.92	92.56	92.29
	(1.08)	(1.56)	(1.41)	(1.64)	(1.52)		(0.69)

1. All figures are averages over the 25 replications. Figures in parentheses represent the estimated standard error of percent correctly classified. Details of replication results are available from the authors.

2. Results for the RPA 1 are from using 10-fold cross-validation. RPA models were based on the symmetric Gini criterion and no linear combinations were used.

discriminant rule was, therefore, applied and the results are summarized in Exhibit 1. The classification accuracy has improved in the case of the bootstrap sample to 90.74% compared to 88.89% using the linear rule. However, the resubstitution loss rates are slightly higher. This may be a result of the overfitting that accompanies a quadratic rule. Replication results do not reveal any significant variation over the bootstrap trials.[12]

Relative influence of variables was ascertained using a forward stepwise selec-

[12] Note than even though the data do not satisfy some assumptions of the MDA models, the bootstrap procedure can be argued to, at least partially, offset this bias.

tion rule. We found that the QR, NITA and NWTD were consistently significant in all the 25 samples. Variables that were selected less frequently were CR and PAY. Remaining variables were not found significant univariate discriminators in the stepwise selection rule used.

Results using the unordered logit model are also presented in Exhibit 1. Overall as well as within group classification results are better than the linear MDA models and comparable to the quadratic MDA. Note the relatively significant variation in classification results for HR customers in the bootstrap sample. Analysis of the replication results does not, however, reveal any particular replication to have caused the high variation. Logit model results reveal that CR, QR, NITA and PAY were consistently significant at the 0.05 level in all of the 25 replications.

LPI, which allows deviation only at the group level, yields an overall classification accuracy of 87.96% and 84.11% for the bootstrap and holdout samples, respectively (Exhibit 1). The results are comparable to the linear MDA but not as good as the results using logit. LPII does not appreciably improve classification results implying that the data are relatively well behaved and allowing the deviation variable at the observation level did not have a great influence. Overall classification accuracies using LPII are 88.89% and 86.92% for the bootstrap and holdout samples, respectively. Variances from both LPI and LPII models are comparable to other models.

In the case of the RPA, overall classification accuracy for the estimation sample was 94.44% (Exhibit 1). The cross-validated accuracy using 10-fold cross-validation was 92.52%. These results are slightly superior to the results obtained using logit. Sensitivity analysis was conducted to determine if the results were dependent on the number of folds used in the cross-validation process. Folds were varied from 1 to 10 and the results are presented in Exhibit 1. As is apparent, the results using RPA are not very sensitive to changes in the number of folds (1 to 10). We present the optimal cross-validation tree, based on 10-fold cross-validation, in Exhibit 2. The tree has 8 terminal nodes and the node membership is also indicated for each node. The optimal RPA tree considers the variables NITA, CR, PAY, NWTD, and LOGTA. Each node has either been assigned to the HR or the NHR category based on node membership. Note further that the tree in Exhibit 2 is a pruned tree reflecting the tradeoff between tree complexity and resubstitution loss. The complex tree, T_{fin}, will classify all the observations but will, of course, have less generalizability.

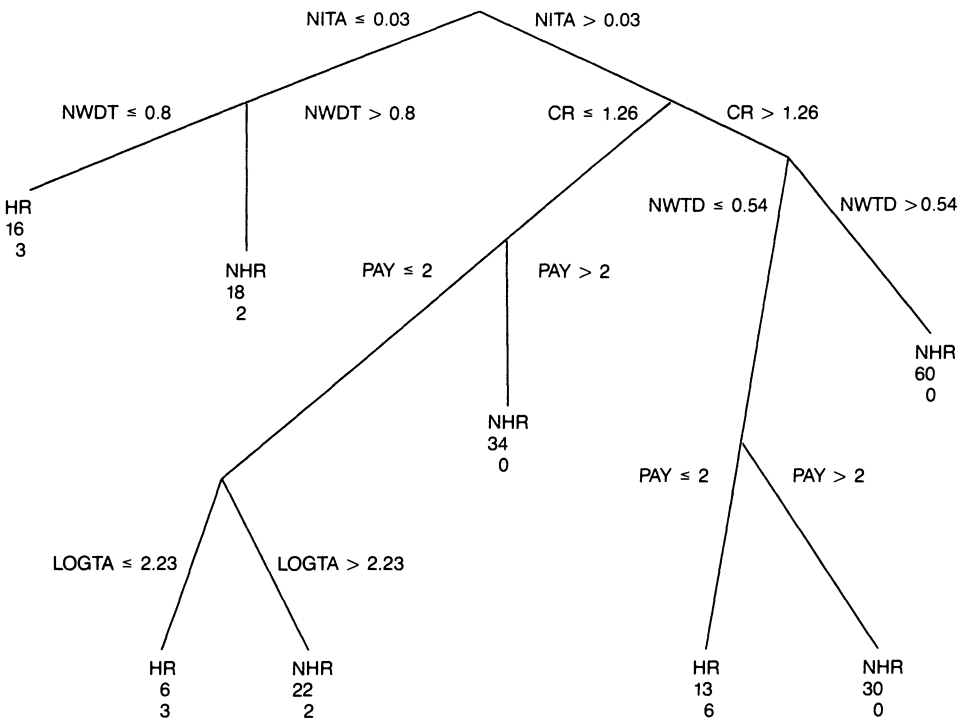
In comparison, RPA appears to be slightly superior as a classification model to replicate the expert's judgements in this study. The classification results obtained using RPA are better than logit, MDA and GP. Besides, they appear robust to changes in the cross-validation scheme used.

B. Credit Limits[13]

We conducted a judgemental input elicitation process for stage IV customers. All judgements were provided by the senior credit manager ('expert') who is also responsible for classifying stage IV customers as HR customers. After many

[13] We present the AHP model first and subsequently provide a comparative assessment of the methodologies in determining credit limits.

Exhibit 2
Optimal RPA Tree



The numbers below each node indicate node membership. The top number is the frequency of the group to which the node has been assigned. The tree is based on a 10-fold cross-validation.

HR = High Risk
NHR = Non-High Risk

PAY:
Good = 2
Fair = 1
Poor = 0

rounds of deliberations (the process is described in Srinivasan and Kim, [23]), five factors were considered relevant: customer background, payment record, geographical location, business potential and frequency and financial soundness.[14] The pairwise reciprocal matrix for the five factors is presented in Exhibit 3. L_{max} is 5.2146 with a consistency ratio of 0.05. The resulting global weights for

[14] Note that within each factor, the expert would consider all criteria that in the expert's opinion, were relevant. Not all of the criteria are quantifiable and, therefore, *ceteris paribus*, we can expect some divergence between the results using AHP's subjective probabilities and objective probabilities from MDA, logit, RPA and GP.

EXHIBIT 3
AHP MODEL ESTIMATION
Reciprocal Judgement Matrix for Stage IV:

Factor	1	2	3	4	5
1	1	0.33	9	3	.125
2	3	1	9	3	.125
3	0.11	0.11	1	.125	.111
4	0.33	0.33	8	1	.125
5	8	8	9	8	1

Lmax = 5.2146 C.I. = 0.054

Weights:	0.0295	0.1236	0.0016	0.022	0.823

Factor Definition:

1 : Customer Background
2 : Payment Record
3 : Geographical Location
4 : Business Potential
5 : Financial Soundness

each of the factors for stage IV customers is also shown in Exhibit 3. Each customer was then evaluated on the five dimensions of interest (corresponding to the five factors) and a set of subjective probabilities was developed. The firm provided us estimates of benefits and losses that are actually used (albeit subjectively) currently in setting limits. The contribution margin on sales was estimated at 10% before-tax and the variable costs were estimated at 90% of sales. Most of the sales by the firm involve a credit period of 15 days only and as a matter of convenience, we ignored the present value effect without loss of any generality.

Subjective probabilities for each customer were then integrated along with the estimates for benefits and costs in formulation (1) or (2) to determine credit limits. Similarly, objective probabilities using MDA, logit, RPA, and GP were also integrated. The results are presented in Exhibit 4. The results reveal that the AHP limits were the closest to actual limits, which is not totally surprising. Consistency, for the purposes of comparison, was defined to refer to cases where the model suggested limits and actual limits were within 10% of each other. Further, since the operational decision is to grant the credit requested or less, positive NPV situations are considered consistent if the actual limit granted is equal to the full amount of credit requested.[15] Some of the cases (=6) in which

[15] If a customer requests for a credit of $100,000, then the operative decision for the firm is to determine if the customer is worthy of being granted full credit. In such cases, it is possible that the tradeoff function specified in formulation may imply a limit in excess of $100,000. However, this is

EXHIBIT 4
CREDIT LIMIT COMPARISON RESULTS

	Consistent	Exceeds Actual in HR cases	Less Than Actual in Non-HR Cases
MDA (Linear)	84.65%	5.12%	10.23%
MDA (Quadratić)	87.44	3.72	8.84
Logit	88.37	2.79	8.84
LPI	85.12	4.65	10.23
LPII	86.05	4.19	9.77
RPA	94.42	1.86	3.74
AHP	96.28	0.00	3.72

suggested limits are less than actual are the exceptions where extraneous factors have had an influence on the decision. Elimination of these cases will further improve the overall consistency of the results from all the models.

The above results provide some insight on the ability of statistical and judgemental models to replicate expert decisions. As stated earlier, we have not addressed the broader issue of which factors are relevant for corporate credit granting. In our opinion, this issue motivates a complementary relationship between judgemental and statistical models. Statistical models like the RPA can perhaps be used with significant success to isolate from a universe of potential factors those few variables that have a significant bearing on the default risk of the customer. This information can then be used in a judgemental model, like the AHP, where objective measurements are considered along with subjective measurements that are difficult to quantify, to provide an optimal setting for corporate credit granting decisions.

V. Summary

The corporate credit granting process has not received adequate attention in the literature, perhaps because of the lack of publicly available data. We provide comparative analysis of the ability of statistical classification models to replicate the decisions of a corporate credit granting expert. It is shown that sequential partitioning procedures provide slightly superior classification results. The superiority of sequential partitioning procedures is likely to be a function of the complexity of the multivariate data being analyzed. A complementary role is suggested for statistical classification models in the corporate credit granting process. Such models can serve the useful purpose of reducing and/or identifying

not to be treated as an inconsistent case. Formulation (2) only specifies the maximum credit limit to grant.

a feasible set of objective information that the credit manager can analyze. Future research is being directed toward examining more explicitly the role of sequential and simultaneous partitioning methodologies in understanding the underlying economics of the credit granting process.

REFERENCES

1. E. I. Altman, R. B. Avery, R. A. Eisenbeis and J. F. Sinkey, Jr., *Application of Classification Techniques in Business, Banking and Finance*, CT: JAI Press, Inc., 1981.
2. H. Bierman and W. Hauseman. "The Credit Granting Decision," *Management Science* (April 1980), 519–532.
3. L. Breiman, J. H. Freidman, R. A. Olshen and C. J. Stone, *Classification and Regression Trees*, Wadsworth: California, 1984.
4. S. Chatterjee and S. Chatterjee. "Estimation of Misclassification Probabilities by Bootstrap Methods," *Communications in Statistics*, (1983), 645–656.
5. S. Chatterjee and V. Srinivasan, "Graphical Analysis and Financial Classification: A Case Study," Working Paper, Northeastern University, Revised, February 1987.
6. W. J. Conover and R. J. Iman. "The Risk Transformation As A Method of Discrimination with some Examples," *Communications in Statistics Theory and Methods*, 9, (1980), 465–487.
7. T. G. Diettrich and R. S. Michalski, "Learning and Generalization of Characteristic Descriptions: Evaluation Criteria and Comparative Review of Selected Methods," *Sixth International Joint Conference on Artificial Intelligence*, IJCAI, Tokyo, Japan, (1979), 223–231.
8. B. Efron, "Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation," *Journal of the American Statistical Association*, (June 1983), 316–331.
9. R. A. Eisenbeis. "Selection and Disclosure of Reasons for Adverse Action in Credit Granting Systems," *Federal Reserve Bulletin*, September, (1980), 727–736.
10. G. Emery, "Positive Theories of Trade Credit," In Y. H. Kim (Ed.), *Advances in Working Capital Management*, Vol. 1, (1987), forthcoming.
11. N. Freed and F. Glover, "Resolving Certain Difficulties and Improving the Classification Power of LP Discriminant Analysis Formulations," *Decision Sciences*, (Fall (1986), 589–595.
12. N. Freed and F. Glover, "Simple But Powerful Goal Programming Models for Discriminant Problems," *European Journal of Operations Research*, 7, (1981), 44–60.
13. H. Frydman, E. I. Altman and D. Kao, "Introducing Recursive Partitioning for Financial Classification: The Case of Financial Distress," *Journal of Finance*, XL, (1984), 269–291.
14. M. L. Marais, J. M. Patell and M. A. Wolfson, "The Experimental Design of Classification Tests: The Case of Commercial Bank Loan Classification," *Journal of Accounting Research*, 22, (1985), Supplement.
15. E. P. Markowski and C. A. Markowski, "Some Difficulties and Improvements in Applying Linear Programming Formulations to the Discriminant Problem," *Decision Sciences*, 16, (1985), 237–247.
16. D. Mehta. "The Formulation of Credit Policy Models," *Management Science*, (October 1968), B30–B50.
17. J. R. Quinlan, "Learning Efficient Classification Procedures and Their Application to Chess End Games," In R. S. Michalski, J. F. Carbonell and T. M. Mitchell (Eds.), *Machine Learning*, Palo Alto, CA: Tioga Publishing, 1983.
18. T. L. Saaty, *The Analytical Hierarchy Process*, New York: McGraw-Hill, 1980.
19. T. L. Saaty, "The Axiomatic Foundation of the Analytic Hierarchy Process," *Management Science*, (July 1986), 841–855.
20. T. L. Saaty and L. G. Vargas, *The Logic of Priorities: Applications in Business, Energy, Health and Transportation*, Boston: Kluwer-Nijhoff, 1982.
21. T. L. Saaty and R. S. Mariano, "Rationing Energy to Industries: Priorities and Input-Output Dependence," *Energy Systems and Policy*, (Winter 1979).
22. V. Srinivasan and Y. H. Kim, "Financial Applications of the Analytic Hierarchy Process," *TIMS International Conference*, Australia, (July 1986).
23. V. Srinivasan and Y. H. Kim, "Modeling the Credit Granting Process in a Fortune 500 Corpo-

ration," *Symposium on Cash, Treasury and Working Capital Management,* Montreal, Canada, (June 1986).
24. V. Srinivasan and Y. H. Kim, "The Credit Granting Decision: A Note," *Management Science,* forthcoming.
25. V. Srinivasan, P. V. Medury, Y. H. Kim and A. F. Thompson, "Predicting Bond Ratings: A Comparative Analysis of Alternative Classification Procedures," *Financial Management Association Meetings,* New York, (October 1986).
26. *Statistical Analysis System,* Version 5 Edition, SAS Institute Inc., Cary: NC, 1985.
27. Y. Wind and T. L. Saaty, "Marketing Applications of the Analytic Hierarchy Process," *Management Science,* (July 1980), 641–658.
28. F. Zahedi, "The Analytic Hierarchy Process—A Survey of the Method and its Applications," *Interfaces,* (July–August 1986), 96–108.

DISCUSSION

ROBERT A. EISENBEIS*: This tutorial paper examines the performance of six different classification techniques (linear and quadratic discriminant analysis, unordered logit analysis, goal programing, a recursive partitioning algorithm, and an analytic hierarchy process) as a tools in credit granting problems. The procedures are compared using data on loans extended by a nonfinancial corporation.

The first main section sets forth the authors' view of the credit granting process and the problem of setting credit limits. Following others, Kim and Srinivasan indicate that in extending credit, the objective should be to maximize the expected payoff from making a loan. That payoff depends not only on the returns from the loan under consideration but also all future credit extensions as well.[1] Formulated in this way, credit extension is a dynamic programing problem. However, as the authors point out it is essential intractable.[2] To circumvent this dead end, they collapse the problem, as others have done before them, into a single period problem where the period is the maturity of the loan, and the net future benefits from future transactions are ignored.[3] Given the purpose of the paper, the authors should have started with the single period model which would have simplified their discussion considerably. Their approach is further muddled by not providing a clear link between the structure of the single period credit granting problem presented in equation 2 and the problem of setting of credit limits described in equation 3.

Setting credit limits is irrelevant for certain types of loans but may be very important in providing trade credit (or in issuing credit cards to individuals or

* University of North Carolina at Chapel Hill

[1] See, for example, Mehta (1968, 1970), Bierman and Hausman (1970) and Dirickx and Wakeman (1976).

[2] Here the authors employ cumbersome and confusing notation, and the model they specify does not fully capture the payment performance over the life of the loan as compared with the value of future credit extensions.

[3] Even this formulation, however, does not recognize that different types of credit with different payment patterns may be requested (eg. single payment v revolving lines, personal v corporate, mortgage v credit card, short term working capital v long term borrowings, etc.), and that the expected probabilities of repayment may be functions of different characteristics of the firm for different types of credit.