THE JOURNAL OF FINANCE • VOL. XL, NO. 1 • MARCH 1985

Introducing Recursive Partitioning for Financial Classification: The Case of Financial Distress

HALINA FRYDMAN, EDWARD I. ALTMAN, and DUEN-LI KAO*

ABSTRACT

The purpose of this study is to present a new classification procedure, Recursive Partitioning Algorithm (RPA), for financial analysis and to compare it with discriminant analysis within the context of firm financial distress. RPA is a computerized, nonparametric technique based on pattern recognition which has attributes of both the classical univariate classification approach and multivariate procedures. RPA is found to outperform discriminant analysis in most original sample and holdout comparisons. We also observe that additional information can be derived by assessing both RPA and discriminant analysis results.

IN THE LAST DECADE and a half, numerous statistical classification models have been constructed for finance-related purposes. These models typically link a set of "independent" variables to a "dependent" variable; the latter can take on two or more discrete values. Classification problems in finance are usually based on two groups whereby observations are assigned to one of the groups after data analysis (for example, commercial loans classified into default vs. nondefault groups, convertible bonds into conversion-nonconversion groups, public vs. private ownership of firms, bankrupt vs. nonbankrupt groups, etc.). Studies involved with a dependent variable which can take on more than two values, i.e., multigroup models, are less common, e.g., bond or common stock ratings. The relationship of the dependent variable can be ordered or not, although in finance there is usually some implicit, if not explicit, hierarchy of groupings, e.g., bond ratings. Even if there is an implicit ranking, the intervals between the rankings are not specified.

With respect to the explanatory variables, multivariate techniques have replaced univariate approaches as data sources, and computer programs to manipulate and examine data have become available and more powerful. Researchers have found that greater information content and accuracy can be derived from more complex model specifications. The most commonly used multivariate classification technique has been discriminant analysis with a number of other parametric techniques suggested to overcome some of the perceived shortcomings in the discriminant methodology. For example, multigroup logit models have

* Assistant Professor of Statistics, Professor of Finance, Graduate School of Business Administration, New York University, and Senior Investment Analyst, General Motors Corporation, respectively. We are grateful to Professor Jerome Friedman for making the recursive partitioning computer program available to us and for a number of helpful suggestions. We are indebted to Burton Singer for stimulating discussions about the recursive partitioning approach. We thank the Salomon Brothers Center at the Graduate School of Business Administration, New York University, for financial support. We also thank an anonymous referee for constructive comments.

been used to provide a more realistic and explicit interval measure between adjacent groups (e.g., Kaplan and Urwitz [17]), in bond rating studies. Probit models attempt to provide explicit probabilities of group membership. Two-group regression analysis can provide a more accepted method for evaluating an individual variable's relative contribution.

One thing that all of these models have in common is the parametric quality of the linkage between the explanatory variables and the groupings. While this enables explicit linkages, a host of statistical problems have been cited as rendering the results somewhat problematic. The potential problems can be categorized as (1) violations of the underlying normality and independence assumptions of the classical linear regression or discriminant approaches, (2) reduction of dimensionality issues, (3) interpretation of the relative importance of individual variables, (4) specification of the appropriate classification algorithm, and (5) time-series prediction test interpretation.[1] Several of these issues will be explored later in Section II.

Despite these problems and the perceived lack of a formal theoretical underpinning of most of the models, multivariate procedures have flourished in a number of areas with increased acceptance by both academics and financial practitioners. Perhaps the best illustration of this flourishing, distinguished by a number of attempts over time to improve classification, is the area of financial distress prediction. The emphasis put on models of bankruptcy classification and prediction is partly the result of the dramatic increase in business failures all over the world but also of the relative success that these models have enjoyed since the original multivariate *Z*-Score approach (Altman [1]) which followed the more traditional univariate method of Beaver [6]. The specter of statistical validity questions as well as attempts to improve classification and prediction accuracy have motivated a steady stream of "new models." Scott [21] reviews many of these empirical models and finds that some of the more well-known ones conform to his notion of a theoretical distress scenario.

The purpose of this paper is to present a new classification procedure called Recursive Partitioning Algorithm (RPA) for financial analysis and to compare it to discriminant analysis (DA). The essence of RPA was originally presented by Friedman [13]. A more general treatment of RPA was given in Breiman and Stone [8] and its statistical properties were discussed in Gordon and Olshen [15]. An excellent and comprehensive exposition of RPA methodology can be found in the recent book by Breiman et al. [9]. RPA is a computerized, nonparametric classification technique, based on pattern recognition. It has attributes of both the classical univariate approach to classification and multivariate procedures. The model which results from RPA is in the form of a binary classification tree which assigns objects into selected *a priori* groups.

RPA has been applied in the area of medical decision making (Goldman et al. [14], Levy et al. [18]), in mass spectra classification (Breiman [7]), and in a recent paper by Marais et al. [19] to commercial loan classification for public and privately held firms. The latter compare RPA results with a polytomous probit function in order to replicate a commercial bank loan review department's

[1] See Altman et al. [3] for a detailed discussion of these points.

classifications into four problem loan rating classes. Some of the interesting features of this study are the use of both financial statement ratios and nonratio indicators, public and private firm data comparison, an examination of the results using uniform vs. bank-specific loss functions, and use of a bootstrap procedure to estimate the expected cost of misclassification.[2]

Our vehicle for illustrating RPA is the classification of financial distress of firms. We will show that, while both RPA and the classical discriminant analysis techniques lead to rather accurate classification results on a data set of bankrupt and nonbankrupt firms, the RPA usually dominates the DA. The magnitude of dominance, which depends on the specification of costs of misclassification and prior probabilities, varies from slight to large. At the same time, RPA's simple, unambiguous binary classification scheme does not have a precise scoring system that is associated with discriminant analysis. It should be made clear at the outset that the purpose of our paper is not to produce the most efficient financial distress prediction model to date. We are more concerned with the qualities of an alternative classification technique and its potential for future applications in finance.

This paper concentrates mainly on RPA and its unique qualities. Although a number of techniques have been put forth to tackle the bankruptcy issue, the standard for comparison is still the DA structure. Since the discriminant analysis format is well known, little time will be spent on its description except when we contrast it with RPA. Sections I and II present RPA and these contrasts. Sections III and IV present the sample properties and empirical results of our comparison tests. Section V discusses the implications of these results, including an analysis of the information benefits derived from utilizing both RPA and DA models for evaluating the performance of firms. Section VI concludes with a summary.

I. The Recursive Partitioning Algorithm

The purpose of this section is to provide a brief but self-contained description of the essential features of RPA. To facilitate the exposition we restrict the analysis to a two-group case and use classification of financial distress of firms as an illustrative context. We find it instructive to describe first the nature of the RPA classification model and next the steps in the construction of the model.

A. The RPA Classification Tree

The inputs to RPA include an original sample consisting of observed data of N objects, together with their actual group classification as well as specification of prior probabilities and costs of misclassifications. We will denote the prior probability of the object belonging to group i by π_i and the cost of misclassifying a group i object to group j by c_{ij}.

The model is in the form of a binary classification tree. As an illustration, in Figure 1 we present an actual tree, constructed by RPA from financial data of

[2] We were made aware of this working paper by the *Journal*'s referee and, to our knowledge, our study and the Marais et al. paper [19] are the only applications of the RPA methodology in finance.

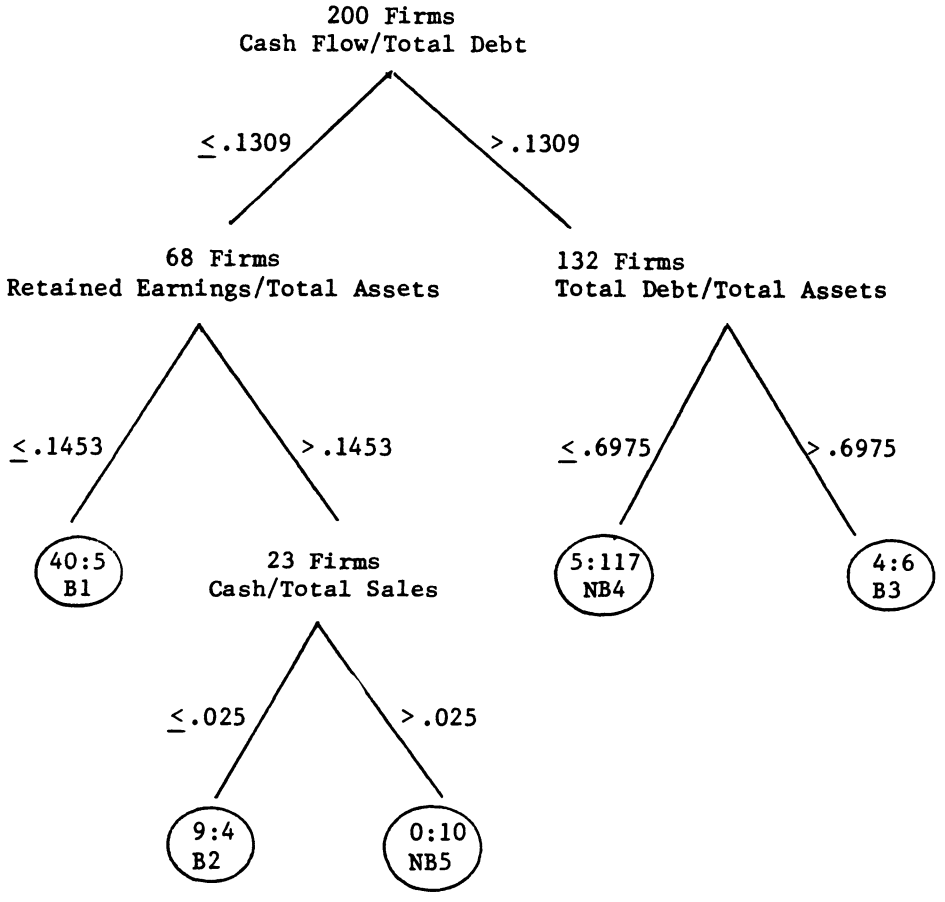

Figure 1. Tree (1). Classification tree based on financial data on 200 firms, prior probabilites of bankrupt and nonbankrupt groups (π_1, π_2) = (0.02, 0.98), and misclassification costs c_{12} = 50, c_{21} = 1. The firms with the value of the partitioning variable higher than the cutoff value go right. The terminal nodes are circled. The leftmost terminal node has 45 firms of which 40 are group 1 and five are group 2 firms. This is a group 1 (bankrupt) node which is denoted by letter B in the circle. The group 2 terminal nodes are denoted by letters NB. The numbers following the letters B and NB can be ignored in this section. This tree misclassifies five bankrupt and 15 nonbankrupt firms. Its resubstitution risk $R(T)$ = 0.19.

200 bankrupt (group 1) and nonbankrupt (group 2) firms, based on the stated prior probabilities and misclassification costs. This tree has five terminal nodes, which are the circled nodes in Figure 1. These represent the final classification of all firms. The 200 firms are distributed among terminal nodes according to their financial characteristics, e.g., the firms with Cash Flow/Total Debt ≤ .1309 and Retained Earnings/Total Assets ≤ .1453 fall into the leftmost terminal node.

In essence, the RPA model partitions variable space into several rectangular regions defined by the characteristics of terminal nodes. All objects falling into a given terminal node, that is, a given region of variable space, are assigned to the same group, e.g., the leftmost node of Tree (1) in Figure 1 is a group 1, distressed firm classification, node. A new object to be classified descends down

the tree and is assigned to the group identified with the terminal node into which it falls.

B. Resubstitution Risk of RPA Classification Tree

The terminal nodes of a classification tree are assigned to groups in such a way as to minimize the observed expected cost of misclassification of each assignment. Another term for the observed expected cost of misclassification, which we will use in what follows, is resubstitution risk.[3] Consider a terminal node t which has $n_i(t)$ objects from group i and let N_i be the size of the entire original sample for the i^{th} group, $i = 1, 2$. The risk of assigning node t to group 1 is defined as $R_1(t) = c_{21}p(2, t) = c_{21}\pi_2 p(t \mid 2) = c_{21}\pi_2 n_2(t)/N_2$. Here $p(2, t) =$ probability that an object is from group 2 and falls into node t and $p(t \mid 2) = n_2(t)/N_2 =$ conditional probability of a group 2 object falling into node t. Similarly, $R_2(t) = c_{12}\pi_1 n_1(t)/N_1$.

The terminal node t is assigned to a group corresponding to the minimum risk; thus the assignment rule is a Bayesian rule. The resulting Bayes risk of node t is $R(t) = \min(R_1(t), R_2(t))$. The risk of the entire tree T, denoted $R(T)$, is the sum of risks of its terminal nodes. Note that if $c_{12} = c_{21} = 1$ and $\pi_i = N_i/N$, $i = 1, 2$, i.e., the prior probabilities are the original sample proportions of group 1 and group 2 objects, then $R_i(t) = n_i(t)/N$ is a sample proportion of group i objects falling into node t. In this case, the assignment rule is particularly simple: each terminal node is assigned to a group which has majority representation in this node, and the risk of the tree is simply the overall misclassification rate of the tree.

C. RPA Model Construction

Step 1: *Constructing small resubstitution risk trees.* The objective in Step 1 is to construct a classification tree with a small resubstitution risk. The following procedure for tree construction used in RPA usually achieves this objective. First, the original sample, thought of as located at the root (top) of the tree, is split into two subsamples according to the "best" splitting rule defined below. The two subsamples "land" in the left and right subnodes of the root node. Next, the process is repeated for each subsample, i.e., each subsample is split further according to the best splitting rule for the subsample. The process continues until the termination criterion, described below, is satisfied.

The best splitting rule for a subsample is selected from the class of univariate splitting rules, i.e., rules which involve splitting an axis of one variable at one point.[4] The criterion for selecting the best splitting rule is based on the measure of sample impurity. The general measure of sample impurity used in RPA is described in Appendix A.

[3] Risk is another term for expected cost of misclassification, and "resubstitution" refers to the fact that risk is evaluated with the original sample's data. When the term risk is used alone it should be understood as resubstitution risk. Similarly, when the term probability is used it should be understood as the probability estimated from the original sample's data.

[4] For this study we restrict the class of splitting rules to univariate rules. It is possible to utilize RPA with the class of rules which involve linear combinations of variables, but the resulting model can be extremely cumbersome and difficult to interpret.

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 (see Appendix A). 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.[5]

The impurity measure has a simple definition when misclassification costs are equal to unit costs and prior probabilities are given by sample proportions of group 1 and group 2 objects. Suppose that group 1 objects in the sample are assigned numerical value 1 and group 2 objects the value 0. The measure of impurity is then given by the sample variance of the numerical values. The sample impurity is equal to zero if and only if all objects in the sample belong to the same group and attains its maximum value when the sample contains the same number of group 1 and group 2 objects.

The splitting procedure terminates when it is impossible, by further splitting, to decrease the impurity of a current tree (defined as the sum of the impurities of all the terminal nodes). The obtained tree will be referred to as $T_{\max}$.

Step 2: Selecting the correct complexity for a tree by cross-validation tests. $T_{\max}$ trees are usually complex trees which overfit the data. That is, their resubstitution risks usually underestimate, often by much, their true risks. To put it differently, $T_{\max}$ trees involve splits specific to the original sample characteristics but not necessarily representative of the population structure. The objective of the second step is to select a tree of correct complexity, from the subtrees of the $T_{\max}$ tree, according to the following procedure. Consider all possible trees obtained by arbitrarily terminating the splitting procedure described in Step 1. For each tree T in this set compute,

$$R(T) + K \times \text{Number of terminal nodes of } T \tag{1}$$

where K is a nonnegative constant interpreted as a penalty for complex trees. For a given K, choose tree T which minimizes the function in (1). Clearly, when $K = 0$, the optimal tree, i.e., the one which minimizes (1), is $T_{\max}$. For any positive value of K, the optimal tree is a subtree of $T_{\max}$. As K increases, the optimal tree becomes less complex but has a larger resubstitution risk. For a sufficiently large value of K, the optimal tree is a no-split tree. For example, Tree (1), in Figure 1, is optimal for small values of penalty K. When K is increased, the split on Total Debt/Total Assets is eliminated. In this case, there is no value of the penalty parameter for which the tree with two splits is optimal; that is, further increase in K results in a no-split tree.

Thus, criterion (1) gives rise to a nested sequence of classification trees of decreasing complexity and increasing resubstitution risk. The final classification tree is selected from this sequence as the one with the smallest cross-validated risk. The cross-validated risk of tree T is computed using a V-fold cross-validation

[5] For each ordered variable X, the original sample contains at most N distinct values $X_1 < X_2 < X_3 < \cdots X_N$. The candidate splits are of the form "Is $X \leq c$?", where c is the midpoint between consecutive distinct values of variable X. Thus, for each ordered variable there are at most $N - 1$ candidate splits. For a categorical variable taking on L distinct values, there are 2^{L-1} candidate splits.

procedure. In this procedure, all observations 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 K 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 cross-validated error rates and risk of tree T are obtained by averaging the resubstitution error rates and risks from all cross-validation trials. Typically, the minimum cross-validated risk tree is less complex than $T_{\max}$.

In addition to the V-fold procedure, other criteria can be used to choose the final tree. These include bootstrap analysis, expert judgment concerning the appropriate complexity of the model, holdout samples, etc. Our rationale for adopting the minimum V-fold cross-validation approach is that it yields reliable estimates of the true risk of classification trees when the sample size is large, as it is in our study.[6]

II. Comparison of the Recursive Partitioning Algorithm and Discriminant Analysis

RPA and DA are both Bayesian procedures with their classification rules derived in such a way as to minimize the expected cost of misclassification. To be more specific, recall that the general Bayes rule in the two-group case is of the form:

Assign an observation with a vector of attribute variables x to group 1 if

$$f_1(x)/f_2(x) \geq \pi_2 c_{21}/\pi_1 c_{12}$$

otherwise assign it to group 2. Here f_1 and f_2 represent the multivariate probability density functions of variables for groups 1 and 2.

DA makes the Bayes rule operational by assuming that densities f_1 and f_2 are multivariate normal. If we also assume that covariance matrices of the two densities are equal, Bayes rule reduces to the:

Linear discriminant rule: Assign observation x to group 1 if

$$x'\gamma - \alpha \geq \ln[c_{21}\pi_2/c_{12}\pi_1]$$

[6] Simulation studies have shown (Breiman et al. [9]) that estimates resulting from V-fold cross-validation, with $V = 5$ or 10, are satisfactorily close to the true risk of classification trees. V-fold cross-validation is preferred to the holdout sample method of estimating risk if the sample size is not large. It is also preferred to the bootstrap method of estimating true risk (see footnote 11 for a brief description of the bootstrap method) because theoretical considerations and simulation studies show that bootstrap estimates of the true risks of classification trees are substantially downwardly biased for some data structures. This downward bias is especially serious for complex trees. Cross-validated estimates are only slightly biased but have larger variances than bootstrap estimates. Thus, unless variances of cross-validated estimates are unacceptably large, which is likely to happen when the sample size is small, these estimates are preferred to bootstrap estimates and are used in Breiman et al. [9] in all of the empirical examples. But Marais et al. [19] use the bootstrap procedure to estimate true risks of their models because the effective size of their samples is very small (less than 10). For a recent analysis of these issues in the context of general classification techniques and an introduction to modified bootstrap procedures, see Efron [11].

otherwise assign it to group 2, where $\gamma = \Sigma^{-1}(\pi_1 - \pi_2)$, $\alpha = (\mu_1 + \mu_2)'$ $\gamma/2$, μ_1, μ_2 are group means, and Σ is a common covariance matrix.

RPA is a nonparametric technique which minimizes the expected cost of misclassification by the univariate splitting procedure described in Section I.

One of the primary differences in the two classification techniques is the manner in which they partition the variable space into classification regions. In order to illustrate this difference, suppose that variable space is two-dimensional and consider a hypothetical RPA model (see Figure 2a) constructed from a sample consisting of objects from two groups. This model partitions the variable space (A, B) into four rectangular regions as shown in Figure 2b. Objects falling into regions 1 and 2 are classified as group 1 and those falling into regions 3 and 4 as group 2. Now consider a hypothetical discriminant function $f(x) = x'\gamma - \alpha$, where $x = (A, B)$, constructed from the same data, and a cutoff point c. The line $f(x) = c$ divides variable space into two half-plane regions with the region $\{x: f(x) \geq c\}$ assigned to group 2 and the region $\{x: f(x) < c\}$ to group 1 (see Figure 2b). The essence of the space partitioning by the two models for higher dimensional problems is the same.

The RPA classification rule partitions variable space, in general, into a number of rectangular regions. The two-group DA classification rule, on the other hand, partitions the variable space into only two half-plane regions. Again referring to Figure 2b, we can observe that differences in the final classification of RPA and DA are denoted by observations falling into the shaded areas.

Another important difference in the two classification techniques is the manner in which prior probabilities and costs of misclassification impact the resulting model. DA models are established first by group separation criteria, i.e., by maximizing the between-group to within-group variance, and only then is the classification rule established for assigning observations into the specified groups based on specified error costs and prior probabilities. RPA, on the other hand, simultaneously determines the variable selection and group assignments with the costs and priors helping to determine the specific variables and splitting score. Changing the costs and priors might very well change the variable selected for splitting; this is not so for DA. Hence, RPA models appear to be more sensitive to costs and priors than DA models.

We have also noted that RPA is a nonparametric technique which eliminates many of the statistical problems attributed to DA (see Eisenbeis [12]). Three key assumptions of the linear DA model are that (1) variables describing the members of the group observations are multivariate normally distributed within each group, (2) group covariances are equal across all groups, and (3) groups are discrete, nonoverlapping, and identifiable. Only the latter assumption is also necessary for RPA.

In fact, deviations from the normality assumption in studies dealing with financial and economic variables are usually the rule. For example, many common ratios are lower bounded by zero but possess no theoretical upper bound and can get quite large, e.g., Sales/Total Assets. Nonnormal distributions can bias the classification results by impacting the assignment rule away from the optimal value. While results may be suboptimal, accuracies still can be acceptable despite the lack of normality (see Altman et al. [5]). RPA is not limited by any

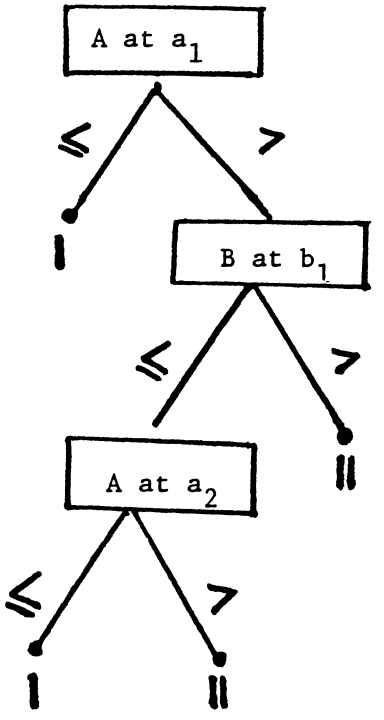

Figure 2a. "Hypothetical" Tree for Two-Group Classification

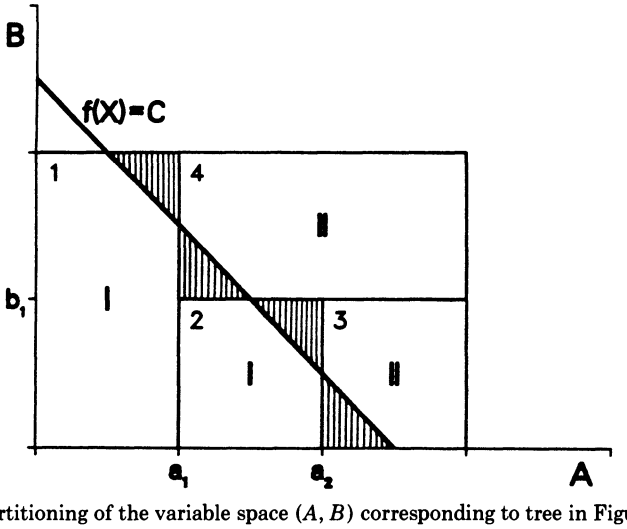

Figure 2b. Partitioning of the variable space (A, B) corresponding to tree in Figure 2a and to the discriminant function constructed from the same hypothetical data. The numbering of the regions is indicated in the upper left corner of the region.

distributional assumptions and in this respect represents an attractive alternative to DA. RPA is also particularly suited for analyzing discrete variables.

The assumption of equal group covariance matrices is also often violated. Relaxation of this assumption in the linear case affects the significance tests for

differences in group means and also the appropriate form of classification rules. Tests of covariance are now quite readily available. When covariances are found not to be equal, the theoretical prescription is to apply a quadratic structure which does not rely on the equal covariances between group assumption. Quadratic models, however, while quite accurate on original samples, are uniformly disappointing vis-à-vis linear models in holdout sample tests (see Altman et al. [5] and Martin [20]). Tests have also shown that violations of the equality of group covariances are a function of the number of variables and the relative sizes of the groups. Smaller number of variables and equal group sizes seem to lead to less biased results. Again, RPA is completely unaffected by group covariance attributes since the splitting point is not influenced by outliers.

With respect to the interpretation of the relative importance of individual variables, both the DA and RPA models present some unique aspects which could lead to interpretation problems. It is fair to say that there is a distinct lack of agreement as to which test is most appropriate in determining the relative importance of discriminant variables. Indeed, we are aware of as many as six different tests as discussed in Joy and Tollefson [16] and Altman and Eisenbeis [4]. While we are not concerned with the individual variable significance compared to the overall profile of variables, it is instructive to assess the univariate contribution.

RPA has a similar quality that is found in most forward stepwise approaches to variable selection. Once a variable is selected as the first splitting variable, the tree is constrained to include that measure first. But, as illustrated in Figure 2a, in RPA the same variable may be used more than once as the partitioning variable. Similar constraints are manifest as additional variables are chosen. Since the RPA is not looking forward beyond the splitting it is currently performing, the resulting model is not an optimal one. Therefore, in both DA and RPA models, individual variable contributions are not completely unambiguous.

In Section V we will compare a different aspect of the two techniques, namely, their scoring systems.

III. Sample Characteristics and Choice of Variables

Our sample has 58 bankrupt industrial companies which failed during 1971–1981.[7] We randomly selected 142 nonbankrupt manufacturing and retailing companies from the COMPUSTAT universe. The financial years investigated were also randomly chosen from the period 1971–1981 and not matched to the exact years of the bankrupt group. This 200-firm sample is available from the authors.

We used 20 financial variables which had been found significant in predicting business failure by three relevant studies (Altman [1], Deakin [10], and Altman et al. [5]) as our variable set. An adjustment for the capitalization of leases was

[7] The bankrupt group includes 16 companies that filed for bankruptcy under Chapter 11 but still were listed on the regular COMPUSTAT tape. The *Wall Street Journal Index* was used to find the date that the companies filed for bankruptcy.

applied to all the variables. Firms have been required to capitalize financial leases since 1980, so models which purport to have applicability in the post-1980 period should adjust past data based on the latest reporting standards (see Altman et al. [5]). The variable set is described in Appendix B.

IV. Development of Models and Classification Results

A. Prior Probabilities and Misclassification Costs

In this section we empirically compare recursive partitioning with discriminant analysis models. The comparison is based on a sample of 200 firms and carried out for misclassification cost c_{12} ranging from 1 to 70, in the way indicated in Table Ia, while $c_{21} = 1$. The prior probabilities of the bankrupt and nonbankrupt groups were fixed at $(\pi_1, \pi_2) = (0.02, 0.98)$.

B. Recursive Partitioning Models (RPA)

For each cost specified in Table Ia, we chose two classification trees, from the sequence of trees resulting from RPA, as RPA models for comparison with DA models. For the first RPA model (RPA1), we chose a relatively complex tree. For the second RPA model (RPA2), we chose the tree with the smallest fivefold cross-validated risk. The smallest cross-validated risk tree was always less complex than the first model tree and with one exception it never had more than three splits.

We next illustrate the sensitivity of the obtained RPA models to change in misclassification costs. The classification trees in Figures 1 and 3 are based on $c_{12} = 50$ and 20, respectively. These trees were chosen from the more complex RPA models. We observe that the first split in both trees involves the variable Cash Flow/Total Debt, but the actual splitting point on the axis of this variable is different for the two trees. In fact, the inspection of all the models, including those not presented here, shows that Cash Flow/Total Debt is the most important discriminator between bankrupt and nonbankrupt groups for the cost c_{12} ranging from 10 to 70.[8] While the trees are stable in terms of the first split, further splits result in somewhat different variables, number of variables, sequencing, and split points.

We next observe, for costs 20 and 50, the trees with the smallest cross-validated risk. For $c_{12} = 20$, the smallest cross-validated risk tree is obtained from Tree (2), shown in Figure 3, by eliminating splits on variables Total Debt/Total Assets and Log (Interest Coverage + 15). For $c_{12} = 50$ it is a one-split subtree of Tree (1).

Table Ia illustrates the dependence of classification results on the cost structure. Typically, specification of costs would depend on the context in which bankruptcy prediction is made and may involve subjective judgment.[9] Table Ia

[8] Cash Flow/Total Debt was also Beaver's [6] best univariate measure. For equal misclassification costs ($c_{12} = 1$), however, the first split involves the variable Cash/Total Sales.

[9] Many studies have implicitly assumed that the costs of misclassifications are equal. However, Altman [2] estimated that, for commercial bank lending officers, classification of a bankrupt firm in the nonbankrupt group is 32 to 62 times more costly than the reverse misclassification.

Table Ia

Misclassification of Firms in the Original Sample by the Models for Different Misclassification Costs*

	c_{12}																	
	1			10			20			30			40			50			60			70		
Model	I	II	T	I	II	T	I	II	T	I	II	T	I	II	T	I	II	T	I	II	T	I	II	T
RPA1	18	0	(18)	14	3	(17)	9	2	(11)	6	11	(17)	6	11	(17)	5	15	(20)	5	12	(17)	2	18	(20)
RPA2	58	0	(58)	19	3	(22)	14	2	(16)	9	19	(28)	9	19	(28)	9	19	(28)	9	9	(18)	4	25	(29)
DA1	48	0	(48)	25	5	(30)	17	11	(28)	13	14	(27)	11	15	(26)	9	17	(26)	8	18	(26)	7	22	(29)
DA2	54	1	(55)	33	6	(39)	18	12	(30)	12	16	(28)	11	21	(32)	8	27	(35)	6	29	(35)	6	30	(36)

* For each model and cost, we list the number of type I misclassifications (a bankrupt firm classified as nonbankrupt), type II misclassifications (a nonbankrupt firm classified as bankrupt), and a total number of misclassifications (in parentheses).

Table Ib

Resubstitution Risks of the Model*

	c_{12}							
Model	1	10	20	30	40	50	60	70
RPA1	0.006	0.069	0.076	0.138	0.159	0.190	0.186	0.173
RPA2	0.020	0.086	0.111	0.224	0.255	0.286	0.248	0.269
DA1	0.017	0.121	0.193	0.231	0.255	0.272	0.290	0.321
DA2	0.022	0.155	0.207	0.235	0.297	0.324	0.324	0.352

* Resubstitution risk = $\pi_1 c_{12} n_1/N_1 + \pi_2 c_{21} n_2/N_2$ where n_i = total number of type i misclassifications, N_i = sample size of the i^{th} group, $i = 1, 2$.

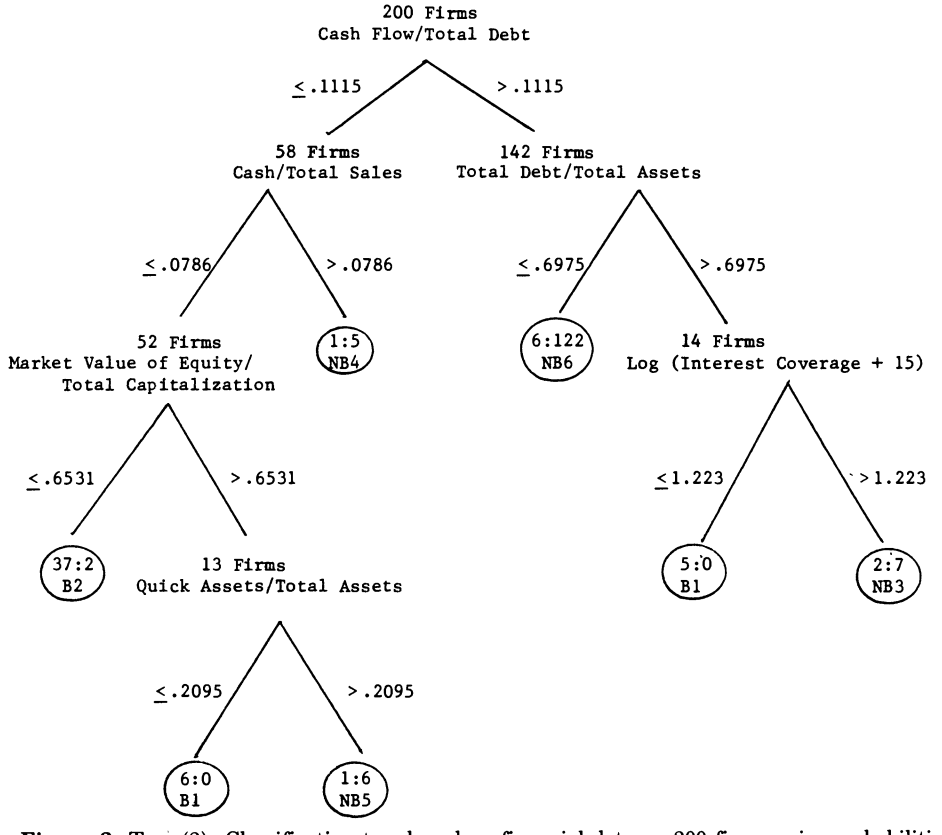

Figure 3. Tree (2). Classification tree based on financial data on 200 firms, prior probabilities $(\pi_1, \pi_2) = (0.02, 0.98)$, and misclassification costs $c_{12} = 20$, $c_{21} = 1$.

allows assessment of the consequences in terms of error rates and risks of different specifications of misclassification costs and may lead to a reevaluation of any particular specification.

C. Discriminant Analysis Models (DA)

We constructed two discriminant functions. The first (DA1) was obtained by a forward stepwise method. It included 10 of the 20 variables considered. The variables and their coefficients are presented in Table IIa. We also constructed a second discriminant function (DA2 model) using only the four most important variables according to the forward stepwise method. The coefficients of the 4-variable discriminant function are presented in Table IIb.

D. Comparing RPA and DA Models

The resubstitution classification results of the RPA1 models dominate the results of both DA1 and DA2 models for all costs. Recall from Section I that, for every cost, RPA2 tree is a subtree of RPA1 tree, hence RPA2 has a larger resubstitution risk. This can be seen clearly in Figure 4.

Table II

Discriminant Analysis Models

Variable	Unstandardized Coefficient	Standardized Coefficient
a) DA1 Model: Stepwise Forward Discriminant Function		
Net Income/Total Assets	5.452	0.527
Current Assets/Current Liabilities	1.758	1.419
Log (Total Assets)	0.505	0.348
Market Value of Equity/Total Capitalization	1.850	0.409
Current Assets/Total Assets	−6.292	−1.068
Cash Flow/Total Debt	−1.021	−0.331
Quick Assets/Total Assets	8.970	1.062
Quick Assets/Current Liabilities	−1.995	−1.112
Earnings Before Interest and Taxes/Total Assets	3.482	0.350
Log (Interest Coverage + 15)	−1.033	−0.165
Constant term = −1.761		
b) DA2 Model: 4-Variable Discriminant Function		
Net Income/Total Assets	5.322	0.514
Current Assets/Current Liabilities	0.622	0.528
Log (Total Assets)	0.712	0.491
Market Value of Equity/Total Capitalization	1.149	0.328
Constant term = −4.041		

The RPA2 model outperforms the DA2 model in terms of fewer number of misclassifications for all costs except $c_{12} = 1$ and is tied for $c_{12} = 30$. For this cost, the RPA2 model coincides with the "naive" model, which classifies all firms to the nonbankrupt group.

The comparison of RPA2 with DA1 for costs ranging from 10 to 70 shows that either RPA2 dominates DA1 substantially (by at least eight correct classifications) or DA1 slightly dominates RPA2 (by at most two correct classifications). For the cost $c_{12} = 50$, the number of type I misclassifications is the same for both models. Finally, the comparison of the results of DA1 and DA2 shows that DA1 dominates DA2 in terms of the total number of correct classifications for all costs, but for costs 30 and above, DA2 classifies correctly more bankrupt firms than DA1.

In Table Ib we compute the risks for all models and all costs. RPA1 models dominate the other models for all costs. RPA2 models are second best for all costs except costs 1 and 50; for these costs DA1 models have slightly lower risks than RPA2 models. For cost 40, the risks of RPA2 and DA1 models are identical, but note (in Table Ia) that classification results are not. In general, RPA2 and DA1 models have very similar risks in the middle range of costs, ($c_{12} = 30$ to 60) while the risk of RPA2 is smaller for the low and high costs (except for $c_{12} = 1$).

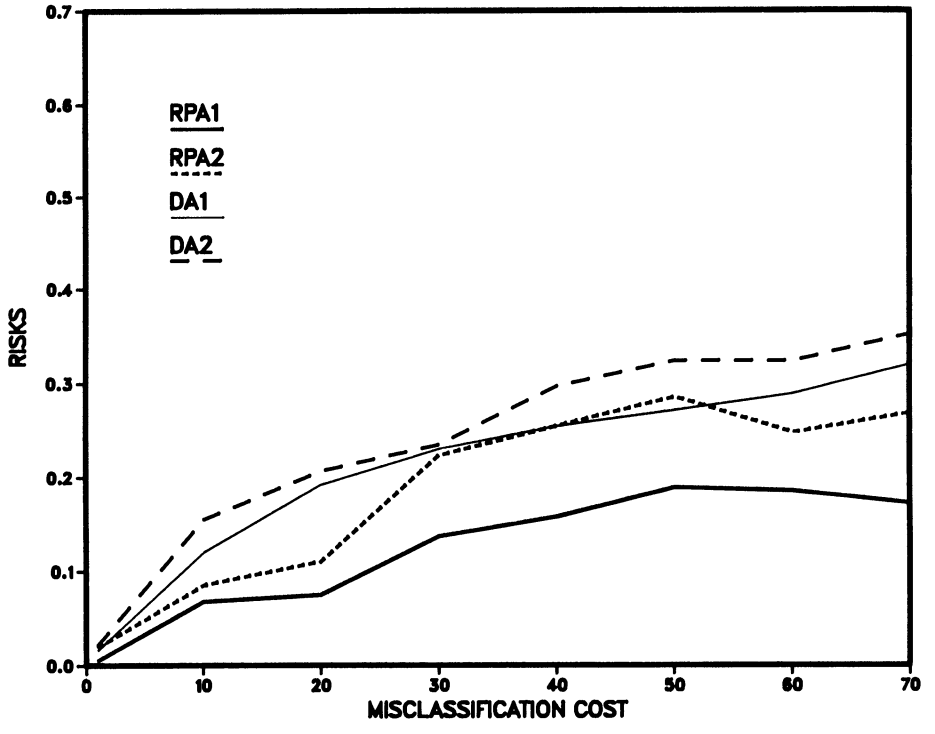

Figure 4. Resubstitution Risks

As expected, risks of DA2 models are somewhat larger than those of DA1 models.

The performance of the models can also be evaluated in the following way. For each model and cost, compute the ratio of the model's risk to the "naive" risk, i.e., the Bayes risk of assigning all objects in the original sample to one group (Table IIIa). The resulting ratio indicates what proportion of "naive" risk is saved by the construction of the model. It is interesting to observe that for all types of models, except RPA1 models, the saving in risk increases monotonically as a function of c_{12}, until $c_{12} = 50$ or 60, after which it monotonically decreases.[10]

Up to this point, we have compared the models based on their resubstitution error rates and risks. In Table IVa we report the cross-validated risks of the models based on the fivefold cross-validation procedure. The DA1 model, constructed on each cross-validation trial by the forward stepwise method, was neither constrained to include the variables chosen by the original model nor to be a 10-variable model. The set of variables was chosen based on negligible increased explanatory power beyond a derived set of variables. DA2 cross-validation models were constrained to be 4-variable models, but not constrained to include the same variables as the original 4-variable model.

We observe, in Table IVa, that cross-validated risks of RPA2 models are smaller than those of DA models for all costs except 50 for which risk of DA1 is the same as that of RPA2 and cost 70 for which the DA2 model has a smaller risk. The RPA1 models, which were by far the best in terms of resubstitution

[10] The last part of this statement is based on the results obtained for costs $c_{12} > 70$, not reported in Table IIIa.

Table IIIa

The Ratios of the Model's Risk to the Risk of the "Naive" Classification Rule*

	Model			
c_{12}	RPA1	RPA2	DA1	DA2
10	0.34	0.43	0.60	0.78
20	0.19	0.28	0.48	0.52
30	0.23	0.37	0.38	0.39
40	0.20	0.32	0.32	0.37
50	0.19	0.29	0.28	0.33
60	0.19	0.25	0.29	0.33
70	0.18	0.27	0.33	0.36

Table IIIb

The Ratios of the Model's Cross-Validated Risk to the Risk of the "Naive" Classification Rule*

	Model			
c_{12}	RPA1	RPA2	DA1	DA2
10	0.62	0.62	0.89	0.77
20	0.59	0.55	0.74	0.62
30	0.63	0.45	0.45	0.52
40	0.55	0.41	0.41	0.46
50	0.51	0.39	0.39	0.41
60	0.51	0.39	0.44	0.40
70	0.54	0.46	0.53	0.38

* The naive classification rule assigns all firms to the nonbankrupt group for cost $c_{12} < 49$, to the bankrupt group for cost $c_{12} > 49$, and is indifferent with respect to the assignment for cost $c_{12} = 49$. Thus, for costs 10 through 40 "naive" risk is computed as $\pi_1 c_{12} = 0.02c_{12}$ and for costs 50 through 70 as $\pi_2 c_{21} = 0.98c_{21} = 0.98$. We exclude the results for $c_{12} = 1$ because for this cost either the actual or the cross-validated risk of each model is lower than the risk of the "naive" classification rule.

risks, are the worst of all models in terms of cross-validated risks for all costs except 10 and 20. These relations can be seen clearly in Figure 5.

The DA2 models, which had larger resubstitution risks than DA1 models, have smaller cross-validated risks than DA1 models for low and high costs. Thus, for low and high costs, the comparison between DA1 and DA2 is somewhat analogous to that between RPA1 and RPA2. Both DA1 and RPA1 are complex models which do better in terms of resubstitution results and, in most cases, worse in terms of cross-validated results than less complex RPA2 and DA2 models.

In Table IIIb we illustrate, estimated by cross-validation, the potential of RPA and DA models to reduce the risk of the "naive" classification rule. The ratios in Table IIIb behave similarly, as a function of c_{12}, to the ratios in Table IIIa. For cost $c_{12} = 50$, the estimated saving in risk for RPA2, DA1, and DA2 models is about 60%, which is about 10% less than the saving in risk estimated by

Table IVa

Cross-Validated Risks of the Models*

	c_{12}							
Model	1	10	20	30	40	50	60	70
RPA1	0.091 (0.020)	0.124 (0.025)	0.235 (0.036)	0.376 (0.049)	0.442 (0.058)	0.497 (0.069)	0.497 (0.077)	0.531 (0.087)
RPA2	0.020 (0.000)	0.124 (0.021)	0.221 (0.035)	0.269 (0.043)	0.324 (0.053)	0.386 (0.063)	0.386 (0.069)	0.455 (0.080)
DA1	0.056 (0.022)	0.179 (0.016)	0.296 (0.019)	0.272 (0.042)	0.329 (0.036)	0.386 (0.080)	0.427 (0.079)	0.518 (0.097)
DA2	0.035 (0.016)	0.155 (0.010)	0.249 (0.021)	0.307 (0.075)	0.365 (0.036)	0.402 (0.058)	0.397 (0.090)	0.369 (0.064)

* Standard deviations of the estimated risks are in parentheses.

Table IVb

Bootstrapped Risks of the Models*

	c_{12}							
Model	1	10	20	30	40	50	60	70
RPA1	0.041 (0.003)	0.114 (0.009)	0.154 (0.016)	0.236 (0.020)	0.269 (0.031)	0.333 (0.037)	0.355 (0.035)	0.353 (0.042)
RPA2	0.023 (0.008)	0.132 (0.011)	0.190 (0.008)	0.267 (0.023)	0.328 (0.025)	0.310 (0.052)	0.374 (0.045)	0.381 (0.045)
DA1	0.037 (0.002)	0.172 (0.008)	0.222 (0.031)	0.286 (0.009)	0.318 (0.018)	0.393 (0.017)	0.458 (0.029)	0.409 (0.021)
DA2	0.020 (0.002)	0.165 (0.012)	0.207 (0.018)	0.274 (0.022)	0.323 (0.029)	0.374 (0.025)	0.376 (0.041)	0.426 (0.025)

* Standard deviations of the estimated risks are in parentheses.

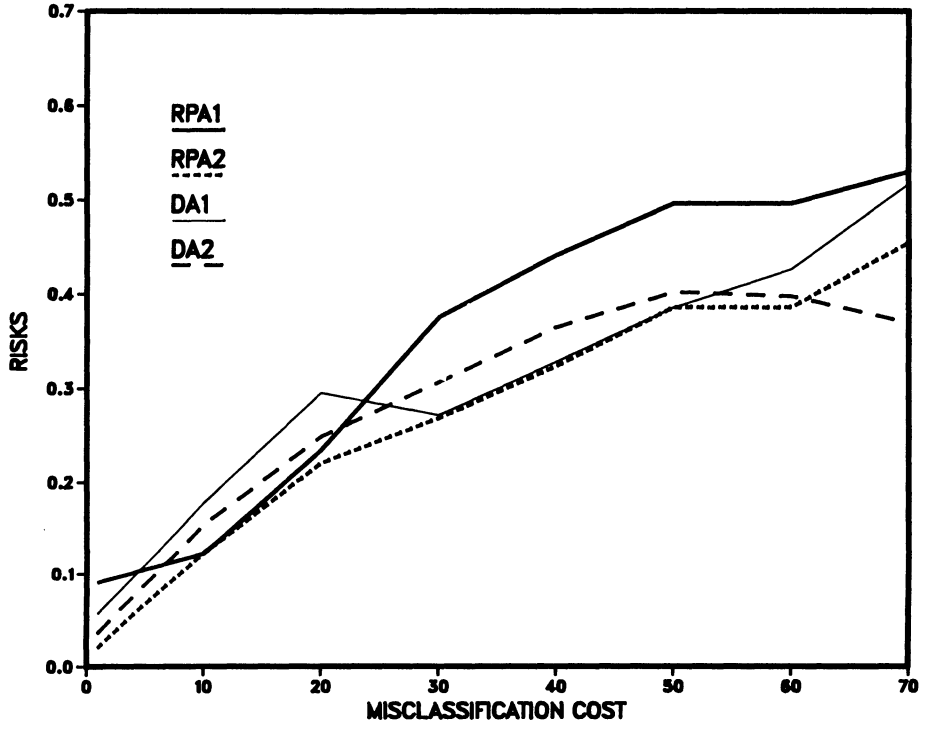

Figure 5. Cross-Validated Risks

resubstitution. For RPA1 the cross-validated saving in risk is about 50%, which is 30% less than the saving in risk estimated by resubstitution.

E. Cross-Validation and Bootstrap Procedure Comparisons

Finally, we have also estimated the risks of our models using a bootstrap procedure.[11] As discussed in Section I and footnote 6, for classification trees the V-fold cross-validation procedure is preferred to the bootstrap procedure, especially when, as is the case in our study, the effective size of the sample is relatively large. Still it is of interest to compare the two procedures' results. The estimates in Table IVb were obtained using five bootstrap samples. The bootstrap results confirm the cross-validated results in terms of the comparisons between RPA

[11] The boostrap estimate of the "true" risk of tree T corresponding to the penalty parameter K is $R(T)$ + bias correction, where the bias correction is obtained as follows: Choose a sample, with replacement, from the original sample of the same size as the original sample and refer to it as a bootstrap sample. Using the bootstrap sample, construct a tree corresponding to penalty K and call it the bootstrap tree. Then classify the bootstrap sample by the bootstrap tree and the original sample also by the bootstrap tree. Typically the risk resulting from the first classification is smaller than that of the second. The difference between these risks is one iteration on bias correction. Now obtain a new bootstrap sample, repeat the process and average the differences in risks. This average is the estimated bias correction. The number of iterations required to obtain stable estimates of the bias correction depends on the sample size and the number of variables. For a more extensive description of the bootstrap procedure in the context of risk estimation of tree classifiers, see Marais et al. [19] and for a general introduction to the bootstrap method of estimating the bias of a variety of statistics and their standard errors, see Efron [11].

and DA models. However, the striking feature of the results in Table IVb is that bootstrapped risks of RPA1 models are much lower than cross-validated risks of these models and usually lower than bootstrapped risks of RPA2 models, making them "better" models, as judged by bootstrap estimates, than RPA2 models. Most likely these results are due (see footnote 6) to the large downward bias of bootstrapped risks for complex trees. Note that, for less complex RPA2 trees, the cross-validated and bootstrapped risks are generally similar. We also observe, as expected, that cross-validated risks have somewhat higher variances than bootstrapped risks.

In summary, our empirical results show that in most cases RPA2 models do better than DA models both in terms of actual, cross-validated, and bootstrapped results. RPA1 models, which are most accurate on resubstitution, are worse than DA models on cross-validation. The noteworthy feature of RPA2 models is their lack of complexity, especially vis-à-vis a 10-variable DA model. RPA models are usually 1-, 2-, or 3-variable models, and in only one case, for cost 20, is RPA2 a 4-variable model.

V. Comparison of the Scoring Systems of RPA and DA

In this section we briefly discuss RPA's and DA's ability to assess the performance of a firm, relative to its prior condition or relative to other firms. While accuracy is the most important attribute of a classification technique, the information from a relative comparison is valuable from a practical viewpoint. The DA scoring system is well known. It is very precise in that it allows for relative ranking between firms and within a firm over time.

As discussed in Section II, the RPA models usually partition the variable space into several "bankrupt" and "nonbankrupt" regions. With each region there is associated a probability of bankruptcy given by $p(1 \mid t)$ as defined in Appendix A, i.e., the probability that a firm with the observation vector corresponding to this region will go bankrupt. We propose that this probability (or the probability of nonbankruptcy) be considered as a "score" of this region. All firms with the observation vectors in a given region have the same score. Thus, the RPA scoring system is discrete with the number of scores equal to the number of regions and therefore cannot be used to compare firms classified within the same region. In this respect, it appears to be less sensitive than DA's continuous scoring system. However, it may be argued that all that one may want is a grouping of firms into different "risk" categories, which RPA provides.

In order to illustrate RPA vs. DA scoring systems, we consider four recently bankrupt firms with the DA and RPA classifications and scores for five years prior to bankruptcy[12] (see Figure 6). To arrive at the classifications and scores, we used the RPA1 and DA1 models for cost $c_{12} = 50$ (see Section IV).[13] For this

[12] Only one of these four companies, namely White Motor Corp., was used in the construction of the DA and RPA models. The DA and RPA scores for years 1–5 prior to bankruptcy are based on the models constructed from data one year prior to failure. We prefer this methodology to the alternative of using a completely different model for each year prior to bankruptcy.

[13] For the purpose of illustration, we have chosen RPA1 and DA1 models rather than less risky (as estimated by cross-validation) RPA2 and DA2 models because the RPA2 model for cost $c_{12} = 50$ is a one-split model and as such does not provide an interesting illustration of the scoring system of recursive partitioning models.

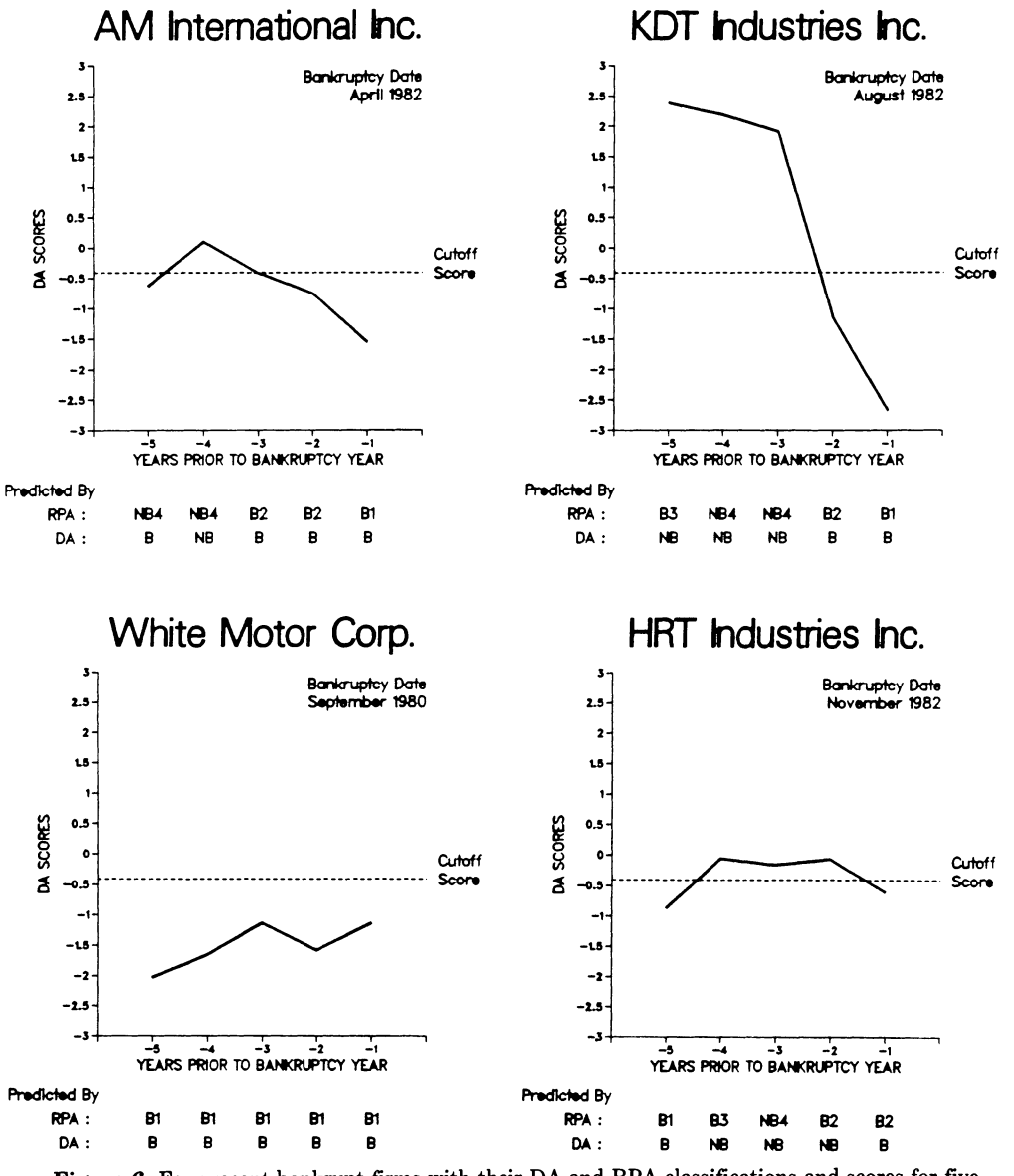

Figure 6. Four recent bankrupt firms with their DA and RPA classifications and scores for five years prior to bankruptcy. The DA cutoff score is −0.41.

cost, the RPA1 model has five terminal regions which can be ordered according to their scores. The ordering is indicated in Figure 1 by the number following "group designation." The region corresponding to B1 is the most "risky," i.e., the firms in this region have the highest probability (0.29) of going bankrupt. Region B2 is the second most "risky," with bankruptcy probability equal to 0.11, etc. The remaining regions correspond to B3, NB4, and NB5.

Inspection of the trend in "scores" in Figure 6 shows that RPA1 usually picks up the deteriorating condition of the firm by "moving" it from a less "risky" to a more "risky" category as the number of years to bankruptcy decreases. For example, we can observe that AM International was classified as bankrupt by RPA and DA for three years prior to its April 1982 bankruptcy filing date. The DA score for AM International graphically shows its deterioration. RPA classified AM International in the second risk category (B2) three and two years prior to its bankruptcy and in the most risky category (B1) one year prior to its bankruptcy, thus also showing the deterioration trend.

The relative performance of firms is more clearly demonstrated using DA since a population of companies can be partitioned into percentiles or other categorizations. The RPA technique's partition of firms into "risk" categories does not allow for comparisons among the firms in the same risk category. For example, one year prior to their bankruptcies RPA and DA models classify all four firms in Figure 6 in the bankrupt group. At that time the DA score for AM International is −1.6, for KDT Industries −2.7, for White Motor −1.1, and for HRT Industries, −0.6. RPA classified HRT Industries in the second risk category (B2) and the remaining three firms in the first risk category (B1), thus not allowing for relative comparisons among the firms classified as B1. However, the fact that categorization according to "risk" is objectively determined by RPA and not a decision variable for a user may be an appealing feature in some application contexts.

VI. Summary of Results

We have explored the qualities of a new statistical classification methodology applied to the corporate distress issue. Recursive Partitioning Algorithm was found to possess the joint positive attributes of multivariate information content and univariate simplicity. Since this methodology is nonparametric, RPA is not vulnerable to the criticisms ascribed to parametric techniques and, in our example, the classification accuracy actually is superior to the more traditional discriminant framework. We are not claiming that RPA will always outperform the numerous other statistical classification techniques, nor does it have the continuous scoring system qualities of discriminant analysis. From a rigorous statistical standpoint, as well as its likely appeal to practitioners, RPA does present several very attractive features.

Classification techniques applied to financial issues continue to be refined and improved. We feel that the attributes of new techniques like RPA can be presented and evaluated in a rigorous framework without the necessity of proving its absolute superiority over existing procedures. Indeed, as we have shown, a technique like RPA can be utilized as a practical decision tool in tandem with another procedure. In our future research, we plan to examine the possibilities and benefits of utilizing the combination of RPA and DA techniques for classification and prediction in other areas of finance. For example, an extension of distressed firm classification is to classify those firms, already in bankruptcy, according to their potential for successful reorganization.

Appendix A: Criterion for Selection of the Optimal Splitting Rules in RPA

In RPA a measure of impurity of node t, $I(t)$, is defined as

$$I(t) = R_1(t)p(1 \mid t) + R_2(t)p(2 \mid t) = 2(c_{12} + c_{21})p(1 \mid t)p(2 \mid t)p(t)$$

where $p(i \mid t) = (\pi_i n_i(t)/N_i)/(\sum_{k=1}^{2} \pi_k n_k(t)/N_k)$ = conditional probability that an object in node t belongs to group i, and $p(t) = \sum_{k=1}^{2} p(k, t) = \sum_{k=1}^{2} \pi_k n_k(t)/N_k$ = probability of an object falling into node t.

$I(t)$ can be interpreted as the expected cost of misclassification when objects in node t are randomly assigned to two groups with the probability of assigning an object to group i being equal to $p(i \mid t)$. $I(t) = 0$ if and only if t is a pure node, i.e., all objects in node t belong to the same group. It attains maximum value at $p(1 \mid t) = p(2 \mid t) = 0.5$ and is symmetric in $p(1 \mid t)$ and $p(2 \mid t)$. The impurity of tree T, denoted $I(T)$, is defined as the sum of impurities of its terminal nodes. Consider a tree T and tree T' obtained from it by splitting a sample in a terminal node t of T into two subsamples which land in left and right subnodes, t_L and t_R, of node t. Then

$$\Delta I(t) = I(T) - I(T') = I(t) - I(t_L) - I(t_R)$$

is the decrease in the impurity of tree T' over tree T. $\Delta I(t)$ is nonnegative and its magnitude depends on the choice of the split. The best split for the current terminal node t is defined as the allowable split which maximizes $\Delta I(t)$.

Appendix B: List of Variables

Cash/Total Assets
Cash/Total Sales
Cash Flow/Total Debt
Current Assets/Current Liabilities
Current Assets/Total Assets
Current Assets/Total Sales
Earnings Before Interest and Taxes/Total Assets
Log (Interest Coverage + 15)
Log (Total Assets)
Market Value of Equity/Total Capitalization
Net Income/Total Assets
Quick Assets/Current Liabilities
Quick Assets/Total Assets
Quick Assets/Total Sales
Retained Earnings/Total Assets
Standard Deviation of (Earnings Before Interest and Taxes/Total Assets)
Total Debt/Total Assets
Total Sales/Total Assets
Working Capital/Total Assets
Working Capital/Total Sales

REFERENCES

1. E. I. Altman. "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy." *Journal of Finance* 23 (September 1968), 589–609.
2. ———. "Commercial Bank Lending: Process, Credit Scoring, and Costs of Errors in Lending." *Journal of Financial and Quantitative Analysis* 15 (November 1980), 813–32.
3. ———, R. B. Avery, R. A. Eisenbeis, and J. F. Sinkey, Jr. *Application of Classification Techniques in Business, Banking and Finance.* Greenwich, CT: Jai Press, 1981.
4. ——— and R. A. Eisenbeis. "Financial Applications of Discriminant Analysis: A Clarification." *Journal of Financial and Quantitative Analysis* 13 (March 1978), 185–95.
5. ———, R. G. Haldeman, and R. Narayanan. "ZETA Analysis: A New Model to Identify Bankruptcy Risk of Corporations." *Journal of Banking and Finance* 1 (June 1977), 29–54.
6. W. H. Beaver. "Financial Ratios as Predictors of Failure." *Empirical Research in Accounting: Selected Studies 1966, Journal of Accounting Research,* Supplement to Volume 4 (January 1967), 71–111.
7. L. Breiman. *Automatic Identification of Chemical Spectra: Technical Report.* Santa Monica, CA: Technology Service Corporation, 1981.
8. ——— and C. J. Stone. *Parsimonious Binary Classification Trees: Technical Report.* Santa Monica, CA: Technology Service Corporation, 1978.
9. ———, J. H. Friedman, R. A. Olshen, and C. J. Stone. *Classification and Regression Trees.* Belmont, CA: Wadsworth, 1984.
10. E. B. Deakin. "A Discriminant Analysis of Predictors of Business Failure." *Journal of Accounting Research* 10 (March 1972), 167–79.
11. B. Efron. "Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation." *Journal of the American Statistical Association* 78 (June 1983), 316–31.
12. R. A. Eisenbeis. "Pitfalls in the Application of Discriminant Analysis in Business, Finance, and Economics." *Journal of Finance* 32 (June 1977), 875–900.
13. J. H. Friedman. "A Recursive Partitioning Decision Rule for Nonparametric Classification." *IEEE Transactions on Computers* (April 1977), 404–09.
14. L. Goldman, M. Weinberg, M. Weisberg, R. Olshen, F. Cook, R. K. Sargent, G. A. Lamas, C. Dennis, L. Deckelbaum, H. Fineberg, R. Stiratelli, and the Medical Housestaffs at Yale-New Haven Hospital and Brigham and Women's Hospital. "A Computer-derived Protocol to Aid in the Diagnosis of Emergency Room Patients with Acute Chest Pain." *New England Journal of Medicine* 307 (1982), 588–96.
15. L. Gordon and R. A. Olshen. "Asymptotically Efficient Solutions to the Classification Problem." *Annals of Statistics* 6 (May 1978), 515–33.
16. O. M. Joy and J. O. Tollefson. "On the Financial Applications of Discriminant Analysis." *Journal of Financial and Quantitative Analysis* 101 (December 1975), 723–39.
17. R. Kaplan and G. Urwitz. "Statistical Models of Bond Ratings: A Methodological Inquiry." *Journal of Business* 52 (April 1979), 231–61.
18. D. E. Levy, J. J. Caronna, B. H. Singer, R. G. Lapinski, H. Frydman, and F. Plum. "Prognosis in Hypoxic-Ischemic Coma Based on Early Neurologic Findings." Working Paper, Cornell Medical Center, 1983.
19. M. L. Marais, J. M. Patell, and M. A. Wolfson. "The Experimental Design of Classification Tests: The Case of Commercial Bank Loan Classifications." *Journal of Accounting Research* 22 (1984), Supplement, March-April 1985.
20. D. Martin. "Logit Analysis and Early Warning Systems for Bank Supervision." *Journal of Banking and Finance* 1 (November 1977), 249–76.
21. J. Scott. "The Probability of Bankruptcy: A Comparison of Empirical Predictions and Theoretical Models." *Journal of Banking and Finance* 5 (September 1981), 317–44.