Quantitative Methods in Credit Management: A Survey

Eric Rosenberg; Alan Gleit

QUANTITATIVE METHODS IN CREDIT MANAGEMENT: A SURVEY

ERIC ROSENBERG

AT&T Bell Laboratories, Middletown, New Jersey

ALAN GLEIT

Citicorp, New York, New York

Many static and dynamic models have been used to assist decision making in the area of consumer and commercial credit. The decisions of interest include whether to extend credit, how much credit to extend, when collections on delinquent accounts should be initiated, and what action should be taken. We survey the use of discriminant analysis, decision trees, and expert systems for static decisions, and dynamic programming, linear programming, and Markov chains for dynamic decision models. Since these models do not operate in a vacuum, we discuss some important aspects of credit management in practice, e.g., legal considerations, sources of data, and statistical validation of the methodology. We provide our perspective on the state-of-the-art in theory and in practice.

A few statistics, gathered from various sources, show the enormous economic incentive for better techniques for credit management: From 1976–1979, the Bell System's bad debt for residence telephone users doubled, going from $128–$256 million. More troubling, bad debt as a percentage of billing rose by 50% over the same period. This problem led to the development and deployment of credit management techniques by AT&T (Showers and Chakrin 1981, Kolesar and Showers 1985). Turning to other industries, in 1978, consumer receivables totaled $276 billion for installment credit, $64 billion for noninstallment credit, and $760 billion for single family mortgage credit (while the U.S. government's debt that year was $780 billion) (Chandler and Coffman 1979). In 1985, the combined VISA/MasterCard worldwide charge volume was $183 billion, up 2% from the previous year (Gist 1986), and credit cards were believed to account for about 2% of consumer spending (up from 0.35% in 1976) (Matthews 1985). Credit card losses in 1985 are estimated at $1.8 billion in the U.S. (1.5% of volume) and $2.1 billion worldwide, and are predicted to be $4.7 billion for the U.S. and $6.3 billion worldwide by 1990. In the third quarter of 1990

credit card delinquency increased by 16% from the previous quarter, and were 27% higher than 5 years ago (*Wall Street Journal* 1991). Lastly, in 1991 about $1 billion of Chemical Bank's $6.7 billion in real estate loans are delinquent, and the bank holds $544 million in foreclosed property; Manufacturers Hanover's $3.5 billion commercial property portfolio is burdened with $385 million in nonperforming loans (Hammer and Shenitz 1991).

To minimize credit losses, a variety of credit management techniques have been developed. The decisions addressed by these techniques fall into two categories. The first category is the decisions of whether or not to extend credit, and how much credit to extend. The second category is those decisions pertaining to an existing account, including: raising or lowering the credit limit; authorizing a specific charge (for a charge card); how long a period to reissue a new charge card for when the cardholder's current card expires; how the account should be treated with regard to promotional/marketing decisions; and deciding when action should be taken on a delinquent account (i.e., determining the "start treatment level") and what action should be taken. Typical collection

589

strategies include: do nothing; use a regular statement to include a reminder; send a computer generated message; send a personal letter; make a telephone call; or assign the account to an outside collection agency. While some of these decisions may apply only to charge cards, in general these decisions apply to many types of credit, including mortgages, retail installment credit, commercial loans, and consumer loans.

Many advantages accrue through the use of quantitative methods for credit management. First, there are obvious benefits from optimally making the decisions specified above: More creditworthy applicants are granted credit (or additional credit), thus increasing profits; more noncreditworthy applicants are denied credit (or given reduced credit), thus decreasing losses; and optimal collections policies minimize the cost of administering collections or maximizing the amount recovered from the delinquent account. In addition, there are indirect advantages, including: applications can be processed quickly; the decisions are objective and not based upon human biases or prejudices (this fairness is crucial in view of antidiscrimination laws in credit granting); the profitability of the lending institution can be tied explicitly to the credit decisions; management has easy control over the system, so that changes in policy can easily be incorporated into the software rather than disseminated through meetings and paper; and fewer people are needed to administer credit granting, and the more experienced people can concentrate on difficult cases (Galitz 1983).

Most of the techniques used in practice utilize a "score" computed for each applicant or existing account to determine the decision. Therefore, much of the credit literature deals with "application scoring" (for the first category of decisions) or "behavior scoring" (for the second category). The intent of applicant scoring is to forecast the future behavior of a new credit applicant; behavior scoring tries to predict the future payment behavior of an existing account. The terms applicant score and behavior score are traditionally used in the context of discriminant analysis, which is, by far, the most common quantitative technique in credit management and which, accordingly, receives the most attention in this survey. Other techniques used and surveyed here are decision trees, expert systems, neural networks, dynamic programming, integer programming, linear programming, and Markov chains.

Credit management is currently as much of an art as a science. While the accept/reject decision for a new applicant is well defined and most amenable to

quantitative analysis, the other decisions are not as easy to formulate and are much less studied, and subjective judgment rather than empirical models appears to be the norm. Indeed, Coffman and Chandler (1983) observed that behavior scoring lacks the widespread use and the industry acceptance that applicant scoring enjoys. Another 1983 paper reports on a survey showing that only one third of the companies surveyed use credit scoring for other than application scoring (Nelson 1983).

Scoring systems utilize information relating to the traditional 5Cs of credit: (1) character (the willingness to repay debt), (2) capacity (the financial ability to repay debt), (3-4) capital and collateral (possessions or equities from which payment might be made), and (5) conditions (reflecting the general economic environment, or special conditions applying to the borrower or the type of credit) (Savery 1977, Sparks 1979, Galitz 1983). The data for scoring systems are obtained through questions on, e.g., the length of time at current address or with current employer, present salary, number of dependents, other loan commitments, and occupation. In addition, behavior scoring typically utilizes information on delinquency during the performance period (a specified period over which performance is observed, e.g., the past 12 months), account activity during the performance period, account balance during the performance period, amount past due, returned checks, age of account, new applicant credit score, and credit bureau data (e.g., past due balances, derogatory information, inquiries) (Coffman and Chandler).

While there have been some excellent papers on various aspects of credit management methods, such as those by Eisenbeis (1977, 1978) on discriminant analysis, these papers assumed a prior knowledge of discriminant analysis. This paper assumes no prior knowledge. Moreover, there are no previous comprehensive surveys of quantitative methods other than discriminant analysis (i.e., surveys that span decision trees, expert systems, and the several dynamic methods that have been proposed). Also, previous mathematically-oriented papers did not delve into such practical issues as sources of data, validation of systems, and regulatory requirements.

The literature search started with a search of the on-line Management Contents and ABI Inform data bases and led to more than 100 papers. The subject of credit screening has even appeared in the theoretical economics literature (Stiglitz and Weiss 1981, Bester 1985). Despite the best of intentions, we clearly could not review every related paper: Lachenbruch's (1979)

classic book on discriminant analysis has 579 references. Many important papers were listed in Zanakis, Mavrides and Roussakis (1986).

This paper is intended to be a comprehensive review of every major mathematical technique in the literature, and to provide some links between the theory and practice of credit management. This review's sole theme is to bring the area of quantitative methods in credit management to the attention of operations research professionals. We have tried to include all important contributions without excessive subjective commentary, preferring each reader to decide for himself or herself what techniques will ultimately be most useful.

This paper is organized as follows. In Section 1 we review the theory and application of discriminant analysis in credit management, including basic theory, choosing the cutoff point, problems in applying discriminant analysis for credit scoring, a review of some interesting case studies using discriminant analysis, and an extension to multiple discriminant analysis models. This section is extensive for two reasons. First, discriminant analysis is, by far, the major quantitative tool in credit analysis. Second, it is not part of the traditional operations research or introductory statistics curriculum. Section 2 provides an example of applying discriminant analysis. In Section 3 we describe integer programming approaches to scoring. Section 4 briefly reviews decision trees, and Section 5 considers some expert systems and neural networks developed for credit management. Section 6 reviews dynamic models for credit decisions, including Markov chains models for account aging, methods for deposit policies, acceptance and credit limit decisions, and start treatment level and collection decisions. Section 7 examines some aspects of credit scoring in practice, including validating the scoring system, sources of credit information, and legal considerations. Section 8 provides an example of the Markov chain approach. Concluding remarks are in Section 9.

1. DISCRIMINANT ANALYSIS

In this section we will review the classic theory of discriminant analysis (DA) and discuss problems in implementing DA, especially when applied to the credit classification problem.

1.1. Basic Theory

Let the population consist of two groups G and B; a member of G or B is called an *observation*. In the credit granting decision G and B are the sets of

"good" and "bad" customers. Let p_G (respectively, p_B) be the proportion of G (respectively, B) in the population. Let c_G be the cost of misclassifying a member of G (i.e., incorrectly assigning it to B) and let c_B be the cost of misclassifying a member of B (i.e., incorrectly assigning it to G). Let x be the vector of independent variables to be used to decide whether an observation is in G or B.

Consider the important special case where the density $f_G(x)$ is multivariate normal with mean μ_G (where $\mu_G \in R^N$), $f_B(x)$ is multivariate normal with mean μ_B, and $f_G(x)$ and $f_B(x)$ have the same N by N covariance matrix Σ. Thus,

$$f_G(x) = (2\pi)^{-N/2} (\det\Sigma)^{-1/2}$$
$$\cdot \exp[-1/2(x - \mu_G)'\Sigma^{-1}(x - \mu_G)].$$

Usually we do not know μ_G or μ_B. We can estimate μ_G by $\overline{x_G}$, where $\overline{x_G}$ is the (componentwise) average of a sample known to belong to G. Similarly, we can estimate μ_B by the average $\overline{x_B}$, and Σ by the matrix S computed using the samples known to belong to G and B. Then the sample-based classification rule is: Assign x to G if

$$L(x) \equiv [x - \tfrac{1}{2}(\overline{x_G} + \overline{x_B})]'S^{-1}(\overline{x_G} - \overline{x_B})$$
$$> \ln((c_B p_B)/(c_G p_G)) \qquad (1)$$

and assign x to B otherwise. Since this rule is linear in x, the technique is called linear discriminant analysis (LDA). Figure 1 illustrates LDA: With $N = 2$, we see the subsets of the good and bad populations accepted at two cutoff scores. Figure 2 illustrates the good/bad tradeoff: For each possible percentage of "goods" accepted (corresponding to some cutoff score), there is a smaller percentage of "bads" accepted.

Clearly, the choice of c_B and c_G has a major effect on the classification results of DA. However, the computation of these costs is often the most difficult problem in applying DA (Morrison 1969). Even today, many banks are only now beginning to incorporate costs into the cutoff score calculation (Fair Isaac Companies 1988). The calculation of c_B and c_G is also tied to the problem of defining which are the "good" and "bad" observations, because these definitions may involve calculating the profit of each account. A consistent methodology should be used for both of these issues.

There is little formal published methodology on computing c_B and c_G, especially with regard to the multiperiod consequences of misclassification (as discussed in Section 6 on dynamic models, the costs should reflect downstream consequences of erroneous decisions, rather than simply the immediate
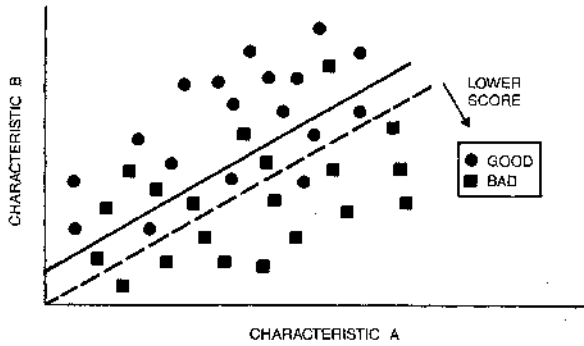
Figure 1. Discriminant analysis.

costs). Also, any cost methodology would necessarily be industry specific, including such features as different tax treatment. In addition to the difficulty of determining costs, the lack of published formal methodology is probably due, in large part, to the fact that a detailed cost analysis is likely to be proprietary.

Another perspective to DA is provided by Beranek and Taylor (1976). Let $I(x) = a'x$, where $a \in R^N$. Let $\lambda_G(\lambda_B)$ be the marginal revenue (cost) per dollar of credit for good (bad) accounts (they also consider a third category of delinquent accounts that is treated similarly and which we ignore for simplicity). Let $P(G|I)$ (respectively, $P(B|I)$) be the conditional probabilities of an account with characteristics x being good (respectively, bad) given $I(x) = I$. Here $P(G|I)$ (respectively, $P(B|I)$) is assumed to be increasing (respectively, decreasing) in I and $P(G|I) + P(B|I) = 1$ for all I. Their rule is to accept all customers scoring I as long as the expected profit per dollar of credit is positive, i.e., if

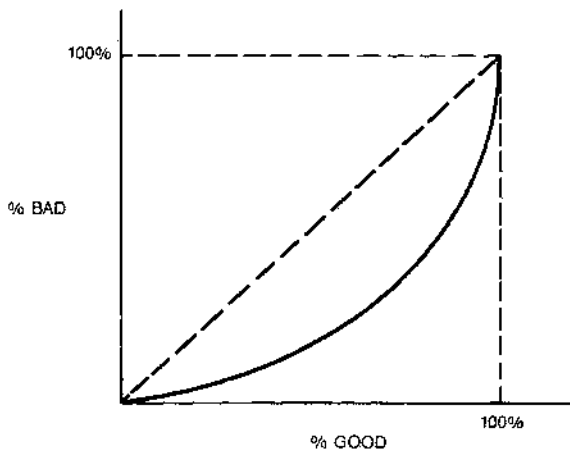$$P(G|I)\lambda_G - P(B|I)\lambda_B \geq 0. \tag{2}$$

Figure 2. Good/bad tradeoff.

The cutoff I^* is the value of I for which the left-hand side of (2) is zero; thus I^* solves

$$P(B|I^*) = P(G|I^*)(\lambda_G/\lambda_B).$$

Neither one of the cutoff rules (1) or (2) is totally satisfactory. They both suffer from the fact that the costs are assumed to be constant for all x. This is not true in general: The lost revenue from misclassifying a good customer is not the same for all customers. Also, the cost of misclassifying a bad customer is not constant: Part of this cost is the collection cost of past due accounts, and different collection strategies, with different costs, may be optimal for different subsets of the population. For example, based on the characteristics x (and possibly the past due amount) it might be more appropriate to use a mail reminder rather than a more expensive collection agency; the difference in these collection costs yields a different c_B in (1) and λ_B in (2). Another example of nonconstant λ_B occurs when x contains categorical data (i.e., $x_j = 1$ if the person has a telephone and $x_j = 0$ otherwise); in the latter case a telephone collection strategy is not available.

The optimal cutoff score can also be determined indirectly by determining the number of applicants that should be accepted or determining the maximum allowable probability of default. Greer (1967) presents a model for determining the optimal number x of credit applicants that should be accepted by a creditor, rather than the usual problem of determining which credit applicants to accept. He presents profit and opportunity cost models that require substantial data, including specifying the proportion of credit customers that default as a function of x. Fixed costs are included. The opportunity cost is convex in x for the sample data presented, and the x minimizing opportunity cost is shown to maximize credit profits. An example is presented for which profits equal $49.74x - 19.91(10^{-5})x^{2.2} + 16.95(10^{-11})x^{3.4} - 4.63(10^{-17})x^{4.6} - 1.97(10^{-24})x^{5.8}$.

A second strategic model by Greer (1968) casts the problem as determining a retailer's maximum acceptable probability of default. He expresses the expected profits from accepting a credit applicant as the sum of 11 costs or revenues, 7 of which are linear in p, the expected probability of default of an applicant. Theoretically, retail creditors can then solve for that p value above which the expected economic profit is negative, and applicants having an estimated probability of default above this value are rejected. The probability of default for an account can either be

formed subjectively or through credit scoring, because the number of bad accounts in each score interval can be used to compute the probability of default for that score interval. Greer notes that one drawback of this approach is that the parameters used in determining the cutoff p may in fact vary with p. A heuristic suggested in this case is to segment the accounts by a given range of values of a priori estimates of default probability (e.g., one group might have a priori estimates of 0.4 to 0.5) and compute a cutoff p value for each group.

A program for penetrating a new market may call for lowering the cutoff score if management is willing to accept more short-term losses in the hope of a market share in the future. Often, two cutoff points are used: If the score falls below the lower cutoff, the applicant will be rejected; if the score exceeds the upper cutoff, the applicant is provisionally approved (pending only a credit bureau check). If the score lies between, additional information (usually a credit bureau report) is obtained and the decision is re-evaluated (Day 1978, Galitz 1983). It is interesting to note that, although this two-cutoff point method appears to be popular, no theoretical justification has been published. Harter (1973), in developing an LDA rule for accepting mortgage loans, notes that, given a fixed amount of money available at a bank for mortgages, the cutoff score can be adjusted so that the total loan amount approved does not exceed the available money. For bank credit cards, offering a low interest rate to attract customers also reduces bank profits; this can be compensated for by lowering risk (e.g., raising the cutoff score) (Gist 1986).

A detailed discussion of the computation of c_B (but not c_G) for commercial bank loans is given in Altman (1980). Stone (1972) also considers the detailed economics of bank loans and shows that the cost of a bank loan to a firm is a step function of the loan size. Gentry (1974) notes that charge-offs on credit can rank behind payroll costs and interest on borrowed funds as the highest expense to a firm. Concerning c_G, for retail credit the opportunity cost for rejecting a good credit risk can include the potential increase in spending over current levels due to increased store loyalty, finance charges on store spending, and a possible decrease in spending due to customer irritation at being denied credit. For bank credit, c_G represents lost interest. Boyes, Hoffman and Low (1989) develop a model of credit card lending that shows how expected earnings on revolving credit card loans depend both on maintained balances and the probability of default. They estimate default probabilities using Manski and Lerman's "exogenous sample maximum likelihood estimator," which maximizes a weighted log likelihood function with weights determined by comparing sample proportions with corresponding population frequencies.

1.2. Problems in Applying Discriminant Analysis for Credit Scoring

Several authors have expressed sharp criticisms regarding the use of discriminant analysis in credit scoring. Many of these criticisms are really problems inherent in credit scoring. Capon (1982) cites several severe methodological problems: Since the scoring system is developed from a sample of people given credit, it is not unbiased when applied to people seeking credit (this problem is discussed in detail below); development of scoring systems with too small samples; and the use of arbitrary judgment when credit scorers assign an applicant to a category (e.g., is an executive assistant clerical or managerial). Galitz observes that, because scoring systems usually treat people with similar characteristics identically, important exceptions may be missed. For example, renting a home is generally considered less creditworthy than owning, but some occupations (e.g., police) may provide rented housing as part of the job; these people should not be penalized. This problem has often been addressed through the use of interaction variables.

Eisenbeis (1977, 1978) critically reviews many of the credit scoring methods reviewed in this paper as well as others. The 1977 paper (with 82 references) also delves into advanced statistical concepts relevant to DA (see also Eisenbeis and Avery). The 1978 reference list of 63 entries is a valuable guide to further nontechnical or moderately technical reading on DA in general, and especially the use of DA in credit decisions. In both papers Eisenbeis identifies and discusses seven types of statistical problems in credit scoring models (an eighth problem is mentioned in the 1978 paper). In the remainder of this section we discuss the views of Eisenbeis as well as other researchers on these problems.

1.2.1. Group Definition

The most severe problem, according to Eisenbeis (1977), is in group definition. Discriminant analysis procedures are appropriate under the assumption that groups are discrete and identifiable (e.g., good or bad customers). Eisenbeis argues that groups should be defined only if natural breaks or discontinuities appear in some variable. Otherwise, segmenting the variable destroys valuable predictive information.

If no natural grouping occurs, it is possible to use regression procedures by selecting a value S and classifying a case as bad if the estimated profitability is less than S, and good otherwise. This has the inherent weakness of being subjective: Different managers may select different cutoff values S even if presented with exactly the same economic conditions. Furthermore, as those economic conditions change, S needs to change, but there seems to be no general objective procedure to do this.

If natural breaks occur, it is not likely that the good and bad categories will change dramatically. (Will creditors change their definitions because costs have risen slightly?) Thus, DA methods are relatively robust, which is a very good feature because otherwise new rules need to be created with each small change in conditions.

1.2.2. Population Priors

The second most severe problem cited by Eisenbeis is that discriminant analysis assumes known a priori population probabilities (e.g., p_G and p_B in Section 1). Most models simply assume equal population probabilities; some use the sample proportions as estimates of population priors, which works if the sample is a random sample from the population. If the sample is not a random sample of the overall population, using sample proportions minimizes the classification errors for the sample, but is biased for the population.

The problem of determining population probabilities becomes more severe when data from a single time period are used to estimate group membership in a future time period: If the population priors vary over time it is not clear what the population priors should be or how they should be estimated (e.g., the number of banks in financial trouble recently has increased greatly).

Experiments by Wagner, Reichert and Cho (1983) indicated that the use of population priors minimized the number of misclassifications, compared to equal priors: The overall accuracy increased from 70.8% to 73.2% for their 3-group model. Thus, a real-world model can be improved by the use of better estimates of the a priori group probabilities.

1.2.3. Unequal Covariance Matrices

The third most severe problem cited by Eisenbeis is that LDA also assumes that the covariance matrices of the two distributions (e.g., $f_G(x)$ and $f_B(x)$ in Section 1) are equal. If they are unequal, quadratic rules are necessary. Several studies comparing LDA and QDA on distributions with unequal dispersions showed that significant differences can occur; results

of the two procedures diverge as the difference in the dispersions and the number of variables increase.

1.2.4. Measuring the Effectiveness of Discriminant Analysis

The fourth most severe problem cited by Eisenbeis concerns measuring the effectiveness of DA. Predicting model performance is generally accomplished using the holdout method: The LDA rule is developed with some fraction of the data, and performance is estimated with the remaining data, known as the holdout sample. Theoretical studies on the effectiveness of holdout samples suffer from scant data (Scott 1978).

Another way to study the discriminating power is through the proportional reduction in error (PRE) measure. Imagine a game in which we randomly draw people from a population and guess whether they are G or B. We can do this knowing nothing (chance) about them or knowing their scores using LDA. If LDA is of value, the cost (or probability) of error using LDA should be smaller than using chance. We define the cost reduction as:

PRE =

$$\frac{\text{the cost of error by chance} - \text{the cost of error by LDA}}{\text{the cost of error by chance}}.$$

Two chance methods stand out.

Method 1. Guess B with probability p_B and guess G with probability p_G. This has the expected error cost

$$C_P = c_B p_G p_B + c_G p_B p_G = (c_B + c_G) p_B p_G$$

and is related to the statistical measure tau.

Method 2. Guess B if $c_B p_B \geq c_G p_G$ and guess G otherwise. Here $c_B p_B$ (respectively, $c_G p_G$) is the expected cost if everyone is classified as G (respectively, B). This has the cost of error $C_M = \min (c_B p_B, c_G p_G)$ and is related to the statistical measure lambda.

For LDA, the cost of error is

$$C_{LDA} = c_B p_B \frac{\text{the number of } B \text{ put in } G}{|B|}$$
$$+ c_G p_G \frac{\text{the number of } G \text{ put in } B}{|G|}.$$

It is instructive to apply these measures to Altman's classic study of corporate bankruptcy (Altman 1968), which used equal population priors and equal costs. Altman and Eisenbeis (1978) revised this study using $p_G = 0.99$, $p_B = 0.01$, $c_G = 2$, and $c_B = 70$, yielding $C_{LDA} = 0.0630$, $C_P = 0.7128$,

and $C_M = 0.70$, which shows that LDA is superior to the two chance methods. In fact, $PRE_P = 0.912$ and $PRE_M = 0.910$, showing a 91% reduction in the cost of errors by using LDA rather than either chance method.

If costs are ignored by setting $c_G = c_B = 1$ and the sole objective of LDA is to maximize the percentage correctly classified, then the results of DA should be compared to C_M: If LDA does not do better, than it is better to classify everyone as belonging to the larger of the two groups. However, since LDA is typically used to classify observations into both groups, the results should generally be compared against C_P (Morrison). Morrison also observes that when one group is much larger than the other, and sample proportions are used in the LDA rule, fewer will be classified in the smaller group than actually belong in it. Thus, since there is often substantial interest in the smaller group, interpreting the results of LDA is difficult with groups of very different size. (See also Joy and Tollefson 1975, 1978 and Lachenbach and Mickey.)

1.2.5. Nonnormal Distributions

The multivariate assumptions in LDA are often violated, especially with categorical data; hence, the validity of using LDA is open to question. Violations of normality may bias the estimated error rates. Testing whether the distribution is multivariate normal is hindered by the fact that most tests are for univariate normality only.

One known way to overcome the problem of nonnormal data is to abandon discriminant analysis for a generalized linear model known as the logit model (McCullagh and Nelder 1983, Nikbakht and Tafti 1989). Given the vector x of credit data for an applicant, the probability p of default is computed by

$$\log \left(\frac{p}{1 - p} \right) = b_0 + \sum_{i=1}^{N} b_i \log (x_i).$$

One advantage of this model over LDA is that maximum likelihood estimation can be used to estimate the parameters b_i, $0 \le i \le N$. Logit models have rarely been used in credit management; one study by Wiginton (1980) found the logit model to be slightly superior to LDA, but found both to have poor predictive ability for the data studied.

Eisenbeis (1977) discusses several studies on the robustness of DA under different distributional assumptions. Logarithmic transformations have been used to make more symmetric the skewed distributions of such data as loan size or firm size. Wagner, Reichert and Cho used natural log transformations for the five continuous variables in their 3-group (good/bad/rejected) model. The accuracy rate for the transformed model is slightly lower than for the untransformed model. The transformation had negligible impact on the good group prediction rate, but dramatically affected the bad/rejected classification rates. For this one study, it appears that the problem of nonmultivariate normal data can be ignored.

1.2.6. Testing for Significance of Individual Variables

There are no statistical procedures analogous to those used in regression analysis to test for the significance of individual variables. In the linear case, it is possible to test for the conditional significance of individual variables, but such tests are not easy to use. Eisenbeis notes that the coefficients in LDA are not unique and only their ratios are unique, while in regression analysis the coefficients are unique (see also Reichert, Cho and Wagner). Thus, we cannot test if the coefficient of a given variable is zero or any other value. On the other hand, various methods have been proposed to determine the relative importance of individual variables. One method used in some studies is to compare, for each variable, the results of LDA with all n variables against the results of LDA with that variable ignored; this method requires the most work but appears to be Eisenbeis's method of choice (no statistical drawbacks are discussed).

1.2.7. Dimension Reduction

Statistical problems can arise in reducing the number of variables in LDA. Dimension reduction is important in practice, because in credit and other economic problems a large number of variables is often present. There are two chief means of reducing dimension. The first is to compute eigenvectors of the equation $|T - \gamma W| = 0$, where T is the total deviation sums of squares matrix and W is the pooled within-groups deviation of squares matrix. The dimension reducing transformation, using the matrix of eigenvectors, preserves relative Euclidean distance among observations and leaves the significance tests and classification results unaffected. However, this property holds if and only if the group dispersion matrices are equal. The second class of methods determines whether a variable contributes significantly to Wilk's lambda or related statistics used in testing equality of group means. These tests have assumed equal group dispersions.

1.2.8. Truncated Samples

The definitions of good and bad groups are partitions of the general population, whereas credit performance data to compute an LDA rule corresponds to the set of people granted credit. Technically, such a model should not be applied to the entire population but instead should be used for loan review of existing accounts (i.e., for behavior scoring).

This problem is highlighted by Harter (1974), who makes the cogent argument that scoring systems perpetuate an institution's loan policy, because people who never applied for a loan, as well as people who are rejected for credit, are not considered in developing systems to separate good risks from bad. Even if applicant data were available for people who did not apply, we would not know if they are good or bad risks. As mentioned in Gentry, some studies have tried to predict the likely behavior of those rejected. Galitz (1983) and Gentry (1974) also recognize the screening bias. The best credit system would grant credit to every applicant during some time period to gain credit information for the entire population, but few companies do this because the credit loss would be prohibitive. We learned of one mail order company that initially grants everyone a small amount of credit, having found that the resulting sales outweigh the credit losses (Fair Isaac). Also, in the development of a scoring system at Standard Oil, a group of marginal accounts that would have been rejected by human appraisal were granted credit to mitigate this screening problem; however, the group of accounts deemed poor was not granted credit (Klingel and Press 1976).

One way of addressing the problem of the missing data on rejected applicants is to use 3-group discriminant analysis to distinguish the good accepted applicants, the bad accepted applicants, and the rejected applicants. Wagner, Reichert and Cho compared the 3-group model to a 2-group (good/bad) model using data from 405 closed commercial loans and 243 rejected applicants. Using equal group prior probabilities for the models, they concluded that the introduction of the rejected group in the 3-group model did not improve the ability to detect good loans or reduce the number of bad loans predicted to be good. In fact, good/bad classification accuracy (ignoring the rejected) for the 3-group model was 57%, compared with 70.9% for the 2-group model (the total accuracy, for all three groups, for the 3-group model is 70.8%).

If credit is not granted to all applicants, the LDA analysis must be performed with a "truncated" sample. Avery (1977) has shown that this can lead to the

conclusion that the two populations have unequal covariances, even if the covariances of the underlying multivariate normal populations are equal, and hence the conclusion that QDA and not LDA should be used. Also, biased estimates of the cutoff and error rates result. Moreover, even granting credit to every applicant yields a bias, because some people may not be aware of the opportunity to apply or may decide not to apply. Eisenbeis mentions three statistical procedures which may lead to acceptable solutions to this problem: Two are known to yield biased estimates of the predictive ability of the scoring model, while a method by Avery can yield unbiased estimates. Eisenbeis concludes that this area requires additional investigation.

1.3. Case Studies in Discriminant Analysis

In this subsection we discuss several studies that utilize LDA. These studies are representative of the extensive investigation in the 1970s, primarily in universities, into the applicability and effectiveness of LDA. This discussion is intended to provide some insight into the wide range of applications of LDA (e.g., consumer loans, commercial bankruptcy, personal bankruptcy, active versus nonactive bank card holders, bank card profit and charge-off, and second mortgage evaluation), the methodological differences (e.g., LDA versus regression analysis, calculation of the one or two cutoff scores used, and functions of ratios of variables rather than of the variables themselves), and the various conclusions reached. Be warned, however, that sample sizes are small in general, and care should be taken in generalizing the results presented.

The earliest development of scoring rules was due to Durand (1951), who studied good and bad personal loans from commercial banks, finance companies, industrial banking companies, and auto finance companies. The analysis showed that good loans contained higher percentages of borrowers with many years on the job; stability of residence was also associated more with good loans than bad, but the difference was not as great as for employment. Durand found that age, car ownership, marital status, and the number of dependents were not correlated to risk (Savery).

In 1958 William Fair and Earl Isaacs of the Stanford Research Institute developed a scoring system used by American Investment Finance Company and Montgomery Ward (Nelson and Illingworth 1989). Another classic early paper was by Myers and Forgy (1963), who used discriminant and regression analysis to predict credit risk using retail credit application

data. They found that, for the most predictive variables, using equal weights yielded similar results as weights obtained from LDA or multiple regression analysis. This is known as the "flat maximum effect" (see Section 2).

In 1968 Altman published his famous "Z-Score" discriminant analysis model for predicting bankruptcy of commercial firms. His model uses five variables: working capital/total assets ratio, retained earnings/total assets ratio, earnings before interest and taxes/total assets, market value of equity/total liabilities, and sales/total assets. (A discussion of the use of financial ratios in commercial credit decisions is found in Kelley 1986.) The model was over 85% accurate in classifying bankrupt firms one statement prior to failure. Scherr (1982) criticizes Altman's model for not including variables representing firm size, firm age, and economic conditions, stating, e.g., that it is well known that younger firms are more prone to failure. Note that the Z-score model cannot accurately predict the time of bankruptcy, because bankruptcy is a legal term and a company may not be closed down even though it is financially insolvent. The Z-scores of Chrysler and International Harvester, which both experienced financial distress, are discussed in Aspinwall and Eisenbeis (1985). A proprietary refinement of the model, utilizing seven variables (overall profitability, size, debt service, liquidity, cumulative profitability, capitalization, and stability of earnings over 10 years), has been applied to a broader range of companies, including manufacturers, retailers, wholesalers, airlines, and some service firms (Haldeman 1977a, b). However, because of major differences in industry characteristics, it is not applied to financial, real estate, utility, or railroad companies (see also Altman 1986).

Orgler (1970) developed a model for commercial loan review (rather than application scoring). He notes that it is difficult to apply consumer loan methodology to commercial loan review because commercial borrowers are, in comparison, a small heterogeneous group (so gathering sufficient data is hard) with large variation in size, terms, collateral, and payment terms for the loans. Also, accurate current data on small commercial loans, especially those that defaulted, are difficult to obtain. Ideally, individual loan evaluation models should be developed for each industry. (A similar opinion by Scherr is that in forecasting firm failure the models should consider differences between industries, e.g., by using dummy variables or by developing separate models for each industry.) Orgler classifies loans as good or bad, and uses multivariate regression analysis.

Lane (1972) used LDA to study classification of consumers filing under chapter 13 of the Bankruptcy Act, debt counselees (both of whom enter programs for repayment of debt), and personal bankrupts (who do not intend to repay debt). Her model used data on approximately 250 members of each group. Two-group discriminant analysis was used with 17 variables and equal population priors. She found that the Chapter 13 filers and the debt counselees (as a group) could be separated from the personal bankrupts, but that the Chapter 13 filers and the debt counselees could not be separated from each other.

A study by Apilado, Warner and Dauten (1974) of consumer loans from banks and finance companies obtained data on 950 loans; half were paid and half were charged off. Through discriminant analysis their major conclusions are that a small number of variables can be used to construct an effective model (only 10 of the 13 possible variables had an F statistic exceeding two), and risk can be reduced without affecting profitability (one third of all bad loans can be eliminated without eliminating any good loans). They developed both univariate analysis models, which ignored interactions among variables, and multivariate models; the multivariate models, as expected, performed better.

Awh and Waters (1974) used discriminant analysis to see how economic and demographic variables (age, income, education, occupation) and attitudinal variables (attitudes toward credit and bank charge cards, both rated as either −1, 0, or +1) could be used to discriminate between 25 active and 57 nonactive bank card holders.

Fitzpatrick (1976) used regression analysis in an exploratory study of the probable determinants of bank card profit and charge-off rates in 1972. The four dependent variables in the regression equations are net profit rate on average annual cardholder outstandings, net profit rate on gross retail volume, net credit loss on average annual cardholder outstandings, and net fraud on average annual cardholder outstandings. No surrogates for general economic or financial market variables were included, which may increase the variability of the regression equations. Using data from 59 banks, some results obtained are: bank size is a significant determinant of net profit and fraud rates; net profit rates on cardholder outstandings are negatively correlated with bank size (total assets); and increasing bank experience with credit cards decreases credit losses and increases profits.

Long and McConnell (1977) developed a model to evaluate applications for second mortgages using data from 394 loans. A good loan was defined to be one

never more than 30 days delinquent. A loan was bad if any payment was 90 or more days late. Of the 117 variables that were considered, maximum predictive power was obtained using only 9 variables: length of time at current address, largest previous amount of credit, employment classification, if the applicant has a telephone, monthly income less committed payments, and four credit payment experience variables. Unlike the usual case for consumer finance loans, where a cutoff score for granting credit is determined from the costs, a single cutoff is not optimal for second mortgages. Instead, the score is used to determine the amount of the loan (as a function of property value) rather than whether the loan will be granted; details of this method are not given.

Recently, Moses and Liao (1987) developed models for bankruptcy prediction of firms using data on 26 failed and 26 nonfailed small, privately-held government contractors. Using seven financial variables (assets, liabilities, net worth, working capital, sales, earnings before taxes, and earnings before interest and taxes (EBIT)), 21 ratios and their 21 reciprocals could be used. Using textbook stepwise discrimination, they classified 79% correctly. They note that the LDA approach suffers in that the ratios chosen may not be standard financial ratios, ratios selected may be highly correlated, and the ratios chosen may not have any relation to economic theory. (Scherr echoes the concern that the models may bear no relation to theory.) Then factor analysis was used to select four of seven basic independent financial factors and the best of the 21 financial ratios that could represent each factor. The factors and ratios chosen are return on investment (EBIT/sales), leverage (net worth/liabilities), liquidity (working capital/assets), and turnover (assets/sales). However, using DA with these four ratios yielded only 73% success. They note that firms with extreme values for some ratios could introduce errors in LDA. Finally, a novel approach was considered in which firms were rank ordered on a set of ratios and a cutoff maximizing the number of firms correctly classified was determined for each ratio. An optimal cutoff was determined: A firm was classified healthy if at least two of its ratios exceeded the cutoff values for the ratios, otherwise it was classified unhealthy. This approach yielded a 79% classification rate.

1.4. Multiple Discriminant Analysis Models

In applying LDA to a large population, in particular a wide geographic region, it may be desirable to use multiple LDA models. For example, since many people moved to the southern U.S. in recent years, the variable representing years on the job is less indicative of creditworthiness there than in the northeast. In 1976 Montgomery Ward used 28 different (manual) scoring systems (Paniello 1976), and Sears has 700 different credit scoring models (almost one per store) (Updegrave 1987). The decision to use multiple models must weigh the costs of developing and maintaining the models (a topic we visit only briefly in Section 6) and smaller sample size against the expected benefits.

The only published study concerning multiple models for credit management is by Churchill, Joyce and Channon (1977), who discuss the use of clustering techniques to determine if more than one discriminant analysis model should be built to serve a population.

2. AN EXAMPLE: CITICORP MORTGAGE INC.

This section demonstrates some of the principles described above by discussing some of the credit policies used by Citicorp Mortgage Inc., among the nation's largest (by number of originations, servicing portfolio, or dollar volume) mortgage banks.

Assume that we have an applicant for a $100,000 mortgage on a home with a purchase price of $125,000. Assume that the applicant has elected a 30-year Adjustable Rate Mortgage indexed off the one-year Treasury index. The current note rate is 7.5% with an interest margin of 3% (i.e., once a year the note rate will be changed to 3% over the then-current one-year Treasury bill interest rate, unless the adjustment or life cap on the note rate has occurred).

The models discussed below will only be applicable to the class of 20% down, 30-year Adjustable Rate Mortgages indexed off the one-year Treasury index. As Citicorp is such a large originator, we can justify such a specific model.

The first problem is to determine the probabilities p_G and p_B and the associated costs c_G and c_B. From past data we know that about 3% of the loan portfolio goes into default and that 15% of all applicants are turned down. Assuming that a good job is done screening good from bad credit risks, we can guess that two thirds of the 15% turn downs are bad. Consequently, we estimate that

$$p_B = (0.03)(0.85) + 0.10 = 0.1255$$

$$p_G = 1 - 0.1255 = 0.8745.$$

As to the costs for incorrect decisions, we again use past data. To estimate the cost of default we need to add together costs of carrying the home for 3 or 4

months prior to foreclosure (paying taxes and insurance as well as forgoing principal and interest), carrying the home for 3–9 months during foreclosure, repairing the home to make it marketable, and the difference between market price and sale price when sold as a distressed property. Partially offsetting these expenses is the equity (downpayment and principal repayment of the borrower) recovered by the bank. All in all, the average loss for a default on a mortgage with 20% down is 23% of the mortgage. So we estimate c_B = $23,000 in this example.

The cost c_G is the lost opportunity to make additional profit. The profit Citicorp makes is built into the margin they charge over the Treasury index. The 3% margin covers a multitude of expenses: The normal cost of obtaining the funds to lend to the borrower is 0.9% above the Treasury index; the cost of the embedded cap options (the note rate may not go up by more than 2% each year and no more than 6% during the lifetime, no matter how much the index changes); and operating and administrative costs of the firm. Of the interest paid approximately 1% is profit. Thus, the approximate profit the first year is $1,000. Adding the profit over the lifetime of the mortgage and present valuing of the cash flow, we estimate that c_G = $4,000.

Hence, the cutoff value for the discriminant analysis is

ln $((c_B/p_B)/(c_G/p_G))$

= ln $((23,000 \cdot 0.1255)/(4000 \cdot 0.8745))$

= ln (0.8252)

= -0.192.

To perform the discriminant analysis, we consider only two variables: income and years on the job. We need to estimate the mean values for these two variables for the good and bad loans as well as the common covariance. The data from those loans that were accepted will produce biased results, because applicants with low salaries were denied loans. Based on loan officers' opinions, we estimate that: "good" have means of $36,000 and 2 years; "bad" have means of $32,500 and 1.5 years; and standard deviations of $2,000 and 0.2 and correlation of 0.9. Then (1) of subsection 1.1 simplifies to: accept the application if

$L(x)$ = 0.000003(income − 34,250)

+ 0.0242(years − 1.75) > −0.192

and otherwise reject the application. As an example, if an applicant has an income of $35,000 and 1 year on the job, $L(35,000, 1)$ = −0.016, so the applicant

should be accepted. The values $P(G|I)$ are not estimated and cannot even be guessed.

3. INTEGER PROGRAMMING APPROACH

The second major approach to making yes/no credit decisions on an individual basis is the integer programming approach of Showers and Chakrin (1981) and Kolesar and Showers (1985). They developed a model to determine which AT&T telephone customers should be required to leave a deposit. The advantage of deposits is that they provide protection against bad debt and also serve to deter risky customers; on the other hand, there is a cost of administering a deposit policy and they deter some profitable customers. While they wanted a simple scoring rule, because the customer data were all binary, they felt that classical DA would not be appropriate. They also wanted all weights on the variables to be 0 or 1 for public policy reasons. The binary data yield a finite set of possible customer profiles, and they formulated a 0–1 knapsack problem to determine which profiles should pass. The knapsack constraint is a bound on the probability of misclassifying a good customer. (They discuss other possible objectives and constraints for this decision problem and discuss relationships between the optimal solution set for the different approaches.) They note two drawbacks with this approach: Some profiles have very few people, so the results for these rules may be unreliable, and implementing the rules requires a table lookup which is difficult in practice. An integer program is then proposed which forces the rules to be linear (i.e., accept if $a'x \geq b$). They then require the scoring weights in the integer program to be 0 or 1, which leads to the rule that no deposit is required if the customer "passes" N of the J questions; a nesting property is used to reduce the computation in finding the 0–1 weights by using enumeration. Finally, the authors compare the N of J rules with both the knapsack solution and an LDA package. Applying the rules to a holdout sample, they observe that the knapsack rule does poorly (due to the small samples for some profiles as discussed above), the integer programming rules somewhat better, but both are more sensitive to random fluctuations in sample data than LDA rules. Thus, although the authors originally considered other rules in large part because LDA was felt to be inappropriate for binary data, it outperformed the other methods.

The fact that Kolesar and Showers observed fairly similar behavior for the integer programming and LDA models would seem to lend support to the "flat maximum effect" discussed by Lovie and Lovie

(1986), who observed that in credit scoring different linear models are often indistinguishable in their predictive ability. In particular, a model with unit or equal weights on predictor variables will often perform as well as one with weights obtained from least squares or LDA. This effect was observed in the Myers and Forgy study discussed in subsection 1.3. Lovie and Lovie state that conditions that seem to yield a flat maximum are choosing mainly collinear predictor variables, aligning all predictor variables in the same direction (preferably positively) as the outcome variable, and choosing a binary outcome variable.

4. DECISION TREES

The third major technique for credit decision making is decision trees. Decision trees were developed in the early 1960s by H. Raiffa and his colleagues at the Harvard Business School (Raiffa and Schlaifer 1961). In 1972 David Sparks at the University of Richmond used a decision tree to build a credit scoring model. Decision trees have gained some popularity and received official recognition when the Federal Reserve Board, in its published interpretation of the Equal Credit Opportunity Act, called decision trees an "empirically derived, demonstrably and statistically sound credit system." A detailed mathematical discussion of decision trees is given in Breiman et al. (1984).

A decision tree is illustrated in Figure 3 (taken from Makowski 1985), where the number in each node represents good account probabilities. The root node represents the universe of all accounts under consideration. Based on the vector x, accounts at a node at level l of the tree are partitioned into two or more nodes at level $l + 1$. The rule used to partition node n_1 at level l may be different from the rule to partition node n_2 at level l. In practice, a single variable is typically used for each branching with the most discriminating variables appearing at the top of the tree and the least discriminating at the leaf nodes. Branching may be done using continuous variables (e.g., income less than or more than $20,000) or discrete (dichotomous) variables (e.g., owns a home or not). A binary tree with L levels will have at most 2^L leaf nodes.

One way to apply a decision tree is to associate with each node either the probability of nonpayment or the profit for the set of people represented by the node (Makowski). The probability at node n will thus be the weighted average (with the weights determined according to the number of people at a node) of the probabilities of the children of node n. To make a decision on an observation (account), we trace down the tree from the root node, choosing the appropriate branches for the observation, until we reach the proper leaf node; comparing the probability of nonpayment or profit at the node to a chosen cutoff yields the decision.

Mehta (1968, 1970) considers the cost of information in the credit granting decision. The following is our abstraction and generalization of his model. Consider a tree, where a node represents either a state of nature (e.g., high probability of nonpayment) or an "action" (grant credit, deny credit, or investigate
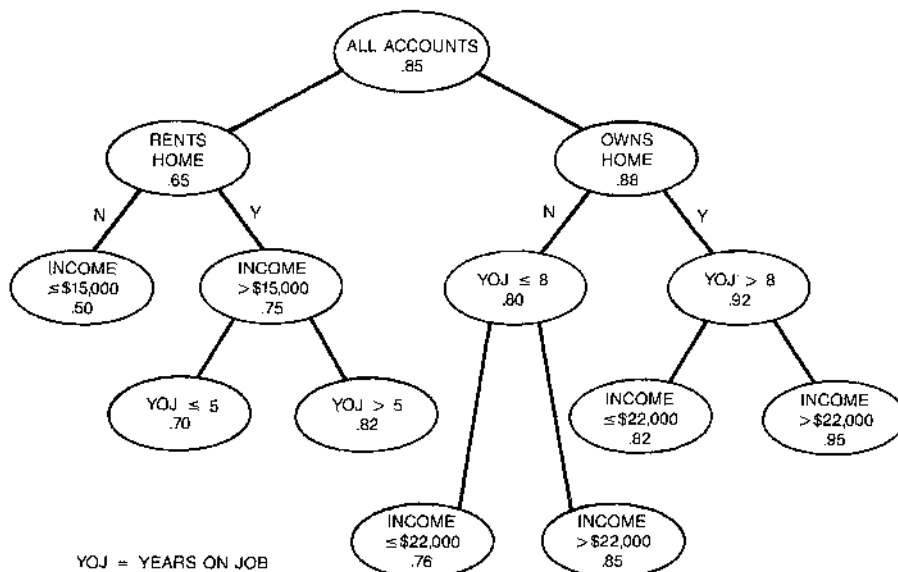

YOJ = YEARS ON JOB

Figure 3. Example decision tree.

further). Each state of nature node always has the three possible action nodes as children. The root node of the tree is defined to be a single state of nature (i.e., a partition with one element). The "grant credit" or "deny credit" action nodes have no children, and the "investigate" node has a set of children representing the possible states of nature for that "investigate" node. Different investigate nodes will, in general, lead to different states of nature. There is a cost associated with each "action" node; this cost is a function of the node and the amount of credit being considered. There is also a probability associated with each state of nature node. We introduce arcs from each "grant credit" or "deny credit" node to an artificial "sink" node. The problem is to find a minimum cost path from source to sink by choosing one of the three action nodes for each possible state of nature node (thus, an action is specified for each possible outcome). Thus, learning more about a potential applicant, by choosing the investigate node, incurs an expense but allows increased revenues (if we learn that the applicant is likely to be good) or reduced losses (if we learn the applicant is likely to be bad). The important point here is that all the possible information on an applicant should not be obtained before making a decision. This notion is analogous to the use of two cutoff scores in LDA: A credit bureau report is obtained only if the score lies between the cutoff values.

The major disadvantage of decision trees is the increased sample sizes needed to obtain statistically sound probability estimates at each node (the nodes at the lowest level have the smallest populations and present the most problem).

The advantages of decision trees are discussed by Makowski. One advantage is that they can reflect the impact of combinations of factors, not just one at a time. In contrast, linear scoring rules consider variables one at a time. He gives as an example the scoring of an applicant for having/not having a credit bureau report, and argues that not having a report is fine for a person who is 20 years old, but is suspicious for a person who is 40. Linear rules would assign a score for the report/no report variable independent of age, while decision trees can contain probabilities for combinations of factors. We have seen that profitability can also be computed using discriminant analysis; the advantage of using decision trees to calculate profitability is that the cost structure can vary from node to node. Other advantages claimed by Makowski, such as ease of use and understanding, are also shared by linear scoring rules and are not compelling arguments for using decision trees.

Coffman (1986) compares decision trees to discriminant analysis. He considers two statistical concepts, intercorrelation and interaction, both of which must be addressed adequately in any sound scoring system (including DA or decision trees). Intercorrelation occurs when some variables x_i and x_j are correlated with each other and also with the quantity of interest (e.g., credit risk). He states that DA is designed specifically to deal with intercorrelation but decision trees, because they do not generally have a node for each possible combination of variables, cannot handle intercorrelation. Interaction occurs when the correlation between some x_i and the quantity of interest (e.g., credit risk) depends on the value of some other variable x_j. When interactions exist among the x_i variables, then their effects are not additive. Coffman then states that tree analysis was designed specifically to handle interactions and can, in fact, be used to test for them. On the other hand, LDA cannot handle interactions unless special variables are included in the model. He states that it is generally much more important to deal adequately with intercorrelation than with interaction, because the former is more prevalent in scoring models. Since LDA deals well with intercorrelation, and decision trees can test for and deal with interactions, both can be used in developing a credit scoring system. These statements are not elaborated on by Coffman.

5. EXPERT SYSTEMS AND NEURAL NETWORKS

The last techniques we mention for making a credit decision for an individual account are expert systems and neural networks.

5.1. Expert Systems

An expert system relies on knowledge and reasoning of human experts to perform a difficult task. Expert systems contain three main components (Nelson and Illingworth): A knowledge base containing all the facts and rules, an inference engine that combines the facts and rules to obtain conclusions, and an interface which allows users to understand the reasoning behind a decision and add or update information on-line. Recently, a number of expert systems have been built to aid in commercial and consumer credit evaluation. For example, Citicorp Mortgage employed a consultant to develop an expert system to do the routine underwriting of loans. Since almost all expert systems allow users to query them about the reasoning used to reach the decision, they can be used to train credit grantors (Zocco 1985, Davis 1987).

At American Express, an expert system called the Authorizer's Assistant was developed to assist in purchase authorization (Davis 1987, Piketty et al. 1987). Each transaction is analyzed by a statistical model to look for charges that fall outside normal patterns. Most charges are normal and approval is automatic. If a charge is abnormal (because of, e.g., payment delinquency, a lost or stolen card, or frequent transactions suggesting fraud) it cannot be automatically approved, and an analysis is required to determine if the person using the card is the true cardholder and if the bill is likely to be paid. To assist in this process, the Inference Corporation built an expert system to advise human authorizers by displaying all relevant information, advising to deny or accept the charge, explaining the advice, suggesting questions to be asked of the cardholder, and suggestions/comments to be noted concerning the account. To build the system, they reviewed hundreds of cases and their resolutions, and interviewed authorizers. Since authorization rules vary somewhat between authorizers, it was necessary to evolve, through discussions, a standard set of rules. The deployable system will contain about 1,500 rules.

LEE (Loan Evaluation Expert) (Bravos 1987) is a knowledge-based system for commercial loan analysis developed at Xerox. LEE works with asset-based lending (also known as collaterized lending), a relatively new service provided by financial institutions which permits customers to apply for a loan on the basis of their assets. LEE works by first identifying the collateral type (usually accounts receivable but also possibly inventory, equipment, plant, etc.). Then the financial status and financial trend of the customer is assessed, yielding a status of good, fair, or poor and a trend of improving, no-change, or deteriorating. In the case of accounts receivable (AR) and deteriorating credit, four variables (quality of AR, recovery potential of AR, control of AR, and source/documentation of AR), whose values depend on the loan application, the customer, and the industry, are used to determine the quality of AR, which is then used to calculate an allowable collateral. In turn this is divided by the requested loan amount to produce the collateral ratio. This ratio is combined with three other variables, and general current economic variables to determine if the collateral is adequate and the loan should be approved, rejected, or further investigated. The number of rules employed, as well as the details of the financial calculations, are not discussed. In 1988 LEE was still in the early stages of testing and evaluation.

5.2. Neural Networks

Neural networks (Gallant 1988, Eberhart and Dobbins 1990, Nelson and Illingworth 1990), which model information processing in the human brain, consist of input, hidden, and output layers of interconnected neurons. Neurons in the one layer are combined according to a set of strengths and fed to the next layer. These strengths allow the network to learn and store associations.

The development of a neural network for credit analysis requires a training stage in which, for example, the network is given actual information about loan defaults and successes along with the support credit application data (i.e., income, occupation, etc.). This information is used to obtain a best set of strengths. Neural networks have been used successfully in corporate credit decisions and in fraud detection; though not yet applied to consumer credit, they are actively being studied and show great promise. Maves (1991) observes that as markets, products, and economics change neural networks can be "retrained" much more quickly than discriminant analysis-based techniques.

The motivation for using neural networks is that, because DA assumes variables are multivariate normal distributed, when this assumption is not satisfied the results obtained by DA may be erroneous (Wilson and Sharda 1991). Neural networks also are applicable when explicit decision rules are unavailable and information is partially correct (Jensen 1992). Several researchers have compared neural networks to classical techniques.

Dutta and Shekhar (1988) apply neural networks to the problem of bond rating. They review past approaches based on multiple regression and note that they are correct at only about 60% even when as many as 35 financial variables and numerous iterative regressions are considered. They attribute this limited success to the inability to accurately define a mathematical model for bond rating. In contrast, a neural network does not require a model, but rather attempts to learn the underlying model from the raw data. They selected 47 companies at random to study, used 30 of them to train the network, and used the remaining 17 to compare regression and the neural network approach. The success rate during the testing phase for a 2-layer network was 88.3% compared to 64.7% for the regression model. They also observed that adding additional layers to the network decreases the total error in the training phase but has little impact on the testing phase.

Wilson and Sharda compare the success rates of discriminant analysis and neural networks on bankruptcy prediction on 129 firms which were either in operation or went bankrupt between 1975 and 1982. Commercial PC-based statistical packages were used for both DA and neural networks. To achieve a better measure of predictive accuracy, they used Monte Carlo resampling techniques to generate multiple subsamples, where each subsample consisted of a training and a testing set. They used the same five financial ratios as Altman (1968) and used five input neurons (one for each ratio), ten hidden neurons in the hidden layer, and two output neurons (one indicating bankrupt, the other indicating nonbankrupt). They trained the network until all firms in the training set were classified correctly (this was possible in all 180 subsamples generated). The percentage of successes depend on the fractions f_1 and f_2 of bankrupt cases in the training and testing sets; when $f_1 = f_2 = 0.5$, the neural networks correctly classified 97% compared to 88% for DA. Neural networks outperformed DA for every value of f_1 and f_2 studied. The greatest improvement was in the classification of bankrupt firms; the authors note that this is important because it is widely accepted that it is more costly to classify a failed firm as nonfailing than the converse. The authors also survey several financial applications of neural networks (e.g., in rating corporate bonds, credit card fraud, and commodity trading).

Jensen mentions two recent studies on how neural networks might be applied to credit granting; unfortunately, neither study presented any statistics on classification accuracy. He then uses commercial PC-based neural network software to analyze data on 125 loan applicants, whose loans were classified as delinquent, charged-off, or paid-up. Since the delinquency rate is only 9.6% and the charged-off rate is 11.2%, there is insufficient data to apply credit scoring. The network has 24 input neurons, two hidden layers with 14 neurons each, and three output neurons. The network misclassified 16% of the applications as good when they were bad and 4% as bad when they were good. In comparison, a credit scoring model misclassified 8% as good when they were bad and 18% as bad when they were good.

6. DYNAMIC MODELS

Almost all credit decisions are dynamic and not single period. This is clearly true for consumer retail or bank credit cards, for both revolving or nonrevolving credit. It is also true for both commercial and consumer loans that are paid in installments. Finally,

even loans to be paid back in full at a specified time become dynamic if the debtor defaults and collections extend over time. Surprisingly, there is scant published literature treating the dynamic credit problem, and this literature is in the academic realm. Judging from the published literature, dynamic models have had considerably less impact in practice than static models. In this section we consider the use of Markov chains for projecting account behavior into the future, acceptance and credit limit decisions, and start treatment level and collections decisions. Such "flow modeling" is routinely used by collection managers even though many have never heard of a Markov process.

6.1. Account Aging and Markov Chains

In 1962 Cyert, Davidson and Thompson (1962) used techniques of Markov chains to study the long-term, expected uncollectible amount in each age category. Here age means how long past due is the account. Define B_j as the amount j periods (e.g., months) past due at some period for $j = 0, 1, 2, \cdots J$. Here j is the state variable, $j = 0$ represents the paid in full state, and $j = J$ represents those accounts J or more periods past due and is the bad debt state. Let B_{ij} be the amount in state j in some period which came from state i in the previous period. Both B_i and B_{ij} are assumed stationary or time-independent. Then the transition probability P_{ij} of a dollar in state i at some period transiting to state j in the next period is given by

$$P_{ij} = \frac{B_{ij}}{\sum_{m=0}^{J} B_{im}}.$$

The states $j = 0$ and $j = J$ are absorbing states: $P_{0,m} = 0$ and $P_{J,m} = 0$ for all m. The authors note that there are two methods available for aging: In the total balance method, all dollars in the account are put in the age category corresponding to the oldest dollars, and in the partial balance method, the dollar balance is allocated among the age categories on the basis of the age of each dollar in the account.

Reorder the states so that the absorbing states 0 and J come first, and then the other states 1, 2, ..., $J - 1$. Partition P into the form

$$P = \begin{pmatrix} I & 0 \\ R & Q \end{pmatrix},$$

where I is the 2 by 2 identity matrix. Let $N \equiv (I - Q)^{-1} = I + Q + Q^2 + \cdots$. Then the entries of

the $(J - 1)$ by 2 matrix NR give the probabilities of absorption in each of the absorbing states 0 and J.

Now let S_0 be the vector whose jth component is the initial amount in the jth age category $j = 1, 2, \ldots, J - 1$. Then $S_t = S_0 Q^t$ is the vector whose jth component is the amount outstanding for the jth age category at the beginning of the tth period for $t = 1, 2, \ldots$. The vector $S_0 NR$ (with two components) is the expected payment and bad debt from the accounts receivable (the authors provide a formula for the variance).

We can generalize this model to include installment loans. As above, let $j = 0$ represent the paid in full state. Let state $j = 1$ represent the current state, i.e., those who are up to date in payments. Let states $j = 2, \ldots, J - 1$ be the states of payment $1, 2, \ldots, J - 2$ periods in arrears. Finally, state $j = J$ is default. For monthly payments of $1, moving from state i to state j involves payment of $i - j + 1$ dollars for $1 \le i, j \le J - 1$. This may be written as a matrix M:

$$M = \begin{pmatrix} 1 & 0 & \cdot & 0 \\ 2 & 1 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ J-1 & J-2 & \cdot & 1 \end{pmatrix}.$$

Let T_t be the multiplication of M with Q_t element by element (e.g., $T_{ij} = M_{ij} Q_{tij}$). Then T_t represents the expected received payments in the sense that $S_0 T_t$ is the vector of expected payments from states $1, 2, \ldots, J - 1$ (excepting last payments and prepayments) at time t and $S_0 T_t e$ (where e is the $J - 1$ column vector with all entries 1) gives the total expected payments.

The Cyert-Davidson-Thompson model assumes the total balance method of account aging in which the age of an account is the age of the oldest unpaid dollar in the account. Thus, for example, if at some time a bill for $20 is one month past due and a bill for $10 is two months past due, the method assumes that $30 are two months past due. The problem with this method is that if a payment of $10 is made, the method then indicates that $20 are two months past due, whereas if we assume the payment applies to the oldest dollar, then we actually have $20 only one month past due. The steady-state distributions computed using total balance aging tend to underestimate the actually paid dollars. Van Kuelen, Spronk and Corcoran (1981) present a simple modification of the CDT method to make the accounting more realistic.

Corcoran (1978) notes that the transition matrices in the CDT model can become unstable, and stability can be enhanced if accounts were first grouped according to size and then a transition matrix computed

for each group; otherwise if large and small accounts were mixed, a large payment would drastically affect the transition probabilities. Corcoran modifies the CDT model to use the partial balance method of aging, and uses the data to compute the monthly transition matrices, which are weighted to arrive at an average or exponentially smoothed matrix. When applied to the aging data for a month, this average matrix yields the estimated data for the next month. This modification avoids the stationarity assumption in the CDT model: See Frydman, Kallberg and Kao (1985) and Mehta (1970) for other extensions of the CDT model.

6.2. Acceptance and Credit Limit Decisions

The first dynamic model for determining credit acceptance or a credit line appears to be by Bierman and Hausman (1970) who consider the question of whether or not to grant credit. They provide the following insightful example of why it is necessary to consider multiple time periods. Consider a product that costs $c = 62$ to produce and sells for $s = 100$. We will consider granting credit as long as the expected k period return is positive.

First, for $k = 1$ let p be the probability that the item, when sold on credit, is paid for in full. If $p = 3/5$, then the expected profit from period 1 to the end of the horizon (period 1 in this case) is $(3/5)(100 - 62) + (2/5)(-62) = -2$, so credit should not be offered in period 1.

Now assume that there is a prior probability distribution on the probability of collection. This Bayesian approach assumes that the credit grantor's prior feelings about the probability of payment can be represented by letting p be a random variable with a Beta distribution with parameters r and n (the expected value is r/n). Suppose that credit is extended n' times and full payment is made r' times. Then the parameters of the Beta distribution are revised according to: $r \leftarrow r + r'$ and $n \leftarrow n + n'$. The new distribution is called the posterior distribution.

Suppose that the parameters of the prior are $r = 3$ and $n = 5$. As shown above, the expected 1-period revenue is -2, so we would not extend credit. Now consider two periods $(k = 2)$. With probability $2/5$ the customer defaults in period 1, in which case he defaults in period two with posterior probability $3/6$. Thus, the expected profit in period 2 is $(3/6)(100 - 62) + (3/6)(-62) = -12$. Since the expected profit from period 2 to the end of the horizon (period 2 in this case) is negative, credit for a defaulted customer should not be granted for period 2. With probability $3/5$ the customer pays in period 1, in which case he

pays in period 2 with posterior probability 4/6. Thus, the expected profit in period 2 is $(4/6)(100 - 62) + (2/6)(-62) = 4.67$. Since this is positive, credit would be granted for a paying customer. Putting this all together, the expected 2-period profit is $(3/5)$ $[(100 - 62) + 4.67] + (2/5)(-62 + 0) = 0.8$. Thus, the expected gain from two periods is positive, so we should offer credit initially.

Bierman and Hausman indicate that determining the prior distribution parameters r and n may not be easy, and that the expected k period profits will, in general, vary with r and n, even if their ratio (the mean of the distribution) remains constant.

We now present Bierman and Hausman's dynamic programming formulation of the multiperiod credit granting problem. Let c_1 be the single period profit if credit is given and payment is made, c_2 be the single period loss if credit is given and no payment is made, (r, n) be the state variable, where r and n are the parameters of the current prior Beta distribution on the probability of payment, $f_i(r, n)$ be the maximum expected discounted payoff from stage i to ∞ given that the current stage is (r, n), $d_i(r, n)$ be the optimal action at stage i, where $d_i(r, n) = 1$ means grant credit and $d_i(r, n) = 0$ means deny credit, and α be the discount rate. Then we have

$$f_i(r, n)$$

$$= \max \{(r/n)(c_1 + \alpha f_{i+1}(r + 1, n + 1))$$

$$+ (1 - (r/n))(c_2 + \alpha f_{i+1}(r, n + 1)); 0\}.$$

The term to the left of the semicolon in the "max" expression is the expected profit if credit is granted, and the 0 represents the expected profit if no credit is granted. Here r/n is the expected probability of payment in the current period. If payment is made, the state (r, n) is updated to $(r + 1, n + 1)$; if no payment is made the state is updated to $(r, n + 1)$.

Bierman and Hausman also consider the dynamic problem of how much credit to grant (i.e., setting the credit line). They assume that, if an amount y of credit is extended in the start of a period, the probability of collection (payment in full) at the end of the period is p^y, where p is the probability of payment in full for a unit of credit. (Partial payments are not considered.) The prior density function of collection is also assumed to be a Beta distribution with parameters r and n and is updated by setting $(r, n) \leftarrow (r + y, n + y)$ if y is extended and collected. Since the updating of the prior turns out to be complicated if collection is not made, they assume that the expected discounted payoff after a period of no collection will be zero.

Since $(r/n)^y = [E(p)]^y$, with the previous notation, we now have

$$f_i(r, n) = \max_{0 \leq y < \infty} \{(r/n)^y[c_1 y + \alpha f_{i+1}(r + y, n + y)]$$

$$+ [1 - (r/n)^y]c_2 y\}.$$

The Bierman-Hausman model has been extended by Dirickx and Wakeman (1976) and Srinivasan and Kim (1987). The major result of Dirickx and Wakeman is that it is not necessary to assume that the expected future payoff from period i on is 0 if no collection is made; although Beta distributions are no longer always preserved, the computations can still be performed. Srinivasan and Kim make a simple modification relaxing the implicit assumption in the Bierman-Hausman model that the credit grantor simultaneously makes collections and extends credit; a simple timing modification is presented.

6.3. Start Treatment Level and Collection Decisions

Mitchner and Peterson (1957) consider the problem of how long to continue to pursue collections for a defaulted loan. The problem is the tradeoff of collections cost against the expected recovery if collections activities are continued. They compare the results of the optimization to historical data and observe that use of the optimal strategy would lead to premature abandonment of only a small number of loans that actually converted to paying status; similarly, use of the strategy would lead to early abandonment of many loans that did not convert to paying status.

Their model assumes a cost c of collections pursuit per loan-month, and a probability $p(t|t_0)dt$ that a loan converts between times t and $t + \delta t$ given that the loan has remained nonpaying for time t_0. Thus, the probability that a new loan will convert between t and $t + \delta t$ is $p(t|0)dt$, and infinite pursuit of a loan yields a probability $\int_0^\infty p(u|0)\, du$ of eventual conversion. Let z be the average fraction of the recovery when collections yields a payment, and let A be the amount owed. Then the optimal amount of time T to pursue a loan satisfies $p(T|T) = c/(zA)$, and the maximum expected profit for a loan of age t_0 is

$$N(T|t_0) = \int_{t_0}^T [zA - c(u - t_0)]p(u|t_0)\, du$$

$$- c(T - t_0)\left[1 - \int_{t_0}^T p(u|t_0)\, du\right].$$

The authors discuss the use of maximum likelihood estimation to estimate the probability of conversion

$p(t|0)$ for a new loan. They also provide a simple heuristic method for estimating the probabilities and show that the heuristic and MLE estimators are equal if the fraction of loans of age t that convert or close-out in r time units is independent of t.

Liebman (1972) developed an infinite-horizon Markov chain model to determine the optimal collections strategies. (Liebman considers the strategies of no action, a letter requesting payment, and a telephone call.) Accounts are divided into subpopulations, depending upon such variables as age, past payment experience, and past and current activity. Let $P^s_{ikm,jln}(c^s_{ikm,jln})$ be the probability (cost resulting from) of an account moving from age class i, charge volume class k and previous experience class m to age class j, charge volume class l and previous experience class n in one time period when strategy s is used. Drawing upon the results of Howard (1960), the optimal policy with t periods left in the planning horizon satisfies

$$v_{ikm}(t) = \min_s \left\{ \sum_{jln} P^s_{ikm,jln}[c^s_{ikm,jln} + \beta v_{jln}(t - 1)] \right\},$$

where beta is the discount rate and $v_{jln}(t)$ is the minimum expected total cost with t stages left starting in state jln. This can be solved using policy iteration (Howard), which consists of solving a set of linear equations and making a series of comparisons.

Liebman also shows that determining the optimal policy can be formulated as a linear program: Let x^s_{ikm} be the steady-state probability that the system is in age class i, charge volume k, previous experience class m and s is selected. The LP is

minimize $\displaystyle\sum_{ikms} x^s_{ikm} \sum_{jln} p^s_{ikm,jln}, \; c^s_{ikm,jln}$

subject to

$$\sum_{ikms} x^s_{ikm}(\delta_{ikm,jln} - \beta P^s_{ikm,jln}) = (1 - \beta)/t \quad \text{all } j, l, n$$

$$\sum_{ikms} x^s_{ikm} = 1, \quad x^s_{ikm} \geq 0.$$

The linear program has the advantage over policy iteration by providing sensitivity analysis of the optimal policy to changes in the costs and transition probabilities.

Pye and Tezel (1974) present a dynamic programming approach to determine the value of a collateralized loan subject to default. Let a series of equal payments be scheduled for the next T periods. Without loss of generality, we can assume that each payment is $1. A payment not received at the end of the scheduled period is one period late, and a foreclosure decision is made. If the payment is two periods late,

another foreclosure decision is made; this continues until foreclosure or full payment is received.

To formulate the model, let n be the number of periods that payment is late, and let $V_t(n)$ be the maximum expected present value of loan payments subsequent to t, given that an optimal foreclosure policy is used and the payment scheduled for period t is n periods late. If the payment n periods late is received in period t, a new loan terminating at time T is negotiated, with an expected present value of $n + 1 + V_t(0)$. If foreclosure occurs in period t, the lender receives collateral with an expected present value of b_t, plus a "deficiency judgment" on back payments with expected present value $\xi_n n$. We assume that $b_t \leq 1 + V_t(0)$ for all t, which means that the value of the collateral does not exceed the value of the loan when payments are on time (otherwise the borrower should sell the collateral). The probability of receiving payment in the next period when payment is currently n periods late is γ_n. The present value this period of a dollar received in the next period is denoted by α.

The dynamic programming recursion is then

$$V_t(n) = \alpha \gamma_n (n + 1 + V_{t+1}(0)) + (\alpha - \alpha \gamma_n)$$
$$\cdot \max \{b_{t+1} + \xi_{n+1}(n + 1), \; V_{t+1}(n + 1)\},$$
$$V_T(n) = \alpha \gamma_n n + (\alpha - \alpha \gamma_n)b_{T+1}.$$

These equations are derived as follows: Given n and no foreclosure in period t, in the next period either payment is received with an expected value of $n + 1 + V_{t+1}(0)$ or payment is not received, in which case the lender can foreclose and receive $b_{t+1} + \xi_{n+1}(n + 1)$, or not foreclose and receive $V_{t+1}(n + 1)$. By working backwards from the horizon, the value of a collateralized loan is $V_0(0)$.

Pye and Tezel consider the special case in which $b_t - b_{t+1} = \lambda > 0$ for all t (the unearned premium decreases by a constant amount each month), $b_{T+1} = 0$ (the payments are completed when the value of the unearned premium is zero), and $\xi_n = 0$ (no collection of defaulted payment upon foreclosure). In this case, they derive the closed-form solution for $V_t(0)$ and a constant v^* such that foreclosure is always optimal when $n = 1$ (whenever a payment is one period late) for any t if and only if $v_1 \leq v^*$.

7. CREDIT SCORING IN PRACTICE

Credit scoring is gaining acceptance: A 1990 survey reported that 82% of banks using expert systems employ credit scoring for commercial, consumer, and mortgage loans, even though the cost of developing a credit scoring model is estimated to be $50,000–

OWN/RENT	OWNS 45	RENTS 18	ALL OTHER 24			
YEARS WITH EMPLOYER	UNDER 1 YEAR 16	1 TO 2 YEARS 22	3 TO 4 YEARS 26	5 TO 9 YEARS 28	10 TO 12 YEARS 29	13 YEARS & OVER 36
CREDIT CARD	CARD 19	NO CARD 0				
FINANCE COMPANY	YES 0	NO 36				
BANK ACCOUNT	CURRENT +DEPOSIT 50	CURRENT 31	DEPOSIT 32	NONE GIVEN 15		
OCCUPATION	PROFESSIONAL/ EXECUTIVE 29	MANAGER/ SEMI-PROFES- SIONAL 28	OFFICE STAFF 25	PRODUCTION 15	SALES 22	ALL OTHER 15
PREVIOUS ACCOUNT	UNSATISFACTORY 0	NEW 55	SATISFACTORY 87			
CREDIT BUREAU	NO FILE 15	DEROGATORY RATINGS −33	1 OR 2 SATISFAC- TORY RATINGS 24	3 SATISFACTORY RATINGS & UP 30		

Figure 4. Example scoreboard.

$100,000 (Jensen). This section discusses the practical aspects of credit scoring, including development of a system, studies comparing human experts to credit scoring, validating the system, sources of data, and legal considerations.

7.1. Development

Consumer scoring development typically begins with tests of 60–80 questions that are narrowed down to the 9 or 12 questions that prove to be the best predictors (Main 1977). An example of an LDA rule for credit scoring (known as a "scorecard") is shown in Figure 4 (taken from Fair Isaac). In consumer scoring, a minimum of approximately 300 bad accounts is needed for statistically significant results (Chandler

1985); having enough good or rejected accounts is not generally a problem. (As noted above, data can be scarce for commercial credit scoring.) Scorecards can be used for both applicant scoring and behavior scoring. Figure 5 (taken from Fair Isaac) illustrates scoring for credit limit adjustment.

In implementing a scoring system, questions may be asked and not scored, and the items which are scored and their weights may not be available to the scorers to reduce the vulnerability to fraud (Day). The questions not scored provide demographic data valuable for financial and marketing purposes.

Some interesting clues to scoring systems are provided by Updegrave. A perfect history of bill payments can yield 25% or more of the points needed for

ACTIONS: O = NO CHANGE TO LIMIT
+ = INCREASE LIMIT
— = DECREASE LIMIT

ACCOUNT STATUS SCORE	CURRENT			30 DAYS PAST DUE		
	<75% OF LIMIT	75 TO 100 % OF LIMIT	>100% OF LIMIT	<75% OF LIMIT	75 TO 100% OF LIMIT	>100% OF LIMIT
BELOW 500	O	O	O	—	—	—
500-599	O	+	O	O	O	O
600 & UP	O	+	+	O	+	+

Figure 5. Using performance score for credit limit adjustment.

approval. Bank credit card delinquencies hurt more than retail card delinquencies. Lenders ignore some derogatory entries, such as small unpaid medical bills, because they usually indicate a dispute over the bill. However, personal bankruptcy usually spells definite rejection. Typically, a monthly income of $1,500 might earn 10 points, $2,500 monthly earns 12 points, and no additional points are earned after $3,000 monthly. Some scoring systems place applicants into one of 8 job categories: professional, managerial, blue collar supervisor, clerical, sales, self-employed, skilled trade, and unskilled worker. Executives and professionals typically earn the most points (about 30), while unskilled workers may earn only 5 (this contradicts Savery).

Sears has found that occupation has no predictive value among its customers (Main), and keeps its losses below 1% by using a high cutoff score. Diner's Club also has high standards and fails about 50% of its applicants. A store with a high profit margin, and thus an ability to withstand heavy losses, might fail only 10%. In today's competitive environment, a bank typically fails only 20–40%. Gist reports that credit management needs careful attention: one San Diego bank offered preapproved credit cards to people who were unemployed or even dead (then lost $63 million and sold its credit card business). Lastly, Anonymous (1986) gives a frightening story (from a creditor's point of view) of the fraudulent acquisition and use of 32 credit cards; the moral of this tale is that ideally all the information on credit applications should be verified; however, this is generally prohibitively expensive in practice.

Gist provides other data on bank card credit operations: one estimate is that a new portfolio of bank credit card users takes from 1–2 years to mature and that losses of 4%–5% can be expected during that time, especially if direct mail solicitation is used; upon maturity, losses of 2–3% can be expected.

We now review the relatively scant literature on the use of behavior scoring (i.e., existing account management). Connors (1988) and Coffman and Darsie (1986) discuss how a collection strategy can be determined using a behavior score. Each range of behavior scores can be assigned to a collection strategy, defined as a sequence of collection actions. The strategy would specify the action to be taken each month if no payment is received.

Behavior scoring is very effective when used in conjunction with experiments to determine the best action (Weingartner 1966, Klingel and Press 1976, Kaye 1981, Coffman 1986). For instance, to test the effectiveness of collection actions, accounts can be split into groups, with a different range of behavior scores for the accounts in each group. Each group is then subdivided into subgroups with a different action assigned to each subgroup. After a period of, say six months, the accounts are examined to determine the effectiveness of each action. By examining the cost/benefit of each action, the best strategy can be identified for each range of behavior scores. Connors notes that continuous testing of alternative strategies allows pro-active credit management, rather than reacting based on signs of deterioration of the scoring system.

A variety of credit systems can be purchased which may offer features beyond credit analysis such as the ability to automatically generate denial letters, obtain credit reports, set up accounts on the credit grantor's mainframe, and integrate with card embossing systems. Since banks may receive as many as 21,000 credit card applications per month, automated systems can considerably reduce costs. A survey of such systems is provided by Cohen (1985).

7.2. Subjective Versus Empirical Scoring

Several early studies pointed to the fallibility of human judgment making in scoring. Gentry mentions a 1964 study in which, of 17 variables, only the interviewers appraisal did not significantly discriminate between the goods and bads. Also, a 1972 study indicated that when credit managers approved people with low scores, this group became the highest delinquency group. Sears studied what happened when credit managers approved a loan that the scoring system rejected and found that 95% of the loans were hard or impossible to collect (Main). A Fair, Isaac Companies survey found that 22% of users stated that loans scoring below the cutoff were never approved; the rest allow human override. Also, 44% of the companies surveyed tracked the performance of those approved due to human override: Of these companies, 84% rated the accounts as unsatisfactory (Nelson). Chandler and Coffman discuss an experiment in which 25 good and 25 bad accounts were scored by a model and a group of experienced analysts. When the cutoff was set to approve only 25 accounts the model made 4 errors (4 known bad accounts were accepted). When the analysts were asked to approve the best 25, only one analyst did as well. Finally, prior to their development of scoring rules to determine deposit policies for residential telephone users, public utility commissions viewed judgmental screening (then in effect) as neither uniform or objective, and not effective in identifying the nonpaying

customer with a reasonable degree of accuracy (Showers and Chakrin).

The above data suggest that a good scoring system outperforms human experts. If true, since expert systems are intended to mimic human experts, the recent interest in expert systems should be confined to those credit management decisions not amenable to empirical methods. For example, fraud detection, but not consumer or commercial loan evaluation, is probably best performed with an expert system.

Not everyone extolls the advent of scoring systems: Harter (1974) claims that scoring systems represent a composite of the loan officer's judgments and loan officers must continue to use their judgments to pre-screen applicants before using a scoring system. Harter then claims that "it is unlikely that credit scoring systems will become widely adopted ... " and " ... will be relegated to the academic world and regulatory agencies ... " (so much for predictions). Hall (1983) discusses the different effectiveness of closed questioning, in which the applicant selects from a set of alternatives or responds to a specific question (the method used by scoring systems), versus open questioning, in which the applicant has a conversation with the credit grantor.

7.3. Validating the Scoring System

An important practical issue is whether existing scoring formulas continue to be effective as economic conditions or the stream of applicants change (Klingel and Press). For example, people today often possess many more bank credit cards than previously. Another example is that, for professionals, the length of time on a job or at a residence may be poor predictors of risk. If changes are necessary, then the entire model and not just the point values need to be updated as new data become important.

Long (1976) developed a model to determine the optimal schedule for updating a scoring system. The time between updatings is shown to be a decreasing function of the decay (performance) rate and the growth rate of good and bad accounts, and an increasing function of the cost of updating the system and the discount rate. Long also discusses a method for determining the decay function (e.g., $f(t) = 1 - \alpha t$ or $f(t) = 1 - \alpha t^2$ for some parameter α), using data on the performance of the system over time.

A survey by Fair, Isaac (Nelson) indicated that only 16% of the companies surveyed had credit scoring development dates before 1979, and 44% reported validations during 1980–1981. Sears, with 60 million cardholders and 700 different credit scoring models (almost one per store), updates scoring systems every

three years (Updegrave). Of the 61% who reported a problem during installation and operation of the systems, difficulty in tracking system performance was identified as a major concern: Only 22% of the companies surveyed are able to track, by score and time on the books, good accounts versus delinquents.

7.4. Sources of Credit Information

The chief source of consumer credit information is the three U.S. credit bureaus that serve a wide geographic area with their own computer data base containing accounts receivable information from credit grantors and public record information. The data bases are vast: TRW estimates it has information on 133 million consumers. The bureaus obtain ledger information from credit grantors, verify employment with employers, and obtain credit-related public records (such as bankruptcy, lawsuit, judgment, or divorce data). The standard format of a credit report contains information on who is responsible for paying the account, the type of business reporting the information, the date of the information (major creditors send updated information monthly), the date the account was opened, the date of the last payment, the highest amount of credit extended, whether the account is open charge, revolving, or installment, explanatory remarks (e.g., dispute following resolution, dispute pending resolution, moved with no forwarding address, repossession, card stolen or lost). People or firms with a legitimate need for credit reports can purchase them from the credit bureaus; the Fair Credit Reporting Act lists permissible purposes for obtaining consumer credit reports. When joining a bureau, a credit grantor usually agrees to supply a list of its current customers and their payment histories.

In accordance with antidiscrimination laws, retailers can supply a bureau with requirements such as income or time on job, and receive a list of potential new customers. Bureaus also may provide computerized lists of people with good payment records; these lists may be purchased by retailers for solicitation purposes. Credit bureaus often offer a retail debt collection service. Consumer rights with respect to credit bureaus are discussed in Criscuoli (1985) and Cole (1988).

Bureaus also offer commercial credit data including current and previous payment information, payment trends, industry payment profiles, public record information, financial information from Standard and Poors, basic company data (e.g., sales, number of employees), and information on debt to government agencies.

In addition to general credit information on consumers, credit bureaus also maintain national lists of delinquents. One such list has data on 10 million people (Hicks 1975). Derogatory information is kept for five years, and the data are updated about twice monthly. Credit card companies and other firms can submit lists of names to determine delinquencies, at a nomimal cost per inquiry. Such a process is called prescreening. For example, the Associated Credit Services allow a financial institution to choose from a variety of "standard" prescreening criteria; ACS then selects, from its 80 million names, individuals who match the requirements. Typical requirements are no bad debts, no suits or judgments against the individual, the individual has good standing with other creditors, and the individual has a job. A variety of prescreening options are discussed by Jarvis (1986).

Names to solicit for new accounts can be obtained from a variety of sources, including recent utility turn-ons (Kane 1982). The Claritas Company provides demographic information on neighborhoods within a zip code, describing neighborhoods by their reaction to promotions and presumed payment characteristics (Rossi 1982).

7.5. Legal Considerations

We mention lastly the complicated and important area of credit scoring systems and laws prohibiting discrimination. The key pieces of legislation are the 1974 Equal Credit Opportunity Act (ECOA) which prohibited discrimination in the granting of credit on the basis of sex or marital status and the implementation of the act through the Federal Reserve Regulation B; and the Amendments of March 1976 that additionally prohibited discrimination on the basis of race, color, religion, national origin, age, receipt of public assistance benefits, and the good faith exercise of rights under the Consumer Credit Protection Act. Regulation B describes criteria that scoring systems must satisfy to ensure they are methodologically and statistically sound. Three criteria are specified (Wagner, Reichert and Cho): the credit data for system development should be either the institution's entire population or a properly drawn sample with both accepted and rejected applicants; prior to implementation, the system must be validated using actual data to ensure that it can distinguish creditworthy from noncreditworthy applicants in a statistically significant manner (no specific statistical test or level of significance is mandated); the system must be periodically revalidated at appropriate time intervals (no interval is mandated).

The subject of credit scoring and discrimination law is well beyond the scope of this survey; for additional information the reader may consult Galitz, and a very detailed and careful analysis in the Yale Law Journal (Anonymous 1979) replete with additional references.

8. AN EXAMPLE: CITICORP MORTGAGE INC.

Continuing the example of Section 2, we can also look at the loan portfolio as it ages and estimate the transition probabilities for the Markov chain approach. We have these states:

state 0: paid off either by prepaying or coming to full term;
state 1: owes the current month;
state 2: one month in arrears;
state 3: two months in arrears;
state 4: three months in arrears;
state 5: in default.

For our assumed 20% down typical adjustable rate mortgage borrower, we estimate the transition matrix P to be

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0.01 & 0 & 0.9625 & 0.0275 & 0 & 0 \\ 0.01 & 0 & 0.9 & 0.05 & 0.04 & 0 \\ 0.001 & 0 & 0.649 & 0 & 0.05 & 0.3 \\ 0 & 0.85 & 0.13 & 0 & 0 & 0.02 \end{pmatrix},$$

Then

$$NR = \begin{pmatrix} 0.97 & 0.03 \\ 0.96 & 0.04 \\ 0.70 & 0.30 \\ 0.13 & 0.87 \end{pmatrix}.$$

Thus, 3% of an initial portfolio will default. The eventual defaults among those in arrears one month is only 4%; the percentage jumps to 30% of those 60 days in arrears. Once a borrower reaches 90 days in arrears, the likelihood of an eventual default is overwhelming (87%).

9. CONCLUDING REMARKS

The 1960s saw the limited use of discriminant analysis in credit scoring, as well as the development of many of the classic multiperiod methods, such as the Cyert-Davidson-Thompson Markov chains model and the Bierman-Hausman dynamic programming approach. The 1970s was a period of widespread experimentation with scoring, and a deeper understanding of the statistical issues involved, as well as refinements on the multiperiod methods. The 1980s have seen limited theoretical progress on both discriminant

analysis or multiperiod methods. Instead, the major themes of the 1980s have been expert systems, multiple scorecards (or, in the "limit" decision tree approaches), and the use of experimental strategies to determine optimal policies.

By far the most mature branch of quantitative methods is in deciding whether to accept or reject a credit applicant (or in loan review). The use of statistical methods and experiments to determine optimal start treatment levels and collections strategies is less well established, and is much more of an art. While Markov chain transition matrices have been used in these methods, the more sophisticated techniques of linear or dynamic programming appear to be unused in practice. For many of the other important credit decisions, such as adjustment of the credit limit, reissue period, and promotions strategy, there is no published evidence of quantitative methods in use and little theory: Only the Bierman-Hausman model (and its refinements) consider the credit limit, and no theory exists for the reissue period and promotional strategies.

ACKNOWLEDGMENT

We benefited from many useful discussions with Mike Connors, Gary Kochman, Brad Murphy, Deb Sarkar, and Gary Ulrich, who also made valuable suggestions on early drafts of this paper.

REFERENCES

ALTMAN, E. I. 1968. Financial Ratios, Discriminant Analysis, and the Prediction of Corporate Bankruptcy. *J. Fin.* **23**, 589–609.

ALTMAN, E. I. 1980. Commercial Bank Lending: Process, Credit Scoring, and Costs of Errors in Lending. *J. Fin. and Quant. Anal.* **15**, 813–832.

ALTMAN, E. I. (ED.). 1986. *Handbook of Corporate Finance*. John Wiley, New York.

ALTMAN, E. I., AND R. A. EISENBEIS. 1978. Financial Applications of Discriminant Analysis: A Clarification. *J. Fin. and Quan. Anal.* March, 185–195.

ANONYMOUS. 1979. Credit Scoring and the ECOA: Applying the Effects Test. *Yale Law J.* **88**, 1450–1486.

ANONYMOUS. 1986. Confessions of a Credit Card Fraud. *The Credit World* Sept./Oct., 32–36.

APILADO, V. P., D. C. WARNER AND J. J. DAUTEN. 1974. Evaluative Techniques in Consumer Finance—Experimental Results and Policy Implications for Financial Institutions. *J. Fin. and Quant. Anal.* March, 275–283.

ASPINWALL, R. C., AND R. A. EISENBEIS (EDS.). 1985. *Handbook for Banking Strategy*. John Wiley, New York.

AVERY, R. B. 1977. Credit Scoring Models With Discriminant Analysis and Truncated Samples. Unpublished paper.

AWH, R. Y., AND D. WATERS. 1974. A Discriminant Analysis of Economic, Demographic, and Attitudinal Characteristics of Bank Charge-Card Holders: A Case Study. *J. Fin.* **29**, 973–980.

BERANEK, W., AND W. TAYLOR. 1976. Credit-Scoring Models and the Cut-Off Point—A Simplification. *Dec. Sci.* **7**, 394–404.

BESTER, H. 1985. Screening vs. Rationing in Credit Markets With Imperfect Information. *Am. Econ. Rev.* **75**, 850–855.

BIERMAN, H., JR., AND W. H. HAUSMAN. 1970. The Credit Granting Decision. *Mgmt. Sci.* **16**, 519–532.

BOYES, W. J., D. L. HOFFMAN AND S. A. LOW. 1989. An Econometric Analysis of the Bank Credit Scoring Problem. *J. Economet.* **40**, 3–14.

BRAVOS, A. G. 1987. LEE: Loan Evaluation Expert, An Expert System for the Assessment of Loan Applications. *Proc. Expert Systems Day in New Jersey*, TIMS/ORSA.

BREIMAN, L., J. H. FRIEDMAN, R. A. OLSHEN AND C. J. STONE. 1984. *Classification and Regression Trees*. Wadsworth International Group, Belmont, California.

CAPON, N. 1982. Credit Scoring Systems: A Critical Analysis. *J. Market.* **46**, 82–91.

CHANDLER, G. G. 1985. Credit Scoring: A Feasibility Study. *Credit Union Exec.* **25**, 8–12.

CHURCHILL, G. A., JR., J. R. NEVIN AND R. R. WATSON. 1977. Credit Scoring—How Many Systems Do We Need? *The Credit World* **66**, 6–11.

CHANDLER, G. G., AND J. Y. COFFMAN. 1979. A Comparative Analysis of Empirical vs. Judgmental Credit Evaluation. *Retail Bank.* **1**, 15–26.

COFFMAN, J. Y. 1986. The Proper Role of Tree Analysis in Forecasting The Risk Behavior of Borrowers. *MDS Reports*, Management Decision Systems, Atlanta. 3, 4, 7, and 9.

COFFMAN, J. Y., AND G. G. CHANDLER. 1983. Applications of Performance Scoring to Accounts Receivable Management in Consumer Credit. Working Paper No. 46, Credit Research Center, Krannert Graduate School of Management, Purdue University, W. Lafayette, Indiana.

COFFMAN, J. Y., AND J. S. DARSIE. 1986. Collection Scoring: Effective Tool for Determining Collection Priorities. *The Credit World* **75**, 26–31.

COHEN, J. M. 1985. Handling the Quest for Credit. *Fin. Comput.* March/April, 28–32.

COHEN, K. J., AND F. S. HAMMER. 1966. Critical Comments on 'Measuring Risk on Consumer Installment Credit' by Paul F. Smith. *Mgmt. Sci.* **12**, 743–744.

COLE, R. H. 1988. *Consumer and Commercial Credit Management*. Irwin, Homewood, Illinois.

CONNORS, M. 1988. Credit Scoring: Aid to Efficiency and Professionalism. *Bus. Credit* **April**, 51–53.

CORCORAN, A. W. 1978. The Use of Exponentially-Smoothed Transition Matrices to Improve Forecasting of Cash Flows From Accounts Receivable. *Mgmt. Sci.* **24**, 732–739.

CRISCUOLI, E. J., JR. 1985. None of Your Business. *Assoc. Mgmt.* **37**, 63–66.

CYERT, R. M., H. J. DAVIDSON AND G. L. THOMPSON. 1962. Estimation of the Allowance for Doubtful Accounts by Markov Chains. *Mgmt. Sci.* **8**, 287–303.

DAVIS, D. B. 1987. Artificial Intelligence Goes to Work. *High Tech.* **April**, 16–17.

DAY, J. 1978. Customized Credit Scoring: The Seattle First System for Bank Card Applicants. *The Credit World* **66**, 12–14.

DIRICKX, Y. M. I., AND L. WAKEMAN. 1976. An Extension of the Bierman-Hausman Model for Credit Granting. *Mgmt. Sci.* **22**, 1229–1237.

DURAND, D. 1951. Risk Elements in Consumer Installment Financing. Study No. 8, National Bureau of Economic Research, New York.

DUTTA, S., AND S. SHEKHAR. 1988. Bond Rating: A Non-Conservative Application of Neural Networks. In *Proceedings IEEE International Conference on Neural Networks*, 443–450.

EBERHART, R. C., AND R. W. DOBBINS. 1990. *Neural Network PC Tools: A Practical Guide*. Academic Press, New York.

EISENBEIS, R. A. 1977. Pitfalls in the Application of Discriminant Analysis in Business, Finance, and Economics. *J. Fin.* **32**, 875–900.

EISENBEIS, R. A. 1978. Problems in Applying Discriminant Analysis in Credit Scoring Models. *J. Bank. and Fin.* **2**, 205–219.

EISENBEIS, R. A., AND R. B. AVERY. 1972. *Discriminant Analysis and Classification Procedures*. Lexington Books, Lexington, Mass.

FAIR, ISAAC COMPANIES CONFERENCE. 1988. Scoring: The Fourth Decade. San Francisco.

FITZPATRICK, D. B. 1976. An Analysis of Bank Credit Card Profits. *J. Bank Res.* **7**, 199–205.

FRYDMAN, H., J. G. KALLBERG AND D.-L. KAO. 1985. Testing the Adequacy of Markov Chains and Mover-Stayer Models as Representations of Credit Behavior. *Opns. Res.* **33**, 1203–1214.

GALLANT, S. I. 1988. Connectionist Expert Systems. *Commun. ACM* **31**, 152–169.

GALITZ, L. C. 1983. Consumer Credit Analysis. *Mgmt. Fin.* **9**, 27–33.

GENTRY, J. W. 1974. Discriminate Analysis in the Credit Extension Decision. *The Credit World* **63**, 25–28.

GIST, D. M. 1986. The Credit Card Game: The Right Controls Give Savings Institutions a Winning Hand. *Sav. Inst.* **107**, 90–94.

GREER, C. C. 1967. The Optimal Credit Acceptance Policy. *J. Finan. and Quan. Anal.* **Dec.**, 399–415.

GREER, C. C. 1968. Deciding to Accept or Reject a Marginal Retail Credit Applicant. *J. Retail.* **43**, 44–53.

HALL, J. H. 1983. An Argument Against Credit Scoring Systems. *Credit Mgmt.* **34**, 28–31.

HALDEMAN, R. G., JR. 1977a. ZETA: A New Scoring System. *Credit and Finan. Mgmt.* **79**, 14–15.

HALDEMAN, R. G., JR. 1977b. ZETA: A Credit Evaluation for Pension Managers. *Pension World* **13**, 28, 30–31.

HAMMER, J., AND B. SHENITZ. 1991. The New Giant on the Block. *Newsweek*, July 29, 1991.

HARTER, T. R. 1973. Measuring Risk on Mortgage Credit. *Mortg. Banker* **33**, 71–80.

HARTER, T. R. 1974. Potentials of Credit Scoring: Myth or Fact. *Credit and Finan. Mgmt.* **76**, 27–28.

HICKS, J. 1975. Tracking Debt Delinquents by Pre-Screening Applicants. *Direct Market.* **37**, 24–28.

HOWARD, R. A. 1960. *Dynamic Programming and Markov Chains*. MIT Press, Cambridge, Mass.

JARVIS, K. M. 1986. Smaller Scale Pre-Screening Options. *Credit* **12**, 34–36.

JENSEN, J. L. 1992. Using Neural Networks for Credit Scoring. *Manag. Fin.* **18**, 15–26.

JOY, O. M., AND J. O. TOLLEFSON. 1975. On the Financial Implications of Discriminant Analysis. *J. Finan. and Quant. Anal.* **10**, 723–739.

JOY, O. M., AND J. O. TOLLEFSON. 1978. Some Clarifying Comments on Discriminant Analysis. *J. Finan. and Quant. Anal.* **March**, 197–199.

KANE, A. T. 1982. Developing New Accounts. *Retail Control* **4**, 34–41.

KAYE, G. H. 1981. Behavior Scoring—Bringing the Future into Focus. *Credit World* **69**, 10–15.

KELLEY, A. A. 1986. Business Credit Skills: Part 2—Ratios and Decision Making Models. *Credit and Finan. Mgmt.* **88**, 9–14.

KLINGEL, A. R., AND M. M. PRESS. 1976. Experiments in a Credit Card Center. *Interfaces* **6**, 38–44.

KOLESAR, P., AND J. L. SHOWERS. 1985. A Robust Credit Screening Model Using Categorical Data. *Mgmt. Sci.* **31**, 123–133.

LACHENBRUCH, P. A. 1975. *Discriminant Analysis*. Hafner Press, New York.

LACHENBRUCH, P. A., AND J. R. MICKEY. 1968. Estimation of Error Rates in Discriminant Analysis. *Technometrics* **February**, 1–11.

LANE, S. 1972. Submarginal Credit Risk Classification. *J. Finan. and Quant. Anal.* **7**, 1379–1385.

LIEBMAN, L. H. 1972. A Markov Decision Model for Selecting Optimal Credit Control Policies. *Mgmt. Sci.* **18**, 519–525.

LONG, M. S. 1976. Credit Screening System Selection. *J. Finan. and Quant. Anal.* **11**, 313–328.

LONG, M. S., AND J. J. McCONNELL. 1977. Credit Scoring System Aids Second Mortgage Makers. *Mortg. Banker* **37**, 29–34.

LOVIE, A. D., AND P. LOVIE. 1983. The Flat Maximum Effect and Linear Scoring Models for Prediction. *J. Forecast.* **5**, 159–168.

MAIN, J. 1977. A New Way to Score With Lenders. *Money* **February**, 73–74.

MAKOWSKI, P. 1985. Credit Scoring Branches Out. *The Credit World*, **75**, 30–37.

MATTHEWS, J. 1985. Credit and Credit Cards. *Credit Mgmt.* **November**, 28–31.

MAVES, G. 1991. Perfecting Prediction. *Marketing* **August**, 24–26.

McCULLAGH, P., AND J. A. NELDER. 1983. *Generalized Linear Models*. Chapman and Hall, New York.

MEHTA, D. 1968. The Formulation of Credit Policy Models. *Mgmt. Sci.* **15**, 30–50.

MEHTA, D. 1970. Optimal Credit Policy Selection: A Dynamic Approach. *J. Finan. and Quant. Anal.* **5**, 421–444.

MITCHNER, M., AND R. P. PETERSON. 1957. An Operations Research Study of the Collection of Defaulted Loans. *Opns. Res.* **5**, 522–545.

MORRISON, D. G. 1969. On the Interpretation of Discriminant Analysis. *J. Market. Res.* **6**, 156–163.

MOSES, D., AND S. S. LIAO. 1987. On Developing Models for Failure Prediction. *J. Commer. Bank Lend.* **69**, 27–38.

MYERS, J. H., AND E. W. FORGY. 1963. The Development of Numerical Credit Evaluation Systems. *J. Am. Statis. Assoc.* **58**, 799–806.

NELSON, O. D. 1983. Credit Scoring: Do AFSA Members have Faith in its Answers. *Credit* **9**, 14–16.

NELSON, M. M., AND W. T. ILLINGWORTH. 1990. *A Practical Guide to Neural Nets*. Addison Wesley, New York.

NIKBAKHT, E., AND M. H. A. TAFTI. 1989. Application of Expert Systems in Evaluation of Credit Card Borrowers. *Manag. Fin.* **15**, 19–27.

ORGLER, Y. E. 1970. A Credit Scoring Model for Commercial Loans. *J. Money, Credit & Bank.* **2**, 435–445.

PANIELLO, R. J. 1976. Scoring and Authorization in a Successful Credit Operation. *Retail Control* **45**, 21–31.

PIKETTY, L., ET AL. 1987. The Authorizer's Assistant: A Large Commercial Expert System Application. In *Proceedings Expert Systems Day* in New Jersey. (TIMS/ORSA), Nov. 5.

PYE, G., AND A. TEZEL. 1974. Optimal Foreclosure Policies. *Mgmt. Sci.* **21**, 141–147.

RAIFFA, H., AND R. SCHLAIFER. 1961. *Applied Statistical Decision Theory*. Colonial Press, Mass.

REICHERT, A. K., C.-C. CHO AND G. M. WAGNER. 1983. An Examination of the Conceptual Issues Involved in Developing Credit-Scoring Models. *J. Bus. and Econ. Statis.* **1**, 101–114.

ROSSI, H. W. 1982. Improving Bottom-Line Performance Through Credit Management and Screening. *Zip* **5**, 60–63.

SAVERY, B. A. 1977. Numerical Points Systems in Credit Screening. *Mgmt. Dec.* **15**, 36–50.

SCHERR, F. C. 1982. Estimating and Using Failure-Forecasting Functions: Some Problems and Some Proposed Solutions. *Baylor Bus. Studies* **12**, 19–33.

SCOTT, E. 1978. On the Financial Applications of Discriminant Analysis: Comment. *J. Finan. and Quant. Anal.* **March**, 201–205.

SHOWERS, J. L., AND L. M. CHAKRIN. 1981. Reducing Uncollectible Revenue From Residential Telephone Customers. *Interfaces* **11**, 21–31.

SMITH, P. F. 1964. Measuring Risk on Consumer Installment Credit. *Mgmt. Sci.* **11**, 327–340.

SPARKS, D. L. 1979. Credit Scoring: A Banker's Tool. *U.S. Banker* **90**, 32–33, 37.

SRINIVASAN, V., AND Y. H. KIM. 1987. The Bierman-Hausman Credit Granting Model: A Note. *Mgmt. Sci.* **33**, 1361–1362.

STIGLITZ, J., AND A. WEISS. 1981. Credit Rationing in Markets With Imperfect Information. *Am. Econ. Rev.* **71**, 393–410.

STONE, B. K. 1972. The Cost of Bank Loans. *J. Finan. and Quant. Anal.* **7**, 2077–2086.

UPDEGRAVE, W. L. 1987. How Lenders Size You Up. *Money* **16**, 145–154.

VAN KUELEN, J. A. M., J. SPRONK AND A. W. CORCORAN. 1981. On the Cyert-Davidson-Thompson Doubtful Accounts Model. *Mgmt. Sci.* **27**, 108–112.

WAGNER, G. M., A. K. REICHERT AND C.-C. CHO. 1983. Conceptual Issues in Credit Scoring Models. *Credit World* **71**, (May/June) 22–25 (part 1), (July/August) 22–28, 41 (part 2).

WALL STREET J. 1991. Consumer Debt, January 8, 1991, p. 1.

WEINGARTNER, H. M. 1966. Concepts and Utilization of Credit-Scoring Techniques. *Banking* **February**, 51–53.

WIGINTON, J. C. 1980. A Note on the Comparison of Logit and Discriminant Models of Consumer Credit Behavior. *J. Finan. and Quan. Anal.* **15**, 757–768.

WILSON, R. L., AND R. SHARDA. 1991. Bankruptcy Prediction Using Neural Networks. *Dec. Support Syst.* (to appear).

ZANAKIS, S. H., L. P. MAVRIDES AND E. N. ROUSSAKIS. 1986. Applications of Management Science in Banking. *Dec. Sci.* **17**, 114–128.

ZOCCO, D. P. 1985. A Framework for Expert Systems in Bank Loan Management. *J. Commer. Bank Lend.* **67**, 47–54.