THE DECISION TO GRANT CREDIT

by

Brian Edward Boyle

B.A., Amherst College
(1969)

S.B., S.M., E.E., Massachusetts Institute of Technology
(1971)

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY IN
OPERATIONS RESEARCH

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May, 1974

Signature of Author...
Department of Electrical Engineering, May 3, 1974

Certified by...
Thesis Supervisor

Accepted by...
Chairman, Departmental Committee on Graduate Students

THE DECISION TO GRANT CREDIT

by

Brian Edward Boyle

Submitted to the Department of Electrical Engineering on May 3, 1974 in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Operations Research.

ABSTRACT

This thesis describes a set of decision models designed to aid a consumer lending institution in performing its credit granting function. The four models presented incorporate most of the factors relevant to the decision to grant credit. In order to support this contention, decision rules are formulated for both installment loan and revolving credit instruments. In addition, the installment loan formulation is applied to actual loan data supplied by the National Shawmut Bank of Boston. Model performances are compared with an existing credit scoring system and found to result in significantly more profitable decisions.

In the course of this research, particular emphasis is placed on the theoretical issues of the underlying pattern recognition and multivariate estimation problems. An information maximizing procedure for feature selection is presented that requires no assumptions about the multivariate probability distribution of feature vectors. A number of pattern recognition algorithms of varying complexity are evaluated within the context of default probability estimation. Their performances are presented both in terms of probability of misclassification and economic expected net present value.

The descriptive Markov process model of delinquent payment behavior requires the estimation of a transition probability matrix $\underline{\underline{P}}(\underline{x})$ as a function of a feature vector $\underline{x}$. This novel estimation problem, which has application in other fields of interest, is addressed in detail. It is shown to be solvable by general multivariate estimation techniques.

The results obtained here suggest that multiperiod decision rules with Bayesian probability updating and detailed outcome spaces can be designed to yield improved performance

over "state-of-the-art credit scoring" decision rules. Moreover, these decision models provide the necessary framework for evaluating policy level decisions. Finally, this class of decision models is shown to be applicable to problems in areas other than credit granting.

THESIS SUPERVISOR: G. Anthony Gorry
TITLE: Associate Professor of Electrical Engineering

ACKNOWLEDGMENTS

I wish to express my gratitude to Professors G. Anthony Gorry and Warren H. Hausman for their continued advice and encouragement throughout the course of this thesis. Their enthusiastic support contributed immeasurably to my efforts. I also would like to thank the other members of my committee, Professors Richard C. Larson and Ian T. Young for their constructive and illuminating criticism of my research.

I am especially indebted to Mr. Charles A. Hunt and Mr. Albert J. Baillargeon of the National Shawmut Bank of Boston. Their interest and support of my work far exceeded the computer time and data which they made available.

Mrs. Rita Hardy and Ms. Sue Sager typed the drafts and final copy of this thesis, and to them I owe particular thanks.

Finally, I wish to acknowledge the support of the National Science Foundation and the M.I.T. Research Laboratory of Electronics.

B.E.B.
Cambridge, Massachusetts
May, 1974

TABLE OF CONTENTS

	Page
Abstract	2
Acknowledgements	4
1. Introduction	10
PART I - Model Development	19
2. Previous Research	20
3. Credit Granting Decision Models	27
3.1 Model 1 - Single Period, Two Outcome	29
3.2 Model 2 - Multi-period, Two Outcome	33
3.3 Model 3 - Single Period, Detailed Outcome	40
3.4 Model 4 - Multi-period, Detailed Outcome	45
3.5 Characteristics of the Models	49
3.5.1 Models 1 and 2	51
3.5.2 Models 3 and 4	61
4. Pattern Recognition Techniques for Default Probability Estimation	78
4.1 Problem Statement and Complexity Considerations	78
4.2 Pattern Recognition Algorithms	83
4.3 Evaluating Algorithm Performance	92
5. Pattern Recognition Techniques for Transition Probability Estimation	96
PART II - Application to Installment Loan Credit Granting	103
6. Introduction to the Case Study	104
6.1 Requirements for Application	105
6.2 Data Description	108
7. Feature Selection - Theory and Application	116
8. Pattern Recognition Results for Default Probability Estimation	129

		Page
9.	Model 1 Results	142
9.1	Estimating the A Priori Probability of Default	143
9.2	Benchmark Results	144
9.3	Model 1 Pattern Recognition Results	158
10.	Model 2 Results	166
11.	Transition Probability Estimation	177
11.1	First Order Test	177
11.2	Stationarity Test	186
11.3	The A Priori Transition Probability Matrix	190
11.4	Pattern Recognition Application to Transition Probability Estimation	194
12.	Model 3 Results	201
13.	Model 4 Results	210
13.1	Model 4A Description	211
13.2	Model 4A Results	217
14.	Case Study Summary	222
PART III - Further Discussion of the Models		228
15.	Sampling from a Pre-Screened Population	229
16.	Generalization to New Credit Instruments	238
17.	Implications of the Models	246
17.1	Operational Credit Policy	247
17.2	Organizational Decision Making	251
17.3	Credit Instruments from the Consumer's Viewpoint	254
17.4	Application to Other Problem Areas	257
References		262
Appendix A - Loan Profitability Analysis		266
Appendix B - Feature Descriptions		279
Biographical Note		282

LIST OF FIGURES

		Page
3.1	Single Period Model	32
3.2	Multi-Period Model	34
3.3	Typical Payment/Delinquency State Description	41
3.5.1	4 State Description	63
3.5.2	Addition of a Delinquent State	71
3.5.3	Expansion of the On-Time State	75
4.1	Densities of Varying Complexity and Interleavedness	80
8.1	Plot of Reward Ranges - Maximum Likelihood Decision	139
9.2.1	Cumulative Distribution of Credit Scores	152
9.2.2	Benchmark Performance on Data Set A2	156
9.3.1	Plot of Reward Ranges - Model 1 Decision	164
10.1	A Priori Default Probability Distribution	170
11.1	6 State Description	178
11.2	7 State Description	182
11.3	13 State Description	184
13.1	Example of Model 4A	219
15.1	Decision Making on Pre-Screened Samples	231
15.2	Model Performance Adjusted for Screening	235
A.1	Distribution of Recovery Revenues	277

LIST OF TABLES

		Page
3.5.1	Typical Parameter Values and Ranges	57
3.5.2	Typical Expected Net Present Values of J Loans	58
3.5.3	Sensitivity of Models 1 and 2 to Parameter Changes	59
3.5.4	Typical Eigenvalues of the Delinquent Process	73
6.1	List of Features in Data Set A	109
6.2	List of Features in Data Set B	113
7.1	Information Content of Each Feature in Data Set A	121
7.2	Feature Ordering Obtained Using Second Order Information	124
7.3	Features Selected Using Second Order Information	125
7.4	Feature Orderings Obtained Using Second Order Information and Different Starting Features	127
7.5	Feature Order Using t-Statistic	128
8.1	Pattern Recognition Performance on Data Set A	134
9.2.1	Reward Function $R(\theta,\hat{\theta})$	149
9.2.2	Benchmark Performance on Data Set A1 as a Function of Cutoff Score	150

		Page
9.2.3	Sensitivity of Cutoff Score to A Priori Probability of Default	155
9.2.4	Cuttoff Score Optimization Using Data Sets A1 and A2	157
9.3.1	Model 1 Performance on Data Set A2	160
9.3.2	Model 1 Performance on Data Set A1	163
10.1	Bayesian Default Probability Updating	172
10.2	Model 2 Performance on Data Set A2	174
10.3	Expected Net Present Values for One Loan and Five Loans	176
11.1	6 State Model Transition Count Matrix	180
11.2	A Priori Transition Probability Matrix	193
11.3	Discriminant Function Matrix (ψ=2)	199
11.4	Discriminant Function Matrix (ψ=10)	200
12.1	Model 3 Performance on Data Set B1	208
13.1	Model 4A Performance on Data Set B1	221
A.1	Proportion of LRS Time Spent on Loan of Average Age t	274
A.2	Proportion of Loan Recovered	275

Chapter 1

INTRODUCTION

Consumer credit evaluation, or _credit scoring_ as it is sometimes popularly called, has been a topic of interest to the OR/MS community for the past ten years. As a result of the increasing availability to the consumer of a number of forms of credit and the concurrent widening appeal of management science techniques, credit managers have begun to view quantitative credit evaluation tools as both helpful and necessary.

Cole [6] provides a comprehensive description of the consumer credit function of a lending institution. The description that follows will serve to define the major aspects of the consumer credit function as well as to provide a standard vocabulary for the remainder of this thesis.

Initially, the customer submits a completed application which contains information relevant to the credit granting decision. This information usually includes demographic information (e.g., age, address, occupation), income information, and other information such as years at present address, car ownership, etc. In addition, past and current credit history is included, as well as a mention of other accounts (checking, savings, etc.) he may have with the institution.

The application is usually made for a particular credit instrument (installment loan or revolving credit) and for a given loan amount or credit line. The term installment loan includes any loan where the loan amount is given to the customer who then makes equal monthly payments over the term of the loan. The interest charged depends on the actual type of loan, e.g., signature loans generally are made at higher rates than home improvement or car loans. The term revolving credit includes charge card accounts (e.g., Master Charge, travel cards, gasoline credit cards, etc.). The customer is permitted to make charge purchases up to a predetermined credit limit or credit line. At any time he may make payments to reduce the amount of his outstanding balance at a rate somewhat above that for installment loans.

Given this initial information on the application, a lending officer will conduct whatever credit investigation seems appropriate. He then makes a decision to accept or reject the application based on application and credit investigation information and any additional information he may have. Hereafter, this additional information, since it is not quantified in the application process, will be called <u>subjective information</u>.

Generally, if the application is rejected, the consumer will be lost as a customer, as he will probably find credit elsewhere. At the very least he will be ill-disposed to further any relationship with the institution that rejected his application. If the application is accepted, the customer begins to make the appropriate monthly payments. Should he not make a payment by the prescribed due date, he is considered a one month <u>delinquent</u> account. Should no payment be made by the following month, he is considered two months delinquent, and so on.

When an account becomes delinquent, affirmative collection action is taken to bring it back to a paying or on-time state. Depending on the amount of time delinquent, this action may take the form of a letter, telephone call, personal visit, or legal action. An account which is frequently delinquent can generate more collection expense than the interest received from the account.

Should the account become seriously delinquent (3 to 6 months, depending on accounting practices), it is usually considered to have defaulted and is charged against a reserve for bad debts. The account is nonetheless pursued and on the average about one-fourth to one-half of the charged-off balance is eventually recovered.

Assuming the account does not default, its payment history is periodically reviewed. In the case of installment loans, this review does not occur until the customer applies for a subsequent loan. In the case of revolving credit accounts, this review comes at a predetermined interval. At this time a second decision to grant credit is made. However, here the decision can be made with the additional information of the first loan or first review period actual payment history.

Each credit granting decision affects the customer's use of related services of the institution. For example, if a customer's application is accepted, he will be more likely to open a new savings account at a bank than if his application is rejected. This cross-selling effect plays a significant role in the credit granting decision.

Another often neglected aspect of the process is the evaluation of revenue from revolving credit accounts. The credit line simply places an upper limit on the amount of the outstanding balance. Revenue is generated on the actual average outstanding balance, which, although related to credit line, is more a function of the customer's credit attitudes and needs. For example, it is possible to segment the credit card market by social class and income levels, which, in turn, will be useful in identifying some of the behavioral characteristics of customers in each market segment. Given these characteristics (e.g., degree of apathetic attitude toward shopping, perceived risk in buying decisions, etc.), we can attempt to predict actual credit card <u>usage</u>.

Outline of the Material That Follows

This thesis describes a general quantitative approach to the decision to grant credit. The models presented were developed with a view toward incorporating the factors most relevant to the decision. Among the factors included are the revenues and costs associated with uncertain outcomes of both the loan applied for and subsequent loans. Loan outcomes are first considered to be either default or non-default and are later expanded to include a detailed description of delinquent behavior. Several pattern recognition techniques are used to estimate outcome probabilities as a function of a set of features or attributes describing the applicant. Loan characteristics (term, interest rate, etc.) are used to parametrically determine rewards for correct classification. The resulting credit granting decision rules have as their objective the maximization of expected net present value.

The thesis is presented in three parts. Part I, which includes Chapters 2 through 5, develops the decision models and the accompanying estimation theory.

Chapter 2 examines some of the published literature in the area that is relevant to the credit granting decision.

Chapter 3 presents four models for the credit decision. Model 1 considers only the initial loan, and only the two outcomes of default and non-default. Model 2 considers the same two outcomes but incorporates the probable effects of subsequent loans. Model 3 expands the outcome space to include delinquent payment behavior for the one loan case. Finally Model 4 presents a theoretical treatment of Model 3 extended to include subsequent loans.

Pattern recognition techniques for default probability estimation are discussed in Chapter 4. A set of algorithms of varying complexity is suggested.

Chapter 5 considers the general problem of estimating the transition probability matrix of a first-order stationary Markov chain as a function of a vector of features. The pattern recognition approach to this problem is necessary for the application of Models 3 and 4.

Part II, including Chapters 6 through 14, presents the results of an application of the models for installment loan credit granting. The data for this case study were supplied by The National Shawmut Bank of Boston (NSB). The analysis of Part II serves as an outline of the steps to be followed to implement the set of decision models. In particular, Chapter 7 proposes an information theoretic approach to feature selection.

Chapters 8 and 9 present the results of a pattern recognition approach to default probability estimation. A significant effect of subsequent loan revenues on the credit granting decision is suggested by the analysis of Chapter 10.

Chapters 11 and 12 demonstrate the feasibility of estimating and applying a detailed state outcome decision model. In the course of estimating the transition probability matrix, a sequential state expansion method for removing second and higher order memory within a Markov process is suggested. A heuristic solution to the detailed-outcome multi-loan decision model (Model 4) is presented in Chapter 13, with a discussion of its performance on actual loan cases.

Part III reconsiders the decision to grant credit in the light of the models developed in Part I and the insights gained through the empirical analysis of Part II. Chapter 15 proposes a method that permits the evaluation of alternative decision rules without having to observe the unknown outcomes of previously rejected loans. Chapter 16 demonstrates the adaptability of the models by formulating the revolving credit decision problem in terms of Model 4. This formulation provides the necessary framework for answering the related question of setting credit limits and determining account review periods. The implications of the models for operational and organizational policy decisions are discussed in Chapter 17. A case is made for variable interest rates on consumer loans based on the quantitative assessment of default risk. Finally, the models presented in the thesis are suggested to have applications for decision making in areas other than consumer credit.

PART I

MODEL DEVELOPMENT

Part I presents a set of general models for consumer credit decision making. The relationship of the proposed models to previous results in the area is given in a review of the literature. The models require the estimation of outcome probabilities as a function of applicant attributes. These multivariate estimation problems are addressed in detail.

Chapter 2

PREVIOUS RESEARCH

A number of previous works relating to the decision to grant credit can be found in the MS/OR literature. Most are addressed to a single aspect of the problem, such as the initial decision, the delinquent account process, or the pursuit of defaulted accounts.

The majority of those works dealing with the initial credit granting decision have been largely concerned with discriminating between "good" (non-defaulting) and "bad" (defaulting) credit risks. Myers and Forgy [21] use both discriminant and regression analysis to predict the final state of the account, but do not take costs and revenues into account for selecting an economically optimal trade off between Type I and Type II errors. Bogess [2] and Weingartner [25] give less technical descriptions of such a linear discrimination procedure and indicate a simplified means of optimizing the trade off between rejecting a good account and accepting a bad one. They do not, however, consider the effect of delinquency behavior, post-default recoveries, or potential profit from subsequent loans.

Smith [24] suggests the use of linear discriminant analysis to develop what he calls a "risk-index". The analysis is conducted marginally on a number of applicant variables, and the risk index is then defined to be the sum of the posterior probabilities of default from each variable. In a critical comment on Smith's paper, Cohen and Hammer [5] point out that this computation is not justified since the variables are not equally important as implicitly assumed. They also comment on Smith's failure to validate his estimates on an independent data sample.

Majone [15, 16] reviews the assumptions implicit in linear discriminant analysis, and indicates that the assumption of normally distributed random variables with common covariance matrices is probably not appropriate in the credit-scoring situation. Chatterjee and Barcun [3] employ a nearest-neighbor rule to perform the good/bad loan discrimination based on eight dichotomous variables and a sample of 774 applications. The economics of the situation are inadequately summarized in a single parameter, the ratio of misclassification costs.

Orgler [22] extends the credit-scoring concept to the review of existing accounts. In addition to the variables available initially, he includes the number of overdue notices, number of early or late payments, and a three-valued subjective evaluation of the payment performance. The probability of default is then estimated by multiple regression, but Orgler does not indicate how this is to be used in the review process.

Several works are addressed to the description of the delinquent account process. Cyert, Davidson and Thompson [9] and Cyert and Thompson [10] set forth a Markovian model. Customers are assumed to move among several delinquent states according to a stationary transition probability matrix. They allow for different customer types by defining several different transition matrices and indicate that "A potential technique for determining which risk category a credit applicant should go into is multiple regression". Liebman [14] describes a similar model, but his concern is more with the optimal action to be taken in each delinquent state (e.g., send letter, telephone call, etc.).

Mitchner and Peterson [20] present the results of a study on the recovery of defaulted loans. They derive results which indicate the optimal time for which non-paying accounts should be pursued.

Greer [13] presents a model which allows for the inclusion of future profits after initial credit granting, but this is a function of the number of applicants accepted and assumes that a ranking of customers from good risks to bad risks is available. He also allows the expected total collection expense to be a function of the number of applications accepted.

The value of initial information is also discussed by Greer [12]. A related paper by Mehta [19] derives sequential sampling rules for credit investigation. These works address an interesting although relatively minor aspect of the credit granting decision problem.

Bierman and Hausman [1] present a multi-period analysis of the decision problem which contains a Bayesian updating of default probability as payment experience is gained. They present a dynamic programming formulation of the multi-period problem, and indicate the importance of considering more than a single time period. Considerable information about the applicant's payment behavior is gained during the first period, which, by resolving some of the uncertainty about the applicant, places the institution in a potentially more profitable position for the second period decision.

The usage aspect of revolving credit is discussed in Mathews and Slocum [17, 18, 23], Curtis [8] and Fazio [11]. These works use factor analysis and multiple regression to perform credit and market segmentation and then predict actual credit card usage as a function of an applicant's credit attitudes.

Crane [7] addresses the issue of bank service interaction. He presents a Markov model which describes the probability that the next service applied for will be of a certain type given the last service type. The model as it is presented has several shortcomings, but it serves to point out the importance of interaction among different services.

In summary, previous papers in this field have generally been directed at a single important factor in the credit granting decision. The development of credit scoring formulas has received much attention, but has been limited in application because of assumptions of linearity, inadequate modeling of the economics of the process, the lack of attention to the effects of delinquency behavior, and the failure to consider potential profit from subsequent loans and other bank services. Consequently, several papers have been addressed to these shortcomings.

In particular, one paper has considered the multi-period aspect of the problem after recognizing that significant information about the applicant's payment behavior is gained after each loan period. Several others have modeled the delinquency and post-default recovery aspects, but still leave unanswered the problems of classifying potential customers into delinquency classes and then estimating their transition probabilities among delinquent states. Several authors have investigated with some success the sub-problem of trying to predict actual credit card usage. Finally, this author is aware of one article which tries to model the interaction or cross-sell effect among the services offered by the lending institution, of which the consumer loan is a major one.

Chapter 3

CREDIT GRANTING DECISION MODELS

Model Development

The major aspects of the credit granting decision process are considered to be

1) the initial prediction of default as a function of the applicant's attributes, with the inclusion of associated costs and revenues due to the loan itself;

2) the potential profit from subsequent loans;

3) the prediction of an applicant's delinquency behavior and the inclusion of delinquency related costs; and

4) the combined effect of considering both subsequent loans and delinquency behavior.

In this part of the thesis these aspects will be incorporated into a set of models which represent a significant advance over the models described in Chapter 2. In Part II, the models will be tested at The National Shawmut Bank of Boston (NSB) through the solution of several estimation and classification problems. The nature of these problems will become apparent as the models are discussed in detail. In Part III, we will discuss the more general applications of these types of decision models.

The models will be developed in four increasing stages of complexity.

1. Initially, a single period binary outcome (default/ non-default) model will serve as a basis for further extension.

2. Then a multi-period binary outcome model will be described to incorporate the expected net present value of second and subsequent period outcomes.

3. The third stage of development will replace the two-valued outcome of default/non-default with a state transition history of payment and delinquency behavior. These states will be of the form on-time, 1 month delinquent, 2 months delinquent, etc. At this stage, the state transition model and the associated estimation problems will be discussed in detail.

4. Finally, at the fourth stage, the probability updating aspects of the multi-period case will be discussed in relation to the state transition description of possible outcomes. Although this final stage model captures the important factors of the problem, it will be computationally necessary to develop an approximate solution for practical decision making.

3.1 Model 1 - Single Period, Two Outcome

This first stage model is structurally similar to the two-outcome credit scoring models found in the literature. However, consideration will be given to the nonlinear nature of the classification problem as well as to the profit implications of default and non-default.

A loan applicant will be characterized by a vector $\underline{x}$ of attributes or "features", which has components age, income, years at current address, etc. The loan amount or credit line applied for will be denoted as A. If the loan is rejected, then no loss due to non-payment is possible. On the other hand, if the loan is accepted, the customer will either eventually pay the loan or will default after having paid less than the full amount of the loan. The probability of default is assumed to depend on both the applicant's features $\underline{x}$ and the loan amount A, and will be expressed as $p(\underline{x},A)$.

Given the customer pays the loan, the net present value of his payments will be denoted by $V_1(A,\underline{L})$. The functional dependence on the loan amount or credit line is obvious. The loan vector $\underline{L}$ is included to provide further description of the loan itself, to include the term of the loan (T), the interest rate (r), the expected credit line usage (for revolving credit), etc.

Given that the customer defaults, the net present value of this outcome will be denoted by $V_0(A,\underline{L})$. For example, suppose $\underline{L}$ describes a $2000 installment loan for 24 months at 13.5% interest and we know that historically default occurs (given that it does occur) on the average after about 11 months. Then by discounting the appropriate cash flows at the relevant cost of capital, we might find the expected net present value to be approximately $V_0(A,\underline{L}) = -\$700$.

Thus, apart from future loan considerations and interactions with other services, the expected net present value of accepting the loan is:

$$\hat{V} = p(\underline{x},A)\ V_0(A,\underline{L}) + [1 - p(\underline{x},A)]\ V_1(A,\underline{L}). \qquad (3.1.1)$$

Of course, if the loan is rejected, $\hat{V} = 0$.

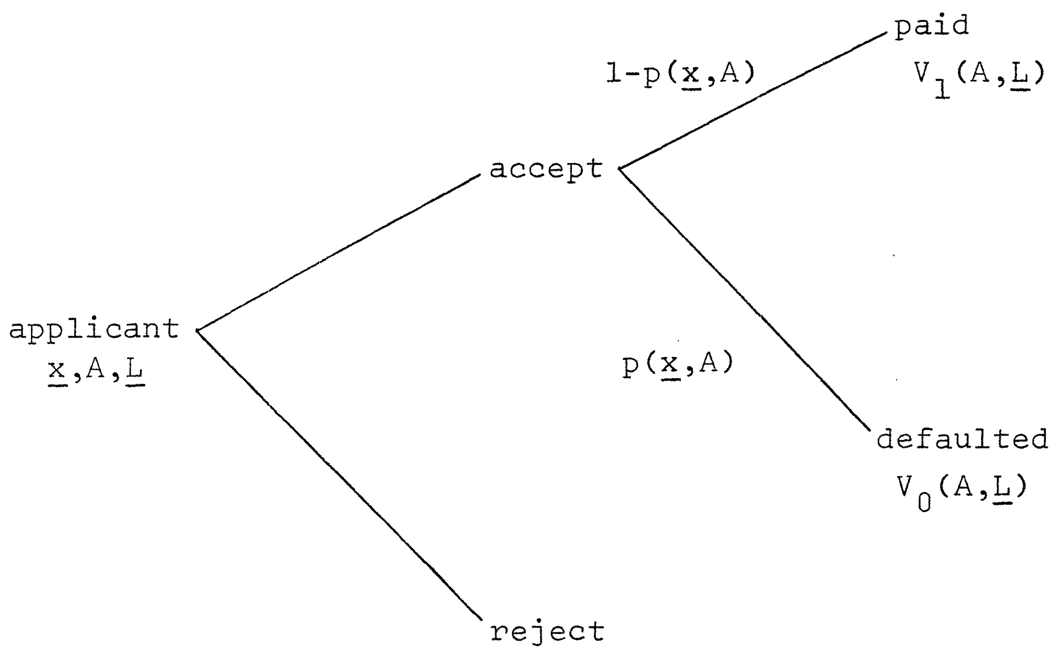
Given this model, the one-loan or one-period decision rule should be:

$$\text{Accept if and only if } \hat{V} > 0. \qquad (3.1.2)$$

Determination of $V_0(A,\underline{L})$ and $V_1(A,\underline{L})$ is essentially straightforward in the case of installment loans (see Appendix A). For revolving credit, however, usage must be predicted, which poses an interesting estimation problem but one that will not be addressed in the thesis. The other significant problem is the estimation of $p(\underline{x},A)$. Since there are strong a priori reasons to believe that default probability may not be linear in $\underline{x}$ and A, this problem will be formulated in

Chapter 4 as a nonlinear statistical classification (pattern recognition) problem. In the case study, the results of this approach will be compared with those obtained by a multiple regression benchmark model in use at NSB.

FIGURE 3.1
Single Period Model



3.2 Model 2 - Multi-Period, Two Outcome

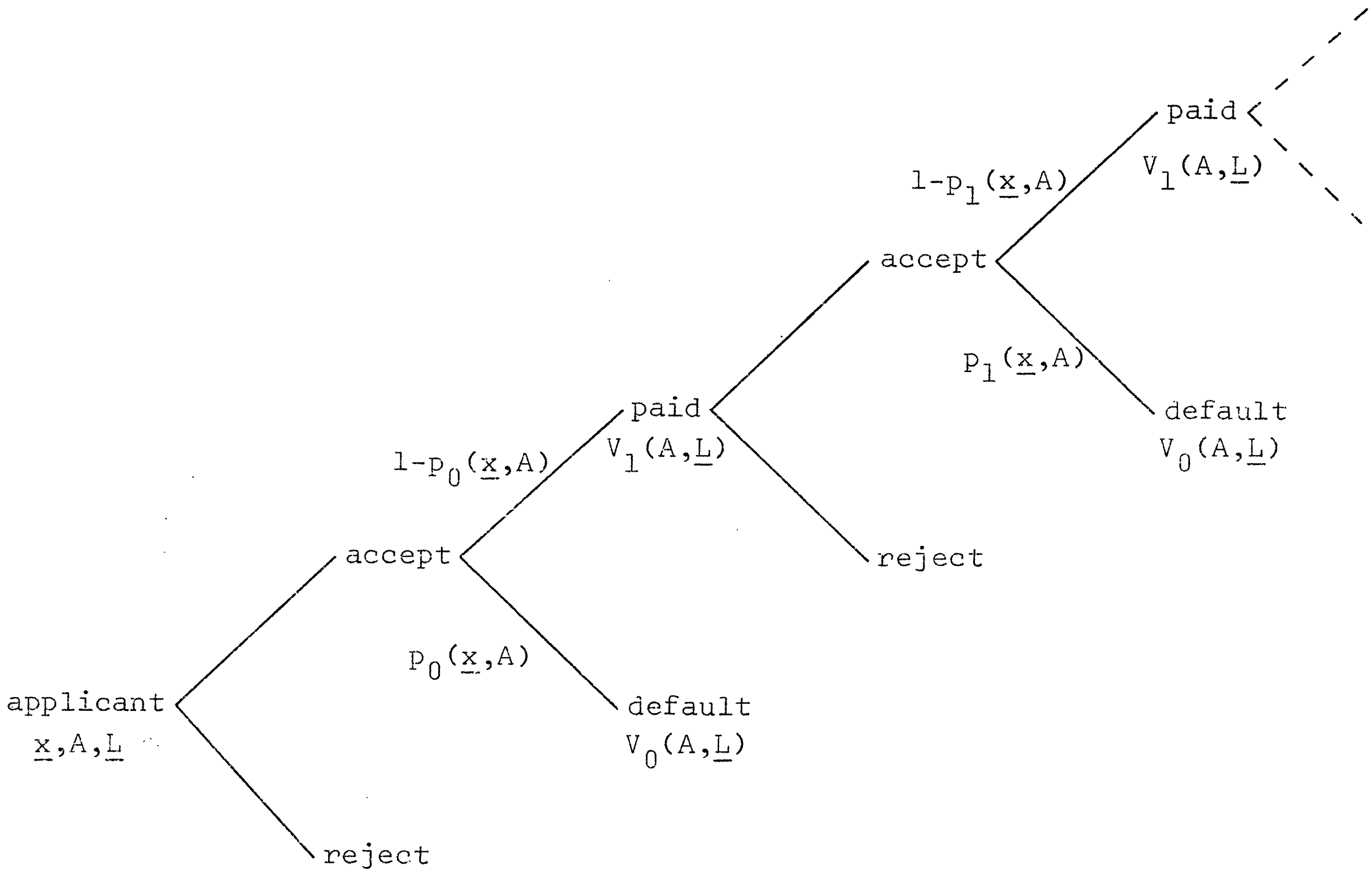
The single period model of Section 3.1 fails to account for the present value of loans that may be made subsequent to the one applied for. If the first installment loan is successfully paid (or if the revolving credit payments are satisfactorily made during the first review period), the lending institution will be in a better position (i.e., a more profitable posture) to re-extend credit for the second period. This reflects the updating of the initial default probability estimate on the basis of the first period outcome. We will assume that once default occurs no further credit will be granted.[1]

In order to extend the single period model to the multi-period case, we need to specify a probability updating rule. Let $p_j(\underline{x},A)$ be the probability of default in period j (the $(j+1)^{st}$ loan) given no defaults in periods 0,1,...,j-1. Thus, the multi-period model requires the additional estimation of second and subsequent period conditional default probabilities. These could be either estimated directly if sufficient data were available (it is usually not available), or estimated

[1]The assumption is justified both by a quantitative analysis in Chapter 10 and by the actual lending policy at most financial institutions.

FIGURE 3.2

Multi-Period Model



via a probability updating rule.[2]

The rule that will be proposed is similar to that of Bierman and Hausman [1]. The initial default probability estimate, $p_0(\underline{x},A)$, will be estimated by a nonlinear pattern recognition technique. A beta natural conjugate prior[3] will be assumed with parameters $r_0(\underline{x},A)$ and $n_0(\underline{x},A)$ such that $r_0(\underline{x},A)/n_0(\underline{x},A) = p_0(\underline{x},A)$. That is, the pattern recognition approach will specify the ratio r/n, leaving only the "diffuseness" (n) of the prior to be determined. To make this determination, the value of actual payment information relative to application information could be investigated. In the absence of such relative information, the diffuseness parameter (n) can be approximately determined by interviewing experienced lending personnel and then testing the sensitivity of the resulting decision rules to different values of a constant n.

Given the above formulation, an optimal decision rule can be found by considering a finite horizon and then solving the appropriate dynamic programming problem.

---

[2]Because our expected net present value criterion includes a discounting operation with discount factor of about 0.9 annually (or about 0.75 for a second loan applied for three years later), accuracy in $p_0(\underline{x},A)$ will be more important than in $p_1(\underline{x},A)$, $p_2(\underline{x},A)$, etc.

[3]See Raiffa and Schlaifer [36] for a discussion of Bayesian probability updating.

Let θ be our decision variable, i.e., $\hat{\theta} = 0$ means do not grant credit, $\hat{\theta} = 1$ means grant credit. If $\hat{V}_J$ is the expected net present value of the next J loans, then our decision rule becomes:

$$\hat{\theta} = \begin{cases} 0 \text{ (reject) if } \hat{V}_J \leq 0 \\ 1 \text{ (accept) if } \hat{V}_J > 0. \end{cases} \tag{3.2.1}$$

$\hat{V}_J$ can be determined in a manner described below.

Let $p_j(\underline{x},A)$ be the probability that the customer defaults in period j <u>given</u> he did not default on the previous j loans. Actually, $p_j(\underline{x},A)$ is considered to be the mean $r_j(\underline{x},A)/n_j$ of the beta prior density describing the updated default probability after j repaid loans, where $r_j(\underline{x},A)$ and n_j are the parameters of the beta density for the random variable $p_j(\underline{x},A)$. Initially, we estimate $p_0(\underline{x},A)$ using pattern recognition techniques. The diffuseness parameter, n_0, for $p_0(\underline{x},A)$ is assumed to be constant and given. $r_0(\underline{x},A)$ is equal to $n_0p_0(\underline{x},A)$ since $p_0(\underline{x},A)$ is the expected value (r/n) of the beta distribution.

If, given $p_0(\underline{x},A)$ and n_0, we observe one repaid loan, the revised default probability density will have parameters $r_1(\underline{x},A) = n_0p_0(\underline{x},A) + 0 = n_0p_0(\underline{x},A)$ and $n_1 = n_0 + 1$. This revised density has expected value:

$$p_1(\underline{x},A) = \frac{r_1(\underline{x},A)}{n_1} = \frac{n_0p_0(\underline{x},A)}{(n_0+1)} .$$

If we observe two repaid loans, the revised default probability density will have parameters

$$r_2(\underline{x},A) = n_0 p_0(\underline{x},A) + 0 + 0 = n_0 p_0(\underline{x},A)$$

and

$$n_2 = n_0 + 1 + 1 = n_0 + 2.$$

In general, if the a priori estimate of the default probability is beta-distributed with parameters $r_0(\underline{x},A) = n_0 p_0(\underline{x},A)$ and n_0, and j loans are repaid, the a posteriori default probability will be beta-distributed with parameters $r_j(\underline{x},A) = n_0 p_0(\underline{x},A)$ and $n_j = n_0 + j$. This j^{th} period estimate is then given by the probability updating rule:

$$p_j(\underline{x},A) = \frac{n_0 p_0(\underline{x},A)}{n_0 + j} \qquad j = 1,..,J-1, \tag{3.2.2}$$

where n_0 is assumed constant and given, and $p_0(\underline{x},A)$ is the default probability estimate obtained by an application of an appropriate pattern recognition technique.

Now define event E_i to be the successful repayment of i loans, with default on loans i+1. Let E_i^* be the successful repayment of i+1 loans. If $V(E_i)$ represents the net present value of event E_i, then

$$V(E_i) = \sum_{j=0}^{i-1} \alpha_j V_1(A,\underline{L}) + \alpha_i V_0(A,\underline{L}) \tag{3.2.3}$$

and

$$V(E_i^*) = \sum_{j=0}^{i} \alpha_j V_1(A,\underline{L}) \tag{3.2.4}$$

where α_j is the appropriate discount factor.

If ℓ is the probability that the customer does not re-apply for a loan in period j, τ is the number of years between periods, and ρ is the cost of capital, then the j^{th} period discount factor is given by

$$\alpha_j = \ell^j (1+\rho)^{-j\tau}. \tag{3.2.5}$$

The probability of events E_i and E_i^* are given by

$$p(E_i) = \left\{\prod_{j=0}^{i-1} [1 - p_j(\underline{x},A)]\right\} p_i(\underline{x},A) \tag{3.2.6}$$

and

$$p(E_i^*) = \prod_{j=0}^{i} [1 - p_j(\underline{x},A)]. \tag{3.2.7}$$

The expected net present value of J loans is then given as

$$\hat{V}_j = \sum_{i=0}^{J-1} V(E_i)p(E_i) + V(E_{J-1}^*)p(E_{J-1}^*). \tag{3.2.8}$$

This is then used in the decision rule (3.2.9):

Accept if and only if $\hat{V}_J > 0$,

that is,

$$\hat{\theta} = \begin{cases} 0 \text{ (reject) if } \hat{V}_J \le 0 \\ 1 \text{ (accept) if } \hat{V}_J > 0 \end{cases} \tag{3.2.9}$$

to obtain a credit granting decision $\hat{\theta}$ which considers the expected net present value due to J loans.

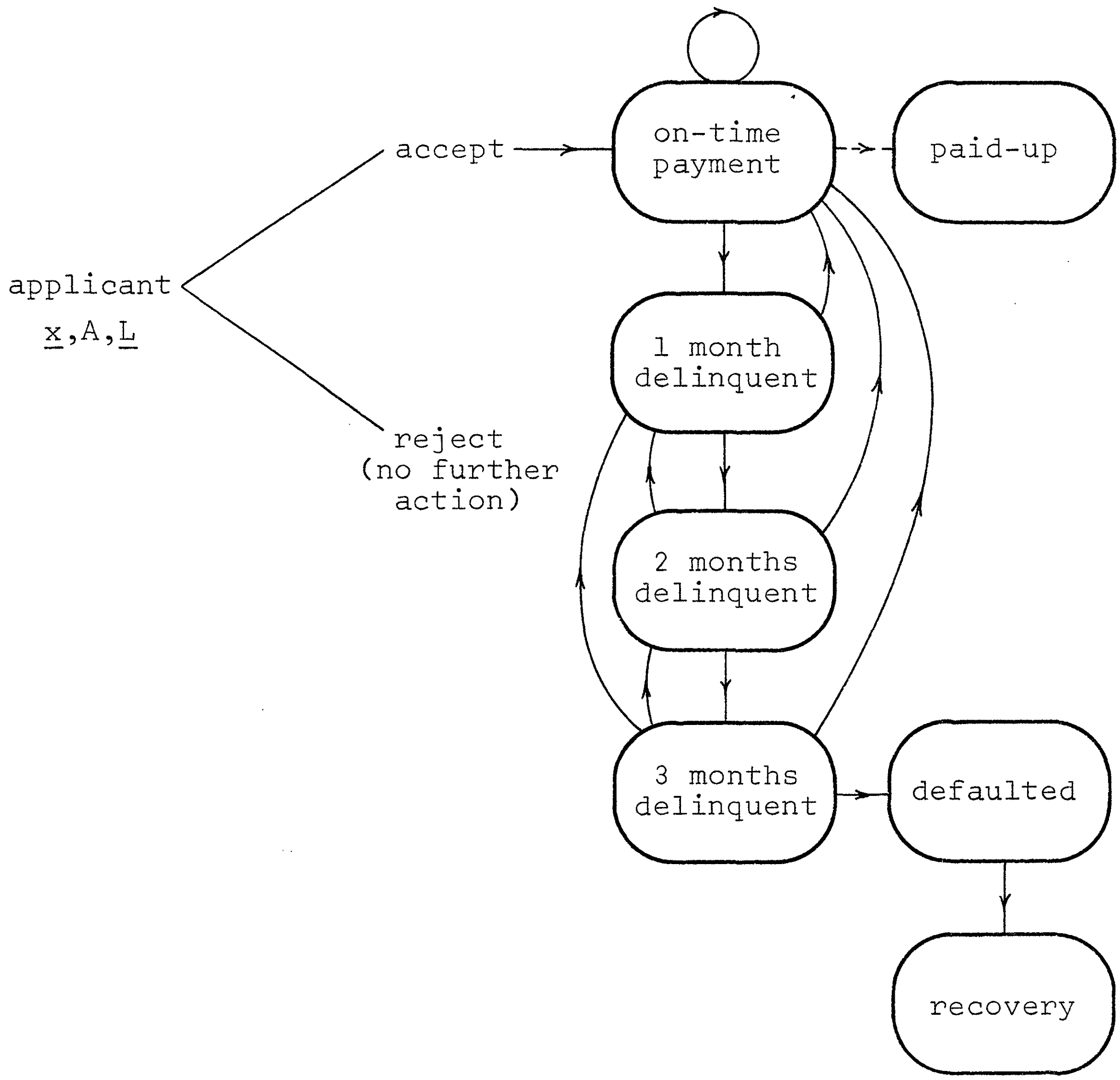
3.3 Model 3 - Single Period, Detailed Outcome

Keeping in mind the basic multi-period model of Section 3.2, we now want to more realistically describe the outcome space for an accepted loan. The binary valued outcome of paid/defaulted is a simplification that ignores an account's delinquency behavior and the associated costs. Control of the outcome of an account does not consist solely of an accept/reject decision, but also includes the degree of collection effort applied to a delinquent account to return it to a paying status. This allocation of collection effort influences the probability of default and the overall profitability of the account.

To sufficiently model this important aspect of the process, the simple outcome description of paid/defaulted will be replaced by the state description shown in Figure 3.3. For installment loans, a paid or non-defaulted loan is one which has made a pre-specified number of payments; a defaulted loan is one which becomes more than a given number of months delinquent. The actual number of months delinquient until write-off (default) is a matter of policy negotiated with a board of federal examiners. At NSB, for example, it is currently three months for all loans except Mastercharge, which is six months. The default probability $p(\underline{x},A)$ is really the probability that the account makes the transition to the default trap state during the term of the loan or the loan review

FIGURE 3.3

Typical Payment/Delinquency State Description



period. $p(\underline{x},A)$ can thus be considered to be a function of the transition probabilities of the delinquent process.

We will assume that the process can be described as a first-order discounted Markov process with stationary transition probability matrix $\underline{\underline{P}}(\underline{x},A)$. That is, the delinquency behavior (and hence the default probability) is seen to be a function of the applicant's feature vector $\underline{x}$ and the loan amount A. The problem of estimating the transition matrix $\underline{\underline{P}}(\underline{x},A)$ as a function of $\underline{x}$ and A will be addressed in detail in Chapter 5. For the case study, the assumptions of the stationarity and the first-order nature of the process will be investigated.

Given the preceding state description, we must identify the costs and revenues associated with the process. Costs associated with the current policy of trying to collect delinquent payments can be straightforwardly obtained. For installment loans, each occupancy of the paying state produces a net revenue of one payment. These costs and revenues are used as elements of a transition reward matrix $\underline{\underline{R}}(A,\underline{L})$, where the transition rewards depend on the loan amount A, and other loan parameters summarized in $\underline{L}$ (e.g., interest rate, term, etc.).

If T is the term of an installment loan, the expected net present value of the loan is given by $V_d(0|T)$, the expectation of total discounted rewards given the account is at the decision state d at time t=0 with T transitions occurring during the loan term. As outlined in Howard [30] this expectation can be computed recursively using the relation (3.3.1):

$$V_i(t|T) = \sum_j p_{ij}[r_{ij} + \beta V_j(t+1|T)] \qquad (3.3.1)$$

where

$V_i(t|T)$ = present value of being in state i after t transitions given the process will terminate after T transitions

p_{ij} = probability of a transition from state i to state j [(i,j)th element of $\underline{\underline{P}}(\underline{x},A)$]

r_{ij} = reward obtained when a transition from state i to state j is made

$\beta = (1 + \rho/12)^{-1}$ = monthly discount factor, where ρ is the annual cost of capital.

If a loan account makes a transition from i to j, the bank will earn the amount r_{ij} plus the discounted amount it expects to earn if the account is in state j after the transition. These earning due to a transition to state j are weighted by p_{ij}, the probability that the transition occurs.

Thus, given $\underline{x}$ and A we estimate $\underline{\underline{P}}(\underline{x},A)$ to obtain the required transition probability matrix. The reward matrix $\underline{\underline{R}}(A,\underline{L})$ is computed from the loan parameters. $V_d(0|T)$ is then computed recursively using (3.3.1). The credit granting decision can then be made using the decision rule:

$$\hat{\theta} = \begin{cases} 0 \text{ (reject) if } V_d(0|T) \leq 0 \\ 1 \text{ (accept) if } V_d(0|T) > 0 \end{cases} . \tag{3.3.2}$$

It should be noted that we are assuming that current collection policy remains unchanged. In fact, for each possible post-default recovery policy and, conditional upon a given recovery policy for each possible collection policy, we could derive a set of optimal credit granting decision rules. In this way we would be solving the larger joint optimization problem of recovery policy, collection policy, and credit grant granting policy given a fixed (over all applicants) collection and recovery policy. This optimization problem provides a topic for future investigation but will not be addressed in any detail in this thesis. However, this model would provide the necessary framework for such an investigation.

3.4 Model 4 - Multi-Period, Detailed Outcome

Ideally, we would like to take the detailed state description of Section 3.3 and allow for multi-period updating in theory, although for practical and computational reasons an approximate solution will be necessary.

The applicant is characterized by his feature vector $\underline{x}$ and loan amount/credit line A. Based on this information, we estimate the transition probability matrix $\underline{\underline{P}}_0(\underline{x},A)$, where the subscript is used to index the macro-period. The term macro-period will be used to indicate a review period, i.e., the decision to grant credit is made at the beginning of each macro-period. The term micro-period refers to the (one month) period of the discrete-time Markov process. For installment loans the macro-period will be on the order of 24 months. For revolving credit, it is quite possible (although costly) to review accounts each month, i.e., a macro-period of one month. In this context, the optimal setting of the review period for revolving credit accounts can be treated as an optimization of the trade-off between the cost of reviewing an account and the expected value of additional information obtained upon review.

The first macro-period transition matrix $\underline{\underline{P}}_0(\underline{x},A)$ is a random variable and will be assumed to have matrix beta prior $\underline{\underline{M}}_0(\underline{x},A)$ which can be estimated by pattern recognition techniques (see Chapter 5). Since the rows of $\underline{\underline{P}}$ are independent, each row of $\underline{\underline{M}}$ corresponds to a multivariate beta or Dirchlet prior that can be estimated independent of the other rows. This reduces to a univariate beta prior for the case where the only two transitions that are possible are to return to paying status or go one more month delinquent.

Conditional upon accepting the application, the account will make T transitions within the paying/collection/default process, where T is the number of months in a macro-period. The transition history can be described by a transition count matrix $\underline{\underline{F}}_1$. Each time a transition from state i to state j occurs, the $(i,j)^{th}$ element of $\underline{\underline{F}}_1$ is incremented by one. $\underline{\underline{F}}_1$ is a matrix random variable with matrix beta-Whittle distribution with parameters $\underline{\underline{M}}_0(\underline{x},A)$ and T.

The transition count matrix $\underline{\underline{F}}_1$ is a sufficient statistic describing the outcome of the first macro-period. Given this observed payment behavior of the account, we are now ready to update this transition probability matrix $\underline{\underline{P}}_0(\underline{x},A)$, and based on this updated estimate, we again face the decision to grant credit for the second macro-period. The updating rule is $\underline{\underline{M}}_1(\underline{x},A) = \underline{\underline{M}}_0(\underline{x},A) + \underline{\underline{F}}_1$ since the transition count matrix $\underline{\underline{F}}_1$ is a sufficient statistic for the matrix-beta natural conjugate prior. This updating rule is the matrix analogy of the univariate beta case where $r'' = r' + r$ and $n'' = r' + n$. Further discussion of the distribution theory involved can be found in Martin [34].

Given the relevant costs and revenues of state occupancies or state transitions, we can compute the net present value accruing during the macro-period for any transition count matrix $\underline{\underline{F}}$. This computation is only approximate, although a rather close one, since $\underline{\underline{F}}$ only summarizes the number of occupancies of any state and does not indicate the micro-periods in which they occurred.

In a manner like the updating for the macro-period 1 decision, we could continue for periods 2,3,...,J. If we considered a finite number of periods, then we could in theory (via dynamic programming) make an optimal intial credit granting decision that accounted for both the detailed payment/ delinquency behavior of the account and the multi-period nature of the decision.

In practice, however, such a model would not be computationally feasible. We must focus our attention toward an approximate solution which captures as much as possible the important aspects of payment/delinquency detail while not neglecting the importance of modeling the multi-period nature of the process. Apart from defaults, lending institutions are particularly concerned with the delinquent collection expenses that accounts create, and should these expenses be too great they will refuse to re-extend credit. That is, in any given macro-period, the detailed transition behavior of the account is quite relevant to the subsequent decision to grant credit. As for the multi-period nature of the process, one hears policy statements of the type, "We're willing to take more risk with new, particularly young customers because if they prove to be a good risk, we have probably gained a lifetime account." A useful decision model should reflect both these concerns.

3.5 Characteristics of the Models

This chapter has presented a set of increasingly complex models for credit granting decision making. Model 1, the single period two outcome model, is representative of the class of models now being employed in the credit industry. Although the structure of this model is relatively simple, the requirements for its implementation include the estimation of a number of parameters, the somewhat involved computation of outcome rewards, and the multivariate estimation of default probability.

Model 2, the multi-period two outcome model, incorporates the expected net present value of subsequent loans by means of a Bayesian default probability updating rule. In addition to the problems presented by Model 1, this multi-period model requires the estimation of four additional parameters: the number of loans considered, the time between loans, the re-application probability, and the "diffuseness" parameter of the beta-distributed default probability.

Models 3 and 4 differ from Models 1 and 2 in their detailed description of the set of possible outcomes. The delinquency state outcome description permits a more exact treatment of the events preceding default. This treatment allows explicit consideration of both the costs associated with the collection of delinquent accounts and the effect of the term of the loan on the probability of default. These considerations have never been treated to any degree in the published literature of the field. In addition, Model 4 allows for the Bayesian updating of the state transition probabilities after each loan (or after each review period) for the purpose of re-evaluating the decision to extend credit. As previously indicated, Model 4 will not be computationally practical, and approximate solutions for a multiperiod detailed outcome description must be found. One such solution is presented in Chapter 13. As we will discuss in Chapter 16, Model 4 includes all of the important ingredients for revolving credit decision-making.

The remainder of this chapter will focus on what might be considered key properties and sensitivities of the models. In Section 3.5.1 we present a parametric analysis of Models 1 and 2 for installment loan decision-making. In Section 3.5.2 we analyze the structure of the state outcome model with a view toward uncovering those aspects of the process that have most significance for the credit granting decision.

3.5.1 Models 1 and 2

In the case of installment lending, Models 1 and 2 are specified by the structure shown in Figures 3.1 and 3.2, and by the following set of parameters:

$p_0(\underline{x},A)$	probability of default on the initial loan given the feature vector $\underline{x}$ and the loan amount A
A	loan amount
r	interest rate
ρ	cost of capital
T	loan term (number of monthly payments).

In addition to the above single-period parameters the four multi-period parameters are:

J	number of loans considered
ℓ	probability of re-application
τ	time between applications (in years)
n_0	diffuseness parameter for the initial default probability.

As mentioned previously, we are assuming that collection costs and pursuit costs and recoveries are given. These costs and revenues are parametrically specified in Appendix A. We feel that including these additional parameters will not significantly contribute to this analysis.

Both Models 1 and 2 are described by equations (3.2.1) through (3.2.9). The outcome rewards $V_0(A,\underline{L})$ and $V_1(A,\underline{L})$, are parametrically given by equations (A.1) through (A.12).

We begin this analysis of Models 1 and 2 by considering these reward equations, since the key issue is that of the tradeoff between the relatively small non-default profit, $V_1(A,\underline{L})$, with large probability of occurrence and the relatively large default loss, $V_0(A,\underline{L})$, with lesser probability of occurrence. That is, considering just one loan period, the expected net present value of the loan is given by (3.1.1) as

$$\hat{V} = p(\underline{x},A)V_0(A,\underline{L}) + [1 - p(\underline{x},A)]V_1(A,\underline{L}). \qquad (3.1.1)$$

Neglecting the fixed administrative costs, $V_1(A,\underline{L})$ is given by (A.1) and (A.2) to be

$$V_1(A,\underline{L}) = A\left\{-1 + \frac{r'}{\rho'}\left[\frac{1+r'}{1+\rho'}\right]^T \frac{[(1+\rho')^T-1]}{[(1+r')^T-1]}\right\}, \qquad (3.5.1)$$

where $r' = r/12$ and $\rho' = \rho/12$ are monthly rates. Inspection of (3.5.1) reveals that rewards are proportional to loan amount A, and for large T, are positively related to the ratio r/ρ.

$V_0(A,\underline{L})$, the default loss, is given by (A.7) through (A.11). These formulas give the default loss conditional upon the period in which default occurs, and then sum these conditional losses scaled by the probability that default occurs in that period given default occurs. This additional operation does not make $V_0(A,\underline{L})$ amenable to the same compact expression given by (3.5.1) for $V_1(A,\underline{L})$. Our analysis of the expression for $V_0(A,\underline{L})$ reveals that the only insight to be gained is that it has as its components the same terms found in (3.5.1), namely the dependence on A and r/ρ.

The effect of the loan term T on the outcome rewards is less easily seen by inspection. To demonstrate the nature of this T dependence, both $V_0(A,\underline{L})$ and $V_1(A,\underline{L})$ were computed for a range of T values. Typical values were assigned to the other parameters of A = \$2000, ρ = .1, and r = .135. The results of this computation are shown below. Both $V_0(A,\underline{L})$ and $V_1(A,\underline{L})$ increase almost linearly with T, since the principle of the loan earns a net interest (in the approximate ratio of r/ρ) over a longer loan period.

	Net Present Value	
Loan Term T	no default $V_1(A,\underline{L})$	default $V_0(A,\underline{L})$
12	27.22	-755
24	60.74	-677
36	93.39	-617
48	125.14	-562
60	155.93	-509

(A = \$2000, r = .135, ρ = .10)

Thus far, we have seen that the outcome rewards (neglecting fixed administrative costs) are proporational to loan amount A, and are approximately linearly related to the ratio r/ρ and to the loan term T. For the single period case, the rewards for default and non-default, $V_0(A,\underline{L})$ and $V_1(A,\underline{L})$ respectively, are weighted by their respective outcome probabilities, $p_0(\underline{x},A)$ and $1 - p_0(\underline{x},A)$, to determine the expected net present value of granting credit. In general, there is little that we can say regarding the manner in which $p_0(\underline{x},A)$ depends on the feature vector $\underline{x}$ and the loan amount A. The nature of this dependence may vary from location to location, depending, in part, on the characteristics of the local population and the local economy. We will leave further discussion of this issue to Chapter 4, which considers the pattern recognition aspects of default probability estimation.

In the multi-period case, we consider the expected net present value of J loans applied for every τ years. The estimate of the probability of default is decreased after every repaid loan according to updating rule given in (3.2.2)

$$p_j(\underline{x},A) = \frac{n_0 p_0(\underline{x},A)}{n_0 + j} , \qquad (3.2.2)$$

where $p_0(\underline{x},A)$ is the initial estimate of the beta-distributed default probability with diffuseness parameter n_0, and $p_j(\underline{x},A)$ is the updated estimate given j repaid loans. Default and non-default rewards are computed as described above and then discounted by the factor α_j,

$$\alpha_j = \ell^j (1+\rho)^{-j\tau} , \quad (3.2.5)$$

if the reward is obtained in period j. The reapplication probability, ℓ, is included in the discount factor α_j. These discounted rewards for period j outcomes are then weighted by their probability of occurrence ($p_j(\underline{x},A)$ or $1 - p_j(\underline{x},A)$) to obtain the expected net present value of J loans, $\hat{V}_J$.

The diffuseness parameter n_0 determines the rate at which the default probability is updated, as seen by (3.2.2). If n_0 is large (relative to j) little updating results since n_0 appears in both the numerator and denominator of (3.2.2). Of course, a large value for n_0 implies greater weight on initial application information relative to payment performance. If n_0 is small relative to j, repayment of the first loan decreases $p_1(\underline{x},A)$ to nearly $n_0 p_0(\underline{x},A)$, which will be nearly zero if $p_0(\underline{x},A)$ is typically small. The implication here is that "first loan performance tells all".

The parameters ℓ and τ influence $\hat{V}_J$ through the discount factor α_j given by (3.2.5). Increasing τ, the reapplication interval, further discounts the expected net present value of subsequent loans. Increasing ℓ, the reapplication probability, increases the probability of obtaining these cash flows from subsequent loans. The effect of considering one more subsequent loan (by incrementing J) is diminished by the discount factor. From a <u>decision</u> viewpoint, if J (say J=5) loans have positive expected net present value, then it is likely that J+1 loans will also have positive expected net present value, and the decision will not be altered.

So as to provide insight into the relative sensitivities of Models 1 and 2 to these parameters, we consider the following brief example. Parameters are initially set to what might be considered typical values for the average lending institution. The parameter values are then varied over a range centered about these typical values for which the expected net present value, $\hat{V}_J$, is computed from (3.2.2) through (3.2.8) and (A.1) through (A.12). In this manner, the sensitivity of $\hat{V}_J$ to each parameter can be determined.

Typical values and the range of values considered for each parameter are given in Table 3.5.1.

Parameter	Typical Value	Range
$p_0(\underline{x},A)$	.05	[.01, .09]
A	\$2000	[\$1000, \$3000]
r	.135	[.12, .15]
ρ	.10	[.08, .12]
T	24 months	[12 months, 36 months]
ℓ	.7	[.5, .9]
τ	2 years	[1 year, 3 years]
n_0	.5	[.3, .7]

TABLE 3.5.1

Typical Parameter Values and Ranges

For a loan with these typical values the expected net present value of considering J loans, $\hat{V}_J$, is given in Table 3.5.2. The discount factor α_J, for the J^{th} loan is also presented to show the diminishing effect of loans further in the future.

J	$\hat{V}_J$	α_J
1	23.86	1.0000
2	50.48	.5785
3	67.17	.3347
4	77.10	.1936
5	82.92	.1120
6	86.32	.0648
7	88.29	.0375
8	89.44	.0217
9	90.10	.0125
10	90.48	.0073
11	90.71	.0042
12	90.83	.0024
13	90.91	.0014
14	90.95	.0008
15	90.98	.0005
16	90.99	.0003
17	91.00	.0002
18	91.00	.0001
19	91.01	.0001
20	91.01	.0000

TABLE 3.5.2

Typical Expected Net Present Values of J Loans

These typical results reveal that about 90% of the present value is obtained in the first five loans, and virtually all present value is obtained after 15 loans. Thus, the credit granting decision is unlikely to be affected by considering only a five-loan horizon. This is considered to be a fortunate result, since management is often seen to be unwilling to consider longer planning horizons.

Table 3.5.3 presents the sensitivity of Model 1 and Model 2 expected net present values to variations of the parameters over the ranges given in Table 3.5.1. We should point out that changing either the loan amount (A) and the loan term (T) may also affect the default probability. The magnitude of this effect can only be determined by investigation on actual data of the particular lending institution. The entries in Table 3.5.3 for A and T assume that $p_0(\underline{x},A)$ remains unchanged.

Parameter	Typical Value	Parameter Change	Change in expected net present value of 1 loan	Change in expected net present value of 5 loans
$p_0(\underline{x},A)$	.05	.01	-7.37	-10.23
A	\$2000	\$500	9.00	27.50
r	.135	.01	20.10	43.37
ρ	.10	.01	-20.16	-45.84
r/ρ	1.35	.1	15.30	34.50
T	24 months	6 months	17.40	36.60
ℓ	.7	.1	-	16.40
τ	2 years	.5 years	-	-5.43
n_0	.5	.1	-	-1.63
	All parameters at typical values gives $\hat{V}_1$ and $\hat{V}_5$ as		23.86 and	82.92

TABLE 3.5.3

Sensitivity of Models 1 and 2 to Parameter Changes

These results for a typical loan suggest that almost all of the parameters can significantly influence the credit granting decision. The only exceptions seem to be the reapplication interval, τ, and the diffuseness parameter, n_0. Of all the parameters that are relevant to the decision to grant credit, n_0 seems to be the least critical.[1]

The results presented in Table 3.5.3 hold true over the entire parameter ranges given in Table 3.5.1. In fact, these marginal sensitivities are very nearly linear ($R^2 > .98$) over the ranges given.

[1]In practice, n_0 is also found to be one of the more difficult parameters to accurately estimate.

3.5.2 Models 3 and 4

Models 3 and 4 are specified by

1) the state outcome description,

2) the transition probability matrix $\underline{\underline{P}}(\underline{x},A)$,

3) the transition reward matrix $\underline{\underline{R}}(A,\underline{L})$, and

4) the macro-period length, T.

One typical state outcome description for installment loans is shown in Figure 3.3 to contain the decision state, the on-time payment state, three delinquent states, and the default trap state. In general, if a loan is allowed to become at most M months delinquent after which it is considered to have defaulted, then with the decision, on-time, and default states, the state description will have <u>at least</u> N = M+3 states. Since we assume that the state description represents a first order Markov chain, the minimal state description may have to be expanded to more than M+3 states to model any second or higher order "memory" in the process.[2]

The transition probability matrix, $\underline{\underline{P}}(\underline{x},A)$, can be estimated as a function of the applicant's feature vector and loan amount A. In general, we can say little about this dependence. For the purposes of this analysis, we will assume that

[2]This state expansion will probably be necessary for most applications. For example, the case study presented in Part II required an expansion from 6 states to 13 states (see Section 11.1)

either $\underline{\underline{P}}$ has been so estimated or that $\underline{\underline{P}}$ represents a typical (a priori) transition probability matrix for the given state description. The relevant elements of the transition reward matrix, $\underline{\underline{R}}$, will be indicated as we analyze the transition behavior of the process over its duration of T transitions.

We begin by considering the simplified 4 state description in Figure 3.5.1 below. We note that the first loan payment is, in general, not due until one month after credit is granted. This is shown as a one-period delay from state D (the decision state) to the on-time state 0 (0 month delinquent). Since this delay is inherent in the process, we need not include state D in the $\underline{\underline{P}}$ matrix description.

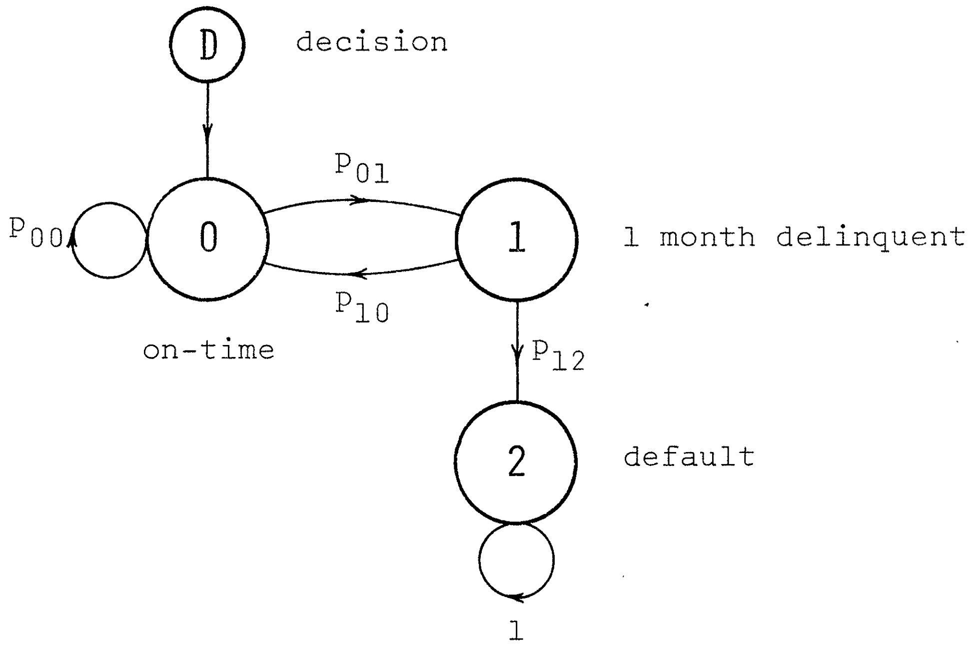
Since the decision to grant credit depends on loan payment revenues, delinquency-related costs and default losses, we might pose the following questions:

What is the probability of default, given P?

How many payments can we expect to obtain before default, should it occur?

How many times will the account be one month delinquent? two months delinquent? and so on.

These questions will first be addressed in the context of the 4 state model, after which we will extend the analysis to show the effect of adding or expanding states.



$$\underline{\underline{P}} = \begin{bmatrix} p_{00} & p_{01} & 0 \\ p_{10} & 0 & p_{12} \\ 0 & 0 & 1 \end{bmatrix}$$

FIGURE 3.5.1
4 State Description

Transform Analysis

The geometric transform or "z-transform" of a Markov chain with transition probability matrix $\underline{\underline{P}}$ is defined as[3]

$$\underline{\underline{\Phi}}^{g}(z) \equiv \sum_{n=0}^{\infty} (\underline{\underline{P}}z)^{n} = [\underline{\underline{I}} - \underline{\underline{P}}z]^{-1}, \qquad (3.5.1)$$

where $\underline{\underline{I}}$ is the identity matrix, and $\underline{\underline{\Phi}}(n) = \underline{\underline{P}}^{n}$ is the n-step transition probability matrix for the first order Markov chain. The inverse transform of $\underline{\underline{\Phi}}^{g}(z)$ is simply the n-step transition probability matrix $\underline{\underline{\Phi}}(n)$. For the analysis of this section, we will find it convenient and insightful to employ transform analysis techniques.

The z-transform of the state description of Figure 3.5.1 can be shown to take the form

$$\underline{\underline{\Phi}}^{g}(z) = \frac{1}{(1-z)(1-az)(1-bz)} \begin{bmatrix} 1-z & p_{01}z(1-z) & p_{01}p_{12}z^{2} \\ p_{10}z(1-z) & (1-p_{00}z)(1-z) & p_{12}z(1-p_{00}z) \\ 0 & 0 & 1-p_{00}z-p_{01}p_{10}z^{2} \end{bmatrix}$$

where $a = \frac{1}{2}[p_{00} + \sqrt{p_{00}^{2} + 4p_{01}p_{10}}]$

$$b = \frac{1}{2}[p_{00} - \sqrt{p_{00}^{2} + 4p_{01}p_{10}}]. \qquad (3.5.2)$$

[3]See Howard [32] for a more detailed discussion of Markov process transform analysis.

The denominator of (3.5.2) is presented in factored form to show that the eigenvalues of the process are

$$\lambda_1 = 1, \lambda_2 = a, \lambda_3 = b.$$

These eigenvalues represent the geometric "decay" factors of the multi-step transition probability matrix $\underline{\underline{\Phi}}(n)$. That is, $\underline{\underline{\Phi}}(n)$ can be expressed as

$$\underline{\underline{\Phi}}(n) = \lambda_1 \underline{\underline{A}}_1 + (\lambda_2)^n \underline{\underline{A}}_2 + (\lambda_3)^n \underline{\underline{A}}_3,$$

where $\underline{\underline{A}}_1$ is the steady-state matrix and $\underline{\underline{A}}_2$ and $\underline{\underline{A}}_3$ are constant matrices.

We are now prepared to answer the questions that were posed above. The probability of default during T+1 periods is equal to the T-step transition probability $\phi_{02}(T)$, where $\phi_{02}(n)$ is the (0,2) element of $\underline{\underline{\Phi}}(n)$. The z-transform of $\phi_{ij}(n)$, namely $\phi^g_{ij}(z)$, is the $(i,j)^{th}$ element of $\underline{\underline{\Phi}}^g(z)$. Thus, the z-transform of $\phi_{02}(n)$ is

$$\phi^g_{02}(z) = \frac{p_{01}p_{12}z^2}{(1-z)(1-az)(1-bz)} . \tag{3.5.3}$$

$\phi_{02}(n)$ can be obtained from tables of the z-transform if we first expand $\phi^g_{02}(z)$ as the sum of partial fractions and then separately invert each term of the sum. Before obtaining the partial fraction expansion we recognize that the numerator of (3.5.3), $p_{01}p_{12}z^2$, represents the probability of immediate default. That is, the probability that the account never makes a payment is $p_{01}p_{12}$. If this is the case, it will enter the default trap state 3 months after credit is granted (the 3 month delay is given by the z^2 term, the z-transform of a 2 period delay, and the 1 month delay from the decision state to state 0). The inverse of $\phi_{02}^{g}(z)$ can then be obtained as

$$\phi_{02}(n) = \begin{cases} 0 & n < 2 \\ 1-p_{01}p_{12}\left[\frac{a^n}{(1-a)(a-b)} + \frac{b^n}{(1-b)(a-b)}\right] & n \geq 2 \end{cases} \qquad (3.5.4)$$

Inspection of (3.5.4) shows that the probability that the account will have defaulted by period n approaches 1 as n increases (if $p_{12} > 0$). The geometric rate at which the default probability increases is determined by the eigenvalue (a or b) with greatest magnitude. The definition of a and b in (3.5.2) reveals (after some algebra) that

$$|a| > |b| \qquad \text{if } p_{00} > 0$$

$$a = -b = \sqrt{p_{10}} \quad \text{if } p_{00} = 0.$$

We would thus always expect a to be the "dominant component" of the transfer function since the on-time self-transition always has positive probability ($p_{00} > 0$).

What is the probability that the account defaults <u>in period n</u>? This "first-passage time" probability, $f_{02}(n)$, can be shown to have z-transform

$$f^g_{02}(z) = \phi^g_{02}(z)/\phi^g_{22}(z). \qquad (3.5.5)$$

From (3.5.2) we obtain

$$f^g_{02}(z) = \frac{p_{01}p_{12}z^2}{(1-az)(1-bz)} .$$

If $|a| > |b|$ and a is nearly 1, then the inverse transform of $f^g_{02}(z)$ is approximately

$$f_{02}(n) \approx p_{01}p_{12}a^{n-2}, \text{ for } n \geq 2.$$

If a is close to one, the ratio

$$f_{02}(n+1)/f_{02}(n) \approx a$$

will be close to one. That is, the probability of default in any one period will be approximately equally likely. (This was indeed found to be the case for a sample of 200 defaulted loans considered in Appendix A.)

Example

The following numerical example should help to clarify these results. Let

$$\underline{\underline{P}}(\underline{x},A) = \underline{\underline{P}} = \begin{bmatrix} .9 & .1 & 0 \\ .8 & 0 & .2 \\ 0 & 0 & 1 \end{bmatrix}$$

represent the delinquency process where there is a .1

probability of delinquency on any given payment, and a .2 probability of default in the next period if the account is delinquent. The eigenvalues of the process are

$$a = .98, \; b = -.08.$$

The n-step transition probability to the default state, $\phi_{d2}(n) = \phi_{02}(n-1)$, and first passage time probability $f_{d2}(n)$ are given below for n=1,...,10. In addition, to demonstrate the effect of the eigenvalues on the geometric rate at which the default probability increases, we have included the second and third terms of $\phi_{d2}(n)$ as given by (3.5.4) (with a one-period delay from state d to state 0).

n	$\phi_{d2}(n)$	$f_{d2}(n)$	a^{n-1} term	b^{n-1} term
1	.000	.000	-	-
2	.000	.000	-.9986	-.0014
3	.020	.020	-.9801	.0001
4	.038	.018	-.9620	-.0000
5	.056	.018	-.9442	.0000
6	.073	.017	-.9267	-.0000
7	.090	.017	-.9096	.0000
8	.107	.017	-.8928	-.0000
9	.124	.017	-.8763	.0000
10	.140	.016	-.8601	-.0000

Note that the first passage time to the default state, $f_{d2}(n)$, is nearly equal in each period. For $p_{00} = .9$ and $p_{10} = .8$, there is one eigenvalue that dominates the transient behavior of the process.

State Occupancy

How many payments can we expect to obtain before default? How many times do we expect the account to be one month delinquent? These questions are equivalent to asking for the expected number of occupancies of states 0 and 1 respectively. If $\overline{\nu}_{dj}(T)$ is the expected number of times state j is occupied (given the process starts in state d, the decision state), then its z-transform $\overline{\nu}^g_{dj}(z)$ is given by

$$\overline{\nu}^g_{dj}(z) = \frac{1}{1-z}\,\phi^g_{dj}(z), \qquad (3.5.6)$$

where $\phi^g_{dj}(z)$ is the z-transform of the multi-step transition probability $\phi_{dj}(n)$. Using (3.5.2) to obtain $\phi^g_{dj}(z)$, we obtain

$$\begin{aligned} \overline{\nu}^g_{d0}(z) &= \frac{z}{(1-z)(1-az)(1-bz)} \\ \overline{\nu}^g_{d1}(z) &= \frac{p_{01}z^2}{(1-z)(1-az)(1-bz)}\,, \end{aligned} \qquad (3.5.7)$$

which when inverted gives

$$\overline{\nu}_{d0}(T) = \frac{1}{p_{01}p_{12}} - \left[\frac{a^{T+1}}{(1-a)(a-b)} + \frac{b^{T+1}}{(1-b)(a-b)}\right] = \frac{1}{p_{01}p_{12}}\,\phi_{d2}(T+2)$$

$$\overline{\nu}_{d1}(T) = \frac{1}{p_{12}} - p_{01}\left[\frac{a^{T}}{(1-a)(a-b)} + \frac{b^{T}}{(1-b)(a-b)}\right] = \frac{1}{p_{12}}\,\phi_{d2}(T+1).$$

For the numerical example where $p_{00} = .9$ and $p_{10} = .8$ we find

$$\overline{\nu}_{d0}(T=10) = 8.6$$

$$\overline{\nu}_{d1}(T=10) = .8\,.$$

Effect of Adding Delinquent States

We will now consider a 5 state description as shown in Figure 3.5.2, which includes an additional 2 month delinquent state. The structure of the transition probability matrix is similar to that of the 4 state description, namely non-zero elements in both the first column ($j \to 0$ transitions) and the upper diagonal ($j \to j+1$ transitions). This structure leads to an analogous z-transform of $\phi_{d3}(n)$, the multi-step transition probability from the decision state d to the default trap state 3, namely

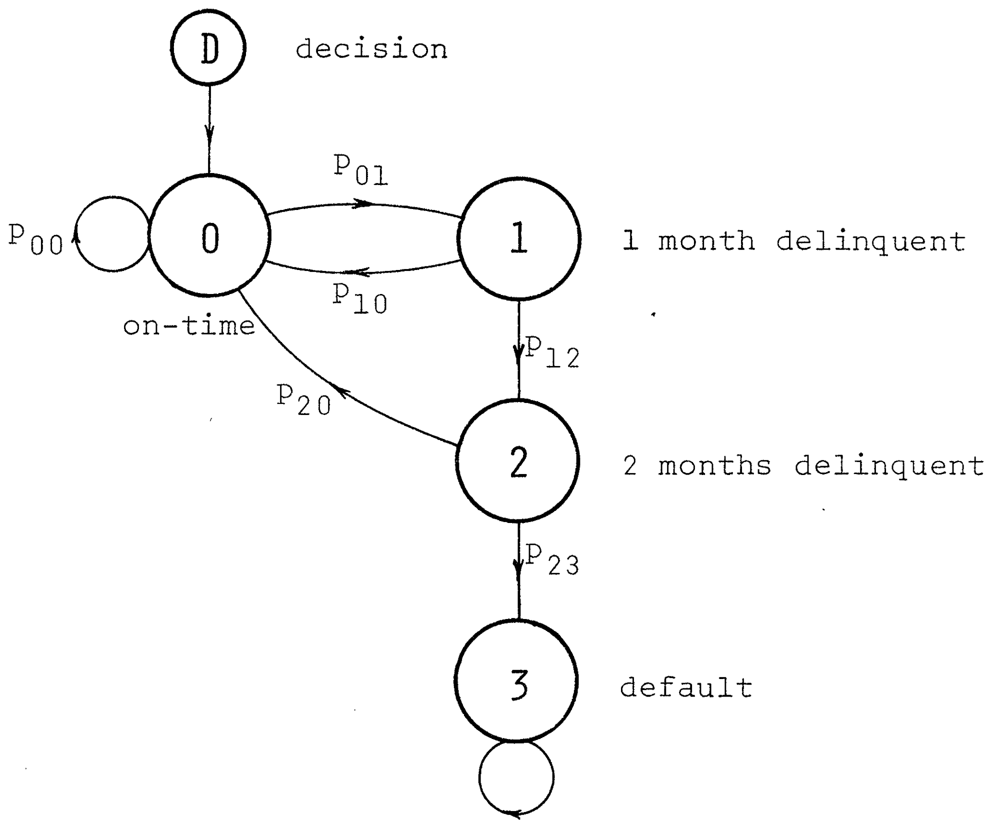
$$\phi^{g}_{d3}(z) = \frac{p_{01}p_{12}p_{23}z^4}{(1-z)(1-az)(1-bz)(1-cz)}, \tag{3.5.8}$$

where a, b, c are the non-unity eigenvalues of $\underline{\underline{P}}$. As before, the numerator represents the probability, $p_{01}p_{12}p_{23}$, that the account never makes a payment and defaults four months after credit granting (z^4 delay term).

The probability of default in T periods will be approximately

$$\phi_{d3}(T) \approx 1 - p_{01}p_{12}p_{23} \frac{a^{T-1}}{(1-a)(a-b)(a-c)} \quad \text{for } T \geq 3, \tag{3.5.9}$$

where a is the dominant eigenvalue (closest to 1). If p_{00} = .9, p_{10} = .8, p_{20} = .7, then a = .9945. The other two eigenvalues are the other roots of the characteristic polynomial $|\underline{\underline{I}}-\underline{\underline{P}}z|$. In general, if there are M delinquent states



$$\underline{P} = \begin{bmatrix} p_{00} & p_{01} & 0 & 0 \\ p_{10} & 0 & p_{12} & 0 \\ p_{20} & 0 & 0 & p_{23} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

FIGURE 3.5.2

Addition of a Delinquent State

(resulting in a (M+2) by (M+2) $\underline{\underline{P}}$ matrix with non-zero upper-diagonal and first column), then $|\underline{\underline{I}}-\underline{\underline{P}}z|$ has the form

$$|\underline{\underline{I}}-\underline{\underline{P}}z| = (1-z)(1-p_{00}z-p_{01}p_{10}z^2-\cdots-p_{01}p_{12}\cdots p_{M-1,M}p_{M,0}z^{M+1}).$$

This result follows directly from a "flow graph" analysis of the process. We also know that the number of positive real roots can not exceed the number of variations in the signs of the coefficients. There is only one variation in the sign of $|\underline{\underline{I}}-\underline{\underline{P}}z|/(1-z)$ so there is only one positive real root. If M is odd there will also be one negative real root and M-1 complex roots. If M is even there will be one positive real root and M complex roots. The complex roots occur in conjugate pairs and determine the magnitude and frequency of oscillations in the multi-step transition probability matrix $\underline{\underline{\Phi}}(n)$. The negative real root represents a decaying oscillation with periodicity n=2.

To demonstrate that the positive real eigenvalue dominates the delinquency behavior, we start with the process of Figure 3.5.1 (M=1 delinquent state) and then add states of increasing delinquency (M=2,3,...). We assume that as we add states, the transition probability from the last delinquent state to the default state increases by .1. That is, we take $p_{01} = .1$, $p_{12} = .2$, $p_{23} = .3$, ..., $p_{67} = .7$. This example gives the eigenvalues shown in Table 3.5.4 for different numbers of delinquent states. Note that the (non-unity)

M number of delinquent states (excluding default)	real roots		magnitude of complex conjugate pairs		
1	.9815	-.08			
2	.9946		.12		
3	.9979	-.15	.16		
4	.9989		.19	.18	
5	.9994	-.21	.23	.21	
6	.9996		.26	.24	.23

TABLE 3.5.4

Typical Eigenvalues of the Delinquent Process positive real eigenvalue (λ_2) is very nearly one and so dominates the transient behavior of the process. Equation (3.5.9) implies that the default probability can be approximated by

$$\phi_{d,M+1}(T) \approx 1 - \left[p_{01} p_{12} \cdots p_{M,M+1} \frac{(\lambda_2)^{T-1}}{(1-\lambda_2)(\lambda_2-\lambda_3)\cdots(\lambda_2-\lambda_{M+2})} \right]$$

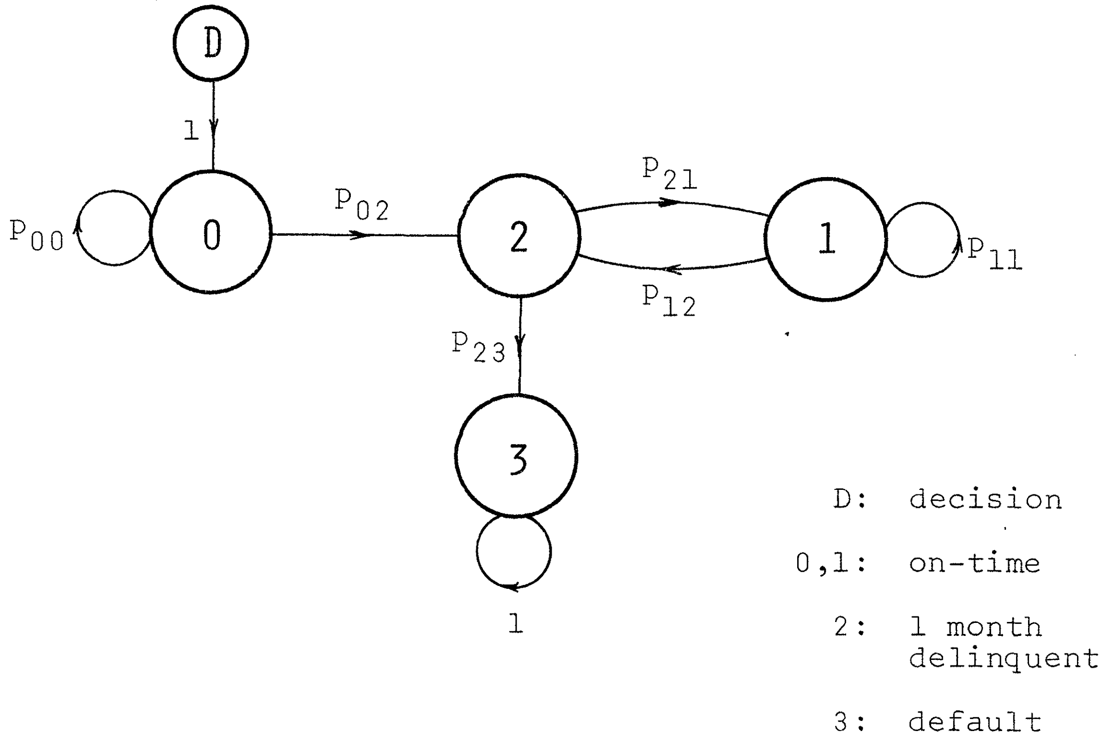
for $T \geq M+1$.

This result provides computationally efficient means of computing the default probability in terms of the "direct transmission" from state d to M+1 (i.e., $p_{01} p_{12} \cdots p_{M,M+1}$) and the eigenvalues $\lambda_2, \ldots, \lambda_{M+2}$. Of course, for the credit granting decision where we estimate $\underline{\underline{P}}(\underline{x},A)$ for each applicant, we would have to resolve for the dominant eigenvalue for each applicant. Computationally, it will prove more economical to compute the T-step transition probability of default recursively from $\hat{\underline{\underline{P}}}(\underline{x},A)$ using the relationship

$$\underline{\underline{\Phi}}(n+1) = \underline{\underline{\Phi}}(n) \hat{\underline{\underline{P}}}(\underline{x},A). \tag{3.5.10}$$

Effect of Expanding States

We now consider the effect of expanding the state description to remove second order memory. Consider the example of Figure 3.5.3 where the on-time state has been expanded as two states, a "never delinquent" state 0 and a "previously delinquent" state 1. The expanded transition probabilitiy matrix $\underline{\underline{P}}'$ can be partitioned as shown in



$$\underline{\underline{P}}' = \left[\begin{array}{c|ccc} p_{00} & 0 & p_{02} & 0 \\ \hline 0 & p_{11} & p_{12} & 0 \\ 0 & p_{21} & 0 & p_{23} \\ 0 & 0 & 0 & 1 \end{array}\right] = \left[\begin{array}{c|ccc} p_{00} & 0 & p_{02} & 0 \\ \hline 0 & & & \\ 0 & & \underline{\underline{P}} & \\ 0 & & & \end{array}\right]$$

FIGURE 3.5.3

Expansion of the On-Time State

Figure 3.5.3. This partitioning implies that the determinant $|\underline{\underline{I}}-\underline{\underline{P}}'z|$ can be expressed in terms of the un-expanded process as

$$|\underline{\underline{I}}-\underline{\underline{P}}'z| = (1-p_{00}z)|\underline{\underline{I}}-\underline{\underline{P}}z|.$$

Hence, an additional positive decay factor of p_{00} affects the transient behavior of the process.

The z-transform of the multi-step default probability $\phi_{d3}(n)$ is

$$\phi^g_{d3}(z) = \frac{p_{02}p_{23}z^3(1-p_{11}z)}{(1-p_{00}z)|\underline{\underline{I}}-\underline{\underline{P}}z|}$$

$$= \frac{p_{02}p_{23}z^3(1-p_{11}z)}{(1-p_{00}z)(1-az)(1-bz)(1-z)}$$

where $a = \frac{1}{2}[p_{11} + \sqrt{p_{11}{}^2 + 4p_{12}p_{21}}]$

$b = \frac{1}{2}[p_{11} - \sqrt{p_{11}{}^2 + 4p_{12}p_{21}}]$.

This can be inverted to give the eigenvalue expansion

$$\phi_{d3}(n) = \begin{cases} 0 & n < 3 \\ p_{02}p_{23}[\alpha_0 + \alpha_1 p_{00}{}^n - \alpha_2 a^n - \alpha_3 b^n] & n \geq 3. \end{cases}$$

where $\alpha_0,\ldots,\alpha_3$ are constant functions of a, b, p_{00}, and p_{11}.

For this case, the dominant components of the transient process are p_{00} and a, where a depends on p_{11} and to a lesser degree on p_{21}.

Summary

This section presents an analysis of the credit granting decision models that considers the important structural elements and parameters of each model. Section 3.5.1 considered the key parameters of Models 1 and 2 as they affect the estimation of expected net present value. The parameters were all seen to be important for this estimation, with the reapplication interval and the diffuseness parameter of the default probability distribution being the least critical.

Section 3.5.2 considered the structure of the Markov process description of delinquent payment behavior. A number of aspects of the process were investigated, including the probability of default, the period in which default occurs, and the occupancy of delinquent states. Transform analysis techniques were employed to determine the relationship of these statistics to the estimated transition probabilities. Finally, the analysis was extended to demonstrate the effect of including additional delinquent states, as well as the effect of expanding the state description to remove second and higher memory in delinquent payment behavior.

Chapter 4

PATTERN RECOGNITION TECHNIQUES FOR DEFAULT PROBABILITY ESTIMATION

The two-outcome models require the estimation of the probability of default, $p(\underline{x},A)$, as a function of a loan applicant's feature vector $\underline{x}$ and the amount of the loan, A. Multivariate estimation problems of this sort frequently occur in "pattern recognition" or classification problems, and consequently research in the pattern recognition field has produced a number of techniques or algorithms that are useful for default probability estimation. In this chapter we will review several of these algorithms and attempt to provide some insight into their essential similarities and differences.

4.1 Problem Statement and Complexity Considerations

Let $p(\theta|\underline{x})$ be the probability that a loan applicant with feature vector $\underline{x}$ is from class θ (e.g., $\theta=0$ for bad loans, $\theta=1$ for good loans). Bayes' theorem gives

$$p(\theta|\underline{x}) \propto p(\underline{x}|\theta)p(\theta). \tag{4.1}$$

That is, $p(\theta|\underline{x})$ can be estimated by first determining the class probability density of $\underline{x}$ conditional on θ, $p(\underline{x}|\theta)$, and then scaling that density estimate by the a priori probability of class θ, $p(\theta)$. This result suggests that we focus on techniques for estimating $p(\underline{x}|\theta)$, and independently resolve the issue of a priori probability determination (this issue is addressed in Chapter 9). The proportionality constant in (4.1) is simply

$$\sum_{\theta} p(\underline{x}|\theta)p(\theta).$$

Complexity Considerations

The choice of a pattern recognition algorithm for estimating $p(\underline{x}|\theta)$ should depend, to some degree, on the "complexity" of the probability density function itself. This concept of complexity and its relation to probability density estimation will be clarified in the discussion that follows.

Consider the following three hypothetical pattern recognition problems characterized by the densities $p(\underline{x}|\theta)$ shown in Figure 4.1. To keep the example simple, we assume equal a priori probabilities and equal misclassification costs. The "well-behaved" unimodal densities of Figure 4.1(a) can be considered less complex than the multi-modal densities of Figure 4.1(b). However, while multi-modal densities are more complex, it is not the density complexity *per se* that contributes to classification errors. For example, the densities of Figure 4.1(c), although exhibiting more modes than those of Figure 4.1(b), present little if any additional problems for deciding class membership.

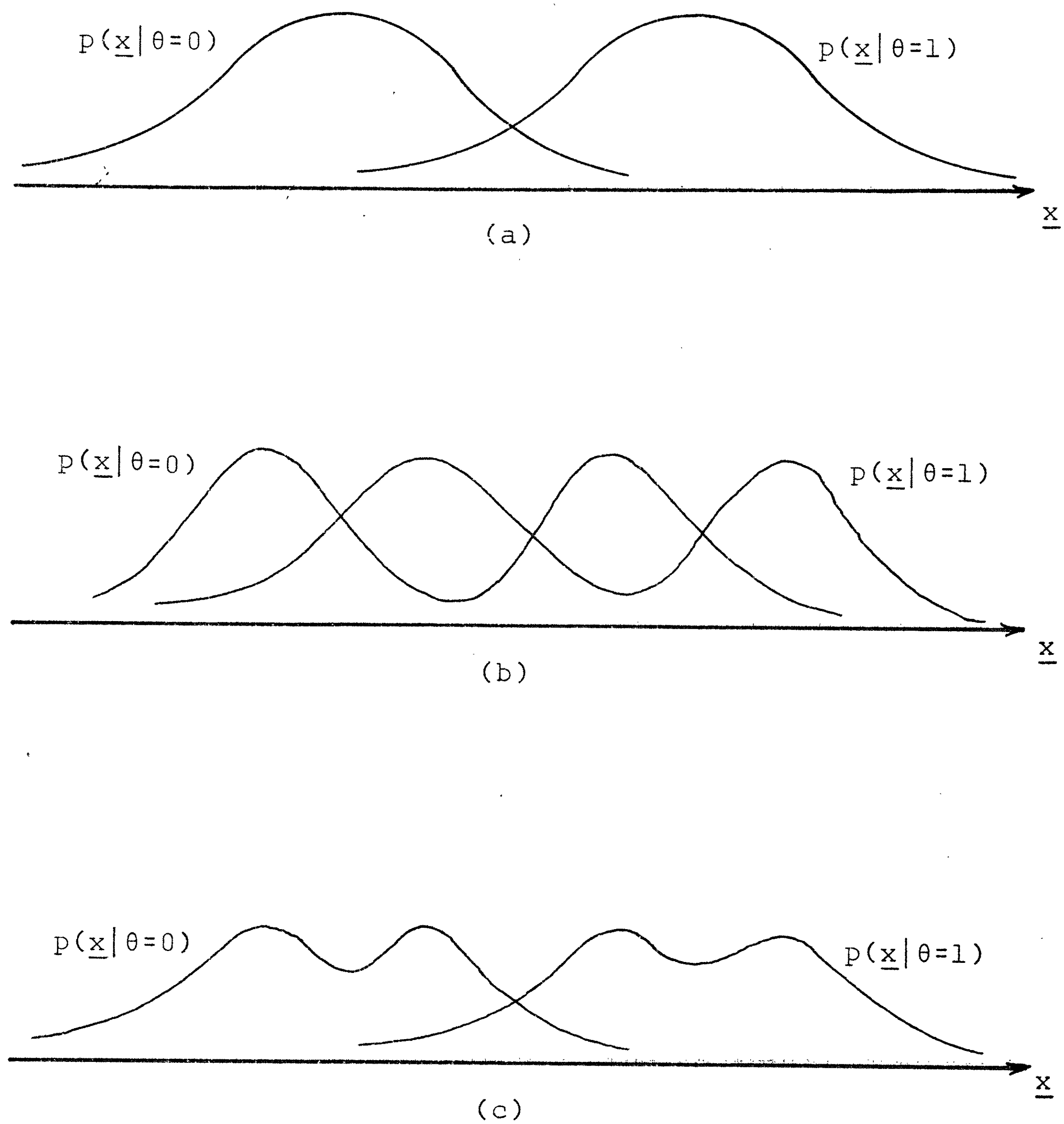


FIGURE 4.1
Densities of Varying Complexity and Interleavedness

This is because the additional modes occur away from the decision boundaries and do not increase the interleavedness of the densities; that is, the number of regions in $\underline{x}$ space needed to specify the decision rule is not increased.

These qualitative notions of individual density complexity and relative density interleavedness are considered to be important factors in the choice of pattern recognition algorithms.[1] Unfortunately, we never have knowledge of the densities themselves, but instead have only a set of samples, $\{(\underline{x}_i, \theta_i): i=1,..,N\}$, for which we know the class membership, θ_i, and the feature vector $\underline{x}_i$.[2] The samples provide the input for a pattern recognition algorithm whose output will be estimates of the class densities.

If the densities $p(\underline{x}|\theta)$ are complex and we are unsure of their degree of interleavedness, we will want to use an algorithm that has the capacity for estimating complex probability densities. This, of course, assumes that we have a sufficient number of samples to distinguish density

[1]Personal communications with Dr. Thomas M. Cover contributed greatly to this discussion of complexity considerations.

[2]We may not even know all of the components of the feature vector. For example, the marital status of a loan applicant may never have been recorded.

complexity from sampling noise. The number of samples required increases with the number of features being considered. Without further research, we can only speculate on how many samples constitute a sufficiently large data base for a particular algorithm.

For the credit granting problem, these uncertainties of complexity, interleavedness, and sample size requirements suggest that we try a number of pattern recognition algorithms of varying complexity. A comparison of the results obtained will provide the best indication of which algorithm most accurately captures the information about θ contained in $\underline{x}$ for the given data set $\{(\underline{x}_i, \theta_i)\}$.

4.2 Pattern Recognition Algorithms

Four different pattern recognition techniques will be used in estimating the probability default $p(\theta=0|\underline{x})$. The techniques are:

K-Nearest Neighbor Rules (K-NNR),[3]

Sebestyen and Edie's algorithm (S & E),[4]

Discriminant Analysis,[5] and

Multiple Regression.

The applicability of each algorithm for probability densities of varying complexity will be discussed, as well as any computational considerations for implementation.

K-Nearest Neighbor Rule (K-NNR)

Let $d(\underline{x},\underline{y})$ be a metric by which we measure the "similarity" between two feature vectors $\underline{x}$ and $\underline{y}$. For example, the standardized Euclidean metric is

$$d(\underline{x},\underline{y}) = \sum_j [(x_j - y_j)/\sigma_j]^2, \tag{4.2}$$

where σ_j is the standard deviation of the j^{th} feature.

Suppose we are trying to estimate $p(\underline{x}|\theta)$ at some point $\underline{x}$ in feature space. K nearest neighbor estimation suggests

---

[3] See Cover [30] and Patrick [37], Chapter 4.

[4] See Sebestyen and Edie [38].

[5] See Anderson [26], Chapter 6, and Patrick [37], Chapter 3.

that we find the K samples of $\{(\underline{x}_i,\theta_i)\}$ that are "nearest" to $\underline{x}$, i.e., those K samples with minimum $d(\underline{x},\underline{x}_i)$. Let k_θ be the number of samples from class θ contained in the tolerance region about $\underline{x}$ out to the K^{th} nearest neighbor. A local estimate (in the neighborhood of $\underline{x}$) for $p(\underline{x}|\theta)$ is then

$$\hat{p}(\underline{x}|\theta) \propto k_\theta/n_\theta, \tag{4.3}$$

where n_θ is the total number of samples from class θ. The proportionality constant is simply the volume of the tolerance region about $\underline{x}$. Since this constant is the same for all classes, it vanishes in the computation of $p(\theta|\underline{x})$ by Bayes' theorem.

If we use K=1 (the 1-NNR), the estimate of $p(\theta|\underline{x})$ will be either 0 or 1 depending on the class of the sample nearest $\underline{x}$. Because $p(\theta|\underline{x})$ is determined by only this single nearest neighbor, the 1-NNR can be considered a very local probability estimation technique. If we use a large value of K (large relative to the sample size), $\hat{p}(\theta|\underline{x})$ will be influenced by samples more distant from $\underline{x}$. This might be considered a more global density estimate, where in the extreme case of K=N we obtain simply the a priori probability $p(\theta)$.

If $p(\underline{x}|\theta)$ is a very complex density we would want to choose a small value of K so as not to "over-smooth" the estimate $\hat{p}(\underline{x}|\theta)$. On the other hand, if $p(\underline{x}|\theta)$ is not very complex (e.g., a normal probability density function) a larger value of K will produce a more accurate estimate $\hat{p}(\underline{x}|\theta)$ by averaging out the sampling noise in the region about $\underline{x}$.

The K-NNR is easily programmed, but unfortunately it is computationally slow. To obtain $\hat{p}(\underline{x}|\theta)$ we must compute $d(\underline{x},\underline{x}_i)$ for all training samples $(\underline{x}_i,\theta_i)$, $i=1,..,N$ and then find the K samples with the smallest values of $d(\underline{x},\underline{x}_i)$. This procedure is particularly slow if the data set is too large to fit in core memory and must be accessed from disk storage.

Performance of the K-NNR for various values of K provides insight into the complexity of the class densities $p(\underline{x}|\theta)$. For this reason, it serves as an appropriate tool for exploratory data analysis.

Discriminant Analysis

Discriminant analysis algorithms assume that the samples are normally distributed in feature space with different class means $\underline{\mu}_\theta$. If the covariance matrices for each class are assumed to be equal, that is $\underline{\underline{\Sigma}}_\theta = \underline{\underline{\Sigma}}$, then the technique is called "linear discriminant analysis" since the decision boundary reduces to a hyperplane in feature space. If the covariance matrices are unequal, then the decision boundary is a quadratic form in $\underline{x}$, and the technique is referred to as "quadratic discriminant analysis."

The probability density functions for both cases given below, where $\underline{\mu}_\theta$ is the mean vector for class θ, $\underline{\underline{\Sigma}}_\theta$ is the covariance matrix for class θ, and $\underline{\underline{\Sigma}}$ is the joint covariance matrix if both classes have equal covariance.

Linear Discriminant Analysis: $\underline{\underline{\Sigma}}_0 = \underline{\underline{\Sigma}}_1 = \underline{\underline{\Sigma}}$

$$p(\underline{x}|\theta) \propto |\underline{\underline{\Sigma}}|^{-1/2}\exp[-(\underline{x}-\underline{\mu}_\theta)^t\underline{\underline{\Sigma}}^{-1}(\underline{x}-\underline{\mu}_\theta)/2] \quad (4.4)$$

Quadratic Discriminant Analysis: $\underline{\underline{\Sigma}}_0 \neq \underline{\underline{\Sigma}}_1$

$$p(\underline{x}|\theta) \propto |\underline{\underline{\Sigma}}_\theta|^{-1/2}\exp[-(\underline{x}-\underline{\mu}_\theta)^t\underline{\underline{\Sigma}}_\theta^{-1}(\underline{x}-\underline{\mu}_\theta)/2] \quad (4.5)$$

The constant of proportionality is $(2\pi)^{-r/2}$ where r is the dimensionality of the feature vector $\underline{x}$.

In order to implement these techniques, we must estimate class mean vectors $\hat{\underline{\mu}}_\theta \equiv \underline{m}_\theta$, class covariance matrices $\hat{\underline{\underline{\Sigma}}}_\theta \equiv \underline{\underline{S}}_\theta$, and joint covariance matrix $\hat{\underline{\underline{\Sigma}}} \equiv \underline{\underline{S}}$. For any $\underline{x}$, the density estimate $\hat{p}(\underline{x}|\theta)$ can be obtained using (4.4) or (4.5). By applying Bayes' theorem we obtain the class probabilities $p(\theta|\underline{x})$.

Discriminant analysis provides optimal density estimates if the samples from each class are actually normally distributed. If the densities are only approximately normal (i.e., not very complex and roughly unimodal) discriminant analysis may still provide relatively good classification results. Since $\underline{\mu}_\theta$ and $\underline{\underline{\Sigma}}_\theta$ (or $\underline{\underline{\Sigma}}$) are estimates for all of feature space, discriminant analysis can be considered a *global* estimation procedure. The algorithms are easily programmed, and density estimates can be computed with little more than one matrix multiplication.

Sebestyen and Edie's Algorithm (S & E)

We have reason to believe that the class densities are perhaps more complex than the multivariate normal densities assumed by discriminant analysis. Market researchers tend to identify several different market segments of the consumer loan population. These segments might be labeled young renters, homeowners, retired persons, and so forth. The implication here is that clusters of loan candidates exist in feature space that are somewhat homogeneous with respect to expected loan payment behavior. If these clusters or market segments do exist, the densities $p(\underline{x}|\theta)$ may be more complex than unimodal multivariate normal densities. It would be desirable to identify these clusters in feature space using a clustering technique, and then estimate a density $p_m(\underline{x}|\theta)$, where m is the cluster index. To obtain $p(\underline{x}|\theta)$ we would find the cluster to which $\underline{x}$ is "nearest" and use that cluster probability density $p_m(\underline{x}|\theta)$.

Sebestyen and Edie [38] suggest such a clustering procedure that "grows" clusters in $\underline{x}$ space, and simultaneously updates cluster densities $p_m(\underline{x}|\theta)$. Training samples of known classification (θ) are introduced sequentially and cluster locations, shapes, and sizes are adaptively determined. The introduction of a new sample point that is "close" to an existing cluster is identified with that cluster. The cluster center is then updated to reflect the influence of the new point $\underline{x}$. If $\underline{c}_m$ is the center of the m^{th} cluster near which $\underline{x}$ falls, and n_m is the number of points previously falling into the cluster, then the updated cluster center becomes

$$\underline{c}'_m = \frac{1}{n_m+1} (n_m \underline{c}_m + \underline{x}).$$

In a similar manner the cluster size and shape (the variance of the points in the cluster along each feature dimension) are also updated.

Training samples that fall far enough away from existing clusters are used to define new clusters with preset initial variance. Some training samples are determined to fall within a "guard zone", that is, neither too distant nor too near existing clusters. These samples are stored until the clustering becomes more well-defined by the introduction of the next batch of samples. At that time they are re-introduced and are more likely to be included in a then existing cluster. At the end of the training process any saved samples are forced to merge with the nearest cluster. Clusters with only a few points can then be merged with nearby clusters to reduce the complexity of the probability density estimate.

This pattern recognition algorithm has the interesting capability of variable complexity. That is, depending on the preset initial variances chosen, "guard zone" thresholds, and post-training cluster merging, probability density estimates of varying complexity can be obtained. In one extreme, if we allow only one sample per cluster, the algorithm functions like the 1-Nearest Neighbor Rule. On the other extreme, if we set the initial variances so large that all samples of a class fall within one cluster, we in effect have the quadratic discriminant algorithm (if distance is measured using the quadratic metric $d(\underline{x},\underline{y}) = (\underline{x}-\underline{y})^t \underline{\underline{\Sigma}}_\theta^{-1} (\underline{x}-\underline{y})$).

The S & E algorithm has the disadvantage of being quite sensitive to these initial parameters, and requires that some degree of art be exercised in its application to obtain density estimates of intermediate complexity (not too many but not too few clusters). Once reasonable ranges of these initial parameters are found, however, it has the advantage of variable complexity, reduced storage requirements relative to K-NNR techniques, and reasonably fast density estimation for samples of unknown classification.

Multiple Regression

Multiple regression techniques can also be used to obtain probability density estimates. If $\{(\underline{x}_i, \theta_i)\}$ is a set of training sample, then a linear model of the form

$$\theta_i = b_0 + \underline{b}^t \underline{x}_i + \varepsilon \qquad (4.6)$$

is solved for minimum mean-square estimates of the coefficients b_0 and $\underline{b}$. Given these estimates, we can obtain a "score", $S(\underline{x})$, for any vector $\underline{x}$ as

$$S(\underline{x}) = b_0 + \underline{b}^t \underline{x}. \qquad (4.7)$$

The training sample is then used with (4.7) to obtain the set of observations $\{[S(\underline{x}_i), \theta_i]\}$. Using this scoring dimensionality reduction, $\hat{p}(\underline{x}|\theta)$ reduces to $p(S(\underline{x})|\theta)$, which can be obtained from the empirical distribution of $S(\underline{x}_i)$ constructed from those samples for which $\theta_i = \theta$.

Because of its simplicity of computation and the familiarity of most researchers with regression analysis, this technique is frequently used to develop linear credit scoring formulas of the form $S = b_0 + \underline{b}^t \underline{x}$. This form represents a linear transformation of feature space to a scalar score variable. Our experience (see Section 9.2) indicates that this score is approximately normally distributed for each class.[6]

Regression analysis bears a close resemblance to discriminant analysis in its linear transformation of feature space and approximately normally distributed score. Both techniques are suitable for estimating probability densities $p(\underline{x}|\theta)$ that are not complex.

[6]Keeping in mind the Central Limit Theorem [of statistics], this might reasonably be expected since score is a weighted sum of a number of features. In fact, if $\underline{x}$ has multivariate normal distribution so will S since it is a linear transformation of $\underline{x}$.

4.3 Evaluating Algorithm Performance

This chapter presents a number of pattern recognition algorithms of varying comlexity that will be used to estimate the a posteriori probability $p(\theta|\underline{x})$ that a sample with feature vector $\underline{x}$ is from class θ. This probability then becomes the input to a decision rule $D[p(\theta|\underline{x})]$ with output $\hat{\theta}$, where $\hat{\theta}$ can be considered a decision-oriented prediction of the outcome θ for the sample $\underline{x}$. For example, $D[\cdot]$ may be the maximum a posteriori probability decision rule of the form

$$\hat{\theta} = D[p(\theta|\underline{x})] = \max_{\theta}[p(\theta|\underline{x})]. \tag{4.8}$$

Or, as indicated in Section 3.2, it could be the multi-period credit granting decision rule given by equation (3.2.9).

Let $R(\theta,\hat{\theta})$ be the reward earned if the decision $\hat{\theta}$ is made and the actual class is θ. We would like to choose a pattern recognition algorithm that gives the maximum expected reward. Letting $\overline{R}_k$ be the expected reward obtained by using the k^{th} pattern recognition algorithm, we would chose the algorithm with maximum $\overline{R}_k$ provided this expected average reward is significantly greater than that obtainable by the other algorithms. If we actually train the set of algorithms on one data set and evaluate them on another data set we will be able to compute average rewards $\overline{R}_k$.

However, if we repeat the experiment using a new data set we may get different average reward estimates due to sampling variations. This implies a degree of uncertainty about any assertion that one particular algorithm will perform better than the rest on independent data. Before going to the expense of implementing one of the algorithms as part of a decision rule, we should want to "statistically test" the validity of this assertion of superior performance. Two tests are proposed below.

Suppose that N samples of known classification are available. Randomly select N/2 samples and label them data set 1, and then label the unselected samples data set 2. Train each of the K algorithms on data set 1 and then compute $\overline{R}_k^{(2)}$, its average reward on the samples of data set 2. Then "criss-cross" the data sets, training on data set 2 and testing on data set 1, to obtain $\overline{R}_k^{(1)}$, the average reward on the sample of data set 1. This procedure gives two sample performance measures for each algorithm, which can be used to test the assertion of superior performance by one particular algorithm.

For one test, we can define the <u>mean reward</u>

$$\overline{R}_k = \frac{1}{2} [\overline{R}_k^{(1)} + \overline{R}_k^{(2)}] \tag{4.9}$$

and the <u>reward range</u>

$$\Delta\overline{R}_k = |\overline{R}_k^{(1)} - \overline{R}_k^{(2)}| \tag{4.10}$$

for each algorithm. Experimental evidence would suggest choosing the pattern recognition algorithm with maximum $\overline{R}_k$. However, if this mean reward exceeds the second greatest mean reward by only a fraction of either reward range, we could not confidently predict that it will outperform the second ranking algorithm on another set of samples.

A second test that might be performed is an Analysis of Variance[7] to test the hypothesis that there is no significant difference among the performance of the set of algorithms. Formally stated, we can test the hypothesis

$$H_0: \quad \overline{R}_1 = \overline{R}_2 = \ldots = \overline{R}_k = \ldots = \overline{R}_K.$$

To apply the analysis of variance test, we will want to use two blocking factors, one for each measure $\overline{R}_k^{(1)}$ and $\overline{R}_k^{(2)}$, and K treatments, one for each algorithm. The F-statistic can be computed for treatments to test H_0. This F-statistic, with (K-1,K-1) degrees of freedom, will call for rejection of H_0 if its value is significantly large. If H_0 is rejected, we can assert that the algorithms do not have the same mean performance.

We should point out, however, that we are not fully justified in using this F-test, since an Analysis of Variance assumes that the reward measurements have the same normal

[7] See Mendenhall [35].

distribution for each algorithm. We know that this assumption will be violated to some degree. A complex algorithm applied to small sample sizes is likely to give more varied rewards from data set to data set than is an algorithm that is less complex. Nonetheless, this analysis may provide some insight into the hypothesis of equal algorithm performance.

Chapter 5

PATTERN RECOGNITION TECHNIQUES FOR TRANSITION PROBABILITY ESTIMATION

The detailed state outcome model requires the estimation of the transition probability matrix $\underline{\underline{P}}(\underline{x},A)$. All Markov process models require transition probability estimation, but, in general, the transition probability matrix is not considered to be a function of a set of variable features. To the best of this author's knowledge, the structure of this estimation problem is not addressed in the published literature of the field. This chapter will demonstrate that this matrix probability estimation problem is solvable using general pattern recognition techniques. In addition, a linear discriminant technique will be shown to yield a computationally efficient means of estimating the transition probability matrix of a Markov process for which it is known that at most two probabilities in each row are non-zero.

Problem Statement

Let $\underline{\underline{P}}(\underline{x})$ be the transition probability matrix of an N-state discrete-time Markov process. Assume that $\underline{\underline{P}}(\underline{x})$ is first-order and stationary for all $x \in X \subset E^r$. Let $p_{ij}(\underline{x})$ be the $(i,j)^{th}$ element of $\underline{\underline{P}}(\underline{x})$.

Let $\hat{\underline{\underline{P}}}(\underline{x})$ be an estimate of $\underline{\underline{P}}(\underline{x})$, with $\hat{p}_{ij}(x)$ an estimate of $p_{ij}(\underline{x})$. We wish to develop a technique for obtaining $\underline{\underline{P}}(\underline{x})$

from a sample $\{(\underline{x}_k,\underline{s}_k): k=1,..,K\}$, where $\underline{x}_k$ is the feature vector of the k^{th} sample, and $\underline{s}_k$ is the corresponding state occupancy vector; that is, $\underline{s}_k$ is a vector whose n^{th} element is the state at which the k^{th} sample was observed to be at time n. The number of elements in $\underline{s}_k$ depends on the number of periods the sample was observed. This number may vary from sample to sample.

Let $c_{ij}^{(k)}$ be the number of state i to state j transitions observed for the k^{th} sample. Let

$$c_{ij} = \sum_{k=1}^{K} c_{ij}^{(k)}$$

and

$$c_i = \sum_{j=1}^{N} c_{ij}.$$

Result 1

$p_{ij}(\underline{x})$ is independent of $p_{k\ell}(\underline{x})$ for $i \neq k$ and all j and ℓ, $i,j,k,\ell = 1,..,N$.

Proof:

Since $\underline{\underline{P}}(\underline{x})$ is assumed to be first-order for all $\underline{x}$, its rows are independent.

This result allows us to consider $\underline{\underline{P}}(\underline{x})$ row by row, so that instead of a single N x N estimation problem we have N independent N x 1 estimation problems. We will show below that this problem can be transformed into N N-class pattern recognition problems. To help make this transformation more clear, we make the notational correspondences:

$$p_{ij} \equiv p(\theta_i=j)$$

$$p_{ij}(\underline{x}) \equiv p(\theta_i=j|\underline{x}). \tag{5.1}$$

That is, the columns of row i are identified with class indices $\theta_i = 1,..,N$. The probability associated with column j of row i then becomes the probability that the row i class index θ_i takes the value j, with this probability a function of the feature vector $\underline{x}$. This identification leads to the following result.

Result 2

$$\begin{aligned} p_{ij}(\underline{x}) &\equiv p(\theta_i=j|\underline{x}) \\ &\propto p(\underline{x}|\theta_i=j)p(\theta_i=j) \\ &\propto p(\underline{x}|\theta_i=j)p_{ij} \end{aligned} \tag{5.2}$$

where $p(\theta_i=j) \equiv p_{ij}$ is simply the a priori transition probability from state i to state j.

Proof:

Result 2 follows directly from (5.1) and Bayes' theorem.

Based on Result 2, an estimate for $p_{ij}(\underline{x})$ is

$$\hat{p}_{ij}(\underline{x}) \propto \hat{p}(\underline{x}|\theta_i=j)\hat{p}_{ij}. \tag{5.3}$$

Given a reasonably large sample we can estimate the a priori transition probabilities p_{ij} by the relative frequency maximum likelihood estimate

$$\hat{p}_{ij} = c_{ij}/c_{i}. \tag{5.4}$$

The quantity $\hat{p}(\underline{x}|\theta_i=j)$ is the estimate of a multi-variate probability density that can be obtained as follows.

Obtaining $\hat{p}(\underline{x}|\theta_i=j)$

For each sample $(\underline{x}_k,\underline{s}_k)$, k=1,..,K let $w_k(i,j)$ be the number of state i to state j transitions given by $\underline{s}_k$. Then consider $\underline{x}_k$ to be a point in feature space with "weight" $w_k(k,j)$, or equivalently consider there to be $w_k(i,j)$ points in feature space at position $\underline{x}_k$. Given this set of weighted samples, $\hat{p}(\underline{x}|\theta_i=j)$ can be obtained by any of a number of pattern recognition (multivariate probability density estimation) techniques. One such technique, namely linear discriminant analysis, will be illustrated below.

Estimating $\underline{\underline{P}}(\underline{x})$ by Linear Discriminant Analysis

Linear discriminant analysis provides a computationally efficient means of estimating $p(\underline{x}|\theta_i=j)$ which can be used in the manner described above to obtain $\underline{\underline{P}}(\underline{x})$. This technique assumes that for given i and j, $\underline{x}$ is normally distributed with mean vector $\underline{\mu}_{ij}$ and covariance matrix $\underline{\underline{\Sigma}}_i$. That is,

$$p(\underline{x}|\theta_i=j) \propto \exp[-1/2(\underline{x}-\underline{\mu}_{ij})^t \underline{\underline{\Sigma}}_i^{-1}(\underline{x}-\underline{\mu}_{ij})] \tag{5.5}$$

Given the set of "weighted" samples $\{\underline{x}_k, w_k(i,j)\}$ we can obtain estimates $\underline{m}_{ij}$ and $\underline{\underline{S}}_i$ for $\underline{\mu}_{ij}$ and $\underline{\underline{\Sigma}}_i$ respectively. Given these estimates we can use (5.5) to compute $\hat{p}(\underline{x}|\theta_i=j)$ and then use this value in (5.3) to obtain $\hat{p}_{ij}(\underline{x})$.

If the transition matrix $\underline{\underline{P}}(\underline{x})$ is known to have at most two non-zero elements in each row, we can further develop this estimation procedure to produce a concise computational formula. Let j1 and j2 be the non-zero columns of row i. Since the matrix $\underline{\underline{P}}(\underline{x})$ is stochastic, $p_{i,j1}(\underline{x}) = 1 - p_{i,j2}(\underline{x})$.

If we now take the natural logarithm of the ratio of $p_{i,j1}(\underline{x})$ to $p_{i,j2}(\underline{x})$ we obtain, using (5.3), (5.4), and (5.5) and simplifying,

$$\ln[p_{i,j1}(\underline{x})/p_{i,j2}(\underline{x})] = (\underline{m}_{i,j1} - \underline{m}_{i,j2})^t \underline{\underline{S}}_i^{-1}\underline{x}$$

$$- \frac{1}{2}(\underline{m}_{i,j1} - \underline{m}_{i,j2})^t \underline{\underline{S}}_i^{-1}(\underline{m}_{i,j1} + \underline{m}_{i,j2}) + \ln[c_{i,j1}/c_{i,j2}]. \tag{5.6}$$

Defining

$$\underline{d}_i^t \equiv (\underline{m}_{i,j1} - \underline{m}_{i,j2})^t \underline{\underline{S}}_i^{-1} \tag{5.7}$$

$$b_i \equiv -\frac{1}{2}(\underline{m}_{i,j1} - \underline{m}_{i,j2})^t \underline{\underline{S}}_i^{-1}(\underline{m}_{i,j1} + \underline{m}_{i,j2}) \tag{5.8}$$

we obtain

$$\ln[p_{i,j1}(\underline{x})/p_{i,j2}(\underline{x})] = \underline{d}_i^t\underline{x} + b_i + \ln(c_{i,j1}/c_{i,j2}) \equiv \ell_i. \quad (5.9)$$

Since $p_{i,j1}(\underline{x}) + p_{i,j2}(\underline{x}) = 1$, it follows that

$$p_{i,j1}(\underline{x}) = \frac{e^{\ell_i}}{1 + e^{\ell_i}} \; ; \; p_{i,j2}(\underline{x}) = \frac{1}{1 + e^{\ell_i}}. \quad (5.10)$$

With N states and $\underline{x}$ an (rxl) vector, a computer program that will estimate these probabilities must store the following numbers of parameters for each row:

$\underline{d}_i^t$	r
$b_i + \ln(c_{i,j1}/c_{i,j2})$	1
	r+1 total parameters.

Thus storing N(r+1) parameters and performing slightly more than one matrix multiplication are all that is required to obtain $\hat{\underline{\underline{P}}}(\underline{x})$.

To summarize this procedure, $\underline{\underline{P}}(\underline{x})$ is estimated row by row. For row i, the set of weighted samples $\{\underline{x}_k, w_k(i,j)\}$ is used to obtain the sample mean vectors $\underline{m}_{ij}$, j=j1,j2 and joint sample covariance matrix $\underline{\underline{S}}_i$. The i^{th} row discriminant vector $(b_i, \underline{d}_i^t)$ is then computed from $\underline{m}_{i,j1}$, $\underline{m}_{i,j2}$, and $\underline{\underline{S}}_i$ using (5.7) and (5.8) For any choice of feature vector $\underline{x}$ and a priori transition counts $c_{i,j1}$ and $c_{i,j2}$, the transition probability estimates $\hat{p}_{i,j1}(\underline{x})$ and $\hat{p}_{i,j2}(\underline{x})$ are determined from (5.9) and (5.10) by only $r+1$ multiplications and one exponentiation, where r is the number of features.

PART II

APPLICATION TO INSTALLMENT LOAN CREDIT GRANTING

Part II presents the results of an application of the decision models developed in Part I. The analysis presents results obtained from installment loan data provided by The National Shawmut Bank of Boston. This case study serves as an outline of the steps to be followed in implementing the set of decision models.

Chapter 6

INTRODUCTION TO THE CASE STUDY

The models presented in Part I are applied to actual case histories of installment loans. The purpose of the case study is threefold. First, the application serves as an outline of the modeling and estimation problems that are encountered in the course of an application of the models. Second, the empirical analysis provides additional insight into the nature of the decision models. And third, the successful application of the models to actual data demonstrates their value for consumer credit decision making.

Section 6.1 outlines the requirements for the practical application of the models for real-time decision making. Particular attention is paid to the multivariate estimation problems that are inherent in the model descriptions.

The data for the study were supplied by The National Shawmut Bank of Boston (NSB). A description of the data sets is given in Section 6.2. The decision models themselves were implemented on NSB's time-shared computer system.

6.1 Requirements for Application

The process of making a credit granting decision requires that consideration be given to

1) outcome rewards,

2) outcome probabilities, and

3) other parameters of the decision model being used.

In the case of Models 1 and 2, outcome rewards are given by the expected net present values of default and non-default. The computation of these rewards for the installment loan decision is outlined in detail in Appendix A. These rewards depend on a number of loan parameters (e.g., loan term, interest rate, etc.) and must be computed for each decision. Decision rules that use the same average reward values for all decisions result in an unnecessary loss of information about individual loan applicants. This is often the case with decision rules of the "credit scoring" variety.

Outcome probabilities are estimated as a function of the applicant's features or characteristics. The determination of which features are most relevant for predicting the outcome is often referred to as "feature selection". One method for selecting a set of features is presented in Chapter 7. Feature selection requires a data base consisting of loans whose outcomes are known. For each loan in the sample, the value of each potentially relevant feature must be ascertained.[1]

Once the feature set is selected, default probabilities can be estimated by any of a number of multivariate estimation techniques. As illustrated in Chapters 8 and 9, this estimation procedure requires a "training data set" whose samples consist of loans whose outcome is known and for which the selected feature values are recorded. In order to test the reliability of the default probability estimates, an independent "test data set" is also required. Typically one set of data is collected and then randomly divided into training and test subsets.

The other parameters that must be considered include the a priori probability of default, and in the multi-period case, the reapplication interval, the reapplication probability, the number of loan periods, and the diffuseness parameter of the Bayesian probability updating rule. A discussion of the relative importance of these parameters was given in Section 3.5.

---

[1]In practice, some of the feature values may be missing. This poses no real conceptual problems for either feature selection or outcome probability estimation. The existence of unknown feature values does, however, require that all statistical routines be modified to account for this possibility.

In the case of Models 3 and 4 outcome rewards take the form of a delinquent state transition reward matrix. Chapter 12 illustrates how these rewards are determined for installment loans. Chapter 16 discusses the important considerations of the reward matrix for revolving credit instruments.

Outcome probabilities are described by a transition probability matrix estimated as a function of the applicant's features. As in the binary outcome case, training and test data sets are required. It is no longer sufficient to know only the default or non-default outcome of the loans in the training data set. As presented in Chapter 5 and illustrated in Chapter 11, the multivariate estimation of transition probability matrices requires that we know <u>actual transition counts</u> for each loan sample. However, since Models 3 and 4 assume that the outcome states represent a <u>first-order</u> stationary Markov chain, <u>actual transition histories</u> are required to verify that the proposed state description satisfies these assumptions. The necessity for actual transition histories is made clear in the analysis of Sections 11.1 and 11.2.

The other parameters required by Models 3 and 4 include the a priori transition probabilities (Section 11.3), the discount factor, and a specification of the macro-period length and number of macro-periods to be considered (Chapters 12 and 13).

6.2 Data Description

Three sets of data are available for testing the credit granting decision models. The first set of data, which will be labeled Data Set A, was collected in 1970 for use in developing NSB's credit scoring formula. The second set of data, Data Set B, is a payment history record from July, 1972, to January, 1974, of a sample of installment loan accounts. The third set, Data Set C, is a complete sample from applications that were processed from June, 1972 to August, 1972.

Data Set A

Data Set A, collected in 1970, includes 672 defaulted or "bad" loans and 1009 non-defaulted or "good" loans. The bad loans are a 100 percent sample of all loans that defaulted from 1967 to 1969; the good loans represent a random sample of accounts that did not default during the same period. Twenty-two features (or variables) are recorded for each loan in the sample. The NSB credit score was also computed and becomes the twenty-third feature. A brief description of each feature is given in the following Table 6.1 with more detailed descriptions given in Appendix B.

TABLE 6.1

List of Features in Data Set A

1. Occupation
2. Years at Occupation
3. Loan Amount
4. Term
5. Purpose
6. Marital Status
7. Age
8. Dependents
9. Own/Rent
10. Years of Residence
11. Income
12. Mortgage/Rent
13. Total Debt
14. Telephone
15. Years at Former Residence
16. Years with Former Employer
17. Other Income
18. Checking
19. Savings
20. Auto
21. Total Monthly Payments
22. Ability to Pay Ratio
23. NSB Credit Score

Unfortunately, many of the sample loans contain missing observations in one or more of the features. When a particular feature was not known (such as years at former residence), it has been coded with a value of 999 so that all statistical routines can process it accordingly.

Data Set A is used in estimation and evaluation of the pattern recognition algorithms used in conjunction with Models 1 and 2. It was randomly divided into two subsets of equal size, Data Set A1 and Data Set A2. Because the nearest neighbor pattern recognition techniques require equal numbers of bad and good loans, and only 672 bad loans are available, both data subsets contain 336 bad loans and 336 good loans.

It is interesting to note that the feature set includes an assortment of feature types. For example, feature 9 (own/rent) is a 0/1 feature. Feature 1 (occupation class) is a grouped variable with a small number of discrete values. And feature 3 (loan amount) is an approximately continuous variable. This combination of feature types makes for an interesting application of general pattern recognition techniques.

Data Set B

Data Set B was collected for the purpose of testing Models 3 and 4. The delinquency behavior of a sample of NSB installment loan accounts was monitored from July, 1972 to January, 1974. The sample includes any loan granted after June, 1972 for which credit scoring information was available (i.e., for which the demographic features listed above were previously recorded). In July, 1972, approximately 15 percent of all loans entering the Collection Department were being monitored. By January, 1974, with more outstanding loans having been made after June, 1972, well over 50 percent of the loans entering the Collection Department were being monitored.

These manually collected payment history records were placed into a computer data file and transformed into a state occupancy history for each account, with the account number retained as a key field for each record.

For the same set of accounts, 15 credit scoring variables, keyed by the account number, were contained in separate data file. These features were recorded and stored at the time the loan was granted by NSB's installment loan management information system. These 15 features, essentially a subset of the features of Data Set A, are listed in Table 6.2.

TABLE 6.2

List of Features in Data Set B

1. purpose group
2. age
3. telephone
4. occupation risk class
5. net monthly income
6. years employed
7. former employer
8. other monthly income
9. checking/savings account
10. auto
11. owner: mortgage x years of residence
12. renter: years of residence
13. ability to pay
14. loan amount
15. loan term (months)

A sort-merge operation was then performed by matching account numbers from the state-occupancy history file and the credit scoring feature file to produce a merged file with both sets of information. The resulting file contains 956 delinquent loan accounts.

Once again, since some of the credit scoring variables may not have been ascertained at credit granting time, the missing observation problem presents itself with Data Set B as well.

As opposed to Data Set A, only discrete (i.e. quantized) features are present in the data (with the exception of loan amount and term) since the NSB credit score worksheet translates all continuous variables into one of a small number of intervals.

Data Set C

Data Set C consists of the 15 credit scoring variables of Table 6.2 for the 2716 installment loans that were applied for during June, July and August, 1972. It was collected for the purpose of estimating the a priori probability of default (Section 9.1). Data Set CR is a sub-sample of the loans in Data Set C.

Data Set CR

Data Set CR consists of the 15 credit scoring variables of Table 6.2 recorded for 63 loan applications that were rejected by NSB during June, 1972. This data set has been collected to complement Data Set B, which includes only loans that had been previously accepted by NSB. Data Set B and CR together will permit an evaluation of Model 4 using accepted loans of good and poor quality as well as rejected loans of unknown quality.

Chapter 7

FEATURE SELECTION - THEORY AND APPLICATION

Any time that the samples of a data set have high dimensionality, the pattern recognition researcher will be considerably slowed in his analysis. This holds true for the credit granting data set in particular. As we have seen, the loan samples have 23 features. This high dimensionality makes for excessive computer storage problems and perhaps loss of performance. Performance can be adversely affected because fewer experiments with 23 features can be conducted than with, say, ten features; and if the ten features are carefully chosen so as to contain most of the information about the sample classes, then by increased experimentation the researcher may be more likely to obtain a superior classification performance.

We then are faced with the problem of which features to select. The pattern recognition literature contains a number of suggested techniques for feature selection. Many, if not most of these techniques are suggested by characteristics of linearly-structured classification models. For example, one can choose those features which have the largest t-statistics, since with normally distributed samples a large t-statistic implies wide separation of the class means. Factor analysis (also known as principal component analysis, eigenvector projection, or Karhunen-Loevè expansion) is another such attempt to reduce the number of features while paying particular attention to the correlations among the original features. Once again this form of analysis is motivated by an underlying hypothesis of normally distributed samples.

The hypothesis of normality is not the only disconcerting aspect of these feature selection techniques. More seriously they do not allow for local behavior, but rather lead one to choose features that contain only global information. To illustrate this point, consider the following two class-two feature problem. Features F1 and F2 are negligibly correlated. For both classes C1 and C2, F1 is marginally normal with moderately different means for each class. Feature F2, on the other hand, is widely distributed such that the class means in the direction of F2 are approximately equal. However, F2 is such that it provides almost perfect discrimination between the classes because there are intervals along the F2 axis that contain either samples from C1 or C2, but not both. This multi-modal nonlinear nature of F2 would not be recognized by the above mentioned feature selection techniques, yet if any feature is to be selected, it should be F2 and not F1.

We mentioned previously that one of the aims of this research is to investigate the credit granting problem without restricting the analysis to linear models. In keeping with this goal, a feature selection technique has been devised which is not restricted to global characteristics of the class probability densities. The criterion for feature selection is not the more common one of separation of classes in feature space, but one of separation of classes in the actual outcome space. That is, we wish to choose features that most correctly classify the samples with respect to their class membership, and to choose them in a manner which is flexible enough to allow for local as well as global density behavior.

The technique is based on the information theory concept of average information. Consider the j^{th} feature X_j which may take on discrete or continuous values. X_j can be used as a predictor of the class (θ) of any particular sample. We would like a measure of how well, on the average, X_j resolves the uncertainty about θ, or in other words, how much mutual information there is between X_j and θ. This measure will be defined below, but first we must partition the range of X_j.

If X_j is continuous (or takes on a large number of discrete values), we first partition it into a manageable number of intervals. Of course, if X_j takes on a small number of discrete values, we have no need to partition it further. Let K be the number of intervals in the partition of X_j. Let $\{(\underline{x},\theta)_i: i=1,..,N\}$ be a set of N samples with known classification θ. Now define $p_{k|\theta}$ to be the conditional probability that a sample from class θ will have the value of X_j fall within the k^{th} interval of its partition. Let p_k be the marginal probability that X_j falls in the k^{th} interval regardless of class, and let p_θ be the a priori probability of class θ.

The mutual information[1] of X_j and θ is defined to be

$$I(X_j,\theta) \equiv \sum_k p_k \sum_\theta p_{k|\theta} \log p_{k|\theta} - \sum_\theta p_\theta \log p_\theta . \tag{7.1}$$

$I(X_j,\theta)$ will be zero if $p_{k|\theta} = p_\theta$ for all k, i.e., if the interval into which X_j falls does not alter the probability of θ given the k^{th} interval (which, by Bayes' rule, is proportional to the probability of the k^{th} interval of X_j given θ). $I(X_j,\theta)$ will take on the maximum value of

$$H(\theta) = -\sum_\theta p_\theta \log p_\theta \tag{7.2}$$

if one of the $p_{k|\theta} = 1$ for each class θ, where $H(\theta)$ is the "uncertainty" existing about θ before X_j is known.

In choosing the intervals for quantitative X_j we should be careful not to make them too wide, thus averaging out any

---

[1]See Fano [31], Chapter 2.
This logarithmic information measure is considered appropriate for the feature selection problem since it provides a measure of statistical constraint. That is, the information provided about θ by the pair of features X_i and X_j is equal to the sum of the information provided about θ by X_i and the information provided about θ by X_j when X_i is known. If X_i and X_j are statistically independent then the information about θ provided by the pair X_i and X_j is simply the sum of the information provided by X_i and the information provided by X_j. These properties hold even in a non-metric feature space, e.g., if X_i and X_j are qualitative variables.

local behavior in the density $f(X_j|\theta)$. For qualitative (nominal) features, we have a natural interval definition already given. Having chosen the intervals, and given a set of samples, we can evaluate $I(X_j,\theta)$ for all features $j=1,..,23$. The features can then be chosen in order of decreasing information content.

For quantitative features, $I(X_j,\theta)$ was computed using intervals obtained by dividing the range of X_j into 16 intervals of equal length. A random sample of 336 bad loans ($\theta=0$) and 504 good loans ($\theta=1$) from Data Set A was used in the computation. So as not to favor the classification of good loans over bad loans, a priori probabilities were set to be $p_\theta = .5$ for $\theta=1,2$. The results are summarized in Table 7.1.

Feature 0 is actually θ, the class index which is a perfect discriminator. Since natural logarithms were used, the maximum information content is observed to be

$$H(\theta) = -\sum_{\theta} p_\theta \log p_\theta = -[.5 \ln(.5) + .5 \ln(.5)] = .693.$$

The above ordering of features might be referred to as a "first-order" selection method, since we selected the features one by one without taking into account feature interactions. Ideally we would like to be able to choose, for any number of features r, the set of features $X_{j1}, X_{j2},..., X_{jr}$ which maximizes the average mutual information between θ and the jointly considered set $X_{j1}, X_{j2},..., X_{jr}$. Since the

j	$I(X_j;\theta)$
0*	0.693
23	0.207
16	0.167
2	0.158
15	0.121
10	0.111
7	0.038
18	0.037
9	0.035
11	0.032
1	0.030
19	0.029
12	0.027
3	0.025
22	0.020
20	0.018
17	0.015
6	0.013
8	0.012
5	0.011
14	0.011
13	0.010
4	0.009
21	0.004

Table 7.1

Information Content of Each Feature in Data Set A

*Feature 0 is the class index.

combinatorial nature of this problem makes its solution computationally infeasible, we must look for alternative heuristic methods.[2]

Consider the following "r^{th}-order" method.

Step 1: Choose the feature j1 with greatest $I(X_{j1};\theta)$.

Step 2: Choose the feature j2 which maximizes $I(X_{j2}, X_{j1};\theta)$; that is, the second feature will be one with the greatest information conditional on the previously chosen features.[3]

Step 3: Choose the feature j3 which maximizes $I(X_{j3}, X_{j2}, X_{j1};\theta)$.

Continue in this manner until r features are selected.

Unfortunately, the number of cells in the joint partition of $X_{j1}, X_{j2}, \ldots, X_{jr}$ grows exponentially with r. This makes computation of $I(X_{jr}, \ldots, X_{j1};\theta)$ impractical.

---

[2]One such heuristic approach presented in Christensen [28,29] closely resembles the methods presented here. Christensen uses an information measure to partition the feature space and then estimate class probabilities.

[3]

$$I(X_{j2}, X_{j1};\theta) \equiv \sum_k p_k \sum_\theta p_{k|\theta} \log p_{k|\theta} - \sum_\theta p_\theta \log p_\theta$$

where k now indexes a region in the (X_{j2}, X_{j1}) plane. The following relationship holds:

$$I(X_{j2}, X_{j1};\theta) = I(X_{j2};\theta|X_{j1}) + I(X_{j1};\theta).$$

As a further approximation to this "r^{th}-order" method consider the following heuristic which uses only "2^{nd}-order" information.

Step 1: Choose the feature j1 with greatest $I(X_{j1};\theta)$.

Step 2: Choose the feature j2 with greatest $I(X_{j2}, X_{j1};\theta)$.

Step 3: Choose the feature j3 with greatest $I(X_{j3}, X_{j2};\theta)$.

And so forth.

This second order heuristic results in the ordering of features in Table 7.2.

The issue now remaining is how many features to choose? From Table 7.1 of first order information, it is clear that we want to include down through feature 10, years of residence. For reasons of making the resulting model more intuitive to its end users, features 11 (income) and 7 (age) are also included. Feature 3 (loan amount) is also included because it appears in the profitability equations.

In their original order, the features selected by the information measure process are summarized in Table 7.3 and are renumbered for reference within the feature sub-set.

The selection of features by this second-order heuristic may be sensitive to the first feature selected. Our confidence in the feature sub-set would be increased if we found that choosing the second most informative feature to start

j	$I(X_j;\theta \mid X_i)$	$I(X_j;\theta)$
23	0.000	0.207
16	0.117	0.167
12	0.055	0.027
2	0.162	0.158
15	0.043	0.121
1	0.041	0.030
10	0.113	0.111
11	0.038	0.032
7	0.054	0.038
17	0.057	0.015
3	0.060	0.025
22	0.060	0.020
8	0.057	0.012
13	0.050	0.010
21	0.047	0.004
9	0.054	0.035
19	0.024	0.029
18	0.030	0.037
20	0.012	0.018
4	0.015	0.009
5	0.034	0.011
6	0.025	0.013
14	0.009	0.011

Table 7.2

Feature Ordering Obtained Using Second Order Information

		Order of Selection	Original Number
1.	Occupation	8	1
2.	Years at Occupation	3	2
3.	Loan Amount	10	3
4.	Age	6	7
5.	Years of Residence	5	10
6.	Income	7	11
7.	Mortgage/Rent	9	12
8.	Years at Former Residence	4	15
9.	Years with Former Employer	2	16
10.	NSB Credit Score	1	23

Table 7.3

Features Selected Using Second Order Information the process led to the selection of very nearly the same sub-set. This experiment was tried by starting the selection process with each of the next four features in Table 7.1. The orderings that result are shown in Table 7.4. It can be seen that very nearly the same sub-set appears at the top of the ordering, thus indicating that the selection process is not very sensitive to the starting feature.

As a further check on the feature selection process, we computed the t-statistic for each of the 23 features. These are summarized in Table 7.5. It is interesting to note that those features with the highest information measure also had a large and statistically significant t-value. The t-statistic, however, is a measure computed in feature space, whereas the information measure is computed in outcome space. The high rank order correlation between features selected using t-values and information measures may lead us to speculate that pattern recognition algorithms that use Euclidean metrics may be appropriate for this feature space.

starting feature:	23	16	2	15	10
	16	23	23	23	23
	12	2	16	16	16
	2	15	12	12	12
	15	1	15	22	2
	1	10	1	10	15
	10	11	10	11	1
	11	7	11	7	7
	7	12	7	17	17
	17	3	17	3	11
	3	22	3	22	22
	22	8	22	8	3
	8	17	8	1	12
	13	5	13	13	8
	21	13	21	21	21
	9	21	9	9	9
	19	9	19	19	19
	18	19	18	18	18
	20	18	20	20	20
	4	20	4	4	4
	5	4	5	5	5
	6	6	6	6	6
	14	14	14	14	14

Table 7.4

Feature Orderings Obtained Using Second Order Information and Different Starting Features

j	t-statistic	% of samples for which feature was not recorded
23	19.4	0
2	11.4	5
10	10.2	5
16	9.4	49
15	8.7	36
18	7.9	1
9	7.5	1
7	7.3	1
19	6.9	1
11	6.8	2
1	6.3	1
20	5.3	1
14	4.3	0
12	4.3	5
5	4.0	0
22	3.1	7
3	1.9	0
17	1.8	1
8	.9	0
21	.6	0
4	.4	0
6	.4	0
13	.2	2

Table 7.5

Feature Ordering Using t-Statistic

Chapter 8

PATTERN RECOGNITION RESULTS FOR DEFAULT PROBABILITY ESTIMATION

The pattern recognition algorithms described in Chapter 4 are used to estimate the default/non-default probability density function $p(\underline{x},A|\theta)$, where $(\underline{x},A)$ represents the ten features listed in Table 7.3, and θ is the class index. As indicated in Chapter 4, the default probability $p(\underline{x},A)$ can be written as $p(\theta=0|\underline{x},A)$ and computed using Bayes' theorem,

$$p(\theta=0|\underline{x},A) \propto p(\underline{x},A|\theta=0)p(\theta=0),$$

where the proportionality constant is the reciprocal of

$$p(\underline{x},A|\theta=0)p(\theta=0) + p(\underline{x},A|\theta=1)p(\theta=1).$$

Thus, given estimates $p(\underline{x},A|\theta=0)$ and $p(\underline{x},A|\theta=1)$, and given the a priori probability of default, $p(\theta=0)$, we can determine the default probability $p(\theta=0|\underline{x},A)$.

This chapter concentrates on pattern recognition performance *per se*. That is, our primary concern here is to make accurate estimates of the probability of default. The decision rules for credit granting that use these estimates will be discussed in Chapters 9 and 10. At that time, a priori probabilities are estimated and misclassification costs are determined so that the decision rules result in economically optimal credit granting decision. Our interest in classification performance with a minimum probability of error criterion arises from the belief that if we cannot make accurate classifications, then it is unlikely that we will be able to make economically profitable decisions.

Experimental Design

Data Set A is randomly divided into two data sets of 336 good loans and 336 bad loans each. These data sets are labeled Data Set A1 and Data Set A2. A sample loan from the data set is of the form $(\underline{x}_i, \theta_i)$, where $\underline{x}_i$ is the 10-feature vector of applicant characteristics, and θ_i is the actual default ($\theta_i = 0$) or non-default ($\theta_i = 1$) outcome of the loan.

Initially, the 672 samples of Data Set A1 are used in conjunction with the set of pattern recognition algorithms to provide estimates, $p(\underline{x}|\theta)$, of the class probability densities. The 672 samples of Data Set A2 are then considered one at a time. For the i^{th} sample $(\underline{x}_i, \theta_i)$, its classification is obtained as the class θ_i with maximum likelihood, that is, using the maximum likelihood decision rule

$$\hat{\theta}_i = D[\hat{p}(\underline{x}_i|\theta)] = \max_{\theta}[\hat{p}(\underline{x}_i|\theta)]. \qquad (8.1)$$

The "reward" or "risk" function for predicting $\hat{\theta}_i$ when the sample is from class θ_i is taken to be

$$R(\theta_i, \hat{\theta}_i) = |\theta_i - \hat{\theta}_i| = \begin{cases} 0 \text{ if } \theta_i = \hat{\theta}_i \\ 1 \text{ if } \theta_i \neq \hat{\theta}_i \end{cases}. \qquad (8.2)$$

This "reward" function actually counts the number of errors. The pattern recognition algorithm that minimizes the average reward on the data set is the algorithm that yields the minimum expected probability of error. The decision rule and reward function were purposely chosen to give this <u>minimum probability of error criterion</u> for algorithm performance. This criterion is equivalent to the maximum a posteriori expected value criterion if we assume equal misclassification costs and equal a priori probabilities. The average reward on Data Set A2 for a given pattern recognition algorithm is then computed as

$$\overline{R}_k^{(2)} = \frac{1}{N_2} \sum_{i=1}^{N_2} R(\theta_i,\hat{\theta}_i),$$

where the sum is taken over all $N_2 = 672$ loans of Data Set A2.

The data sets are then "criss-crossed", and $\hat{\theta}_i$ determined using (8.1) for each loan in Data Set A1. The density estimates $p(\underline{x}|\theta)$ are determined, in this case, from the samples of Data Set A2 using each of the pattern recognition algorithms. The average reward for each algorithm on Data Set A1 is then computed as

$$\overline{R}_k^{(1)} = \frac{1}{N_1} \sum_{i=1}^{N_1} R(\theta_i,\hat{\theta}_i),$$

where the sum is now taken over all N_1=672 loans of Data Set A1.

The average rewards are computed as the misclassification error rates of the three pattern recognition algorithms:

1) K-Nearest Neighbor Rule (K-NNR) (K=1,11,55,99),
2) Sebestyen and Edie's algorithm (S & E), and
3) Discriminant Analysis (Quadratic and Linear).

Multiple regression was not considered since it bears a close resemblance to linear discriminant analysis. The algorithms are listed in an approximate order of decreasing complexity. That is, the 1-NNR might be considered to give the most local estimates of the densities $p(\underline{x}|\theta)$, whereas linear discriminant analysis provides the least complex description of the unknown densities. We first present the classification results for the algorithms, then briefly outline the manner in which the algorithms were applied, and finally evaluate the results obtained.

Results

Table 8.1 presents the average error rates, $\overline{R}_k^{(\cdot)}$, on both Data Sets A1 and A2 for each of the algorithms. In addition to the average error rate, the misclassification rates P_{BG} (bad loans classified as good) and P_{GB} (good loans classified as bad) are presented. After reviewing the manner in which these results are obtained, we address the question of superior performance on the part of any one algorithm.

K-Nearest Neighbor Rule

K-NNR estimation is performed using values of K=1, 11, 55, and 99. This range of "smoothing factors" is chosen to provide insight into the relative complexity of the two densities, $p(\underline{x}|\theta=0)$ and $p(\underline{x}|\theta=1)$. The metric is chosen to be the standardized Euclidean distance

$$d(\underline{x},\underline{y}) = \sum_j [(x_j - y_j)/s_j]^2, \qquad (8.3)$$

where s_j^2 is the training sample variance of the j^{th} feature.

In Chapter 4 we mentioned that the K-NNR provides a tool for exploratory data analysis. Its application to both Data Sets A1 and A2 reveals that the density of good loans ($\theta=1$) appears relatively more complex than the density of bad loans ($\theta=0$). That is, a larger smoothing factor (larger K) favors the correct classification of bad loans at the expense of an increased misclassification of good loans. Equivalently,

	Data Set A2			Data Set A1		
	P_{BG}	P_{GB}	$\overline{R}_k^{(2)}$	P_{BG}	P_{GB}	$\overline{R}_k^{(1)}$
1 NNR	.21	.33	.272	.33	.27	.302
11 NNR	.10	.39	.244	.17	.31	.243
55 NNR	.06	.46	.263	.11	.36	.235
99 NNR	.04	.50	.268	.09	.43	.263
S & E	.22	.21	.219	not evaluated		
Quad. Discr.	.04	.43	.237	.19	.29	.240
Lin. Discr.	.12	.31	.214	.27	.22	.249

TABLE 8.1

Pattern Recognition Performance on Data Set A local density estimates (smaller K) result in better performance for good loans, whereas more global estimates (larger K) result in better performance for bad loans. The relative complexity of $p(\underline{x}|\theta=1)$ is perhaps indicative of the existence of clusters of good loans in feature space (e.g., the existence of homogeneous market segments). The evidence that bad loans are not as complex may suggest that the population of loan applicants who default can not be so conveniently described.

Sebestyen and Edie's Algorithm (S & E)

This clustering approach is further motivated by the K-NNR results. Because these results suggest that the class of good loans ($\theta=1$) is relatively more complex, the S & E algorithm parameters are chosen to favor the growth of smaller clusters for class $\theta=1$ and larger clusters for $\theta=0$. Smaller, more numerous clusters are obtained with smaller guard zone thresholds and smaller preset initial cluster variances.

The results presented in Table 8.1 are obtained after using the samples of Data Set A1 to adaptively define 21 clusters for class $\theta=0$ and 48 clusters for class $\theta=1$. The standardized Euclidean metric (8.3) is used. A large number of clusters for $\theta=1$ results in a more complex description of the density estimate $p(\underline{x}|\theta=1)$. When equal clustering parameters are used for each class, the classification performance is significantly affected—the error rate increases from .22 (as in Table 8.1) to almost .30. The implication here is that the advantages of variable complexity of the S & E algorithm can only be realized by an artful approach to its application. The exploratory results obtained with the K-NNR provide some insight into the approach that should be taken.

Discriminant Analysis

Both linear and quadratic discriminant analysis are also used to estimate the densities $p(\underline{x}|\theta)$. Equations (4.4) and (4.5) indicate that $\underline{m}_\theta$, $\underline{\underline{S}}_\theta$, and $\underline{\underline{S}}$ must be estimated. Estimates of the class mean vectors, $\underline{m}_\theta$, and the class covariance matrices $\underline{\underline{S}}_\theta$ are computed from the 336 training samples of class θ. The estimate of the pooled covariance matrix, $\underline{\underline{S}}$, is computed from all 672 training samples.

To obtain a linear discriminant density estimate, $\hat{p}(\underline{x}_i|\theta)$, for the i^{th} test set sample, the estimates $\underline{m}_\theta$ and $\underline{\underline{S}}$ are used with $\underline{x}_i$ in (4.4). To obtain a quadratic discriminant density estimate, the estimates $\underline{m}_\theta$ and $\underline{\underline{S}}_\theta$ are used with $\underline{x}_i$ in (4.5).

Table 8.1 indicates that quadratic discriminant analysis has slightly poorer performance than linear discriminant analysis (its mean error rate is .011 greater). One explanation for this poorer performance might be that quadratic discriminant analysis requires the estimation of two 10-by-10 covariance matrices from 336 samples each (as opposed to a single pooled covariance matrix from 672 samples). These somewhat insufficient sample sizes may result in spurious estimates of many of the covariance elements.

Performance Evaluation

Section 4.3 suggests two methods for evaluating these performance results of the various algorithms. The first method requires the comparison of mean rewards

$$\overline{R}_k = \frac{1}{2}(\overline{R}_k^{(1)} + \overline{R}_k^{(2)})$$

relative to the reward ranges

$$\Delta R_k = |\overline{R}_k^{(1)} - \overline{R}_k^{(2)}|.$$

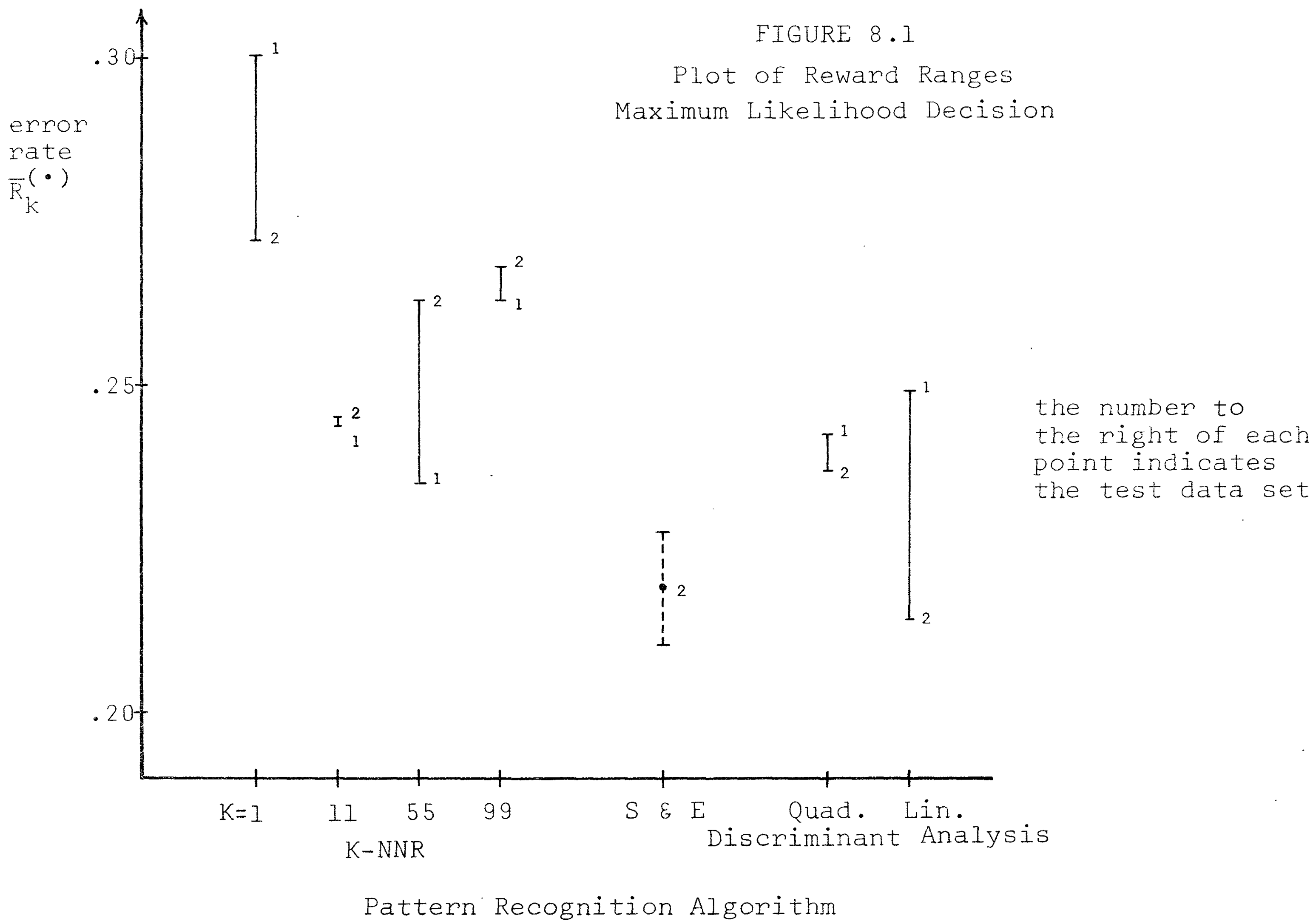
Figure 8.1 graphically presents these reward ranges centered about the mean reward. Since the S & E algorithm was not used on Data Set A1 (due to computer time availability), its reward range is shown as the average of the other six ranges.

The algorithms are presented from left to right in approximate order of decreasing complexity. The performance of all algorithms lies roughly between error rates of .21 and .30, indicating that, in all likelihood, the Bayes' error[1] for the good loan/bad loan pattern recognition problem is quite high (say .15 to .20). Figure 8.1 suggests that the S & E algorithm may slightly outperform linear discriminant analysis (by about .01) but not significantly so.

[1]The minimum classification error, achieved by knowing the class probability densities, is often called the Bayes' error.

FIGURE 8.1

Plot of Reward Ranges
Maximum Likelihood Decision



The analysis of the variance test suggested in Section 4.3 is applied to the results of Table 8.1. The resulting F-statstic for differences in algorithm mean rewards is

$$F_{6,6} = 4.39$$

with (6,6) degrees of freedom.[2] This exceeds the .95 percentile value of 4.28. Although the assumption of equal variance of algorithm performance (i.e., equal reward ranges) does not hold, this result suggests that there is a significant difference among the performance of the various algorithms. The best performance is attained by S & E and linear discriminant analysis.

The F-statistic for differences between mean performance on the two data sets (.250 and .245) is $F_{1,6} = .36$. We thus conclude that there is no significant difference between the samples of the two data sets with respect to our ability to predict class outcomes $\hat{\theta}_i$.

Finally, we should point out that these performance results are obtained with sample sizes of 336 loans of each class. Larger sample sizes might tend to improve the performance of the more complex algorithms relative to the rest,

[2]The S & E rewards were taken to be $\bar{R}^{(1)} = \bar{R}^{(2)} = .219$. This results in a slightly reduced variance of the means and hence a slightly enlarged value for the F-test.

as well as reduce the magnitude of the reward ranges. Since we use all the data available, we are unable to investigate these suggested effects of larger sample sizes. Nonetheless, sample size sensitivity remains a point worth remembering in future pattern recognition experiments of this type.

Chapter 9

MODEL 1 RESULTS

Model 1, the single-period two-outcome model, was presented in Section 3.1. In order to apply the decision model to actual loan data, we need to estimate a number of parameters. Chapter 8 presented the pattern recognition results for Data Set A. These alternative pattern recognition algorithms, along with an estimate of the a priori default probability, will be used to provide estimates for the default probability, $p(\underline{x},A)$. In addition to the default probability estimate, Model 1 requires estimates of $V_0(A,\underline{L})$ and $V_1(A,\underline{L})$, the expected net present value for a defaulted (bad) and repaid (good) loan, respectively. These estimates are given in Appendix A as a function of the loan amount, A, and the loan vector, $\underline{L}$ (term, interest rate, etc.).

In this chapter we evaluate the credit granting performance of Model 1 on Data Set A, and compare its performance to the NSB benchmark credit scoring rule. Before making these evaluations, we must determine an estimate for the a priori probability of default.

9.1 Estimating The A Priori Probability of Default

Since the probability of default given $\underline{x}$ and A, $p(\theta=0|\underline{x},A)$, is proportional to $p(\underline{x},A|\theta=0)\ p(\theta=0)$, we are faced with determining the a priori probability of default $p(\theta=0)$. This determination is non-trivial when we realize that all of the sample loans in Data Set A (good/bad loans) are loans that were once pre-screened by the bank and accepted. This pre-screening is also present in Data Set B (delinquency tracked loans). What is needed then is a set of loan application data that is a true cross-section of all applicants prior to the credit granting decision, and not just loans that were accepted.

For this purpose Data Set C was collected from all installment loans applied for during June, July and August, 1972. These 2716 loans were all scored using the NSB credit scoring formula. Since the probability of default can be expressed as a function of score, S, we have an estimate of $p(\theta=0|S)$ for each of the 2716 loans. From Data Set C we can also obtain the a priori probability, $p(S)$, of observing a score S by using a relative frequency estimate.

Given $p(\theta=0|S)$ and $p(S)$ we can obtain the unconditional probability of default prior to any prescreening as

$$p(\theta=0) = \int_S p(\theta=0|S)p(S)dS. \tag{9.1.1}$$

The estimate obtained using Data Set C is $p(\theta=0) = .048$.

9.2 Benchmark Results

When confronted with a consumer credit granting decision, most large lending institutions apply a quantitative decision rule which takes the form of a weighted credit scoring formula. Currently in use at The National Shawmut Bank is one such scoring formula developed in 1971. A stepwise multiple regression technique[1] was applied to Data Set A to obtain a set of credit score weights for each of 14 features. Weights for selected features are presented in Appendix B, with higher scores indicative of better loan quality (lower default probability).

[1]The regression analysis was conducted in a manner similar to that outlined in Myers and Forgy [21].

When a customer applies for a loan, the values of each of his 14 features are transcribed to a coded credit score worksheet containing scores for each feature. The individual feature scores are then added manually to obtain a total credit score S. The score obtained is then compared to a cutoff score S*, and if S is greater than S* credit is usually granted. If S is less than S*, the application for credit is rejected. The lending officer can override this credit scoring decision, but is required to document his reasons for doing so[2].

This thesis presents a series of decision models that are intended to more accurately model the important aspects of the credit granting decision. In order to consider them to be an improvement, we would expect them to provide decision rules which result in greater overall profit than can be obtained with the best of alternative rules actually in use. The NSB credit scoring formula is one such "state-of-the-art" alternative decision rule. For this reason, it provides a good benchmark model with respect to which Models 1

[2]A high scoring loan, for example, may be rejected because of a derogatory credit bureau report. A low scoring loan may be accepted if it is secured by collateral or is guaranteed by a reliable co-maker.

through 4 can be evaluated. In this section we evaluate the actual performance of this benchmark credit scoring model on Data Set A.

Optimizing the Cutoff Score - Empirical Approach

The weighted sum of the 14 features provides a credit score that falls within the range of -150 to +250. For the loans of Data Set A1 the score is observed to have the following characteristics:

	Bad Loans	Good Loans
mean	-4.2	46.6
standard deviation	34.8	36.7
range	-90 to 96	-55 to 140

The credit scoring decision rule requires a cutoff score S*: the loan application is accepted if its score S exceeds S*, and rejected otherwise. Since the performance of this decision rule is sensitive to the choice of cutoff score S*, the optimal setting of S* was empirically determined using Data Set A1. Cutoff scores were varied over the range -40 to +80, with total profit computed for each cutoff score value. If a loan scores less than or equal to S*, it is rejected ($\hat{\theta}=0$) and has no effect on profit. If it scores greater than S*, it is accepted ($\hat{\theta}=1$) and its "weighted value" is added to total profit. By "weighted value" we mean the value of the loan (which will be explained below) weighted by either $p(\theta=0)$ if the loan was actually a bad loan or by $p(\theta=1)$ if the loan was actually a good loan. Weighting by the a priori probability, $p(\theta)$, provides a realistic total profit figure that is not influenced by relative sample proportions of good and bad loans. As given by the results of Section 9.1, the a priori probability of default, $p(\theta=0)$ was taken to be .048, with $p(\theta=1)$ the complementary probability.

The value of the loan was determined in the manner outlined in Appendix A. In Appendix A the value of a good loan, $V_1(A,\underline{L})$ and the value of a bad loan, $V_0(A,\underline{L})$ are given as a function of the loan amount, A, and the vector of loan attributes, $\underline{L}$ (e.g., term, interest rate, cost of capital, etc.). $V_1(A,\underline{L})$ is simply the net present value of the monthly payment stream discounted at the appropriate cost of capital, minus the loan amount itself and any administrative costs. $V_0(A,\underline{L})$ represents the default loss to the lending institution. The detailed analysis of Appendix A suggests that this loss, appropriately discounted, amounts to about one-third of the original loan amount.

In Data Set A1, for example, the average loan amounts for good and bad loans are \$1368 and \$1112, respectively, with an average term of 20 months. Typical interest rates are currently 13.5%. If we assume a 10% cost of capital and these average loan characteristics, we obtain (using Appendix A) good and bad loan values of:

$$V_1(A,\underline{L}) = \$\ 30.81$$

$$V_0(A,\underline{L}) = -\$435.00$$

To evaluate a given cutoff score S^*, we consider each loan one by one. The two possible values of the i^{th} loan, $V_0(A_i,\underline{L})$ and $V_1(A_i,\underline{L})$, are computed using the actual amount of the loan, an average loan term of 20 months, an interest rate of 13.5%, and a cost of capital of 10%. The loan's score, S_i, is compared with S^*, and if $S_i<S^*$ the loan is rejected $(\hat{\theta}=0)$; if $S >S^*$ the loan is accepted $(\hat{\theta}=1)$. If θ is the actual loan class ($\theta=0$ for a bad loan, $\theta=1$ for a good loan), the effect of the decision $\hat{\theta}$ is given by the reward function $R(\theta,\hat{\theta})$ in Table 9.2.1 below.

		$\hat{\theta}$ = 0	$\hat{\theta}$ = 1
θ	0	0	$V_0(A,\underline{L})$
	1	0	$V_1(A,\underline{L})$

$R(\theta,\hat{\theta})$

Table 9.2.1
Reward Function $R(\theta,\hat{\theta})$ for Deciding
$\hat{\theta}$ When Actual Class is θ

That is, if the credit scoring rule rejects the loan, it has no effect on profit. If it accepts the loan, either a loss of $V_0(A_i,\underline{L})$ occurs if it is a bad loan or a gain of $V_1(A_i,\underline{L})$ is realized if it is a good loan.

Table 9.2.2 presents the results for the range of cutoff scroes from -40 to +100. The two error probabilities P_{BG} (classifying a bad loan as a good loan) and P_{GB} (classifying a good loan as a bad loan) are given, as well as the total profit obtained using S*. Total profit is then normalized to adjust for the discrepancy between the true a priori probabilities

Cutoff Score S*	P_{BG}	P_{GB}	Total Profit	Average Profit per Applicant
-40	.86	.003	3895	11.59
-20	.72	.03	4556	13.56
0	.52	.10	5374	15.99
10	.40	.17	6184	18.41
15	.32	.20	6776	20.17
20	.25	.23	6935	20.64 (max)
25	.21	.27	6817	20.29
30	.17	.31	6738	20.05
35	.13	.36	6704	19.95
40	.09	.41	6750	20.09
50	.04	.52	6171	18.37
60	.02	.61	5046	15.02
80	.003	.82	2595	7.72
100	.000	.93	1045	3.11
perfect information	0	0	10311	30.69
accept all	1.0	0	3168	9.43

TABLE 9.2.2

Benchmark Performance on Data Set A1

as a Function of Cutoff Score

of bad and good loans (.048 and .952) and the actual numbers of bad and good loans in the data set (336 each). This normalized profit will be properly referred to as "average profit per applicant". Average profit per applicant is a more useful measure since the bank can determine total annual profit by multiplying average profit per applicant by the number of loans applied for in one year.

Maximum profit of $20.64 per applicant is achieved with a cutoff score of $S^*=20$. If we had perfect information about each loan (i.e., $\hat{\theta}=\theta$), we would realize an average profit of $30.69. On the other hand, if we accepted all applicants, we would realize an average profit of only $9.43.

Optimizing the Cutoff Score - Analytical Approach

If we know the score probability density for each class, $p(S|\theta)$, and were willing to use an average loan amount, $\bar{A}$, instead of the actual amount for each loan, we could analytically determine S^*. A probability paper plot (Figure 9.2.1) of the cumulative distribution of good loans and bad loans of Data Set A1 indicates that score is approximately normally distributed given θ.

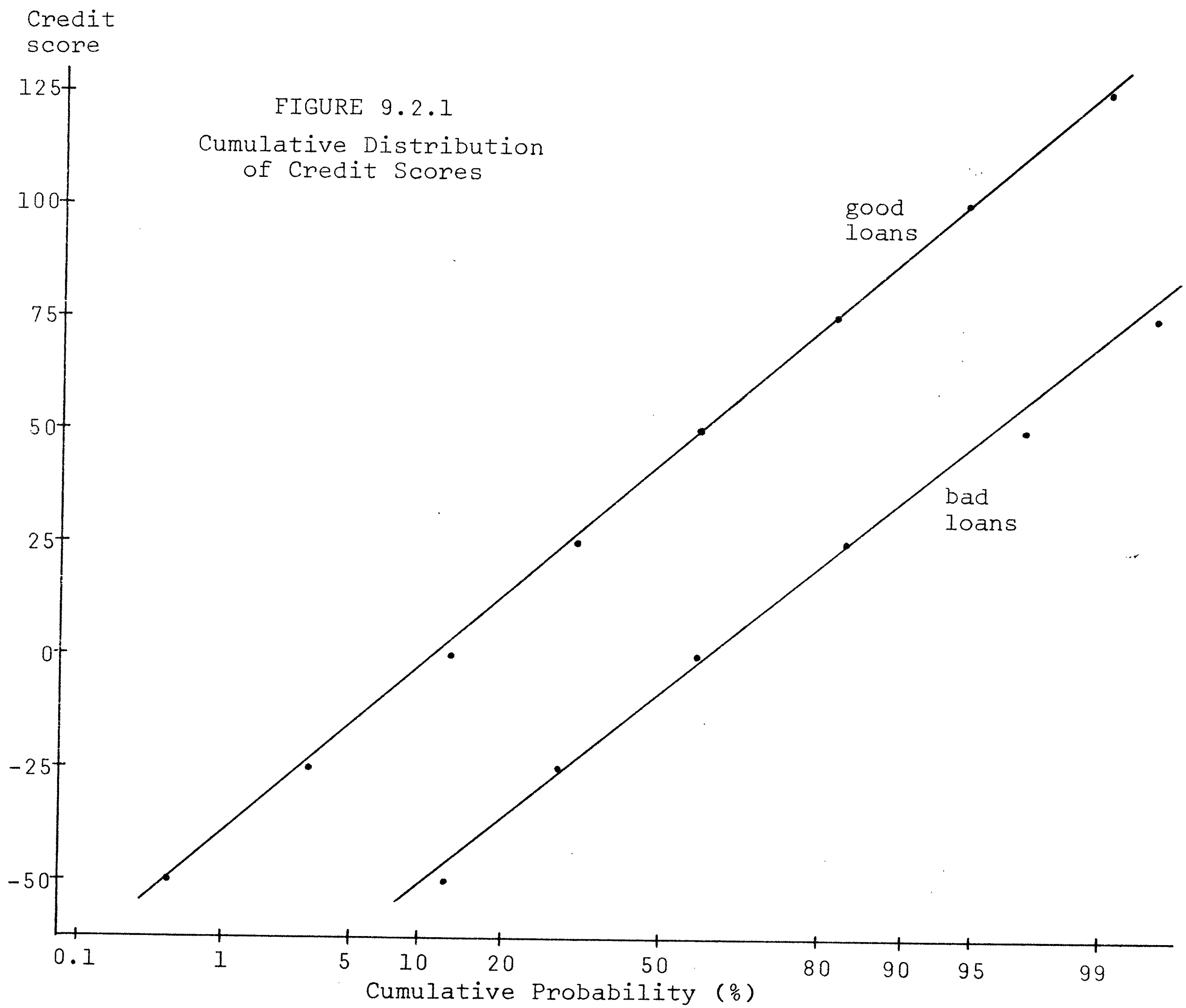


FIGURE 9.2.1
Cumulative Distribution of Credit Scores

Bad loans ($\theta=0$) are observed to have mean score $\mu_0=-4.2$ and standard deviation $\sigma_0=34.8$; good loans have mean score $\mu_1=46.6$ and standard deviation $\sigma_1=36.7$.

For this analysis we will use the following typical loan parameter values:

term	$T = 20$ months
interest	$r = 13.5\%$
cost of capital	$\rho = 10\%$

average bad loan amount $\bar{A}_0 = \$1112$
average good loan amount $\bar{A}_1 = \$1368$

The a priori probability of default will be taken as:

$$p_0 = p(\theta=0) = .048$$

The misclassification costs are:

$$-V_0(A_0,\underline{L}) = \$435.08 \text{ for false acceptance, and}$$

$$V_1(A_1,\underline{L}) = \$\ 30.81 \text{ for false rejection.}^3$$

If $p_i(S) \equiv p(S|\theta=i)$ is the probability density of score given class $\theta=i$ (0 or 1), then then decision rule which minimizes expected costs is:[4]

[3]This form of analysis dictates the use of an opportunity cost for false rejection, since we are assuming no reward is obtained for correct acceptance.

[4]See Anderson [26], Chapter 6.3.

$$\hat{\theta} = \begin{cases} 0 & \text{if } \dfrac{p_0(S)}{p_1(S)} \geq \dfrac{V_1(\bar{A}_1,\underline{L})\ p_1}{V_0(\bar{A}_0,\underline{L})\ p_0} \\ 1 & \text{otherwise} \end{cases} \tag{9.2.1}$$

Substituting the values given above for the right-hand side and assuming $p_0(S)$ and $p_1(S)$ to be the normal density functions with means and variances as given above, this reduces to:

$$\hat{\theta} = \begin{cases} 0 & \text{if } S \leq 13.3 \\ 1 & \text{otherwise .} \end{cases}$$

That is, the analytically determined optimal cutoff score is $S^* = 13.3$.

Because this analytical approach requires that we use an average loan amount instead of actual amounts for each loan in the data set, and because we assumed that scores were normally distributed, we will use the empirically determined cutoff score of $S^* = 20$. The analytical approach, however, does permit a convenient means of analyzing the sensitivity of S^* to the choice of a priori default probability, $p_0 = p(\theta=0)$. Table 9.2.3 gives the cutoff score as a function of $p(\theta=0)$.

$p(\theta=0)$	S^*
.01	-30
.03	1
.04	8
.048	13
.06	19
.08	27
.10	33
.15	44

Sensitivity of Cutoff Score to A Priori Probability of Default

TABLE 9.2.3

Performance on Data Set A2

Using Data Set A1, the optimal benchmark decision rule was determined to be:

$$\hat{\theta} = \begin{cases} 0 \quad \text{(reject) if} \quad S \leq 20 \\ 1 \quad \text{(accept) if} \quad S > 20 \end{cases}$$

This rule was then applied to the loans of Data Set A2, using the same loan parameters of:

$T = 20$ (term in months)
$r = .135$ (interest rate)
$\rho = .10$ (cost of capital)

The results obtained are presented in Figure 9.2.2 below.

Confusion Matrix

Predicted Class ($\hat{\theta}$)

		0	1	Error Rate
Actual Class (θ)	0	252	74	.22
	1	94	261	.28

Total profit on Data Set A2 = $5782.00

Average profit per applicant = $ 17.21

Perfect Information on Data Set A2 yields an average profit of $ 27.37

FIGURE 9.2.2

Benchmark Performance on Data Set A2

Cutoff Score Range

The empirical optimization of cutoff score on Data Set A1 gave the optimal cutoff score as S*=20. If Data Set A2 is used the optimal cutoff score is found to be S*=35. This implies that, for this case study, a sample size of 336 loans from each class only permits the determination of the optimal cutoff score to within a range of 15 points.

When we consider the combined 1344 loans of Data Sets A1 and A2, the optimal cutoff score is determined to be S*=20. The results of this analysis are presented in Table 9.2.4 below.

Cutoff Score S*	P_{BG}	P_{GB}	Profit per Applicant
10	.36	.18	17.28
15	.29	.21	18.69
20	.24	.25	18.92 (max.)
25	.21	.28	18.80
30	.17	.33	18.45
35	.12	.38	18.66
40	.09	.43	18.46
45	.06	.48	18.01
50	.04	.54	17.00

TABLE 9.2.4

Cutoff Score Optimization Using
Data Sets A1 and A2

9.3 Model 1 Pattern Recognition Results

The pattern recognition algorithms described in Chapters 4 and 8 were evaluated with respect to their economic performance using Model 1, the basic single period, two-outcome model summarized in equation (3.1.1):

$$\hat{V} = p(\underline{x},A)\, V_0(A,\underline{L}) + [1 - p(\underline{x},A)]\, V_1(A,\underline{L}). \qquad (3.1.1)$$

$\hat{V}$ is the expected net present value of accepting the loan characterized by its amount A, loan vector $\underline{L}$, and applicant feature vector $\underline{x}$. $V_0(A,\underline{L})$ is the net present value of a defaulted (bad) loan and $V_1(A,\underline{L})$ is the net present value of a non-defaulted (good) loan.

Let θ be the indicator variable for the loan, i.e.

$$\theta = \begin{cases} 0 & \text{if the loan was bad} \\ 1 & \text{if the loan was good,} \end{cases}$$

and let $\hat{\theta}$ be the predicted value of θ, that is, $\hat{\theta}=0$ means we predict that the loan will default and we do not grant credit, and $\hat{\theta}=1$ means we predict the loan will not default and we grant credit. To maximize expected net present value, our decision, as expressed by the value of $\hat{\theta}$, should be:

$$\hat{\theta} = \begin{cases} 0 & \text{if } \hat{V} \le 0 \\ 1 & \text{if } \hat{V} > 0 \end{cases} \qquad (9.3.1)$$

The reward for deciding $\hat{\theta}$ (reject/accept) when the loan belongs to class θ is given by the reward function $R(\theta,\hat{\theta})$ summarized in Table 9.2.1. That is, if we reject the loan ($\hat{\theta}=0$), then there is no profit or loss incurred. If we accept the loan ($\hat{\theta}=1$) and the loan defaults ($\theta=0$), then a loss of $R(0,1) = V_0(A,\underline{L})$ is incurred. If we accept the loan and the loan does not default, then a profit of $R(1,1) = V_1(A,\underline{L})$ is realized.

Chapter 8 described the pattern recognition estimation of $p(\underline{x},A)$, the probability of default ($\theta=0$), for each of the N=672 loans in Data Set A2. Using these estimates we can now evaluate the performance of the various algorithms using profit as a criterion.

Table 9.3.1 presents the performance of the set of pattern recognition algorithms on Data Set A2. The results are given both in terms of total profit on the data set, as well as average profit per applicant. In addition, the two probabilities of error are given, where P_{BG} is the frequency with which a bad loan is predicted to be good (accepted) and P_{GB} is the frequency with which a good loan is predicted to be bad (rejected).

ALGORITHM	P_{BG}	P_{GB}	Total Profit	Average Profit	%	Rank
1 NNR	.16	.42	5095	15.16	55.4	7
11 NNR	.09	.48	5628	16.75	61.2	4
55 NNR	.09	.57	5370	15.98	58.4	6
99 NNR	.08	.61	5432	16.17	59.1	5
S & E	.15	.34	5768	17.17	62.7	3
Quad. Disc.	.03	.51	4708	14.01	51.2	8
Lin. Disc.	.12	.45	5970	17.77	64.9	1
Benchmark (S* = 20)	.22	.28	5782	17.21	62.9	2
Perfect information	0	0	9196	27.37	100.0	-
Accept all	1.0	0	2199	6.55	23.9	-

TABLE 9.3.1

Model 1 Performance on Data Set A2

Because the loan vector $\underline{L}$ is not known for each of the loans of Data Set A, the following typical values were used:

$r = .135$ (simple annual interest rate)

$T = 20$ (loan term in months)

The cost of capital was assumed to be

$\rho = .10$

and the a priori probability of default, as computed in Section 9.1, was taken to be

$p(\theta=0) = .048.$

The column labeled (%) presents the performance as a percent of the average profit that would be obtained if the actual outcome (θ) were known (i.e., if we had perfect information about θ). For comparison purposes, the algorithms are also ranked according to their performance.

As suggested in Chapter 4, we can determine the significance of these results by "criss-crossing" the data sets. This cross-cross experimental design is analogous to that presented in Chapter 8. However, instead of the maximum likelihood decision rule we use the decision rule implied by Model 1. The reward function, $R(\theta,\hat{\theta})$, is given by Table 9.2.1.

Table 9.3.2 presents the results obtained by training the pattern recognition algorithms on Data Set A2 and testing their performance on Data Set A1. The benchmark cutoff score was empirically optimized on Data Set A2, and in this case was found to be $S^*=35$. This implies that for $N=672$ training samples the range of optimal cutoff scores is approximately 15 points ($S^*=20$ on Data Set A1). The reward ranges for the pattern recognition algorithms are plotted in Figure 9.3.1. The rewards are presented as the percent of the reward obtained by perfect information about θ.

Examination of Figure 9.3.1 reveals that no single algorithm significantly outperforms the others. This contention is supported by an anlysis of variance, which gives an insignificantly small F-statistic of $F_{6,6} = .77$ for the hypothesis of no significant difference among the mean rewards. An increased sample size might permit a more accurate training of the more complex algorithms. Judging by the relatively large reward range for the quadratic

Algorithm	P_{BG}	P_{GB}	Total Profit	Average Profit	%
1 NNR	.28	.34	5902	17.57	57.2
11 NNR	.17	.39	6964	20.73	67.5
55 NNR	.14	.43	7114	21.17	69.0
99 NNR	.14	.46	6980	20.77	67.7
S & E		not	evaluated		
Quad. Disc.	.14	.40	7325	21.80	71.0
Linc. Disc.	.24	.37	6693	19.92	64.9
Benchmark (S* = 35)	.13	.36	6704	19.95	65.0
Perfect information	0	0	10312	30.69	100.0
Accept all	1.0	0	3168	9.43	30.7

TABLE 9.3.2

Model 1 Performance on Data Set A1

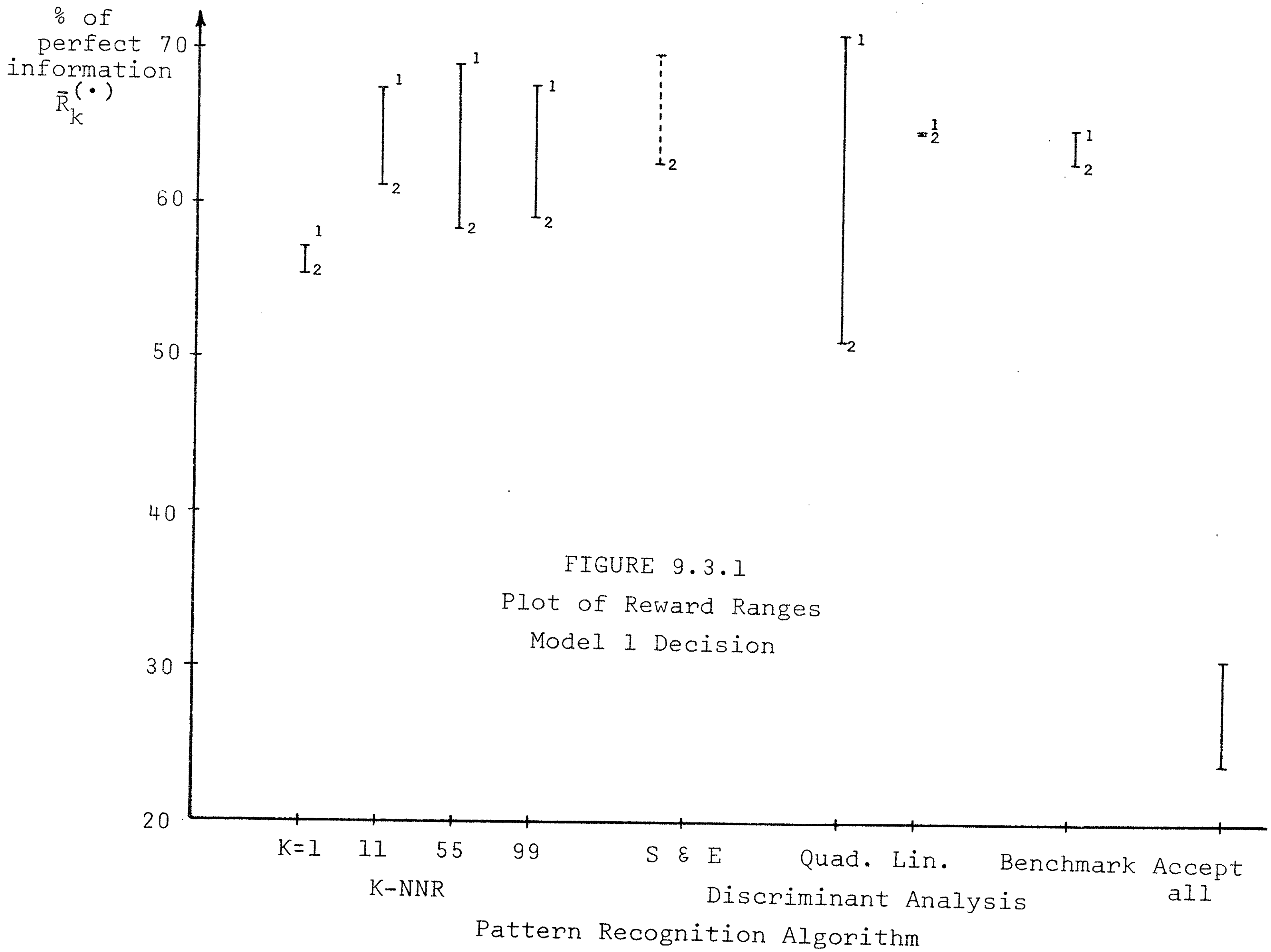


FIGURE 9.3.1

Plot of Reward Ranges

Model 1 Decision

discriminant algorithm, it would appear that 336 loans per class might not be a sufficiently large sample size (given ten features). However, given training data sets of this size, the choice of a pattern recognition algorithm might well be influenced by computational costs and other costs of implementation.

Chapter 10

MODEL 2 RESULTS

The preceding chapter presented the results of a credit granting policy based on a single loan's outcome. However, as pointed out in Section 3.2, the customer's relationship with the lending institution does not end after the first loan if it is successfully repaid. Instead, the customer can be expected to return for subsequent loans. If the first loan was successfully paid, we expect an even greater probability of his repaying a second loan should he apply for it. Similarly, if two loans are repaid, the third is still more likely to be paid. Model 2 is intended to account for these subsequent loans, on a discounted basis, so that the decision to grant credit reflects the applicant's expected multi-loan profit. The assumptions and relationships of Model 2 will be reviewed briefly before presenting the results obtained on Data Set A.

We consider J loans (the loan applied for and J-1 subsequent loans), with loan parameter vector $\underline{L}$ and amount A. $V_0(A,\underline{L})$ and $V_1(A,\underline{L})$ represent the net present value (relative to the time the loan was granted) of a defaulted and non-defaulted loan, respectively. We assume that subsequent loans will be applied for every τ years, with some probability ℓ that the customer will reapply for loan j given he applied for loan j-1. The cost of capital for the lending institution is denoted by ρ.

Let $p_j(\underline{x},A)$ be the probability that the customer defaults in period j given he did not default on the previous j loans. As explained in Chapter 3.2, the probability updating rule is

$$p_j(\underline{x},A) = \frac{n_0 p_0(\underline{x},A)}{n_0 + j} \quad j=1,..,J-1 \tag{3.2.2}$$

where $p_j(\underline{x},A)$ is the probability the customer defaults in period j, and n_0 is the parameter of the beta distribution on $p_0(\underline{x},A)$, the probability of default on the first loan.

Define event E_i to be the successful payment of i loans, with default on loan i+1. Let E_i^* be the successful payment of i+1 loans. If $V(E_i)$ represents the net present value of E_i, then

$$V(E_i) = \sum_{j=0}^{i-1} \alpha_j V_1(A,\underline{L}) + \alpha_i V_0(A,\underline{L}) \tag{3.2.3}$$

and

$$V(E_i^*) = \sum_{j=0}^{i} \alpha_j V_1(A,\underline{L}), \tag{3.2.4}$$

where the discount factor α_j is given by

$$\alpha_j = \ell^j (1+\rho)^{-j\tau}. \tag{3.2.5}$$

The probability of event E_i is given by

$$p(E_i) = \left\{\prod_{j=0}^{i-1} [1 - p_j(\underline{x},A)]\right\} p_i(\underline{x},A) \tag{3.2.6}$$

and

$$p(E_i^*) = \prod_{j=0}^{i} [1 - p_j(\underline{x},A)]. \quad (3.2.7)$$

The expected net present value of J loans is then given as

$$\hat{V}_J = \sum_{i=i}^{J-1} V(E_i)p(E_i) + V(E_{J-1}^*)p(E_{J-1}^*). \quad (3.2.8)$$

The decision rule becomes:

Accept if and only if $\hat{V}_J > 0$,

that is,

$$\hat{\theta} = \begin{cases} 0 \text{ (reject) if } \hat{V}_J \leq 0 \\ 1 \text{ (accept) if } \hat{V}_J > 0 \end{cases} . \quad (3.2.9)$$

Model 2 uses the same initial loan default probability $p(\underline{x},A) \equiv p_0(\underline{x},A)$. This default probability was estimated in precisely the manner described in Chapter 9.3 for the loans in Data Set A2. As described in Chapter 3.2 the first loan default probability is assumed to be beta-distributed with parameters $[r_0(\underline{x},A),n_0]$. We estimate $p(\underline{x},A)$ and consider it to be the expected value of this default probability. This implies that $r_0(\underline{x},A) = n_0 p_0(\underline{x},A)$.

We must now determine the "diffuseness" parameter n_0 of the default probability beta distribution. Since no sample information is available for this determination, we are forced to estimate n_0 using subjective information only. To accomplish this, several installment loan managers were questioned regarding the relative importance of "credit scoring" information (i.e., $\underline{x}$,A) versus loan payment information. The consensus was that information regarding the customer's payment behavior on one loan is "approximately twice as valuable" as knowing his credit scoring attributes. Based on this finding, n_0 was taken to be 0.5.

It is interesting to note that any value of n_0 less than 2 implies a convex beta prior distribution. That is, if $p(\underline{x},A)$, the expected value of the beta default probability distribution, is, for example, .048 (the a priori default probability), and $n_0 = 0.5$ the shape of the beta distribution is roughly as shown in Figure 10.1.

The expected default probability is $p_0(\underline{x},A) = .048$. If one repaid loan is observed the updated default probability becomes

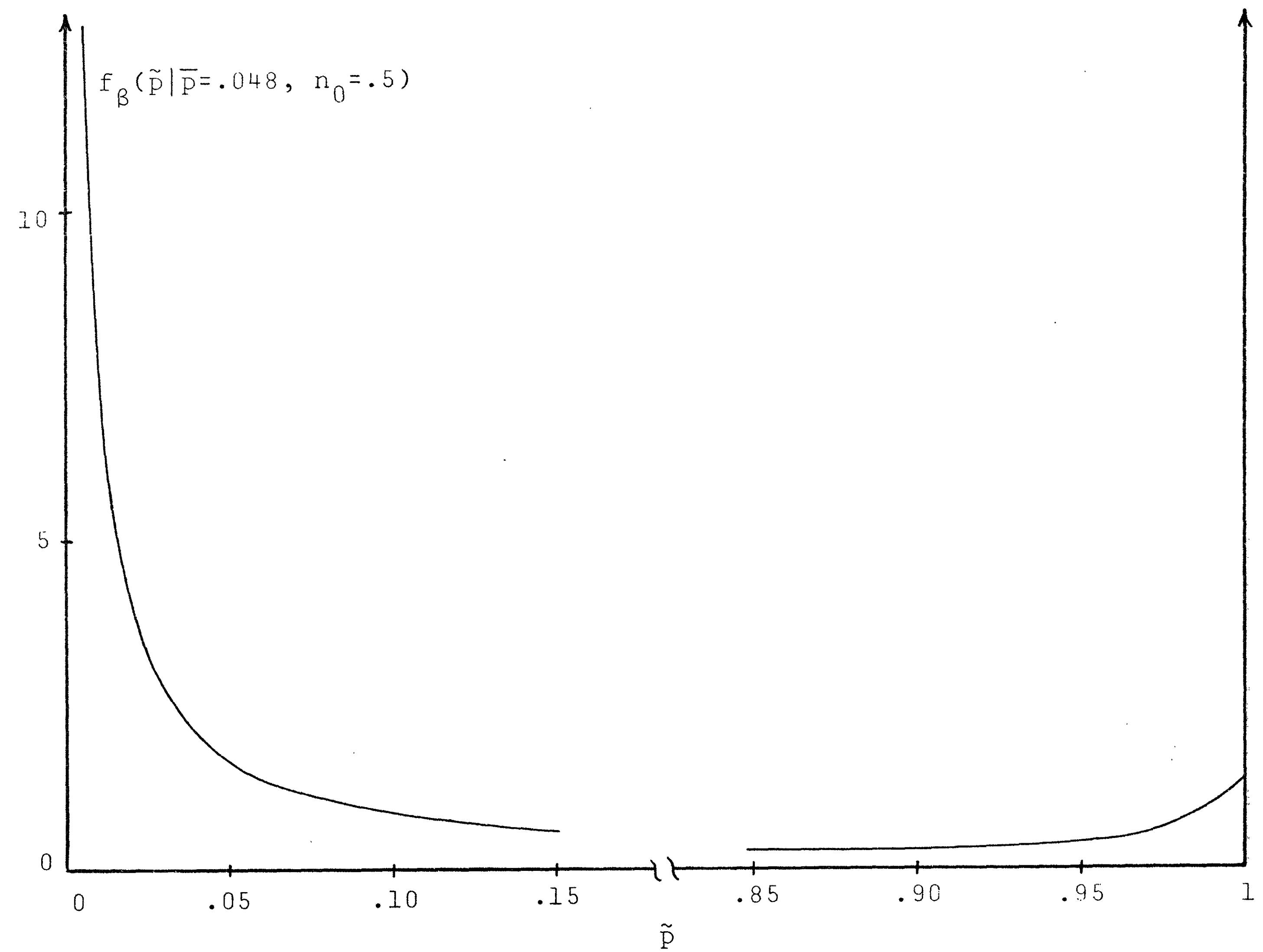
$$p_1(\underline{x},A) = \frac{.5(.048)}{.5 + 1} = .016,$$

but if one defaulted loan is observed the updated default probability becomes

$$p_1(\underline{x},A) = \frac{.5(.048) + 1}{.5 + 1} = .683.$$

Observing the payment performance on one loan causes us to significantly alter our default probability estimate. This is particularly true if a loan is defaulted. No reasonable

FIGURE 10.1 - A Priori Default Probability Distribution



values of $V_0(A.\underline{L})$ and $V_1(A,\underline{L})$ would make a loan with default probability $p(\underline{x},A) = .683$ seem profitable. This result strongly supports the simplifying assumption in Model 2 that once a default occurs, the customer is not considered for subsequent loans. The beta natural conjugate prior to $n_0 = .5$ and $r_0(\underline{x},A) = n_0 p_0(\underline{x},A)$ results in the updated default probabilities $p_j(\underline{x},A)$ (given j repaid loans) presented in Table 10.1.

In order to test Model 2 on Data Set A2, we assumed a five-loan horizon (J=5) with two years between loans (τ=2). A value of ℓ=.7 was taken as the probability that the customer reapplies. Two years between loans is approximately the average reapplication interval experienced at NSB. Five loans were chosen to extend the horizon to a total of 10 years. After 10 years the discount factor (with .10 cost of capital) is $(1.1)^{-10} = .38$, which significantly diminishes the effect of loans 10 years in the future. In addition it was felt that management of most banks would be reluctant to consider more extended horizons. The probability of reapplication of ℓ=.7 represents a subjective estimate by an installment loan executive.

Since we do not know the actual outcome for the loans of Data Set A2 beyond the first loan, we are forced to evaluate Model 2 performance in terms of first loan profit only. Table 10.2 presents the results of applying the Model 2

$p(\underline{x},A)$ \ j	1	2	3	4
.4	.1333	.0800	.0571	.0444
.2	.0677	.0400	.0286	.0222
.1	.0333	.0200	.0143	.0111
.048	.0160	.0096	.0069	.0053
.02	.0067	.0040	.0029	.0022

$p_j(\underline{x},A)$

TABLE 10.1

Bayesian Default Probability Updating

decision rule to Data Set A2 using each of the set of pattern recognition algorithms to estimate $p(\underline{x},A)$. The two error rates for bad loans accepted, P_{BG}, and good loans rejected, P_{GB}, are given, as well as the average profit per applicant on the initial loan. Since we do not know the outcome of loans 2 through 5, we can only give an updated projected figure. The "average expected additional profit per applicant" was determined as follows. If the loan was a good loan, we update our default probability according to the formula

$$p_1(\underline{x},A) = \frac{n_0 p_0(\underline{x},A)}{n_0 + 1} .$$

If the loan was a bad loan, we update using

$$p_1(\underline{x},A) = \frac{n_0 p_0(\underline{x},A) + 1}{n_0 + 1} .$$

We then evaluate the expected net present value of loans 2 through 5 using this updated value of $p_1(\underline{x},A)$. This expected net present value is then considered to be the expected additional profit on loans 2 through 5. The total multi-period profit is then given as the sum of the known first loan profit (or loss) and the expected additional profit of the subsequent loans.

Model 2 simulates the willingness on the part of installment loan managers to take a chance on the first loan with the hope that the customer will prove to be a worthy future loan

Algorithm	P_{BG}	P_{GB}	Average profit per applicant	Expected add'l profit	Total
1 NNR	.16	.42	15.16	32.65	47.81
11 NNR	.17	.40	16.07	37.02	53.09
55 NNR	.19	.43	16.32	37.05	53.37
99 NNR	.20	.44	16.24	36.76	53.00
S & E	.17	.32	17.03	36.85	53.88
Quad. Discr.	.03	.50	14.28	26.21	40.49
Lin. Discr.	.24	.35	16.79	39.99	56.78
Knowledge of first loan outcome	0	0	27.37	48.05	75.65
Accept all	1.0	0	6.55	-	-

TABLE 10.2

Model 2 Performance on Data Set A2

customer in spite of a marginal credit score. In this sense, Model 2 provides a more liberal decision rule than Model 1 because it is able to trade off future possible profit against first loan default risks. Table 10.3 presents the expected net present values given by Model 1 (upper value) and Model 2 (lower value). Two separating lines are shown. The region between the two lines represents initial default probabilities as a function of loan amount for which the account shows an expected one loan loss but an expected five loan profit. The range of default probabilities for which this is the case becomes broader as the loan amount increases because of fixed administrative costs.

	Amount				
p($\underline{x}$,A)	500	1000	2000	3000	4000
0.0	5 11	19 44	50 110	79 176	109 243
.025	-2 2	9 30	30 84	52 138	74 193
.05	-8 -6	-1 15	12 58	26 101	39 144
.075	-14 -14	-12 1	-6 33	-1 64	4 96
.10	-20 -22	-22 -12	-25 8	-28 28	-31 49
.125		-32 -25	-43 -16	-55 -7	-66 2
.15			-63 -41	-82 -43	-101 -44
.175					-136 -90
.20					-171 -135

upper value: 1 loan
lower value: 5 loans

TABLE 10.3

Expected Net Present Values for 1 Loan and 5 Loans

Chapter 11

TRANSITION PROBABILITY ESTIMATION

The 6 state model of Chapter 3.3 will now be applied to Data Set B. However, before we begin to consider delinquency behavior summarized in the transition probability matrix $\underline{\underline{P}}(\underline{x},A)$ as a function of the feature vector $\underline{x}$ and loan amount A, we should determine the appropriateness of the 6 state model itself. In review, the 6 basic states are:

decision

0 months delinquent (on time)

1 month delinquent

2 months delinquent

3 months delinquent

default (more than 3 months delinquent).

The state transition diagram is given below as Figure 11.1.

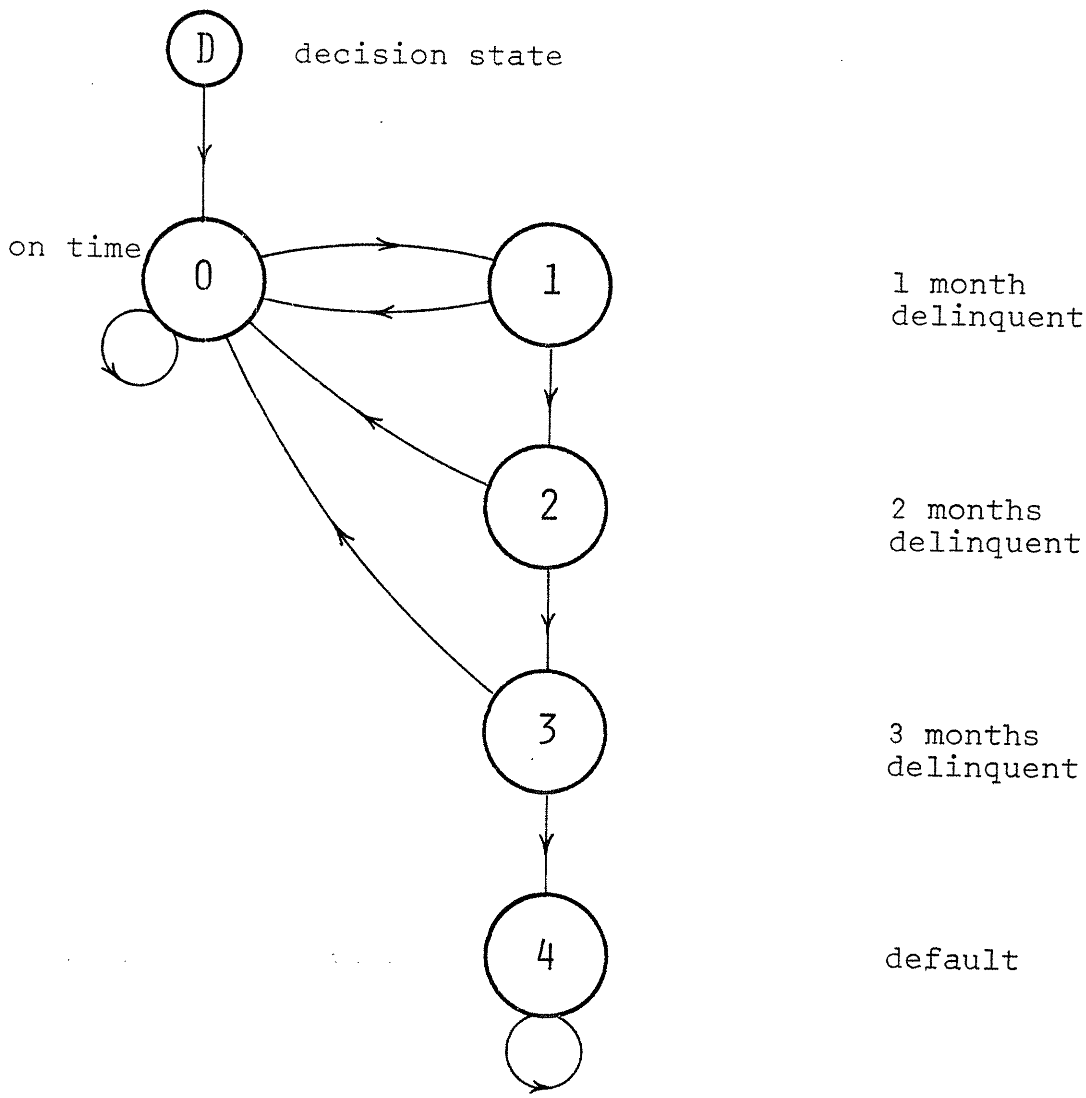
11.1 First Order Test

Model 3 assumes that $\underline{\underline{P}}(\underline{x},A)$ is the transition probability matrix of a first order Markov chain. Let us define $\underline{\underline{P}}$ to be the a priori transition probability matrix. That is, $\underline{\underline{P}}$ characterizes the transition behavior of the average loan unconditional on the feature vector $\underline{x}$ and loan amount A.

We would like to determine if $\underline{\underline{P}}(\underline{x},A)$ is first order for all ranges of $(\underline{x},A)$, but unfortunately this requires a sample considerably larger than that of Data Set B. Consequently, we will test the hypothesis that $\underline{\underline{P}}$ is first-order.

FIGURE 11.1

6 State Description



As an estimate for $\underline{\underline{P}}$ we use the relative frequency, maximum likelihood estimate obtained from Data Set B. Table 11.1 gives the actual transition count matrix and transition probability matrix for the 6 state model. Since we are concerned with the predicted payment behavior if an account is accepted, we need consider only the last five states.

The second order transition count matrix was computed in order to test the first order assumption. The appropriate chi-square test[1] was applied, with a resulting chi-square value of $\chi^2 = 1300$. Such a large chi-square value ($p < .001$) indicates that the 6 state model is not first order, and that, instead, the next state of delinquency depends on more than simply the current state.

This finding suggests that we expand our state description in an attempt to more accurately model the memory retained in the process. However, we would prefer to keep the number of states as small as possible, so that the model is efficient from both an estimation and a computation standpoint. Fortunately, the chi-square test itself provides us with insight into where states should be added.

The chi-square test for equality of two transition probability matrices is performed on a row-by-row basis. For example, the actual results from the above test were:

[1] See Billingsley [27], Chapter 5.

	0	1	2	3	4
decision	956				
0	7714	1805	0	0	0
1	1480	0	311	0	0
2	270	6	0	37	0
3	6	0	0	0	29
4					29

Table 11.1

6 State Model Transition Count Matrix

State i	χ^2_i
0	628
1	384
2	238
3	50
	1,300

There is no contribution from the trap state 4 since the only possible transition is the self-transition.

These χ^2 values indicate that an account's transition from state 0 depends more heavily on its previous state history than its transition from states 1, 2, or 3. An examination of the second-order transition count matrix reveals that a 0 → 1 transition is much more probable if the acount was previously in states 1, 2, or 3. That is, once a delinquency pattern is established it is more likely to reappear even after the account returns to a current paying status. This finding suggests adding state 5, as shown in Figure 11.2, with the interpretation that it is an on-time state given previous delinquency. The right-hand number 0 in the diagram for state 5 indicates that the state is an on-time state (0 months delinquent).

The addition of state 5 gives a first-order test χ^2 value of 703, with state 5 this time the principal contribution. This suggests adding an additional on-time state 6 (see Figure

Number to the right of the state index is the number of months delinquent.

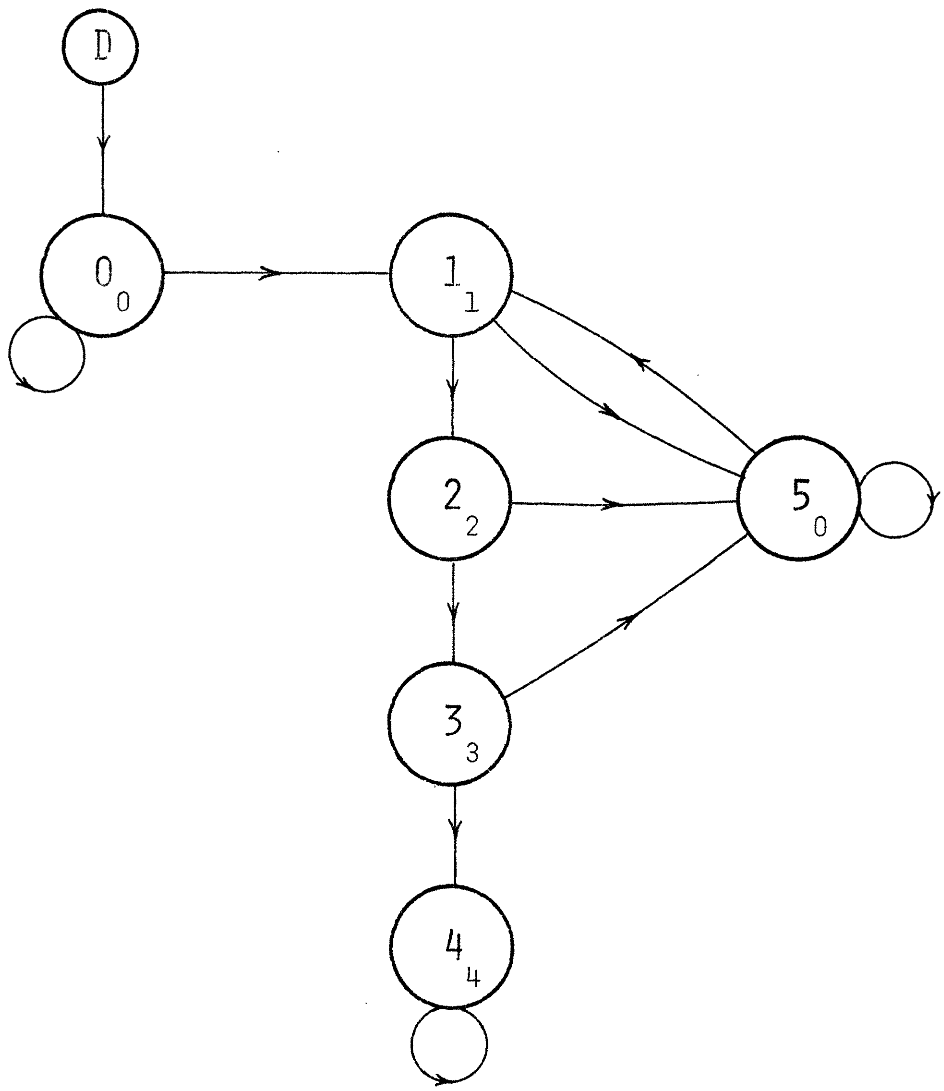


FIGURE 11.2
7 State Description

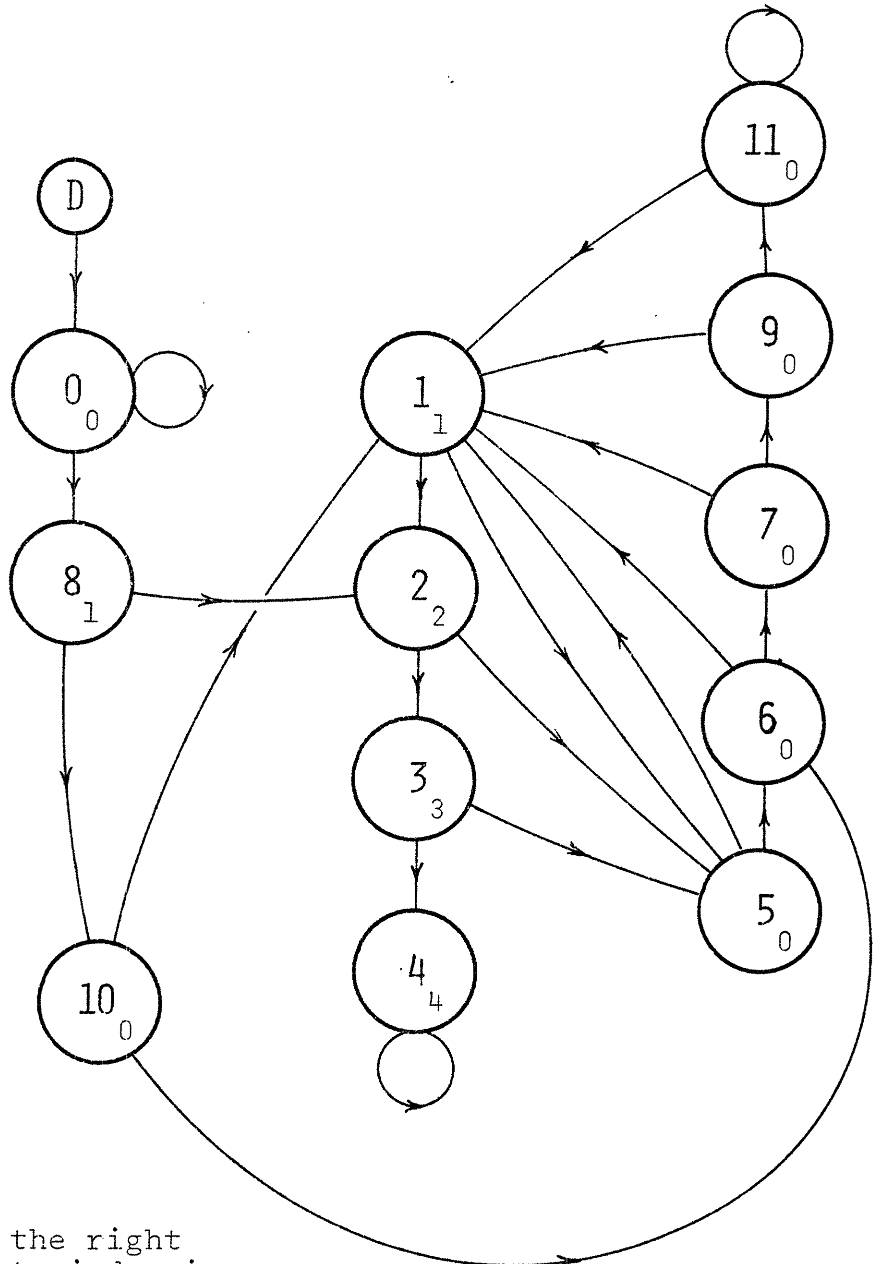
11.3), since, apparently, previous delinquent behavior is "remembered" for more than one on-time payment.

This procedure of adding states where appropriate, so as to model the memory still remaining, was continued, with the following χ^2 results.

State added	χ^2
Initial model	1,300
5	703
6	81
7	53
8	44
9	51
10	20
11	17

After adding state 11 (an on-time state), it became apparent that little further reduction of χ^2 could be obtained by adding another state. No single state seemed to overly contribute to the total χ^2 value, as can be seen by the following breakdown by state.

State i	χ^2_i
1	8.5
2	2.4
5	0.1
6	1.5
11	5.0
	17.5



Number to the right of the state index is the number of months delinquent.

FIGURE 11.3

13 State Description

It was felt that attempting to further improve the model might be an exercise in trying to fit the data rather than modeling the underlying Markov process. With the state description expanded to 12 states (plus the decision state), the second order transition counts become small to the point where diluting them further by adding more states begins to invalidate the large sample requirements of the chi-square test. Moreover, there is little statistical need to reduce the fit of the model. A χ^2 value of 17.5 is not large enough to reject the first-order hypothesis at the .95 confidence level ($\chi^2_{.95}$ = 19.7 with 11 degrees of freedom).

From the point of view of obtaining a descriptive model of delinquency behavior, we are pleased with the actual state description obtained. The model has the capability of describing all of the recognized types of delinquency behavior. For example, the customer who becomes delinquent because he forgot to send in his payment before going on vacation will make the transitions

$$0 \rightarrow 8 \rightarrow 10 \rightarrow 6 \rightarrow 7 \rightarrow 9 \rightarrow 11 \rightarrow \ldots \rightarrow 11.$$

The chronic delinquent who pays every other month will make the transitions

$$0 \rightarrow 8 \rightarrow 10 \rightarrow 1 \rightarrow 5 \rightarrow 1 \rightarrow 5 \ldots$$

11.2 Stationarity Test

The 13 state model resulting from the analysis of the previous section is seen to be approximately first-order. We now wish to test it for stationarity. In other words, we want to know if the transition probability matrix $\underline{\underline{P}}$ remains constant throughout the term of the loan. This amounts to testing the hypothesis

$$H_0: \underline{\underline{P}}(t=1) = \underline{\underline{P}}(t=2) = \ldots = \underline{\underline{P}}(t=T),$$

where $\underline{\underline{P}}(t)$ is the transition probability matrix at time t, for t=1 to the number of payments, T.

Since T can range from 6 to 60 or more monthly periods, to test H_0 we would have to estimate a large number of 13 by 13 transition probability matrices. Many of the elements of the matrices, e.g., $p_{23}(t=15)$, would be estimates based on a very small number of samples (fewer than 5). Since the chi-square test is essentially a large sample test, it was decided to modify H_0 as follows:

$$H_0: P(t=1\text{-}5) = P(t=6\text{-}10) = P(t=11\text{-}15) = P(t=16\text{-}20) = P(t=21+).$$

Rather than consider each payment period separately, group the periods into 5 times frames: t = 1 to 5, t = 6 to 10, t = 11 to 15, t = 16 to 20, and t = 21+. This grouping requires the estimation of 5 transition probability matrices, yet it still provides a reasonable test of the hypothesis that payment behavior does not vary during the course of the loan.

Using Data Set B, the appropriate chi-square test for stationarity[2] was used to test this hypothesis. Chi-square values were computed on a row-by-row basis with the following results.

State i	χ^2_i
0	76.4
1	.6
2	.4
3	.5
4	0.0 (trap state)
5	8.5
6	5.4
7	5.9
8	5.5
9	.7
10	6.5
11	6.8
	117.3

The degrees of freedom for the test are:

$$d.f. = R(C-1)(T-1) = 12(2-1)(5-1) = 48,$$

where R is the number of rows,

C is the number of columns estimated for each row, and

T is the number of time frames considered.

[2]See Billingsley [27], Chapter 5.

The .95 percentile of the χ^2 distribution with 48 degrees of freedom is about 65, which is less than the observed value of 117. This would suggest rejecting the stationarity hypothesis H_0.

However, if we only consider states 1 through 11 and exclude state 0, the total chi-square value would be 40.9 with 44 degrees of freedom. This is less than the .50 percentile of the χ^2 distribution with 44 degrees of freedom ($\chi^2_{.50} = 43.3$ with 44 d.f.). Thus, any non-stationarity in the process is due only to the transition behavior from state 0 (on-time and not previously delinquent).

Taking a closer look at state 0 non-stationarity we suspect that it is not significant enough to affect our model from a decision modeling (as opposed to a statistical modeling) point of view. This belief is supported by the following table of results, which gives the time-dependent estimates for $p_{00}(t)$ and the number of samples used to estimate this on-time self-transition probability.

Estimate	Value	Number of Samples
$p_{00}(t = 1\text{-}5)$	.982	30039
$p_{00}(t = 6\text{-}10)$	.988	24472
$p_{00}(t = 11\text{-}15)$	.993	11620
$p_{00}(t = 16\text{-}20)$	.999	1598
$p_{00}(t = 20+)$	.897	29
$p_{00}(\text{all } t)$	.987	67758

These values indicate that an account that establishes a perfect payment record early in the loan term will tend to maintain that perfect record later in the loan. Nonetheless, this tendency does not appear to be strong enough to warrant the additional modeling complexity as well as estimation and computation expense that would be required to accurately incorporate it into the decision model. To support this contention, Model 3 was actually programmed with a transition probability matrix for which p_{00} (and $p_{08} = 1 - p_{00}$) was time varying according to the $p_{00}(t)$ values given below.

t	$p_{00}(t)$	p_{00}
1-5	.982	.987
6-10	.988	.987
11-15	.993	.987
16+	.999	.987

The standard (stationary) Model 3 was also programmed with the time-averaged $0 \rightarrow 0$ transition probability of p_{00} = .987.

The remainder of the $\underline{\underline{P}}$ matrix was taken from the a priori transition probability matrix estimated in the following section. Using an average loan term of T = 20 months, the default probability was computed as the T-step transition probability from the decision state to the default trap state 4. When the time-varying $p_{00}(t)$ were used, the default probability was computed to be .007993; when the time-average p_{00}

was used, the default probability was computed to be .007996. This close agreement leads us to conclude that what non-stationarity exists is insignificant. The observed discrepancy in default probability of less than 0.04% is much too small to alter the model's credit granting decisions.

11.3 The A Priori Transition Probability Matrix

Given the samples in Data Set B and the 13 state model shown in Figure 11.3, we now are prepared to estimate $\underline{\underline{P}}$, the a priori transition probability matrix. This matrix is of particular importance for estimating $\underline{\underline{P}}(\underline{x},A)$. As outlined in Chapter 5, the $(i,j)^{th}$ element of $\underline{\underline{P}}(\underline{x},A)$ can be written as

$$p(\theta_i=j|\underline{x},A) \propto p(\underline{x},A|\theta_i=j)p(\theta_i=j). \tag{5.2}$$

The a priori probability $p(\theta_i=j)$ is simply the $(i,j)^{th}$ element, p_{ij}, of $\underline{\underline{P}}$.

An estimate for p_{ij} is given by the relative frequency maximum likelihood estimate

$$\hat{p}_{ij} = c_{ij}/c_i, \tag{5.4}$$

where c_{ij} is the number of $i \to j$ transitions observed, and

$$c_i = \sum_j c_{ij}.$$

Unfortunately, Data Set B is a pre-screened data set. That is, it contains only accounts that had previously been accepted by NSB. Since the bank tries to screen out loan applications that are probable delinquent or even defaulting accounts, the relative frequencies seen in Data Set B will tend to give a priori transition probabilities that characterize a better-than-average account.

Ideally, we would like to have transition histories for rejected applicants, and use these transition counts to augment those of Data Set B. Since this is not possible, we are forced to make certain assumptions regarding the probable delinquency behavior of unobserved rejected applicants. A reasonable simplifying assumption is that rejected accounts differ from accepted accounts only in the on-time self transition probability p_{00}. That is, a sample of rejected accounts would have transition counts for states 1 through 11 in the same relative frequency as accepted accounts, but the ratio c_{00}/c_0 would be smaller. The important question is how much smaller?

The results of Section 9.1 indicated that the a priori probability of default is about .048. This would suggest that adding a proportionate number of rejects of Data Set B (and hence lowering c_{00}/c_0) should result in a default state trapping probability of .048. That is, if $\overline{T}$ is the average term of all loans, we want p_{00} to be such that the probability of being in a trap state 4 at period $t = \overline{T}$ to be .048. With $\overline{T} = 20$ months, and $\underline{\underline{P}}$ given by the Data Set B relative frequencies, a value of $p_{00} = .73$ gives the desired .048 default probability. The complete a priori transition probability matrix is given in Table 11.2 (only non-zero entries are shown).

i	$j1$	$P_{i,j1}$	$j2$	$P_{i,j2}$
0	0	.733	8	.267
1	2	.213	5	.787
2	3	.121	5	.879
3	4	.829	5	.171
4	4	1.0	-	-
5	1	.541	6	.459
6	1	.146	7	.854
7	1	.100	9	.900
8	2	.136	10	.864
9	1	.079	11	.921
10	1	.287	6	.713
11	1	.042	11	.958

Table 11.2

A Priori Transition Probability Matrix

11.4 Pattern Recognition Application to Transition Probability Estimation

Chapter 5 developed the correspondence between transition probability estimation of $\underline{\underline{P}}(\underline{x},A)$ and the classification probability determination common to pattern recognition problems. In particular a concise linear discriminant formulation was given in equations (5.7) through (5.10).

This technique is applied to the 12 delinquency states of the Markov process model (see Figure 11.3). Equations (5.7) and (5.8) specify the (r+1) vector (i.e., $[\underline{d}_i^t, b_i]$) that must be determined for each row i:

$$\underline{d}_i^t \equiv (\underline{m}_{i,j1} - \underline{m}_{i,j2})^t \, \underline{\underline{S}}_i^{-1} \tag{5.7}$$

$$b_i \equiv - \frac{1}{2}(\underline{m}_{i,j1} - \underline{m}_{i,j2})^t \underline{\underline{S}}_i^{-1} (\underline{m}_{i,j1} + \underline{m}_{i,j2}), \tag{5.8}$$

where $\underline{m}_{i,j}$ is the weighted mean vector of all samples that made $i \rightarrow j$ transitions (weighted by the number of $i \rightarrow j$ transitions observed), and $\underline{\underline{S}}_i$ is the weighted covariance matrix computed from all samples that made transitions from state i (to both j1 and j2).

In keeping with the feature selection analysis of Chapter 7, the transition probability matrix $\underline{\underline{P}}(\underline{x},A)$ will be estimated as a function of the following r = 10 features.

feature	description
1	occupation
2	years at occupation
3	amount (A)
4	age
5	years of residence
6	income
7	mortgage/rent
8	years at former residence
9	years with former employer
10	NSB score

Note that the loan amount is actually the third component of the feature vector $\underline{x}$.

Unfortunately, Data Set B is not large enough to permit the estimation of the (10 x 10) covariance matrix $\underline{\underline{S}}_i$ for some states i. For example, there were only 35 transitions observed out of state 3. We would be over-fitting the data if we tried to use the $\underline{x}$ vectors of the accounts that made these 35 transitions to estimate the 55 elements of $\underline{\underline{S}}_3$ and the two mean vectors $\underline{m}_{34}$ and $\underline{m}_{35}$.

If instead of considering all 10 features in estimating $p_{34}(\underline{x},A)$ and $p_{35}(\underline{x},A)$, we considered only the three most important features, we would have to estimate only six elements of a (3 x 3) $\underline{\underline{S}}_3$ and the two mean vectors.

A prudent approach to this estimation problem is to let the number of features used to estimate row i of $\underline{\underline{P}}(\underline{x},A)$ depend on the number of samples that are available to estimate $\underline{\underline{S}}_i$ and $\underline{m}_{i,j1}$ and $\underline{m}_{i,j2}$. If we let ψ be an "over specification factor" and there are c_i transitions from state i, then we can safely estimate a total number of mean-covariance parameters of at most c_i/ψ. That is, for a given number of observations c_i, a larger over-specification factor ψ allows more accurate estimation of a smaller number of components of $\underline{m}_{i,j1}$, $\underline{m}_{i,j2}$, and $\underline{\underline{S}}_i$.

The estimation of mean vectors and covariance matrices was performed with two over-specification factors, $\psi = 2$ and $\psi = 10$. The following number of features r were selected for each row i of $\underline{\underline{P}}(\underline{x},A)$.

row i	r (ψ = 2)	r (ψ = 10)
0	10	10
1	10	4
2	4	1
3	3	1
4 (trap state)	-	-
5	10	6
6	8	2
7	5	1
8	8	3
9	3	1
10	10	4
11	4	1

If fewer than 10 features are to be selected we must decide which to choose. An examination of the t-statistics for 0→0 and 0→8 transitions (most frequently observed) suggests the following ranking of features in order of decreasing discriminatory power:

feature	description
10	NSB score
3	amount (A)
4	age
6	income
8	years at former residence
2	years at occupation
9	years with former employer
1	occupation
5	years of residence
7	mortgage/rent

Thus if we have only enough samples to use three features, we would choose features 10, 3, and 4.

This feature reduction and estimation procedure was performed for both over-specification factors $\psi = 2$ and $\psi = 10$. Mean vectors $\underline{m}_{i,j1}$ and $\underline{m}_{i,j2}$ and covariance matrices $\underline{\underline{S}}_i$ were computed, from which (5.7) and (5.8) were used to determine $\underline{d}_i^t$ and b_i. The computed values of $\underline{d}_i^t$ and b_i corresponding to states 0 through 11 are presented in Tables 11.3 and 11.4 for over-specification factors of $\psi = 2$ and $\psi = 10$, respectively. An N x (r+1) discriminant function matrix and the N a priori probabilities determined in Section 11.3 are all that are needed to determine $\underline{\underline{P}}(\underline{x},A)$ for any applicant's features $(\underline{x},A)$.

Equations (5.9) and (5.10) indicate that first we compute

$$\ell_i = \underline{d}_i^t x + b_i + \ln(p_{i,j1}/p_{i,j2}) \tag{5.9}$$

and then obtain $p_{i,j1}(\underline{x},A)$ and $P_{i,j2}(\underline{x},A)$ as

$$p_{i,j1}(\underline{x},A) = \frac{e^{\ell_i}}{1 + e^{\ell_i}} \tag{5.10}$$

$$p_{i,j2}(\underline{x},A) = 1 - p_{i,j1}(\underline{x},A).$$

i	b_i	$\underline{d}_i^t$									
0	-0.008	-0.036	-0.057	0.076	0.008	-0.083	-0.195	-0.077	-0.057	-0.044	0.032
1	-0.002	0.033	-0.057	-0.077	0.003	-0.081	0.254	0.009	0.193	-0.158	-0.028
2	0.401			-0.118	0.013		0.533				-0.118
3	3.486			-0.880	-0.177						-0.115
4											
5	0.141	0.147	0.121	0.010	-0.006	0.255	-0.132	0.007	0.188	-0.120	-0.041
6	0.064	-0.181	0.104	0.028	-0.018		0.176		0.034	0.080	-0.021
7	-0.147			0.254	0.022		-0.269		0.014		-0.007
8	-0.141	-0.085	0.236	0.025	0.003		0.485		-0.158	0.305	-0.076
9	-0.323			0.336	0.021						-0.038
10	0.197	0.163	0.161	0.009	-0.012	0.439	-0.148	0.168	-0.070	-0.039	-0.056
11	0.301			0.216	0.013		-0.200				-0.052

TABLE 11.3

Discriminant Function Matrix ($\psi=2$)

i	b_i	$\underline{d}_i^t$									
0	-0.008	-0.036	-0.057	0.076	0.008	-0.083	-0.195	-0.077	-0.057	-0.044	0.032
1	0.127			-0.072	0.001		0.208				-0.031
2	0.555										-0.075
3	2.746										-0.306
4											
5	0.053		0.119	-0.011	-0.008		-0.082		0.112		-0.012
6	0.130			0.035							-0.023
7	-0.139										0.017
8	0.096			0.057	0.007						-0.026
9	-0.094										0.011
10	0.049			0.004	-0.002		-0.147				0.015
11	0.289										-0.035

TABLE 11.4

Discriminant Function Matrix (ψ=10)

Chapter 12

MODEL 3 RESULTS

Model 3 was developed to more realistically describe the outcome space for an accepted loan. To sufficiently model the delinquency behavior and associated costs of a loan, the simple outcome description of default/non-default is replaced by the state description presented in Chapter 11. Since this 13 state model was seen to be approximately first order and stationary, we can consider a loan applicant to be completely characterized by his expected transition probability matrix $\underline{\underline{P}}(x,A)$, by the loan reward matrix $\underline{\underline{R}}(A,\underline{L})$, and by the term of the loan T. Given these estimates, the single-loan decision to grant credit can be evaluated as a discounted Markov process decision model.[1]

Chapter 5 indicated the transition probability matrix $\underline{\underline{P}}(\underline{x},A)$ could be estimated using pattern recognition techniques. One such technique based on linear discriminant analysis was developed in detail and applied in Chapter 11. It was seen that $\underline{\underline{P}}(\underline{x},A)$ could

[1] See Howard [32], Chapters 2 and 7

be determined as a function of $(\underline{x},A)$ by little more than one matrix multiplication of $(\underline{x},A)$ and a discriminant function matrix.

Transition Reward Matrix

The transition reward matrix $\underline{\underline{R}}(A,\underline{L})$ depends on both the loan characteristics $(A,\underline{L})$ and previously determined costs of collection. Let r_{ij} be the reward associated with an $i \rightarrow j$ transition, where for notational simplicity, the dependence on $(A,\underline{L})$ is assumed. The values of r_{ij} will now be presented on a transition by transition basis.

Credit granting to on-time

When the loan is granted, the account moves to the on-time state and a cash outflow of -A occurs.

On-time to on-time

This includes the following $i \rightarrow j$ transitions: $0 \rightarrow 0$, $5 \rightarrow 6$, $6 \rightarrow 7$, $7 \rightarrow 9$, $9 \rightarrow 11$, $11 \rightarrow 11$, and $10 \rightarrow 6$. When an account makes an on-time to on-time transition a cash inflow of one payment is made. Appendix A shows how payment size is determined as a function of loan amount A, loan term T, and the simple annual interest rate r of the loan.

On-time to 1 month delinquent

This includes the transitions 0→8, 5→1, 6→1, 7→1, 9→1, 10→1, and 11→1. An account becomes delinquent when no payment is received by the payment due date (plus an undisclosed grace period). At this time collection action is initiated at the "first call level". A previous cost study performed by the author indicated that the cost of collection effort at the first call level is approximately $2.40 per month per delinquent account. Thus for these transitions we set $r_{ij} = -2.40$.

1 month delinquent to on-time

This includes the transitions 1→5 and 8→10. Typically the customer must make two payments on the next due date to return to on-time status, and collectors are instructed to telephone the customer and ask that he send his late payment with the current payment that is due. This produces a cash inflow of two payments.

1 month delinquent to 2 months delinquent

This includes 1→2 and 8→2 transitions. The same cost study alluded to above determined the per month cost of 2nd level collection effort to be $3.30 per account. Since no payment is received, we set r_{ij} = -3.30 for these transitions.

2 months delinquent to on-time (2→5)

Here, two delinquent payments plus the currently due payment are received for a total cash inflow of r_{25} = 3 payments.

2 months delinquent to 3 months delinquent (2→3)

The cost study determined that the more intense 3rd level collection effort costs $7.50 per month per account, giving $r_{23} = -7.50$.

3 months delinquent to on-time (3→5)

Here, three delinquent payments plus the currently due payment are received for a total cash inflow of r_{35} = 4 payments. We would note, however, that seldom are four payments simultaneously received. Nonetheless, this minor discrepancy leads to only insignificant effects on the total net present value of the process, and is most unlikely to change the credit granting decision.

3 months delinquent to default (3→4)

When an account becomes more than 3 months delinquest, it is considered to have defaulted. The account is then charged-off against a reserve for bad debts, and a separate recovery department attempts to collect the charged-off balance. Typically, they recover about

37% of the balance (on a discounted basis), but at an average recovery effort cost of $106 per account. Model 3 takes this recovery cost into account as soon as the account makes the transition to the default trap state. This gives $r_{34} = -106$.

Default state to Default state ($4 \rightarrow 4$)

When a loan enters the default trap state at time t, the charged-off balance is $(T-t+3) \cdot a$ where T is the loan term and a is the payment size. The value of 3 is included since the account did not make its last 3 scheduled payments. The results of Appendix A indicate that about 37% of this balance will be recovered.

We can approximately model the post default recoveries by including a positive reward on $4 \rightarrow 4$ transitions of $r_{44} = .37a$ and continuing to run the Markov process model for a full T periods.

Evaluating Model 3

The discriminant function matrix was determined[2] using all loans in Data Set B. A subset, hereafter referred to as Data Set B1, was then used to test the model's performance. Data Set B1 consists of all 28 "bad" loans in Data Set B (i.e., loans that defaulted) and

[2]See Table 11.3.

100 randomly selected "good" loans (i.e., loans that were never delinquent during the 15 month observation period). For each of these 128 loans, $\underline{\underline{P}}(\underline{x},A)$ and $\underline{\underline{R}}(A,\underline{L})$ were estimated and the expected net present value of all rewards for T transitions was computed. T was taken to be the actual term of the loan being tested.

Let $V_d(0|T)$ be the expected net present value of all rewards over T periods, given the loan starts at the decision state d. We can compute $V_d(0|T)$ recursively using (3.3.1.), with the transition reward matrix $\underline{\underline{R}}(A,\underline{L})$ given above and the transition probability matrix $\underline{\underline{P}}(\underline{x},A)$ estimated in the manner outlined in Chapter 11. The monthly discount factor is taken to be $\beta = (1 + \rho/12)^{-1}$, with $\rho = .10$ the annual cost of capital.

The single loan decision rule is then

$$\hat{\theta} = \begin{cases} 0 \text{ (reject) if } V_d(0|T) \leq 0 \\ 1 \text{ (accept) if } V_d(0|T) > 0. \end{cases} \qquad (3.3.2)$$

For the loan that is actually from class θ ($\theta=0$ for bad loans, $\theta=1$ for good loans) and Model 3 giving the decision $\hat{\theta}$, the reward function $R(\theta,\hat{\theta})$ (not to be confused with the transition reward matrix $\underline{R}(A,\underline{L})$) is given in Table 9.2.1.

In a manner similar to that of Chapter 9.3 for Model 1, we can compute the weighted total profit of applying the Model 3 decision rule to Data Set B1. This weighted total profit can then be normalized to give the average profit per applicant. Table 12.1 presents these results, along with the misclassification probabilities for Model 3. In addition, we present the performance of the benchmark decision rule (credit scoring) and the profit given perfect knowledge of θ.

The results in Table 12.1 show that although Model 3 tends to reject a large proportion of good loans, those that are rejected have small average profit relative to those that are accepted ($11.99 compared to $51.99). Further examination of the loans of Data Set B1 reveals a significant number of good loans with small loan amounts and short loan terms (e.g., A = $800, T = 6 months). Even with a perfect payment record, these loans are only marginally profitable (if not actually unprofitable) since the net interest earned barely offsets the administrative

Decision Rule	P_{BG}	P_{GB}	Average Profit per Applicant	Average Profit on Good Loans Rejected	Accepted
Model 3	.07	.73	10.85	11.99	51.99
Benchmark (accept all)	1.00	0	-1.72	-	22.79
Benchmark (S* = 20)	.86	.03	1.76	7.65	23.26
Perfect Information	0	0	21.92	-	21.92

TABLE 12.1

Model 3 Performance on Data Set B1

processing costs.[3] Model 3 tends to reject these small loans when additional expected delinquency costs are considered. On the other hand, the benchmark credit scoring rule considers only the score, regardless of the loan amount or term. This partially accounts for the performance of Model 3 relative to the benchmark decision rule.

The loans of Data Set B1 have all been accepted by the NSB benchmark decision rule. This pre-screening implies that these loans are probably of better quality than the average loan applied for. The a priori probability of default, given the loans were previously screened and accepted, should be somewhat less than the estimate of $p(\theta=0) = .048$. The effect of this pre-screening bias will be considered in detail in Chapter 15.

[3]These loans are more economically processed as a revolving credit cash advance.

Chapter 13

MODEL 4 RESULTS

If we try to extend the detailed outcome description of Model 3 to more than a single loan period, the set of possible outcomes expands so rapidly that even considering a two-loan case is computationally impossible with 13 states. The outcome of any one loan is described by its T month state occupancy history. From this history we can compute a transition count matrix $\underline{\underline{F}}$. Depending on the transitions we observe, we update our estimate of $\underline{\underline{P}}(\underline{x},A)$ and decide whether to re-extend credit. The distribution theory behind this Bayesian updating can be found in Martin [34]. With Model 2, the only outcomes considered were default and non-default. In this case, however, the outcome space becomes the entire range set of the transition count matrix $\underline{\underline{F}}$. For a 13 state model and 20 transitions an extremely large number of different transition count matrices are possible. Still worse, we must then evaluate the decision to re-extend credit <u>conditional</u> upon each of these transition count outcomes. As Martin indicates[1],

---

[1]Martin [34], p. 179

> "It does seem clear, however, that, for problems with a large number of states in which a high degree of accuracy is required, we must think in terms of hours, not minutes, of computer time. This is not to say that the Bayesian method of dealing with Markov chains with uncertain transition probabilities must be abandoned as impractical. But it must be recognized that, for the present state of the art, the Bayesian treatment is probably most practical for problems with two or three states, loose prior distributions, and large differences in the rewards associated with different actions. As problems tend to differ from these criteria, the decision maker must balance increasing computation time against the required accuracy of the solution and choose an appropriate approximation."

Clearly, we must focus our attention toward an approximation to Model 4 which captures as much as possible the important aspects of payment/delinquency detail while not neglecting the importance of modeling the multi-loan nature of the process. One such approximation, Model 4A, is presented in this chapter. Its economic performance is then tested on a sample of loans from Data Set B.

13.1 Model 4A Description

Model 4A is intended to model delinquency behavior with its related costs in the context of a credit granting decision which considers the significant present value contribution of subsequent loans. The Markov process model presented in Chapters 3 and 11 was motivated by the desire to include delinquency-related costs. The previous analysis of Chapter 11 indicates that it is descriptive of delinquency behavior and thus delinquency-related costs. The T-step transition probability from state d (the decision state) to state 4 (default) provides an alternative pattern recognition determined estimate for the probability of default $p(\underline{x},A)$.

Moreover, the Markov process model implicitly gives this default probability as a function of the loan term T. Model 2, on the other hand, does not include loan term as a component of the feature vector $\underline{x}$ because loan term ranked 22nd out of 23 features in terms of information content about the loan class θ (see Table 7.1)

The analysis presented in Appendix A tends to support a dependence of the default probability $p(\underline{x},A)$ on T. The conclusion reached there was that default is equally likely to occur at any time during the term of the loan. This was expressed quantitatively by (A.6) as:

$$p(\gamma) = \begin{cases} T/(T-1) & 0 < \gamma \leq (T-1)/T \\ 0 & (T-1)/T < \gamma \leq 1 \end{cases} \tag{A.6}$$

where γ is the fraction of payments made before default. If t = time of default, then the probability that default occurs before time t (given default occurs) is given by

$$p(\gamma<t/T|T) = \int_0^{t/T} p(\gamma|T)d\gamma = t/(T-1). \qquad (13.1)$$

This result implies that the probability of default increases linearly throughout the term of the loan. This conclusion is indeed consistent with the Markov process model of delinquency behavior. In fact, the Markov process model apparently implies that, given ($\underline{x}$,A), p($\underline{x}$,A) increases almost linearly with T.

In light of this evidence that p($\underline{x}$,A) increases with T, we might reasonably wonder why a lending institution does not discourage longer term loans. The answer lies in the increased profitability to the lender from a longer term obligation. For a given loan amount, interest rate, and cost of capital the net present value of payment cash flows, $V_1(A,L)$, increases with T. For example, the profitability formulas of Appendix A applied to a $2000 loan made at 13.5% interest with 10% cost of capital gives the following net present values (excluding delinquency costs):

Term (Months)	Net Present Value given no default	given default
12	27.22	-755
24	60.74	-677
36	93.39	-617
48	125.14	-562
60	155.93	-509

In fact, these net present values increase almost linearly with loan term T.

Thus as T increases, so does both the lender's risk and his return; that is, $p(\underline{x},A)$, $V_1(A,\underline{L})$ and $V_0(A,\underline{L})$ all increase with T. The Markov process description of payment behavior properly incorporates this dependence on loan term, as evidenced by the following experiment. The payment behavior of this same $2000 loan was simulated using the a priori transition probability matrix described in Chapter 11. The transition reward matrix was computed using the same 13.5% interest rate and 10% cost of capital, but with varying loan terms. Transition rewards were computed to include only delinquency costs. The following T-step default probability and the expected net present value of delinquency costs were observed.

T	Default Probability	Delinquency Costs = D
12	.029	- 9.49
24	.055	-12.23
36	.076	-14.53
48	.097	-16.65
60	.117	-18.57

It thus seems appropriate to use the Markov process description to obtain both an initial estimate of the probability of default and an estimate of expected delinquency costs. As indicated above, the detailed outcome description of a single loan is exploited to obtain a delinquency cost estimate. Having obtained this cost estimate, we then map the set of possible transition histories into one of two outcomes, a transition to the default trap state and no default after T transitions. That is, since there are a multitude of possible (NxN) transition count matrices $\underline{\underline{F}}$, we will partition the range of $\underline{\underline{F}}$ into the two classes

$$\theta(\underline{\underline{F}}) = 0 \quad \text{and} \quad \theta(\underline{\underline{F}}) = 1, \text{ where}$$

$$\theta(\underline{\underline{F}}) = \begin{cases} 0 & \text{if } \underline{\underline{F}} \text{ includes a transition to the default trap state} \\ 1 & \text{otherwise.} \end{cases}$$

This mapping of $\underline{\underline{F}}$ to θ provides the approximation needed to make Model 4 computationally manageable. Instead of estimating $Pr[\underline{\underline{F}}|(\underline{x},A),T]$ for all $\underline{\underline{F}}$ individually, we can now estimate

$$\Pr[\theta(\underline{\underline{F}})=0|(\underline{x},A),T] = \sum_{\forall \underline{\underline{F}} \ni \theta(\underline{\underline{F}})=0} \Pr[\underline{\underline{F}}|(\underline{x},A),T]$$

This probability $\Pr[\theta(\underline{\underline{F}})=0|x,A,T]$ is simply the expected probability of default given $(\underline{x},A)$ and T.

This reduction to a two outcome description permits a consideration of subsequent loans through the use of the multi-period two outcome model described in Chapters 3.2 and 10. The Markov process description provides the necessary term-dependent default probability for the first loan

$$\Pr[\theta(\underline{\underline{F}})=0|(\underline{x},A),T] \equiv p_0(\underline{x},A)$$

which then is updated as in Model 2 as each subsequent loan is repaid. In addition, since the detailed state description provides an estimate of expected delinquency costs, D, we can adjust the net present value of default and non-default to be $V_0(A,\underline{L})-D$ and $V_1(A,L)-D$ respectively.

This approximation assumes that delinquency costs D will be the same for all subsequent loans, when in fact we might expect them to become less with each subsequent loan. However, since delinquency costs of future loans are discounted at the relevant cost of capital, we expect that this assumption will have a negligible effect on the multi-period expected net present value. As presented in Chapters 3.2 and 10, $\hat{V}_J$ will denote the expected net present value of J loan periods.

13.2 Model 4A Results

Model 4A was used to estimate the J-period expected net present value, $\hat{V}_J$, for all 28 defaulted loans and a random sample of 100 never-delinquent "good" loans from Data Set B. These 128 test loans are the same loans used in Chapter 12 to test Model 3.

The transition probability matrix $\underline{\underline{P}}(\underline{x},A)$ was estimated in the manner described in Chapter 11. The transition reward matrix $\underline{\underline{R}}(A,\underline{L})$ was computed in a manner similar to that described in Chapter 12, except that only collection effort costs and "float" costs associated with late payments were considered. For example, the float cost associated with a 1→5 transition (1 month delinquent to on-time) is the difference between the present value of one payment at time t+1 and one payment at time t (when the payment was due). If ρ is the cost of capital and a is the payment amount, this difference is

$$\text{float cost } (1\rightarrow 5) = -a(\rho/12) \; . \qquad (13.2)$$

For a \$2000 loan with 24 month term at 13.5% simple annual interest, the payment size is \$95.55. With $\rho=.10$, the float cost incurred due to this payment received one month late (a $1 \to 5$ or $8 \to 10$ transition) is given by (13.2) as -\$.78.

The default probability $p(\underline{x},A)$ was then computed as the T-step transition probability to state 4 using the pattern recognition estimate for $\underline{\underline{P}}(\underline{x},A)$. Actual sample loan amounts A and loan terms T were used. Expected delinquency costs, D, were obtained from $\underline{\underline{P}}(\underline{x},A)$ and $\underline{\underline{R}}(A,\underline{L})$ using the recursion relation (3.3.1.). Default and non-default expected values, $V_0(A,\underline{L})$ and $V_1(A,\underline{L})$, were computed as indicated in Appendix A, from which the expected loss due to delinquency (D) was subtracted.

The multi-period model described in Chapter 3.2 (Model 2) was then used to estimate $\hat{V}_J$. The number of loans, J, was set to 5 initially and then later to 1 for comparison purposes. The probability of reapplying was chosen to be $\rho=.7$. The typical interest rate of r=.135 was used, and the cost of capital was taken to be $\rho=.10$. Loan reapplications were assumed to occur every 2 years (τ).

Figure 13.1 illustrates the application of Model 4A to a typical installment loan.

FIGURE 13.1. Example of Model 4A

FEATURE VECTOR 25 10 1500 -5 999 15 0 -5 -5 55

P MATRIX						
	.914	.234	.157	.982	1.000	.574
	.139	.075	.153	.070	.335	.041

MONTH	DELINQ COSTS	P(DEFAULT)
0	0.00	.000
1	-0.20	.000
2	-0.48	.000
3	-0.84	.000
4	-1.22	.002
5	-1.63	.004
6	-2.06	.006
7	-2.50	.009
8	-2.95	.012
9	-3.39	.016
10	-3.83	.019
11	-4.27	.023
12	-4.70	.026
13	-5.13	.030
14	-5.54	.033
15	-5.95	.037
16	-6.35	.040
17	-6.73	.044
18	-7.12	.047

DEFAULT PROBABILITY = .0472
AMOUNT = 1500 PAYMENT SIZE = 92.52
DELINQ COSTS = -7.12
NPV OF DEFAULT = -566.84 NPV OF NO DEFAULT = 23.45

EVENT I		P(I)	V(I)
DEFAULT IN PERIOD	1	0.047	-567
NO DEFAULT THRU	1	0.953	23

ENPV = -4.42 (1 PERIOD MODEL)

EVENT I		P(I)	V(I)
DEFAULT IN PERIOD	1	0.047	-567
DEFAULT IN PERIOD	2	0.015	-304
DEFAULT IN PERIOD	3	0.009	-153
DEFAULT IN PERIOD	4	0.006	-65
DEFAULT IN PERIOD	5	0.005	-14
NO DEFAULT THRU	5	0.918	52

ENPV = 14.60 (5 PERIOD MODEL)

Table 13.1 presents the results obtained on the 100 "good" loans and 28 "bad" loans from Data Set B1. When five loans are considered, Model 4A significantly outperforms the benchmark credit scoring model. The loans of Data Set B had all previously been accepted by NSB. This leaves unanswered the question of how well Model 4A performs on loans that NSB rejected. Unfortunately, we do not know the outcome of a rejected loan. Nevertheless, we would expect Model 4A to reject a substantial proportion of the loan applications that NSB rejected. To verify this expectation, Model 4A was used to make credit granting decisions on the 63 rejected loans of Data Set CR. With a five loan horizon, Model 4A rejected 61% of these loans, and with a one loan horizon it rejected 76% of them. Chapter 15 illustrates how performance can be evaluated without actually knowing the outcomes of these rejected loans.

Decision Rule	P_{BG}	P_{GB}	Average Profit per Applicant	Average Profit on Good Loans Rejected	Average Profit on Good Loans Accepted
Model 4A (5 loans)	.32	.61	6.05	6.93	47.60
Model 4A (1 loan)	.18	.71	7.78	10.20	53.60
Benchmark (accept all)	1.00	0	-1.72	-	22.79
Benchmark (S* = 20)	.86	.03	1.76	7.65	23.26
Perfect Information	0	0	21.92	-	21.92

TABLE 13.1

Model 4A Performance on Data Set B1

Chapter 14

CASE STUDY SUMMARY

The installment loan case study of Part II represents a typical application of the decision models to a particular lending institution. In the course of presenting the case study analysis, we have outlined the steps that must be taken to implement the models. Chapter 6 reviewed the data requirements of the models and described the three data sets actually collected. The feature selection problem was discussed in detail in Chapter 7. We should caution the reader that the set of ten features selected represents the result obtained for a particular credit instrument at a given bank. These features may not be the appropriate ones to use for revolving credit instruments, or for installment loan decisions at another institution.

Chapters 8 and 9 presented a general pattern recognition approach to the multivariate estimation of default probability. The algorithms presented do, however, assume that the similarity between loan applicants can be expressed as the weighted Euclidean distance between their feature vectors. Evidence that partially supports this assumption is presented at the end of Chapter 7. Nonetheless, a non-metric approach[1] to this estimation problem is suggested as a possible subject for future investigation.

The clustering pattern recognition algorithm suggested by Sebestyen and Edie [38] was seen to perform at least as well as the other algorithms used. The complexity of the algorithm suggests that its relative performance might improve with larger training sample sizes. If this were found to be the case, we might further investigate the clusters that are defined with a view toward identifying them with consumer sub-populations (market segments).

Chapter 10 demonstrated the significant present value contribution of potential future loans. The mechanics of the Bayesian probability updating rule were illustrated, and shown to agree in spirit with management's belief in the importance of long-term customer relationships. Consumer loan managers often speak of "reaching" to grant credit to new customers. Quantitatively, they are trading-off an increased probability of default on the initial loan for expected future revenues. The results presented in Table 10.3

[1] The theory presented in Chapter 7 provides a basis for such a non-metric approach. The techniques presented by Christensen [28,29] represent a similar non-metric approach based on feature-outcome mutual information.

indicate that for a $2000 installment loan, this tradeoff can be quite significant in terms of the acceptable level of default risk. A default probability of .07 is unacceptable if only the first loan is considered (it has negative expected present value). However, if five loans are considered and the probability of default is updated as each loan is repaid, a default probability of up to .11 still yields a positive expected net present value.

The formulation of Model 2 assumes that the outcome of any loan represents an observation from a Bernoulli process. This assumption permits the use of the beta natural conjugate prior for the specification of the probability updating rule. Although this assumption seems plausible from a theoretical standpoint, its validity should perhaps be verified by further investigation. An investigation of this nature requires that a sample of accounts be observed over the course of several loans. In the absence of such a verification, we nonetheless feel that the results obtained are representative of the significant effect of future loans on the initial credit granting decision.

Chapter 11 considered the problem of estimating the transition probabilities of a Markov process as a function of the applicant's feature vector. Section 11.1 presented evidence that the initial state description was not at all first order. The higher order memory in the process was modeled by expansion of the appropriate states. This expansion was performed in a sequential manner, whereby states are added one at a time so as to maximally reduce the total second order memory of the process. The final state description of Figure 11.3 shows that delinquent/payment behavior contains up to sixth order memory.

Once delinquency occurs, the process is seen to be approximately stationary. The transition probability from the "never-delinquent" on-time state to one month delinquent decreases slightly over time as a perfect payment record is accumulated. However, this small degree of non-stationarity was seen to have a negligible effect on the accept/reject decision.

In general, transitions from a more delinquent to a less delinquent state are possible. These transitions are not observed in Data Set B because the installment loan collection department at NSB attempts to induce a delinquent customer to return to an on-time payment schedule as soon as possible. This type of collection effort is not as prevalent in the case of revolving credit instruments, especially for charge cards, with the result that more than two transitions are possible from delinquent states. However, this possibility poses no conceptual problems for the estimation of the transition probability matrix. Computationally, we must simply estimate an increased number of probabilities.

Chapter 12 presents the results obtained by Model 3 on the 128 loans of Data Set B1. These results indicate that Model 3 significantly outperforms the benchmark credit scoring rule. However, in analyzing the performance of Model 3 we assumed that the loans of Data Set B1 are representative of the applicant population. In fact, we know that they are of a better average quality since they were previously scored and accepted. This pre-screening of Data Set B1 serves to decrease the relative frequency of bad loans. The effect of pre-screening on the results obtained is discussed in Chapter 15.

Chapter 13 develops an approximate solution to the multi-period detailed outcome model. This approximation uses the Markov process model to estimate the default probability and expected delinquency costs. These delinquency costs are then subtracted from the default and non-default outcome rewards to obtain delinquency-adjusted rewards. These estimates are then used as input to the multi-period binary outcome Model 2. As with Model 3, the results presented somewhat overstate the relative performance of Model 4A due to the pre-screening of Data Set B.

In summary, the results of the application of the decision models to the NSB installment loan case study suggest that the proposed models can be efficiently programmed to yield improved credit granting decisions. The presentation of the analysis serves to outline the steps that are required for the application of the models to similar decision problems in other contexts. Finally, the analysis itself has provided a number of insights into the relevant components of the decision to grant credit.

PART III

FURTHER DISCUSSION OF THE MODELS

Part I presented a theoretical development of a general set of consumer credit decision models. Part II presented the results of an application of these models to actual loan histories. Part III reconsiders the models in light of the insights gained by the theoretical and empirical analysis of Parts I and II. In particular, the problems presented by the pre-screening of available data are discussed, and a procedure for removing these biases is presented. In general, the decision models are shown to be readily adaptable to revolving credit instruments. The implications for operational decisions and organizational level policy decisions of the lending institution are discussed. A case is made for variable interest rates on consumer loans based on the quantitative assessment of default risk. Finally, the basic structures of the models are reviewed and shown to have possible application for decision problems in areas other than consumer credit granting.

Chapter 15

SAMPLING FROM A PRE-SCREENED POPULATION

A common characteristic of problems requiring an accept/reject decision rule is that samples, whose outcomes are known, had to have been previously accepted. Had they been previously rejected, they would have never entered the process in the first place. Consequently, the data that is available for evaluating the decision model is of better caliber than the population in general. This bias, due to what will be referred to as "pre-screening", presents problems for model evaluation. Frequently, these biases are simply ignored. This chapter addresses this issue of sampling from a pre-screened population. The nature of the problem is discussed in the context of the credit granting models, and a procedure for explicitly considering the effects of this bias is presented.

Consider the two-outcome description for installment loans that is described by Model 4A of Chapter 13. The outcomes are denoted as $\theta=0$ for bad (default) and $\theta=1$ for good (non-default). A decision rule can be developed as in Part II that gives the decision $\hat{\theta}=0$ for rejection or $\hat{\theta}=1$ for acceptance. To evaluate the decision rule's performance, we can compare its decisions ($\hat{\theta}$) with the actual known loan outcomes (θ) of a given data set. Given $R(\theta,\hat{\theta})$, the reward function for deciding $\hat{\theta}$ when the true outcome is θ, we can compute the average expected reward per applicant by normalizing the total reward on the data set both by the number of loans in the set from class θ and by the a priori probability of class θ.

But here we must be careful to specify what we mean by the a priori probability of class θ. Figure 15.1 indicates that we only observe the outcome θ of loans that were previously accepted ($\delta=1$) by the screening rule in use when the loan application was made. We should also be concerned with our model's performance on loans that the screening rule rejected ($\delta=0$), since if our model is used in place of the screening rule, it must be able to make decisions on these screened rejects as well. The a priori probability of default appropriate for evaluating the model's performance on <u>accepted</u> screened loans ($\delta=1$) is $p(\theta=0|\delta=1)$. If the screening rule has any merit (presumably it does or else it would not be used), then we expect that the default probability for accepted screened loans $p(\theta=0|\delta=1)$ is somewhat <u>less</u> than the default probability $p(\theta=0)$ for the applicant population in general. Conversely, the default probability for rejected screened loans is somewhat greater.

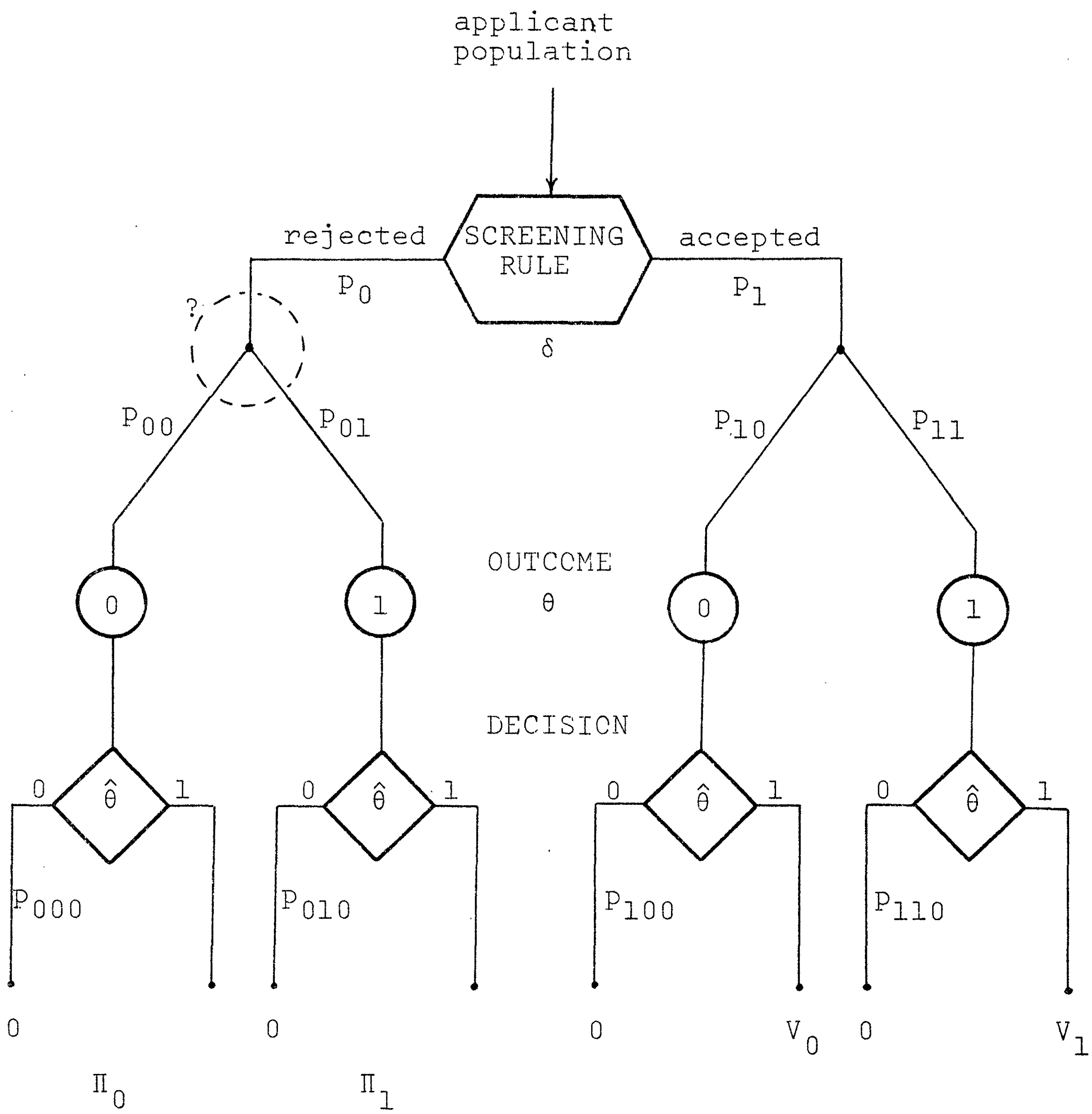


FIGURE 15.1
Decision Making on Pre-screened Samples

We can decompose the a priori default probability as:

$$p(\theta=0) = p(\theta=0|\delta=0)p(\delta=0) + p(\theta=0|\delta=1)p(\delta=1), \qquad (15.1)$$

where

$p(\theta=0)$ is the default probability for the population,

$p(\delta=0)$ is the reject rate of the screening rule,

$p(\delta=1) = 1-p(\delta=0)$ is the accept rate of the screening rule,

$p(\theta=0|\delta=0)$ is the default probability for rejected screened loans, and

$p(\theta=0|\delta=1)$ is the default probability for accepted screened loans.

All but one of these parameters can be obtained directly. The default probability $p(\theta=0)$ for the applicant population can be determined by the method shown in Section 9.1. The reject rate $p(\delta=0)$ of the screening rule is easily ascertained, as is its complement. The default probability for accepted screened loans can be computed as the lending institution's ratio of defaulted loans to total accepted loans. This leaves the a priori default probability for rejected screened loans $p(\theta=0|\delta=0)$ to be computed from (15.1) as a function of $p(\theta=0)$, $p(\delta=0)$, and $p(\theta=0|\delta=1)$. This a priori probability is thus computed without having to observe the outcome of rejected screened loans. Using this probability as the proper normalizing factor, we can evaluate the model's expected performance on rejected screened loans. This evaluation is accomplished in the following manner.

Estimating Performance on Pre-screened Rejected Loans

Since we do not know the outcomes (θ) of these rejected screened loans, we are forced to evaluate performance using average reward values. That is, interpret $R(\theta,\hat{\theta}|\delta=0)$ to be the reward obtained by the decision $\hat{\theta}$ on rejected screened loans ($\delta=0$) if the loan had outcome θ. The average reward for rejected screened loans could be obtained from the sample of rejected loans if we apply the model to each loan to obtain $\hat{\theta}$ and first assume that $\theta=0$. We will denote this average reward as Π_0. We then repeat the procedure assuming $\theta=1$ for each loan to obtain Π_1, the average reward realized by the model on rejected screened loans if the loans had outcome $\theta=1$. The expected average reward on rejected screened loans is then given by

$$\Pi = \Pi_0 p(\theta=0|\delta=0) + \Pi_1[1-p(\theta=0|\delta=0)], \qquad (15.2)$$

where $p(\theta=0|\delta=0)$ is computed using (15.1)

This expected average reward on rejected screened loans is then added to the average expected profit on accepted screened loans (as computed in Part II for each model) to obtain the average expected profit per unscreened applicant. This measure is indicative of the performance we could expect if the present screening rule were replaced by the decision rule being evaluated.

Case Study Results

This bias-compensating procedure was used to evalute the performance of Models 3 and 4 on Data Set B. Data Set B was pre-screened using the NSB benchmark credit scoring rule; that is, the loans in Data Set B were previously accepted by the benchmark screening rule. Because NSB allows its lenders to override the credit scoring decision, the screening rule does not correspond exactly to the strict cutoff score that was optimally determined to be $S^* = 20$. To investigate the performance of the NSB credit scoring decision rule _if_ the cutoff score were strictly adhered to, we include an evaluation of the decision rule "Accept if and only if $S^* > 20$" in the analysis.

Figure 15.2 presents the results obtained after considering the 63 rejected screened loans of Data Set CR. These performance measures represent the average profit per applicant (unscreened) that can be expected on the first loan only. That is, since we do not know the outcome of subsequent loans, we can not make the multi-period evaluation to verify the long-run performance of Model 4 when the planning horizon is extended to J=5 loans.

For the NSB case study, the post-screen default probability, $p(\theta=0|\delta=1)$, was determined to be about .02. This represents a reduction of about 60% from the a priori probability of default $p(\theta=0) = .048$. The screen reject rate, $p(\delta=0)$,

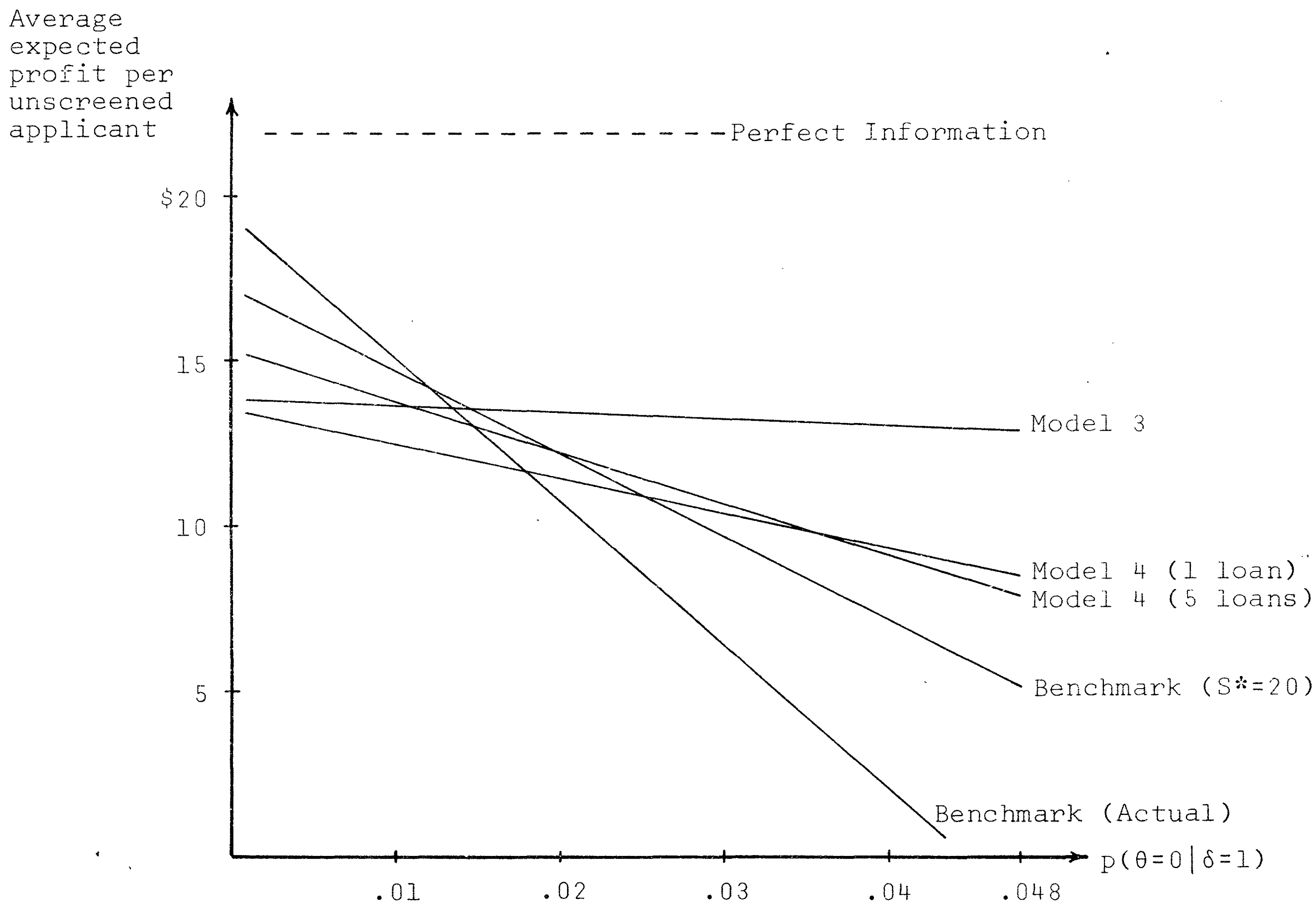


FIGURE 15.2 - Model Performance Adjusted for Screening

is observed to be about 15%. We then evaluated performance for the range of post-screen a priori default probabilities, $p(\theta=0|\delta=1)$, from 0.0 to 0.048. This form of analysis demonstrates the effects of ignoring pre-screened rejected loans in evaluating performance. As can be seen from Figure 15.2, as the screening rule decreases the post-screen a priori default probability, ignoring screening-induced bias overstates the performance of the proposed decision rule.

Figure 15.2 gives the following performance measures for the alternative decision rule given $p(\theta=0|\delta=1) = .02$.

Decision Rule	Average Expected Profit per Unscreened Applicant
Model 3	13.50
Model 4	
(1 loan)	11.40
(5 loans)	12.20
Benchmark	
(S^* = 20)	12.10
(Actual)	10.70
Perfect Information	21.92

Models 3 and 4 appear to significantly outperform the actual benchmark screening rule on the sample of 100 good loans and 28 bad loans in Data Set B1. The strict cutoff score of $S^* = 20$ also seems to outperform the actual screening rule. The implication here is that, on the average, the lending officer's subjective information has negative expected value. Of course, the sample size is not sufficiently large to make these assertions with a strong degree of confidence.

Performance differences on the order of $1.00 per applicant may not seem significant at first, but consumer credit granting is a large volume operation. At NSB, for example, over 30,000 loans are applied for annually. This represents an annual increase in earnings of over $30,000. For larger banks and finance companies, similar performance improvements can realize over one million dollars in additional annual earnings.

Chapter 16

GENERALIZATION TO NEW CREDIT INSTRUMENTS

Historically, consumer credit decision models have been developed to model fixed-term installment loans. Recently, however, credit instruments of a new variety have experienced increasing demand from the consumer population. One such instrument, often referred to as "revolving credit", has progressed from retail charge cards (e.g., Master Charge or Bank Americard) to lines of credit of several thousand dollars that are accessed on demand by simply writing a check on a special account. These "check loans" are predicted to increase in popularity, even to the extent where major purchases such as automobiles may no longer be financed by installment loans.

These new forms of revolving credit cannot be properly modeled by simple credit scoring techniques. The failure of these techniques results from the continuing or "revolving" nature of this form of credit. The definition of a "good" or "bad" loan becomes fuzzy. By "bad" do we mean default within one year after credit granting? Within five years after credit granting? Even if the customer defaults after five years, he may have used his line of credit enough so that the finance charges he has paid more than exceed the amount lost through default.

The revenues from a revolving credit customer depend on the extent to which he uses his line of credit. An account with a high usage factor will be significantly more profitable to the lending institution. In order to make a credit granting decision, it is not sufficient to simply compute loss probabilities. An additional prediction of usage is required to determine expected profit.

For the case of revolving credit, the treatment of delinquent payments warrants reconsideration. On one hand, delinquency is a step in the direction of default and costly collection effort is expended to bring the account up to date. On the other hand, delinquent payments increase the finance charges received and so increase the gross profitability of the account. Consumers are aware of this increased profitability to the lending institution and have been seen to exhibit a more relaxed attitude toward revolving credit delinquency. This behavior makes it highly desirable for credit granting decision rules to properly account for risks and returns associated with delinquency.

Models 3 and 4 can be readily applied to give credit granting decision rules for these newer credit instruments. Delinquency behavior, as it relates to the risk of default, is summarized in the applicant's transition probability matrix $\underline{\underline{P}}(\underline{x},A)$, where now A denotes the ceiling on the line of credit. The revenues from finance charges are summarized in the transition reward matrix $\underline{\underline{R}}(A,\underline{L},U(\underline{x},A))$ where the additional parameter $U(\underline{x},A)$ represents the prediction of usage given the applicant's feature vector $\underline{x}$ and the credit limit A. Operationally, we may define $U(\underline{x},A)$ to be the average fraction of the credit line A that is maintained as a balance in the account. For example: A $2,000 line of credit with U = .5 means that an average balance of $1,000 will be maintained. This balance will result in a monthly finance charge (at 12% annual interest) of $10. If the account becomes one month delinquent, an additional finance charge on the late payment will be due. This additional revenue tends to offset the additional expected loss due to increased default risk. Most banks recognize this aspect of revolving credit and do not initiate delinquent collection procedures until an account becomes two months delinquent.

When a revolving credit customer experiences financial difficulty, he will generally draw on his line of credit until he is no longer able to make the necessary minimum monthly payments. In these cases, default losses tend to be near the credit limit. The state transition Models 3 and 4 can account for this phenomenon through the addition of an appropriate cash outflow as part of the transition reward to the default trap state.

The rapid growth of revolving credit has left managers used to thinking in terms of installment lending with several unanswered questions. The three most frequently asked questions are:

1) How should the initial credit granting decision be made?

2) How should we set the credit limit?

3) How often should the account be reviewed?

In the remainder of this section we will outline how the more complex Models 3 and 4 are well suited to providing these answers.

The Initial Credit Granting Decision

The parameter T, which previously denoted the installment loan term, here represents the planning horizon. The expected net present value $V_d(0|T)$ of T monthly periods provides the appropriate decision criterion. $V_d(0|T)$ is determined recursively using (3.3.1):

$$V_i(t|T) = \sum_j p_{ij}[r_{ij} + \beta V_j(t+1|T)]. \tag{3.3.1}$$

The monthly discount factor, β, can also include the probability that the customer maintains his credit line in an active status. Because revolving credit customers are more varied in their banking habits, this activity factor can play an important role in the credit decision.

Setting Credit Limits

Credit limits must be set on an individual basis, depending on the features of the applicant. A low credit limit reduces the default risk to the lending institution, but a limit below an amount which the applicant can financially maintain as an average balance will unnecessarily reduce finance charge revenues. A limit above the applicant's "financial capability" only serves to expose the lender to additional default risk. The applicant's financial capability can be defined in terms of his feature vector $\underline{x}$. Too high a credit limit (given his features $\underline{x}$) may encourage financial irresponsibility which will be reflected in his transition probability matrix $\underline{\underline{P}}(\underline{x},A)$. Too low a credit limit will reduce the finance charge revenue components of $\underline{\underline{R}}(A,\underline{L},U(\underline{x},A))$. The model's explicit consideration of these aspects permits the formulation of a decision rule for revolving credit of the form:

$$\text{Accept with credit limit } A^* \text{ if } \max_{A} V_d(0|T) > 0,$$

where A^* is the credit limit which maximizes $V_d(0|T)$.

Determining the Review Period

The multi-period extension to Model 4 provides the necessary structure for evaluating the time of next account review. Payment performance is reviewed periodically to determine whether the credit line should be renewed (continued) and whether a credit limit increase is warranted. In terms of our credit granting models, this review process permits an updating of the payment transition probability matrix $\underline{\underline{P}}(\underline{x},A)$ (see Section 3.4). The usage factor $U(\underline{x},A)$ can also be re-estimated in light of observed credit usage, and an updated transition reward matrix $\underline{\underline{R}}(A,\underline{L},U(\underline{x},A))$ can be obtained.

These updated parameters can then be used to re-determine the optimal credit limit. If the new credit limit represents an increase over the old limit, the customer is notified. However, if the new limit is <u>less</u> than that initially granted, it may not be advisable to lower it for reasons of good-will. Nonetheless, the lending institution is made aware that the account may represent a more risky customer and can take whatever action it feels is prudent. This action may take the form of more frequent future reviews, or tighter delinquent collection action. If the account's updated expected net present value is negative, the lender may decide not to renew the loan.

Account review, if done manually, can be costly relative to the gains made by credit limit changes. If automated, these costs may be on the order of from a few cents up to a dollar per account per review, depending on the sophistication of the review process. The more costly the review process, the less frequently will we want to review accounts.

In terms of Model 4, the macro-period corresponds to the review period. The model itself does not restrict the macro-periods to have equal length, T. In fact, we might expect that, initially, reviews should be more frequent until the estimates of payment behavior become more stable (reduced variance in the estimates of $\underline{\underline{P}}(\underline{x},A)$ and $U(\underline{x},A)$). If T_j is the length of the j^{th} review period, we would expect

$$T_1 < T_2 < \dots < T_J,$$

where $T_1 + T_2 + \dots + T_J$ is the length of the planning horizon divided into J macro-periods. In this way, we would be paying decreasing sampling costs for decreasing amounts of information as our estimates become tighter.

The immediate issue is the setting of the next review period T_1. It should be pointed out that this depends conditionally upon the review periods $T_2,\dots,T_J$, and that the optional solution can only be obtained by a computationally involved method of successive approximations.[1] It is clear that approximations to Model 4 must be investigated in order to answer this issue. One such approximation, Model 4A, is suggested in Chapter 13. We believe that the approximation

[1] See Martin [34], Chapter 5. Martin proves that as the macro-periods become longer, the expected net present value during that period approaches the total expected net present value that results from a terminal decision after which no review is taken.

suggested provides an interesting subject for future investigation into solution methods to the problem of review period determination.

Chapter 17

IMPLICATIONS OF THE MODELS

The research presented in previous chapters has wide-ranging implications for lending and other risk-selection industries, as well as for the nature of credit instruments themselves. This chapter is devoted to a discussion of these implications from a variety of viewpoints. For instance, operational credit policy questions can be answered by appropriate parametric analysis of the models presented. These credit granting models further serve as sub-models for decision making at a higher organizational level (e.g., capital budgeting and marketing decisions). Taking the consumer's point of view, we suggest some possible variations on existing credit instruments and credit policies that could increase the availability of credit to the general population. Finally, we indicate that the models presented here have possible application in other industries.

17.1 Operational Credit Policy

Collection Policy

Models 3 and 4 permit the investigation of effects on profitability due to collection policy changes. The transition probabilities of the delinquent Markov process can be partially controlled by the allocation of collection effort at different collection levels or by increased or decreased collection effort in general. The effect on total profitability can be ascertained and weighed against any additional costs associated with these policy changes. On an even more operational level, the models provide the necessary quantitative tools for determining the best allocation of the efforts of an individual collector among the delinquent accounts he is charged with.

Sequential Decision Making

The decision to grant credit can be made in a sequential manner. For example, loans that show very large positive (or negative) expected net present value by Model 1 will also probably be shown to be profitable (or unprofitable) by Model 4. For these cases there is no need to more accurately predict profitability since the decision will not be altered. By applying simple, less costly models first, we may be able to achieve economies within the evaluation and decision process itself. Continuing this sequential approach one step further, we may want to gather feature information about an applicant sequentially. If a credit bureau report, whether favorable or unfavorable, will not change the decision, then it should not be obtained. A savings of over a dollar (on the order of 5% of average expected profit on the first loan) will thus be achieved. This sequential approach can be easily incorporated in a model whose probability estimation procedures have the capability of handling unknown features.

Subjective Information

Although quantitative decision rules are becoming more familiar to credit managers, many find it difficult not to allow the lending officer to exercise some degree of subjective judgement in the final credit granting decision. This "override" option is found to be more prevalent where decision making is decentralized (e.g., a branch banking system). The assumption here is that the human lending officer may have at his disposal information that is not or cannot be captured in a quantiative model. This may be something as insignificant as "The customer seems honest," or something as relevant as "He may have difficulty making the payments now, but in two months he will finish medical school and has a job waiting."

Given that circumstances of this nature are not uncommon, we would want to have the lending officer consider the model's decision in the light of any additional subjective information he might have. If he has none (e.g., a mail order application) he should grant credit strictly on the basis of the objective model's decision (apart from some pattern recognition of his own). If the applicant is extremely atypical (i.e., an outlier in feature space) he might almost completely disregard the model and make his own subjective decision.

Thus, in every credit granting decision there will be this mix of objective and subjective information. We might argue that better (more experienced) lenders should be permitted to more heavily weigh the value of subjective information. An untrained lender, on the other hand, should initially be constrained to follow the model's decision. It is clear that any effective management decision system must also provide methods for measuring both the objective/subjective mix being employed by each lender and the value of his subjective information.

Including Exogenous Economic Conditions

Most credit granting models do not include the effects of exogenous economic conditions on the earnings of consumer credit operations. These effects can be readily modeled by the inclusion of selected economic indicators in the estimation of default probabilities or transition probabilities. Since economic conditions and local demographic characteristics vary over time, any practical credit granting decision system should be able to track these changes in the applicant population.

Modeling Costs

Models 1 through 4 represent four models of both increasing complexity and increasing cost of application. As with any modeling effort, we must remain aware of the cost effectiveness of the modeling process itself. A model, such as Model 4, that is costly to parameterize and implement should provide performance increases that at least compensate for these costs. It seems quite likely that the complexity of the model that will be cost effective will depend on the volume of credit applications processed and the availability of the data needed for model parameterization.

17.2 Organizational Decision Making

Related Consumer Services

The models are also sufficiently flexible to allow for the effect of inter-related consumer services. At a bank, for example, if a customer's credit application is accepted, he will be more likely to open a new savings account than if his application is rejected.

This cross-sell effect of the loan decision on other services will be included as follows. Let $s_i' = 1$ if service type i is subscribed to at the time of the credit granting decision; $s_i' = 0$ if it is not. Similarly let s_i'' be the indicator variable for service type i shortly after the decision. Let $\hat{\theta} = 1$ if credit is granted, and 0 if it is not. Let $p(s_i''|s_i',\hat{\theta})$ be the appropriate probability of service i after the decision given the state before the decision itself. Finally, let Y_i be the expected net present value of service type i for the particular customer in question. Then the cross-sell effect of the loan decision is given by

$$\sum_i \sum_{s_i''} p(s_i''|s_i',\hat{\theta})(s_i'' - s_i')Y_i.$$

Thus, the additional contribution to other service revenue due to acceptance (relative to rejection) is given by

$$\text{Cross-Sell Effect} = \sum_i \sum_{s_i''} [p(s_i''|s_i',\hat{\theta}=1) - p(s_i''|s_i',\hat{\theta}=0)](s_i''-s_i')Y_i.$$

Considering the credit granting decision in the context of related services will have the effect that some applicants that before appeared marginally unprofitable will now be accepted.

Marketing Policy Decisions

These models also provide excellent tools for marketing policy decisions. The demographic characteristics of an applicant are explicitly incorporated through his feature vector $\underline{x}$. The credit granting model computes the expected net present value of an applicant, and if this value is positive credit is granted. Let $V^*(\underline{x})$ be this value for an applicant with features $\underline{x}$ ($V^*(\underline{x}) = 0$ if credit is not granted). Let $N(\underline{x})$ be the number of applicants with features $\underline{x}$. Then total profit will be given by

$$V^*_{total} = \int_X V^*(\underline{x})N(\underline{x})d\underline{x}.$$

If marketing policy is to be re-evaluated (e.g., by advertising to a particular market segment), the effect of alternative policies on consumer credit profits is given by

$$\Delta V^*_{total} = \int_X V^*(\underline{x})[N'(\underline{x})-N(\underline{x})]d\underline{x},$$

where $N'(\underline{x})$ is the number of applicants with features $\underline{x}$ that are expected if the alternative policy is adopted. This formulation permits market planners to evaluate the effect of alternative marketing strategies given the credit granting decision rules being used.

Capital Budgeting

Credit granting decision models of the type presented in this thesis provide the tools needed to evaluate policy alternatives that affect more than just the consumer credit department of the organization. If capital must be rationed, the models can be used to determine which applicants represent the most profitable lending opportunities. If additional sources of capital are available, but at a higher cost, the models can be used to evaluate the expected return on this capital if it is used for consumer loans as opposed to other investment opportunities.

17.3 Credit Instruments From The Consumer's Viewpoint

From an investment theory viewpoint, it appears that decision rules of the type suggested above result in inefficiencies in the consumer loan market. That is, the expected return on a consumer loan is only partially related to the risk inherent in granting the loan. To the extent that different risk categories of loans carry different interest rates (r), the lending institution is compensated for risk. However, within a loan type the compensation for the riskiness of an individual applicant takes only the form of a binary accept/reject decision. Perhaps one can argue that the consumer loan market as a whole is efficient because of the existence of finance companies and other such institutions that charge relatively higher interest rates. However, we feel a case can still be made for an inefficient market on the grounds that the banking industry within a loan risk class only compensates for risk to the extent that it accepts or rejects a particular applicant.

From the consumer's point of view, this inefficiency can have disconcerting consequences. Take, for example, the consumer whose car loan at $r = 14\%$ is rejected because the net present value of expected cash flows is -$10. His disutility of doing without a car or of seeking out other loan sources may be significantly greater than $10. In fact, such a customer may be quite willing to pay a "risk-compensating charge" of $11 so as to raise his expected value to +$1. By such a transfer, the inefficiency inherent in the binary credit granting decision rule would seem to be resolved in a manner whereby both the bank and the consumer would benefit. Of course, the "risk-compensating charge" of $11 could be amortized over the term of the loan by an increase in the interest rate, which is what is likely to happen if the consumer seeks other loan sources.

As the consumer demonstrates his "credit-worthiness" through payment behavior, an updating of his default probabilities (as implied by the multi-period models) would soon reduce his interest rate to the normal level. In this manner, a consumer who was previously denied the chance to obtain credit can, with only a small risk-justified expense, establish credit with which to finance both immediate and future needs.

The risk-adjusting of interest rates could work in the opposite direction for low-risk consumers. Individuals who demonstrate good payment performance over a period of time are in effect being discriminated against by current credit practices. As they demonstrate their reliability for handling credit obligations, they should be charged lower risk premiums in the form of reduced interest rates.

These variable interest rate adjustments can be determined parametrically from the credit granting models presented in Chapter 3. In fact, many banks and other finance institutions feel a social responsibility to make credit available to consumer sub-populations who might otherwise be denied credit. Models such as these would enable the lending institution to quantify the extent to which they, in effect, are subsidizing these high risk groups. Such a quantification might open the door to a government subsidized minority lending program. Such programs already exist on a piecemeal basis, but are generally restricted to educational and home mortgage loans.

The concept of variable interest rates based on the assessed risk of the customer is only novel to consumer credit granting — such practices are commonplace for commercial credit granting. The primary obstacles in the consumer area have been the difficulty of making the risk assessment economically for a large volume of applicants. Automated credit granting decision rules may provide the means for implementing these concepts in the area of consumer credit granting.

17.4 Application To Other Problem Areas

We have previously indicated how the models presented can be parameterized for application to any lending institution. This parameterization is necessary to account for demographic variations due to both geographical location and the particular credit instrument being modeled. With reparameterization, these models can be used for credit granting decisions of banks, finance companies, retail companies, airlines, gasoline credit sales, and so-called "luxury" credit cards. Moreover, the general structure of this class of models finds applications in areas other than credit granting. A brief review of the model structure will be given and its potential applicability to three unrelated areas will be suggested.

General Structure of the Model

Model 4, the multi-period detailed outcome model, is the decision model that includes all the properties of Models 1, 2, and 3. Although computationally impractical, approximate solutions can be devised to capture the most important aspects of a given problem. One such approximation, Model 4A, is suggested in Chapter 13.

The decision maker is faced with a binary accept/reject decision. If he rejects the applicant, no further action takes place. If he accepts him, the applicant makes transitions among the states of a Markov chain. The decision maker's state of knowledge about the applicant is summarized in a transition probability matrix $\underline{\underline{P}}(\underline{x})$ and transition reward matrix $\underline{\underline{R}}(\underline{x})$, where both matrices are conditioned on a vector of features $\underline{x}$ characterizing the applicant, and any other relevant parameters.

The reward matrix $\underline{\underline{R}}(\underline{x})$ is estimated. It may or may not be known with certainty. The transition probability matrix is assumed to have a matrix-beta distribution. The expected value of this distribution is used in evaluating expected discounted rewards. We have shown in Chapter 5 how this expected transition probability matrix $\underline{\underline{P}}(\underline{x})$ can be estimated for a particular applicant by pattern recognition techniques.

The decision maker has the option of observing transitions over a specified macro-period to update his estimates of the transition probability matrix. If the reward matrix is not known with certainty, it too can be updated in the light of the account's behavior during the review period. At this time, the decision maker can re-evaluate his decision.

In addition to the feature vector $\underline{x}$, the transition probability and reward matrices can be conditioned upon a set of policy parameters. Such a parametric formulation permits the investigation and possible optimization of policy alternatives.

Other Areas of Application

A model of this general nature was shown to be appropriate for installment loan credit granting decisions through the case study application of Part II. Chapter 16 showed how it could be readily applied to revolving credit. In addition, it appears that the model has a potential application for insurance underwriting, school admission, and criminal justice-related decisions. The nature of these applications will be briefly discussed.

The insurance underwriting decision is somewhat similar to the credit granting decision. An applicant for automobile, fire, or health insurance is either rejected or accepted by the insurance underwriter. If accepted, the customer makes transitions among different outcome states, which may represent different accident categories. These transition probabilities $\underline{\underline{P}}(\underline{x})$ can be initially estimated as a function of the applicant's features, which may include his age, geographical location, driving record (in the case of automobile insurance), occupation, etc. The reward matrix can be computed as a function of insurance liability limits and premium assessments.

A periodic review of the insurance customer allows for updating his transition probability matrix. In the case of automobile insurance, a successful driving record during the review period might warrant a premium reduction. This class of models might serve in this manner as the basis for a merit rating system.

The school admission problem of determining which applicants to accept and which to reject is suggested as another problem area that may be amenable to analysis by the methods presented here. Candidates can be characterized by a vector of features or attributes that may include previous grades, test scores, reference evaluations, and interview impressions. If admitted, the student "outcome" may be grossly described by either eventual graduation or else failure or dropout. On a more detailed level, we might describe his outcome on a semester-by-semester basis, where his outcome in any one semester could be given by a set of states (e.g., good standing, marginal standing, and failure or dropout). We would then be faced with the problem of predicting his transitions among these states as a function of the features ascertained at the time of application.

Within the criminal justice system, we find decision problems of a nature similar to credit granting. One problem is the decision to grant pre-trial release of an individual arrested for alleged wrongdoing. If he is released on his own recognizance, there is a certain probability that he will not appear at his arraignment. This probability can be estimated as a function of the individual's features, and then used to weigh show/no-show outcome rewards that might be determined by an appropriate cost-benefit analysis.

A second problem is the decision to grant parole. If parole is granted, the individual may later be rearrested for another (possibly different) crime.[1] The possibility of rearrest may be posed as a pattern recognition problem. The multi-period aspects of the problem, namely the potential for subsequent crimes, suggest that a probability updating scheme similar to Models 2 and 4 may be appropriate.

Further discussion of these applications will be left as topics for possible future investigation. It has only been our intent to suggest these particular problems as examples of the applicability of the pattern recognition based Markovian decision model to problem areas other than credit granting.

[1] A. Blumstein and R. Larson ["Models of a Total Criminal Justice System," Operations Research (March-April 1969), pp. 199-232] present an excellent introduction to this problem area.

REFERENCES

Credit Granting References

1. BIERMAN, HAROLD, JR. and HAUSMAN, WARREN, H., "The Credit Granting Decision", Management Science, Vol. 16, No. 8, (April, 1970), pp. 519-532.

2. BOGESS, WILLIAM P., "Screen-test Your Credit Risks, "Harvard Business Review, (Nov. - Dec., 1967), pp. 113-122.

3. CHATTERJEE, S., and BARCUN, S., "A Nonparametric Approach to Credit Screening", Journal of the American Statistical Association, Vol. 65, No. 329, (Mar., 1970), pp. 150-154.

4. COHEN, K. J., and HAMMER, F. S., Analytical Methods in Banking, Richard V. Irwin, Inc., Homewood, Illinois, 1966.

5. COHEN, K. J., and HAMMER, F. S., "Critical Comments on "Measuring Risk on Consumer Instalment Credit" by Paul F. Smith, Management Science, Vol. 12, No. 9, (May, 1966), pp. 743-744.

6. COLE, ROBERT H., Consumer and Commercial Credit Management, Richard D. Irvin, Inc., Homewood, Illinois, 1972.

7. CRANE, DWIGHT B., "Marketing Strategy and Bank Service Interaction: A Probability Model", Journal of Bank Research, (Summer, 1970), pp. 49-56.

8. CURTIS, WILLIAM W., "Social Class or Income?", Journal of Marketing, (Jan., 1972), pp. 67-68.

9. CYERT, R. M., DAVIDSON, H. J., and THOMPSON, G. L., "Estimation of the Allowance for Doubtful Accounts by Markov Chains", Management Science, (April, 1962) pp. 287-303.

10. CYERT, R. M. and THOMPSON, G. L., "Selecting a Portfolio of Credit Risks by Markov Chains", The Journal of Business of the University of Chicago, Vol. 41, No. 1, (Jan., 1968), pp. 39-46.

11. FAZIO, VINCENT J., "Consumer Payment Practices and Preferences in the Boston Metropolitan Area", S. M. Thesis, M. I. T., (1972).

12. GREER, CARL C., "Measuring the Value of Information in Consumer Credit Screening", Management Services, (May-June, 1967), pp. 44-45.

13. GREER, CARL C., "The Optimal Credit Acceptance Policy", Journal of Financial and Quantitative Analysis, (Dec., 1967), pp. 399-415.

14. LIEBMAN, LEON H., "A Markov Decision Model for Selecting Optimal Credit Control Policies", Management Science, Vol. 18, No. 10, (June, 1972), pp. 519-525.

15. MAJONE, G., "Classification and Discrimination in the Analysis of Credit Risks: 1", Management Sciences Research Report No. 120, Carnegie-Mellon University (1968).

16. MAJONE, G., "Classification and Discrimination in the Analysis of Credit Risks: 2", Management Sciences Research Report No. 128, Carnegie-Mellon University (1968).

17. MATHEWS, H. LEE and SLOCUM, JOHN W., JR., "Social Class and Commercial Bank Credit Card Usage", Journal of Marketing, (Jan., 1969), pp. 71-78.

18. MATHEWS, H. LEE and SLOCUM, JOHN W., JR., "Correlatives of Commercial Bank Credit Card Use", Journal of Bank Research, (Winter, 1972), pp. 20-27.

19. MEHTA, D., "The Formulation of Credit Policy Models", Management Science, (Oct., 1968), pp. 30-50.

20. MITCHNER, MORTON and PETERSON, RAYMOND, P., "An Operations-Research Study of the Collection of Defaulted Loans", Operations Research, (1967), pp. 522-545.

21. MYERS, JAMES H. and FORGY, EDWARD W., "The Development of Numerical Credit Evaluation Systems", American Statistical Association Journal, (Sept., 1963), pp. 522-545.

22. ORGLER, YAIR E., "Evaluation of Bank Consumer Loans with Credit Scoring Models", Journal of Bank Research, (Spring, 1971), pp. 31-37.

23. SLOCUM, JOHN W., JR., and MATHEWS, LEE H., "Social Class and Income as Indicators of Consumer Credit Behavior", Journal of Marketing, (April, 1970), pp. 69-74.

24. SMITH, PAUL F., "Measuring Risk on Consumer Instalment Credit", _Management Science_, Vol. 11, No. 2, (Nov., 1964), pp. 327-340.

25. WEINGARTNER, H. MARTIN, "Concepts and Utilization of Credit-Scoring Techniques", _Banking_, (Bef., 1966), pp. 51-58.

Statistical References

26. ANDERSON, T. W., _An Introduction to Multivariate Statistical Analysis_, John Wiley & Sons, Inc., New York, 1958.

27. BILLINGSLEY, P., _Statistical Inference for Markov Processes_, Univ. of Chicago Press, Chicago, Ill., 1961.

28. CHRISTENSEN, R. A., "Entropy Minimax, a Non-Bayesean Approach to Probability Estimation from Empirical Data", _Proc. IEEE Conf. on Systems, Man, and Cybernetics_, Boston, Mass., (Nov. 5-7, 1973), pp. 321-325.

29. CHRISTENSEN, R. A., "Entropy Minimax Processing of Incomplete Data Sets", _Proc. IEEE Conf. on Systems, Man, and Cybernetics_, Boston, Mass., (Nov. 5-7, 1973), pp. 326-327.

30. COVER, T. M. and HART, P. E., "Nearest Neighbor Pattern Classification", _IEEE Transactions on Information Theory_, Vol. IT-13, No. 1, (Jan. 1967), pp. 21-27.

31. FANO, R. W., _Transmission of Information_, The M.I.T. Press, Cambridge, Mass., 1968.

32. HOWARD, RONALD A., _Dynamic Programming and Markov Processes_, John Wiley & Sons, Inc., New York, 1960.

33. KULLBACK, S., _Information Theory and Statistics_, John Wiley & Sons, Inc., New York, 1959.

34. MARTIN, J. J., _Bayesian Decision Problems and Markov Chains_, John Wiley & Sons, Inc., New York, 1967.

35. MENDENHALL, WILLIAM, _The Design and Analysis of Experiments_, Wadsworth Publishing Co., Inc., Belmont, Calif., 1968.

36. RAIFFA, H. and SCHLAIFER, R., _Applied Statistical Decision Theory_, The M.I.T. Press, Cambridge, Mass., 1961.

37. PATRICK, F. A. and FISCHER, F. P., III, "A Generalized K-Nearest Neighbor Rule", Information and Control, Vol. 16, No. 2, (Apr. 1970), pp. 128-152.

38. SEBESTYEN, G. and EDIE, J., "An Algorithm for Non-Parametric Pattern Recognition", IEEE Transactions on Electronic Computers, Vol. EC-15, No. 6, (Dec. 1966), pp. 908-915.

Appendix A

LOAN PROFITABILITY ANALYSIS

This appendix serves to quantify in economic terms the expected net present value to the lending institution of both defaulted and non-defaulted installment loans. The analysis identifies the revenue and cost components as a function of loan parameters $(A, \underline{L})$, where A is the loan amount and $\underline{L}$ is a vector of other loan attributes (e.g., interest rate, term, etc.). The following table defines and summarizes the parameters explicitly considered.

A	loan amount
a	monthly payment
r	simple annual interest rate
ρ	annual cost of capital
T	loan term (number of monthly payments)
t_0	number of months between the time the loan is granted and the first payment is due
t_1	number of months from receipt of last payment to default
r'	$= r/12$ monthly interest rate
ρ'	$= \rho/12$ monthly cost of capital
γ	fraction of the original loan amount (A) that will be charged off *given* the loan defaults
Z	present value of the fraction of the charged-off amount (γA) that is recovered after default

C	present value of costs associated with post-default recovery efforts
$R_0(A,\underline{L})$	expected net present value of cash flows until the loan defaults, given that it defaults
$R_1(A,\underline{L})$	present value of cash flows of a non-defaulting loan
$V_0(A,\underline{L})$	expected net present value of a loan with parameters $(A,\underline{L})$, given the loan defaults
$V_1(A,\underline{L})$	net present value of a loan with parameters $(A,\underline{L})$, given the loan does not default

A.1 Economic Components of Installment Loan Profitability

The primary components of an account's profitability are:

1) net discounted revenues
2) charge-off or default losses
3) recovery revenues after default
4) delinquency-related costs
5) loan processing costs.

A brief description of a possible account history will clarify the meaning of these components.

An installment loan is made for a fixed amount (A) and usually for a given purpose, e.g., car purchase, home improvement, etc. Repayment of the loan is usually in equal monthly installment payments for a specified number of payments called the term of the loan (T), typically 12 to 36 months. The sum of the payments exceeds the loan amount A so as to yield a net positive return to the bank.

During its term the account may be delinquent in making a payment. If the delinquency continues, collection effort in the form of letters and telephone calls is initiated. This collection effort thus incurs costs which detract from the overall profitability of the account. These expected delinquency-related costs will be expressed as D.

Should the account become more than 90 days delinquent it is removed from the books as an asset through an accounting procedure known as "charge-off" or "write-off". It is at this point that the loan is considered to be a bad or defaulted loan. Usually a number of installments will have been paid, the fraction remaining will be denoted by γ and the charge-off balance will be γA.

Although such a charged-off loan is considered a bad debt and has been removed from the active account books, it is still legally recoverable for up to seven years. A separate recovery section then attempts to recover the charged-off balance, γA. Typically, the section will recover a fraction (about one-third) of the charged-off balance. The actual value of these recovery revenues is the present value of the recovered cash flows minus the present value of the costs associated with the recovery effort. The present value of the expected recovered cash flows as a fraction of the charge-off balance, γA, will be denoted by Z, so that the present value of recoveries will be $Z(\gamma A)$. The present value of the expected associated costs will be denoted by C.

A.2 Net Discounted Revenues

Revenue on a loan account is dependent upon the annual interest rate r at which the loan is made. This interest rate fluctuates with the business cycle, and is often quoted as a simple annual rate. This rate is currently seen to be about $r = 13.5\%$ for the average loan.

The lending institution itself is typically highly leveraged (debt /equity ratio in excess of 10.0). Capital is typically borrowed from depositors or in the Federal money market.

The annual cost of capital ρ is the appropriate discount factor to use in computing the net present value of loan related cash flows. Rather than concern ourselves with a detailed estimation of ρ, we rely on an estimate quoted by the bank of $\rho \approx .10$. This estimate seems reasonable in light of the risk characteristics of a large commercial bank.

The cash flows of an installment loan are of the following nature. A cash outflow of $-A$ at time $t=0$ is made when credit is granted. T equal monthly payments of size a are made, with the first payment due one month after the loan is made. The payment size is given by the formula

$$a = \frac{Ar'(1+r')^T}{(1+r')^T - 1} \qquad \text{(A.1)}$$

where $r' = r/12$ is the monthly interest rate.

Cash Flow Given No Default

If the loan does not default, the net present value of its cash flows will be

$$R_1(A,\underline{L}) = -A + (1+\rho')^{-t_0} \sum_{t=0}^{T-1} a(1+\rho')^{-t}, \qquad (A.2)$$

where $\rho' = \rho/12$ is the monthly cost of capital. The factor $(1+\rho')^{t_0}$ in the denominator adjusts the present value to the time of loan granting since t_0 is the average delay from loan granting to the first payment date.

Cash Flow Given Default

If the loan defaults, these revenue cash flows will terminate at the last payment made. If γ is the average fraction of the loan charged off, then the expected number of payments made will be approximately $M = (1-\gamma)T$. The net present value of these M payments as a function of γ is given by

$$R_0(A,\underline{L}|\gamma) = -A + (1+\rho')^{-t_0} \sum_{t=0}^{M-1} a(1+\rho')^{-t}, \qquad (A.3)$$

where $M = (1-\gamma)T$.

If $p(\gamma)$ is the probability that the fraction γ will be charged off given that the loan defaults, then the expected net present value of these cash flows given default is

$$R_0(A,\underline{L}) = \int_{\gamma=0}^{1} R_0(A,\underline{L}|\gamma)p(\gamma)d\gamma. \qquad (A.4)$$

The distribution $p(\gamma)$ will be estimated in the following section on charged-off losses.

A.3 Charge-Off Or Default Losses

Part of the loss due to the loan defaulting can be considered to be the expected opportunity loss $R_1(A,\underline{L}) - R_0(A,\underline{L})$ equal to the expected net present value of payments that should have been received but were not. As pointed out above in equation (A.4) this opportunity loss is dependent upon the distribution of γ, the fraction of payments not received.

In order to obtain an estimate for $p(\gamma)$, a sample of defaulted loans was collected. All 220 installment loans charged off during the period January 1972 through October 1972 were evaluated and the following information was recorded.

1) account number and loan type
2) approved date
3) loan amount (A)
4) charge-off amount (γA)
5) date at which account became delinquent
6) date of charge-off
7) loan term (T)

Given this information we were able to compute the fraction charged off $\gamma = (\gamma A)/A$, the delay from credit granting to first payment (t_0), and the delay from last payment to charge-off (t_1).

We are interested in t_0 and t_1 so as to properly time the cash flows. The average values for these delays were found to be

$$t_0 = .89 \text{ months}, \ t_1 = 3.6 \text{ months}.$$

These delays reflect delays inherent in the loan administration process. Explicitly including them in the model allows for later sensitivity analysis on the effect of administrative policy changes.

The more critical problem is that of estimating $p(\gamma)$. The following linear models were explored:

$$\gamma = \bar{\gamma} \tag{A.5.1}$$

$$\gamma = b_0 + b_1 T + b_2 A. \tag{A.5.2}$$

The coefficients of (A.5.1) and (A.5.2) were estimated by multiple linear regression, with the following results:

$$\gamma = \bar{\gamma} = \underset{(.289)}{.472} \tag{A.5.1a}$$

$$\gamma = .48 - \underset{(.26)}{.084T} + \underset{(.10)}{0.0A} \tag{A.5.2a}$$

$$(R^2 = .07, F = .034)$$

Values beneath the coefficient estimates are standard errors of the estimate.

A visual examination of scatterdiagrams of γ versus T and γ versus A suggest that default is likely to occur uniformly throughout the loan term. That is, conditional upon the occurrence of default, the bank is likely to receive anywhere from 0 to T-1 payments with equal probability. This suggests the distribution:

$$p(\gamma) = \begin{cases} T/(T-1) & 0 < \gamma \le (T-1)/T \\ 0 & (T-1)/T < \gamma \le 1 \end{cases} . \qquad (A.6)$$

The average loan term for the sample was 24.1 months. If γ is uniformly distributed between 0 and T-1 according to (A.6) then the expected value of γ is:

$$E[\gamma] = (T-1)/2T$$

which for T = 24.1 months is $E[\gamma] = .481$. This agreement with $\bar{\gamma} = .472$ given by (A.5.1a) supports the hypothesis that γ is uniformly distributed according to (A.6).

We can now rewrite (A.4) as

$$R_0(A,\underline{L}) = -A + (1+\rho')^{-t_0} \sum_{M=0}^{T-1} \frac{1}{T-1} \sum_{t=0}^{M-1} a(1+\rho')^{-t}. \qquad (A.7)$$

A.4 Recovery Revenues After Default

Two of the required parameters, the average discounted recovery cost per loan (C) and the present value of the average ratio (Z) of discounted recovery revenues to charge-off amount were estimated from information provided by the Installment Loan Recovery Section (LRS) of NSB. The average discounted cost of collecting a defaulted loan is estimated by

$$C = \sum_{t=.5}^{9.5} c_t(1+\rho)^{-t}, \qquad (A.8)$$

where c_t is the annual cost of pursuing a defaulted loan t years after default. The midpoint (t = .5, 1.5, 2.5,...) of the recovery year is used since, for example, 0 to 1 year-old loans have average age of 0.5 years.

Four loan adjusters pursue all defaulted loans, with the following division of labor.

Number of Adjusters	Loan Worked by an Adjuster by Year Charged Off
1	1971
1	1970
1	1965, 1967, 1968 (1/2), 1972 (1/2)
1	1966, 1968 (1/2), 1969, 1972 (1/2)

1964 loans receive about 3 hours effort per month.
1963 loans receive about 1 hour effort per month.

From this table, we can estimate e_t, the proportion of the LRS's effort spent on loans of average age t.

Average Time Loan Has Been in LRS	Proportion of LRS Time Spent on Loan of Average Age t
.5	.30
1.5	.25
2.5	.24
3.5	.10
4.5	.05
5.5	.03
6.5	.02
7.5	.01
8.5	.005
9.5	.003

Table A.1

Proportion of LRS Time Spent on Loan of Average Age t

At the time of this survey, the loan recovery section was pursuing 534 defaulted loans. The annual budget for the section was \$75,000. This implies that on a fully allocated basis the cost of pursuit after t years since default is

$$c_t = \frac{75000}{534} e_t. \qquad (A.9)$$

Since we use (A.9) and the e_t values of Table A.1 to obtain c_t, we can estimate C from (A.8). Using $\rho = .10$ we obtain

$$C \approx \$106.$$

The data of Table A.2 provides a means for estimating Z, the average ratio of discounted recovery revenues to charge-off amount.

Year Charged-Off	Year of Recovery			
	1968	1969	1970	1971
1962	.0019	.014	.013	.004
1963	.0495	.0210	.0074	.0033
1964	.0247	.0167	.0135	.0074
1965	.0403	.0313	.0250	.0187
1966	.0762	.0533	.0302	.0200
1967	.1560	.0806	.0582	.0345
1968	.0930	.1515	.0585	.0387
1969	-	.1420	.0991	.0582
1970	-	-	.0873	.1100
1971	-	-	-	.1238

Table A.2
Proportion of Loan Recovered

This distribution of recovery revenues is plotted in Figure A.1 as the fraction z_t recovered t years after default. If we let z_t be the average fraction recovered t years after default we estimate Z using

$$Z = \sum_{t=.5}^{9.5} z_t(1+\rho)^{-t} \tag{A.10}$$

Using $\rho = .10$ and z_t values from Figure A.1 we obtain:

$$Z = .37.$$

A.4 Delinquency Costs and Processing Costs

Delinquency costs are extremely difficult to estimate on an average loan basis, and are only mentioned in this appendix for the sake of completeness. Model 3 was developed for the express purpose of accounting for delinquency related costs.

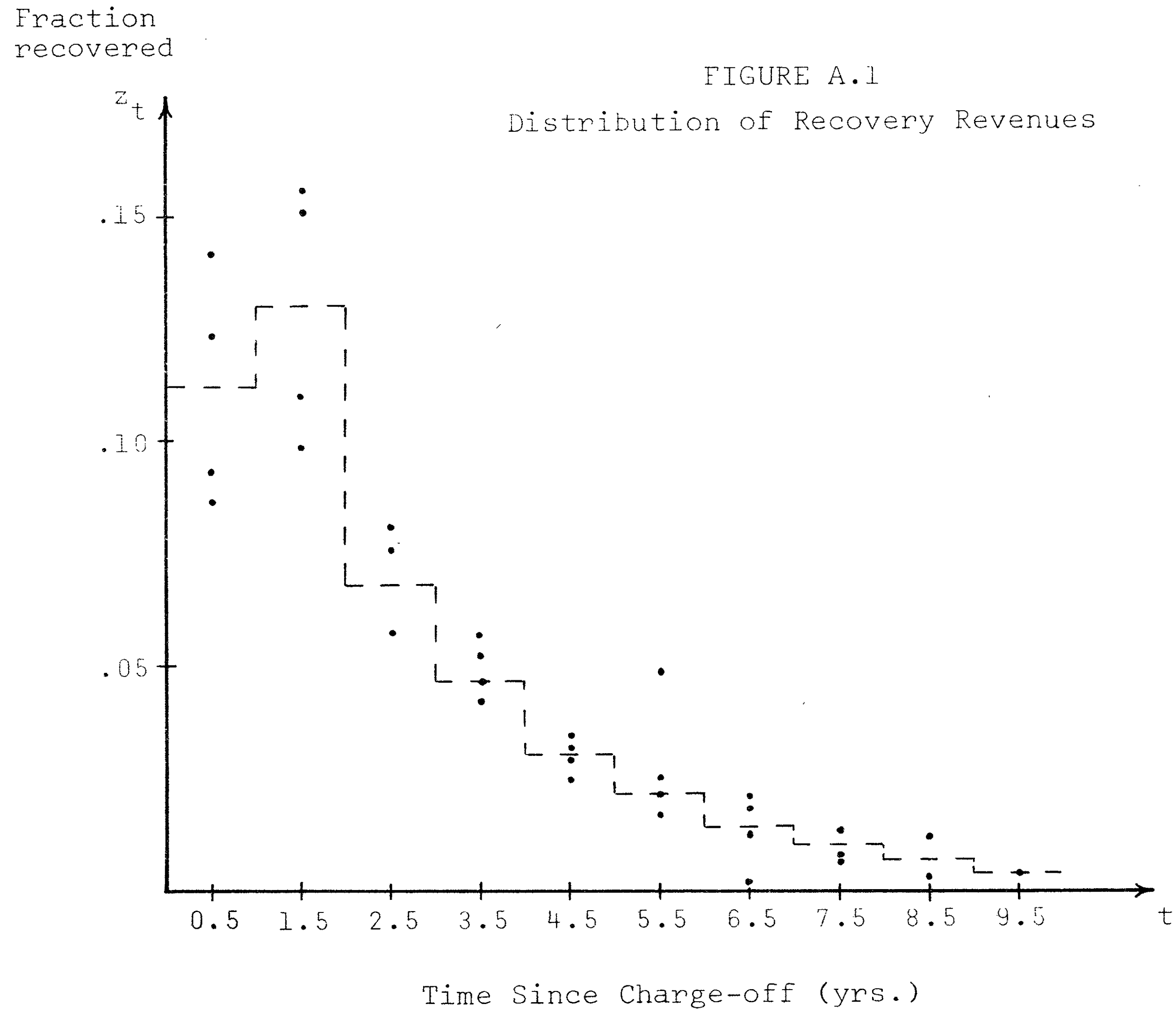
Administrative costs associated with putting the loan onto the books and processing monthly payments are also difficult to identify. For the purpose of this thesis they are estimated at $10 per loan.

A.5 Obtaining $V_0(A,\underline{L})$ and $V_1(A,\underline{L})$

The results obtained in sections A.1 through A.4 are now combined to give $V_0(A,\underline{L})$ and $V_1(A,\underline{L})$, the expected net present value given default and non-default respectively. These are:

FIGURE A.1

Distribution of Recovery Revenues



$$V_0(A,\underline{L}) = R_0(A,\underline{L}) + (Z\overline{\gamma}A-C)(1+\rho')^{-[(1-\gamma)T+t_0+t_1]} - 10, \quad (A.11)$$

where the first term is the expected discounted revenues computed using (A.1) and (A.7); the second term is the recovery revenues net of pursuit costs discounted to the average time at which the loan enters the recovery section, i.e., $(1-\gamma)T + t_0 + t_1$; and the third term is the administrative processing costs.

$$V_1(A,\underline{L}) = R_1(A,\underline{L}) - 10, \quad (A.12)$$

where the present value of T monthly payments, $R_1(A,\underline{L})$ is given by (A.1) and (A.2), and administrative processing costs are taken to be $10.

Appendix B

FEATURE DESCRIPTIONS

1. Occupation

Code		Score
1	sales managers retired foreman professional workers	25
2	management middle management elected officials craftsmen	15
3	lower middle management clerical workers operatives semi-skilled technical workers	5
4	proprietor, self-owner	0
5	laborers, sales workers	0
6	other service workers transportation drivers	0

2. Years at Occupation - (.1 yrs.)

3. Loan Amount - (10$)

4. Term - Months

5. Purpose

Code		Score
1	taxes, boat, education, vacation, auto, insurance	20
2	home improvement, cycle	15
3	business & investment, home furnishings	10
4	funeral, personal	5
5	bills, medical, moving	0
6	renewal	-5

6. Marital Status

1. single
2. married
3. widowed
4. separated
5. divorced

7. Age

Code	Range	Score
1	18-24	0
2	25-36	-5
3	37-42	2
4	43-50	11
5	51-56	15
6	57-62	20
7	62+	10

8. Dependents - (number of)

9. Own/Rent - (residence)

10. Years of Residence - (.1 yrs.)

11. Income

Code	Range	Score
1	0-300	-5
2	301-420	0
3	421-540	5
4	541-700	10
5	700-1000	15
6	1000+	20

12. Mortgage/Rent - ($)

13. Total Debt - (10$)

14. Telephone - 1 Yes; 2 No

15. Years at Former Residence - (.1 yrs.)

16. Years with Former Employer - (.1 yrs.)

17. Other Income - ($)

18. Checking - 1 Yes; 2 No

19. Savings - 1 Yes; 2 No

20. Auto - 1 Yes; 2 No

21. Total Monthly Payments - ($)

22. Ability to Pay

$$\frac{\text{net mo. income} + 1/2\ (\text{other income})}{\text{total mo. payments} + \text{rent or mortgage} + 20\ (\text{dependents})}$$

23. NSB Credit Score

BIOGRAPHICAL NOTE

Brian Edward Boyle was born at Fort Monroe, Virginia, on March 24, 1948. The son of Col. (USA Ret.) and Mrs. Harry F. Boyle, he attended a number of schools, graduating from Pacelli High School, Columbus, Georgia, in May, 1965. He entered Amherst College in September, 1965 where he studied mathematics and economics. In September, 1968 he entered M.I.T. under a combined degree program, and in June, 1971, received a B.A. degree in mathematics from Amherst College, a S.B. degree in electrical engineering and a S.M. degree in operations research from M.I.T. In September, 1971 he received the Electrical Engineer degree.

Mr. Boyle held a National Science Foundation fellowship from 1970 to 1973, and served as a teaching assistant in the Electrical Engineering Department in 1972. During the summer of 1972, he worked as a Management Science Analyst for the National Shawmut Bank of Boston. Since 1972, he has been a consultant to a number of computer technology, financial, and consulting organizations.