MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS
STANDARD REFERENCE MATERIAL 1010a
(ANSI and ISO TEST CHART No. 2)

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Acquisitions and Bibliographic Services

Acquisitions et Services Bibliographiques

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

BANKRUPTCY PREDICTION-A NONPARAMETRIC APPROACH

by

Yong Jiang, B.Sc (HONOURS)

A thesis submitted to
the Faculty of Graduate Studies and Research
in partial fulfilment of
the requirements for the degree of
Master of Management Studies

School of Business
Carleton University
Ottawa, Ontario

June 15, 1993

©copyright
1993, Yong Jiang

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file / Votre référence

Our file / Notre référence

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-315-89882-8

Canada

Name _____

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four-digit code in the spaces provided.

--	--	--	--

U·M·I

SUBJECT TERM

SUBJECT CODE

Subject Categories

THE HUMANITIES AND SOCIAL SCIENCES

COMMUNICATIONS AND THE ARTS

Architecture 0729
Art History 0377
Cinema 0900
Dance 0378
Fine Arts 0357
Information Science 0723
Journalism 0391
Library Science 0399
Mass Communications 0708
Music 0413
Speech Communication 0459
Theater 0465

EDUCATION

General 0515
Administration 0514
Adult and Continuing 0516
Agricultural 0517
Art 0273
Bilingual and Multicultural 0282
Business 0688
Community College 0275
Curriculum and Instruction 0727
Early Childhood 0518
Elementary 0524
Finance 0277
Guidance and Counseling 0519
Health 0680
Higher 0745
History of 0520
Home Economics 0278
Industrial 0521
Language and Literature 0279
Mathematics 0280
Music 0522
Philosophy of 0998
Physical 0523

Psychology 0525
Reading 0535
Religious 0527
Sciences 0714
Secondary 0533
Social Sciences 0534
Sociology of 0340
Special 0529
Teacher Training 0530
Technology 0710
Tests and Measurements 0288
Vocational 0747

LANGUAGE, LITERATURE AND LINGUISTICS

Language
General 0679
Ancient 0289
Linguistics 0290
Modern 0291
Literature
General 0401
Classical 0294
Comparative 0295
Medieval 0297
Modern 0298
African 0316
American 0591
Asian 0305
Canadian (English) 0352
Canadian (French) 0355
English 0593
Germanic 0311
Latin American 0312
Middle Eastern 0315
Romance 0313
Slavic and East European 0314

PHILOSOPHY, RELIGION AND THEOLOGY

Philosophy 0422
Religion
General 0318
Biblical Studies 0321
Clergy 0319
History of 0320
Philosophy of 0322
Theology 0469

SOCIAL SCIENCES

American Studies 0323
Anthropology
Archaeology 0324
Cultural 0326
Physical 0327
Business Administration
General 0310
Accounting 0272
Banking 0770
Management 0454
Marketing 0338
Canadian Studies 0385
Economics
General 0501
Agricultural 0503
Commerce Business 0505
Finance 0508
History 0509
Labor 0510
Theory 0511
Folklore 0358
Geography 0366
Gerontology 0351
History
General 0578

Ancient 0579
Medieval 0581
Modern 0582
Black 0328
African 0331
Asia, Australia and Oceania 0332
Canadian 0334
European 0335
Latin American 0336
Middle Eastern 0333
United States 0337
History of Science 0585
Law 0398
Political Science
General 0615
International Law and Relations 0616
Public Administration 0617
Recreation 0814
Social Work 0452
Sociology
General 0626
Criminology and Penology 0627
Demography 0938
Ethnic and Racial Studies 0631
Individual and Family Studies 0628
Industrial and Labor Relations 0629
Public and Social Welfare 0630
Social Structure and Development 0700
Theory and Methods 0344
Transportation 0709
Urban and Regional Planning 0999
Women's Studies 0453

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES

Agriculture
General 0473
Agronomy 0285
Animal Culture and Nutrition 0475
Animal Pathology 0476
Food Science and Technology 0359
Forestry and Wildlife 0478
Plant Culture 0479
Plant Pathology 0480
Plant Physiology 0817
Range Management 0777
Wood Technology 0746

Biology

General 0306
Anatomy 0287
Biostatistics 0308
Botany 0309
Cell 0379
Ecology 0329
Entomology 0353
Genetics 0369
Limnology 0793
Microbiology 0410
Molecular 0307
Neuroscience 0317
Oceanography 0416
Physiology 0433
Radiation 0821
Veterinary Science 0778
Zoology 0472

EARTH SCIENCES

Biogeochemistry 0425
Geochemistry 0995

Geodesy 0370
Geology 0372
Geophysics 0373
Hydrology 0388
Mineralogy 0411
Paleobotany 0345
Paleoecology 0426
Paleontology 0418
Paleozoology 0985
Palynology 0427
Physical Geography 0368
Physical Oceanography 0415

HEALTH AND ENVIRONMENTAL SCIENCES

Environmental Sciences 0768
Health Sciences
General 0566
Audiology 0300
Chemotherapy 0992
Dentistry 0567
Education 0350
Hospital Management 0769
Human Development 0758
Immunology 0982
Medicine and Surgery 0564
Mental Health 0347
Nursing 0569
Nutrition 0570
Obstetrics and Gynecology 0380
Occupational Health and Therapy 0354
Ophthalmology 0381
Pathology 0571
Pharmacology 0419
Pharmacy 0572
Physical Therapy 0382
Public Health 0573
Radiology 0574
Recreation 0575

Speech Pathology 0460
Toxicology 0383
Home Economics 0386

PHYSICAL SCIENCES

Pure Sciences

Chemistry
General 0485
Agricultural 0749
Analytical 0486
Biochemistry 0487
Inorganic 0488
Nuclear 0738
Organic 0490
Pharmaceutical 0491
Physical 0494
Polymer 0495
Radiation 0754
Mathematics 0405
Physics
General 0605
Acoustics 0986
Astronomy and Astrophysics 0606
Atmospheric Science 0608
Atomic 0748
Electronics and Electricity 0607
Elementary Particles and High Energy 0798
Fluid and Plasma 0759
Molecular 0609
Nuclear 0610
Optics 0752
Radiation 0756
Solid State 0611
Statistics 0463

Applied Sciences

Applied Mechanics 0346
Computer Science 0984

Engineering
General 0537
Aerospace 0538
Agricultural 0539
Automotive 0540
Biomedical 0541
Chemical 0542
Civil 0543
Electronics and Electrical 0544
Heat and Thermodynamics 0348
Hydraulic 0545
Industrial 0546
Marine 0547
Materials Science 0794
Mechanical 0548
Metallurgy 0743
Mining 0551
Nuclear 0552
Packaging 0549
Petroleum 0765
Sanitary and Municipal 0554
System Science 0790
Geotechnical 0428
Operations Research 0796
Plastics Technology 0795
Textile Technology 0994

PSYCHOLOGY

General 0621
Behavioral 0384
Clinical 0622
Developmental 0620
Experimental 0623
Industrial 0624
Personality 0625
Physiological 0989
Psychobiology 0349
Psychometrics 0632
Social 0451

The undersigned recommend to the Faculty of Graduate Studies
and Research acceptance of the thesis

BANKRUPTCY PREDICTION - A NONPARAMETRIC APPROACH

submitted by Yong Jiang, B.Sc. (Honours)
in partial fulfillment of the requirements for
the degree of Master of Management Studies

Thesis Supervisor

Director, School of Business

Carleton University

July 27, 1993

ABSTRACT

This study incorporated nonparametric discriminant analysis into corporate bankruptcy prediction. It is the first study to date that has used nonparametric methods of discriminant analysis in the area of bankruptcy prediction. The reasons for such a study derive from the practical difficulties posed by the distributional assumptions required for a valid use of traditional discriminant analysis and logit analysis.

In this thesis, the differences arising from applications of parametric discriminant analysis, non-parametric discriminant analysis and logit analysis have been studied, and the efficiency of these three methods when applied to bankruptcy data has been evaluated in terms of error count estimates. The results indicate that nonparametric discriminant analysis(the Epanechnikov model) has the best predictive ability among the three methodologies. The research has also examined issues and problems related to the use of matched samples in empirical studies, and methods for properly accounting for matching have been explored.

TABLE OF CONTENTS

ABSTRACT	(i)
TABLE OF CONTENTS	(ii)
LIST OF TABLES	(v)
LIST OF FIGURES	(vii)
LIST OF APPENDICES	(viii)
NOTATION	(ix)
CHAPTER 1 INTRODUCTION	1
1.1 Importance of Bankruptcy Prediction	1
1.2 Bankruptcy Prediction Models	2
1.2.1 Discriminant Analysis	3
1.2.2 Logistic Regression	4
1.2.3 Non-Parametric Consideration	4
1.3 Organization of the Thesis	7
CHAPTER 2 BACKGROUND	8
2.1 Defining Corporate Problems	8
2.2 Bankruptcy Prediction	10
2.2.1 Traditional Ratio Analysis	10
2.2.2 Multiple Discriminant Analysis	11
2.2.3 Logistic Regression Analysis	14
2.2.4 The Functional Equivalence of Discriminant Analysis and Logistic Regression Analysis	17
2.2.5 Sampling Consideration	18
CHAPTER 3 LITERATURE REVIEW	21

CHAPTER 4 RESEARCH QUESTIONS AND HYPOTHESES	30
4.1 Research Questions	30
4.2 Hypotheses	30
CHAPTER 5 METHODOLOGY	32
5.1 Discriminant Analysis	32
5.1.1 Bayes Theorem	33
5.1.2 Parametric Methods	34
5.1.3 Density Estimation and Nonparametric Methods	35
5.2 Logistic Regression Analysis and Logit Analysis	37
5.2.1 The Naive Estimator	38
5.2.2 Definition of the Univariate Kernel Density Estimation	39
5.2.3 Definition of the Multivariate Kernel Density Estimation	40
5.2.4 Choosing the Kernel and the Smoothing Parameters	41
5.2.5 Measures of Discrepancy: Mean Integrated Square Error	42
5.2.6 Choice of Window Width for a Standard Distribution	44
5.3 Classification of Error-Rate Estimates	45
5.4 Parameter Estimation for a Matched Analysis	48
CHAPTER 6 DATA COLLECTION	51
6.1 The Sample of Bankrupt Firm	51
6.2 The Sample of Non-Bankrupt Firm	53
6.2.1 The Matching Debate	53
6.2.2 Selection of Matching Firms	55
6.3 Financial Ratios Used	56
CHAPTER 7 DATA ANALYSIS AND RESEARCH RESULTS	59

7.1	Test of Normality and the Homogeneity of Covariance	63
7.1.1	Test of Normality	63
7.1.2	Test of Homogeneity of Covariance	73
7.2	Normal Discriminant Analysis	74
7.2.1	Descriptive Examination of Group Differences	75
7.2.2	Empirical Results of Quadratic Discriminant Analysis	80
7.2.3	Empirical Results for Linear Discriminant Function	83
7.3	Results of Epanechnikov Kernel Analysis	87
7.4	Results From Logit Analysis	91
7.5	The Predictive Summary for the Total Five Year Period	93
7.6	Comparison With Results From Altman's Studies	97
CHAPTER 8 CONCLUSION		102
8.1	Summary of the Study	102
8.1.1	Comparison Among Models Associated with Three Methodologies	103
8.1.2	Ranking Predictive Ability of All Models	104
8.1.3	Issues Related to Matched Design and Analysis	105
8.2	Value of the Research	105
8.3	Limitations of the Study	106
8.4	Directions for Future Study	107
REFERENCES		109

LIST OF TABLES

Table 5.2.1	The constant $A(K)$, as defined in equation 5.2.14, for various kernels K . Numerical values are given correct to 2 decimal places	45
Table 7.1	Design for Data Analysis	62
Table 7.2.1.1	Variable Means and Test Significance (Data Set DS1)	75
Table 7.2.1.2	Mann-Whitney Test Results of All Five Variables in Data Set DS1 . .	77
Table 7.2.1.3	Kolmogorov-Smirnov Test Results of Variables X_1 to X_5 in DS1 . . .	78
Table 7.2.1.4	The Median Test Results of Variables X_1 to X_5 in DS1	79
Table 7.2.1.5	Multivariate Statistics and Exact F Statistics (DS1)	80
Table 7.2.2.1	Resubstitution Summary Using Quadratic Discriminant Function (DS1)	81
Table 7.2.2.2	Error Count Estimates Using Resubstitution and Quadratic Discriminant Function (Data Set DS1)	81
Table 7.2.2.3	Posterior Probability Error Rate Estimate Using Resubstitution and Quadratic Discriminant Function (Data Set DS1)	82
Table 7.2.2.4	Crossvalidation Summary Using Quadratic Discriminant Function (DS1)	82
Table 7.2.2.5	Error Count Estimates Using Cross-validation and Quadratic Discriminant Function (Data Set DS1)	83
Table 7.2.2.6	Posterior Probability Error Rate Estimate Using Crossvalidation and Quadratic Discriminant Function (Data Set DS1)	83
Table 7.2.3.1	Resubstitution Summary Using Linear Discriminant Function (DS1) .	84
Table 7.2.3.2	Error Count Estimates Using Resubstitution and Linear Discriminant Function (Data Set DS1)	84

Table 7.2.3.3	Posterior Probability Error Rate Estimate Using Resubstitution and Linear Discriminant Function (Data Set DS1)	85
Table 7.2.3.4	Crossvalidation Summary Using Linear Discriminant Function (DS1) .	85
Table 7.2.3.5	Error Count Estimates Using Cross-validation and Linear Discriminant Function (Data Set DS1)	86
Table 7.2.3.6	Posterior Probability Error Rate Estimate Using Crossvalidation and Linear Discriminant Function (Data Set DS1)	86
Table 7.3.1	Resubstitution Results Using Epanechnikov Kernel Density and Unequal Smoothing Parameters for Five Years Predictive Accuracy	90
Table 7.3.2	Resubstitution Results Using Epanechnikov Kernel Density and Equal Smoothing Parameters for Five Years Predictive Accuracy	90
Table 7.4.1	Analysis of Maximum Likelihood Estimates for First Year Data Set . .	92
Table 7.4.2	Classification Table Using Logit Analysis for Five Years Data Versions	93
Table 7.5.1	Summary for Five Years Predictive Accuracy Using Models Derived From Three Different Methodologies	95
Table 7.6.1	A Comparison With Altman's Finding	98
Table 7.6.2	Ranking the Predictive Ability of Models	100

LIST OF FIGURES

Figure 1.2.1	Diagram for Evaluation of Data	7
Figure 7.1.1	Plot of Chi-Square With Mahalanobis Distance on Bankrupt Firms Using First Year Data (Data Set DS1)	67
Figure 7.1.2	Plot of Chi-Square With Mahalanobis Distance on Non-Bankrupt Firms Using First Year Data	67
Figure 7.1.3	Test of Univariate Normality of X_1 in Bankrupt Group Using DS1	68
Figure 7.1.4	Test of Univariate Normality of X_1 in Non-Bankrupt Group Using DS1	68
Figure 7.1.5	Test of Univariate Normality of X_2 in Bankrupt Group Using DS1	69
Figure 7.1.6	Test of Univariate Normality of X_2 in Non-Bankrupt Group Using DS1	69
Figure 7.1.7	Test of Univariate Normality of X_3 in Bankrupt Group Using DS1	70
Figure 7.1.8	Test of Univariate Normality of X_3 in Non-Bankrupt Group Using DS1	70
Figure 7.1.9	Test of Univariate Normality of X_4 in Bankrupt Group Using DS1	71
Figure 7.1.10	Test of Univariate Normality of X_4 in Non-Bankrupt Group Using DS1	71
Figure 7.1.11	Test of Univariate Normality of X_5 in Bankrupt Group Using DS1	72
Figure 7.1.12	Test of Univariate Normality of X_5 in Non-Bankrupt Group Using DS1	72

LIST OF APPENDICES

Appendix A	The List of Bankrupt and Non-Bankrupt Firms	115
Appendix B	B.1 Plot of Chi-Square With Mahalanobis Distance on Bankrupt Group Using Second Year Data DS2	118
	B.2 Plot of Chi-Square With Mahalanobis Distance on Non-Bankrupt Group Using Second Year Data DS2	118
	B.3 Plot of Chi-Square With Mahalanobis Distance on Bankrupt Group Using Third Year Data DS3	119
	B.4 Plot of Chi-Square With Mahalanobis Distance on Non-Bankrupt Group Using Third Year Data DS3	119
	B.5 Plot of Chi-Square With Mahalanobis Distance on Bankrupt Group Using Fourth Year Data DS4	120
	B.6 Plot of Chi-Square With Mahalanobis Distance on Non-Bankrupt Group Using Fourth Year Data DS4	120
	B.7 Plot of Chi-Square With Mahalanobis Distance on Bankrupt Group Using Fifth Year Data DS5	121
	B.8 Plot of Chi-Square With Mahalanobis Distance on Non-Bankrupt Group Using Fifth Year Data DS5	121
Appendix C	C.1 Test of Univariate Normality of X_1 in Bankrupt Group Using DS2	122
	C.2 Test of Univariate Normality of X_1 in Non-Bankrupt Group Using DS2	122
	C.3 Test of Univariate Normality of X_2 in Bankrupt Group Using DS2	123
	C.4 Test of Univariate Normality of X_2 in Non-Bankrupt Group Using DS2	123

C.5	Test of Univariate Normality of X_3 in Bankrupt Group Using DS2	124
C.6	Test of Univariate Normality of X_3 in Non-Bankrupt Group Using DS2	124
C.7	Test of Univariate Normality of X_4 in Bankrupt Group Using DS2	125
C.8	Test of Univariate Normality of X_4 in Non-Bankrupt Group Using DS2	125
C.9	Test of Univariate Normality of X_5 in Bankrupt Group Using DS2	126
C.10	Test of Univariate Normality of X_5 in Non-Bankrupt Group Using DS2	126
Appendix D		
D.1	Test of Univariate Normality of X_1 in Bankrupt Group Using DS3	127
D.2	Test of Univariate Normality of X_1 in Non-Bankrupt Group Using DS3	127
D.3	Test of Univariate Normality of X_2 in Bankrupt Group Using DS3	128
D.4	Test of Univariate Normality of X_2 in Non-Bankrupt Group Using DS3	128
D.5	Test of Univariate Normality of X_3 in Bankrupt Group Using DS3	129
D.6	Test of Univariate Normality of X_3 in Non-Bankrupt Group Using DS3	129
D.7	Test of Univariate Normality of X_4 in Bankrupt Group Using DS3	130
D.8	Test of Univariate Normality of X_4 in Non-Bankrupt Group Using DS3	130
D.9	Test of Univariate Normality of X_5 in Bankrupt Group Using DS3	131
D.10	Test of Univariate Normality of X_5 in Non-Bankrupt Group Using DS3	131

Appendix E	E.1	Test of Univariate Normality of X_1 in Bankrupt Group Using DS4	132
	E.2	Test of Univariate Normality of X_1 in Non-Bankrupt Group Using DS4	132
	E.3	Test of Univariate Normality of X_2 in Bankrupt Group Using DS4	133
	E.4	Test of Univariate Normality of X_2 in Non-Bankrupt Group Using DS4	133
	E.5	Test of Univariate Normality of X_3 in Bankrupt Group Using DS4	134
	E.6	Test of Univariate Normality of X_3 in Non-Bankrupt Group Using DS4	134
	E.7	Test of Univariate Normality of X_4 in Bankrupt Group Using DS4	135
	E.8	Test of Univariate Normality of X_4 in Non-Bankrupt Group Using DS4	135
	E.9	Test of Univariate Normality of X_5 in Bankrupt Group Using DS4	136
	E.10	Test of Univariate Normality of X_5 in Non-Bankrupt Group Using DS4	136
Appendix F	F.1	Test of Univariate Normality of X_1 in Bankrupt Group Using DS5	137
	F.2	Test of Univariate Normality of X_1 in Non-Bankrupt Group Using DS5	137
	F.3	Test of Univariate Normality of X_2 in Bankrupt Group Using DS5	138
	F.4	Test of Univariate Normality of X_2 in Non-Bankrupt Group Using DS5	138
	F.5	Test of Univariate Normality of X_3 in Bankrupt Group Using DS5	139
	F.6	Test of Univariate Normality of X_3 in Non-Bankrupt Group Using DS5	139

	F.7	Test of Univariate Normality of X_4 in Bankrupt Group Using DS5	140
	F.8	Test of Univariate Normality of X_4 in Non-Bankrupt Group Using DS5	140
	F.9	Test of Univariate Normality of X_5 in Bankrupt Group Using DS5	141
	F.10	Test of Univariate Normality of X_5 in Non-Bankrupt Group Using DS5	141
Appendix G	G.1	Variable Means and Test of Significance Using Second Year Data DS2	142
	G.2	The Mann-Whitney Test Results of All Five Variables in Data Set DS2	142
	G.3	Kolmogorov-Smirnov Test Results of Variables X_1 to X_5 in Data Set DS2	143
	G.4	The Median Test Results of Variables X_1 to X_5 in Data Set DS2	143
	G.5	Multivariate Statistics and Exact F Statistics (Data Set DS2)	143
Appendix H	H.1	Variable Means and Test of Significance Using Second Year Data DS3	144
	H.2	The Mann-Whitney Test Results of All Five Variables in Data Set DS3	144
	H.3	Kolmogorov-Smirnov Test Results of Variables X_1 to X_5 in Data Set DS3	145
	H.4	The Median Test Results of Variables X_1 to X_5 in Data Set DS3	145
	H.5	Multivariate Statistics and Exact F Statistics (Data Set DS3)	145
Appendix I	I.1	Variable Means and Test of Significance Using Second Year Data DS4	146
	I.2	The Mann-Whitney Test Results of All Five Variables in Data Set DS4	146

	I.3	Kolmogorov-Smirnov Test Results of Variables X_1 to X_5 in Data Set DS4	147
	I.4	The Median Test Results of Variables X_1 to X_5 in Data Set DS4	147
	I.5	Multivariate Statistics and Exact F Statistics (Data Set DS4)	147
Appendix J	J.1	Variable Means and Test of Significance Using Second Year Data DS5	148
	J.2	The Mann-Whitney Test Results of All Five Variables in Data Set DS5	148
	J.3	Kolmogorov-Smirnov Test Results of Variables X_1 to X_5 in Data Set DS5	149
	J.4	The Median Test Results of Variables X_1 to X_5 in Data Set DS5.	149
	J.5	Multivariate Statistics and Exact F Statistics (Data Set DS5)	149
Appendix K	K.1	Analysis of Maximum Likelihood Estimates And Classification Table For Data Set DS2	150
	K.2	Classification Table (DS2)	150
	K.3	Analysis of Maximum Likelihood Estimates And Classification Table For Data Set DS3	151
	K.4	Classification Table (DS3)	151
	K.5	Analysis of Maximum Likelihood Estimates And Classification Table For Data Set DS4	152
	K.6	Classification Table (DS4)	152
	K.7	Analysis of Maximum Likelihood Estimates And Classification Table For Data Set DS5	153
	K.8	Classification Table (DS5)	153

NOTATION

C_0	constant
C_d	volume of the unit d -dimensional sphere
$\mathbf{m}_t$	the p -dimensional vector containing variable means in group t
S_t	the covariance matrix within group t
$\mathbf{x}$	a p -dimensional vector containing the quantitative variables of an observation
S	the pooled covariance matrix
t	a subscript to distinguish the groups
n_t	the number of training set observations in group t
$ S_t $	the determinant of S_t
q_t	the prior probability of membership in group t
$\text{diag}(S)$	the diagonal matrix of the pooled covariance matrix
$\text{diag}(S_t)$	the diagonal matrix of the covariance matrix within group t
I	the identity matrix.
$\mathbf{x}$	a vector containing the variables of an observation
f_t	the probability density function in group t
$p(t \mathbf{x})$	the posterior probability of $\mathbf{x}$ for group t
e_t	the classification error rate for group t
$\hat{e}$	total error rate estimate
S	the set of observations in the (training) sample
n	the number of observations of S
R_t	the set of observations such that the posterior probability belonging to group t is the largest
R_{ut}	the set of observations from group u such that the posterior probability belonging to group t is the largest.
h	the fixed radius of the kernel
$K(\mathbf{x})$	kernel function
$K_{\mathbf{x}}(\mathbf{x})$	multivariate Epanechnikov kernel function
$d_t^2(\mathbf{x})$	square of distance from $\mathbf{x}$ to group t
$\pi(\mathbf{x})$	the conditional mean of Y given $\mathbf{x}$
$E(Y \mathbf{x})$	the conditional mean of Y given $\mathbf{x}$
$g(\mathbf{x})$	the logit transformation of $\pi(\mathbf{x})$
$P(Y=i \mathbf{x})$	the conditional probability that Y is equal to i
Z	discriminant function
$Z(\mathbf{x})$	quadratic discrimination function

BANKRUPTCY PREDICTION-A NONPARAMETRIC APPROACH

CHAPTER 1

INTRODUCTION

1.1 Importance of Bankruptcy Prediction

Citing data from the Economic Report (1986), reported that a U.S. business failed on average every forty-nine minutes during 1970 while the corresponding average interval was sixteen minutes in 1983. According to the latest Economic Report, recent comparable data from Consumer and Corporate Affairs reveal that in March of this year, a business failed every 35 minutes, on average, in Canada. This increase in both the size and number of industrial and banking failures has reawakened interest in the issue of bankruptcy prediction. The failures will cause managers and employees to lose their jobs; it will also cause governments to lose part of tax base and result in creditors and investors losing their money. Therefore, knowing whether the firm is likely to fail is thus of paramount importance to these parties. Each party has impact on a firm's likelihood of failure, directly or indirectly. Managers and employees, through their participation in a firm's operation, can play a direct role on a firm's likelihood of failure. While creditors and investors do not generally have the ability to directly participate in the operations of the

firm, they can, through their contribution to a firm's finance, have an indirect impact on a firm's failure.

Understanding the determinants of corporate bankruptcy is important both in practice and in theory. There are numerous practical examples of this importance. Regulatory agencies judge whether or not a regulated company is in danger of failing based on the bankruptcy prediction models, it can be exempted from normal antitrust prohibitions and allowed the going bankrupt firm to merge with others under the Failing Company Doctrine (Bankruptcy Act for Canadian firms). Auditors use bankruptcy prediction models to decide whether or not a corporation is to be taken into consideration of bankrupt. "Understanding corporate bankruptcy is important in theory as well. For example, the probability of bankruptcy is an important parameter in valuation formulas for debt and equity" (Scott, 1976). When investors are risk averse, market valuation formulas involve parameters that capture both the probability of bankruptcy and its effect on the utility of the average investor.

1.2 Bankruptcy Prediction Models

This section will discuss some technical preliminaries which underlie the main research issues of this thesis. Most bankruptcy-prediction models are derived using data from a sample of firms that eventually failed together with data from a sample of contemporaneous firms that did not fail.

In general, development of bankruptcy models follows the following steps. First, a number of plausible and traditional financial ratios are calculated from financial statements that were published before failure. Next, the researcher searches for a formula, based either on a single ratio or combination of ratios, that best discriminates between firms in the sample of failure and firms in the sample of that remained solvent. A careful researcher also tests the resulting formula both on the original "training" sample and on a test sample of failure and non-failure that are not used to derive the formula. Finally, in a few cases (for example, Altman, 1968), the model may be revalidated over time based on observations after the model was built.

1.2.1 Discriminant Analysis

Two of the most widely used statistical procedures in empirical studies of corporate bankruptcies are discriminant analysis (DA) and logit analysis. Discriminant analysis is really appropriate and justifiable only under (a) multivariate normality of the explanatory variables (a difficult assumption to satisfy in practice), (b) complete equality of all of the underlying covariance matrices (Press and Wilson, 1978). When the normality assumption is not realized, it is natural to consider data transformation that might induce normality. Such transformations will not typically induce equality of covariance matrices, however. For a linear model, the main goals for transforming data are to achieve the fulfilment of the following assumptions (if the original data fails to fulfil them):

- (1) additivity, i.e. the main effects combine linearly to explain the observations,
- (2) homogeneity, i.e., equal variance matrices among the two groups (bankrupt and

non-bankrupt),

- (3) normality, i.e. the observations are assumed to have a normal distribution within each group.

Draper and Hunter (1969) list methods of finding transformations which may simultaneously satisfy (more or less well) these three assumptions. This can usually be done as long as one does not press too hard for any one of them. But it may happen that through a transformation which leads to one of the desiderata some other desirable characteristic gets lost. In some cases, satisfactory transformations are simply not available. On the other hand, even when transformations exist, various problems arise in presenting results in the original variables when the analysis has been performed on the transformed variables (Grimm, 1980). In practice, therefore, alternative approaches to discriminant analysis may turn out to be necessary.

1.2.2 Logistic Regression

When the assumptions of discriminant analysis are not met, researchers usually opt instead for logistic regression analysis, which relates the logarithm of the ratio of the odds of corporate bankruptcy to a linear function of predictor variables. Efron (1975) states that logit analysis is appropriate under a general exponential family assumption on the conditional distribution of the vector of predictor variables, $\mathbf{x}$. Therefore, if normality is rejected, logit analysis is a natural parametric alternative.

1.2.3 Non-parametric Considerations

Logit analysis is not the only alternative, however, and in situations where the

distributional forms of the predictor variables may not be known, e.g., in a study involving a new population of companies on a new set of predictor variables. it may not be the best alternative. There exist non-parametric methods of estimating the distribution functions of the sample data, and these estimates can then be incorporated into the classification functions to yield non-parametric estimates of the probability of bankruptcy. This approach has not been followed in all of the studies on corporate bankruptcy undertaken in the last few decades. As Zavgren (1985) has noted, the conclusions of most bankruptcy studies are difficult to assess because "they play loose with the assumptions of discriminant analysis". This paper, therefore, will focus on a comparison of non-parametric discriminant analysis with the classical approaches of discriminant analysis and logistic regression for a set of real data. Figure 1.2.1, taken from Grimm (1980), provides a schematic of the approach that should typically be followed in assessing a data analysis problem.

*abandon hope all ye enter here

Figure 1.2.1 DIAGRAM FOR EVALUATION OF DATA (From Grimm,1980)

1.3 Organization of the Thesis

Before the literature on bankruptcy prediction can be described in detail, it is necessary to present an overview of the technical aspects of bankruptcy definition and bankruptcy prediction, including some details on the primary statistical methodologies used to date. This is done in Chapter 2. An historical review of the literature is then presented in Chapter 3, which points to the need for the present investigation. The specific research questions are described in Chapter 4, followed in Chapter 5 by a detailed exposition of the theory of discrimination and classification, leading to a discussion of the Epanechnikov kernel density estimation method which is the principal technical focus of the thesis. The methods for defining and collecting the sample data are described in Chapter 6, followed in Chapter 7 by the results of the data analysis which compares the non-parametric approach to the two classical approaches of discriminant analysis and logistic regression. Conclusions are presented in chapter 8.

CHAPTER 2

BACKGROUND

2.1 Defining Corporate Problems

The unsuccessful business enterprise has been defined in numerous ways in order to depict the formal process confronting the firm and/or to economically categorize the problems involved. Bathory (1985) developed a tripartite analysis of business failure. He classified corporate insolvency into three groupings--each with its own specific characteristics:

- (1). acute insolvency: e.g., insufficient cash in the present or short term to meet financial obligations as they fall due;
- (2). chronic insolvency: e.g., inability to meet financial obligations as they fall due over two or more accounting periods; and
- (3). terminal insolvency: e.g., inability to fund more or less permanent changes in the balance.

This study, however, will adopt definitions for the three generic terms which are commonly found in the literature, namely, failure, insolvency and bankruptcy, as given in the book by Altman (1971). Altman's definitions are follows:

Failure: The situation where the realized rate of return on invested capital,

with due allowances for risk, is significantly and continually lower than prevailing rates on similar investments. Somewhat different economic criteria have also been cited, including insufficient revenues to cover costs and situations where the average return on investment is below the firm's costs and situation where the average return on investment is below the firm's cost of capital.

Legal failure: The situation when a company can no longer meet the legally enforceable demands of its creditors. It should be noted that the term "legal" does not necessarily imply formal court involvement.

Insolvency or Technical Insolvency: The situation when a firm cannot meet its current obligations, signifying a lack of liquidity.

Insolvency in an equity sense: The situation when firm's total liabilities exceed a fair valuation of its total assets, i.e., the real net worth of the firm is negative.

Bankruptcy: This term can be used to describe two specific situations: (1) When the real net worth of the firm is negative; (2) When a firm has made a formal declaration of bankruptcy in a court, accompanied by a petition to either liquidate its assets or attempt a recovery program. All the so-called "bankrupt" selected for use in this study correspond to the second of these definitions.

2.2 Bankruptcy Prediction

2.2.1 Traditional Ratio Analysis

The detection of company operating and financial difficulties is a subject which has been particularly susceptible to financial ratio analysis. "Studies on the use of financial ratios in business failure can be traced to 1930s" (Altman, 1968). A few earlier studies had concluded that failing firms exhibit significantly different financial ratio measurements than do continuing firms. But the theoretical starting point for looking at the analysis of failure is with the work of Beaver (1966). Beaver defined "failure" as legal failure, i.e., the firm's inability to pay its financial obligations as they mature. Thus his study of 79 failed firms included 59 bankruptcies, 16 non-payments of preferred dividends, 3 bond defaults and 1 overdrawn bank account for the period 1954 to 1964. Beaver matched the 79 failed firms with 79 non-failed firms by industry and assets size. By employing a univariate model to differentiate survivors from failures, Beaver arrived at measures which achieved a moderate level of predictive accuracy. The ratios Beaver found useful included cash flow to total debt, net income to total assets, total debt to total assets, working capital to total assets and the firm's current ratio. However, ratio analysis presented in this univariate fashion is susceptible to faulty interpretation and is potentially confusing. For instance, a firm with a poor record of profitability and/or solvency may be regarded as a potential bankrupt. However, if it has above average liquidity, the situation may not be considered serious. Therefore, instead of only testing each ratio independently using a test

of means, we need to build a model, based on financial ratios, that accounts for their combined effects on the firm. The use of discriminant analysis to simultaneously estimate a failure probability using several financial ratios is one way of addressing this issue.

2.2.2 Multiple Discriminant Analysis

Multiple discriminant analysis (MDA) is used for two purposes: (1) Identifying and describing major differences among the groups in question. A preliminary analysis often conducted in such situation consists of a one-way or higher multivariate analysis of variance (MANOVA), along with finding the associated p values and variance-explained values (Huberty & Smith, 1982). (2) classifying units into pre-determined groups on the basis of their scores on a set of predictor variables. In this study, primary use of multiple discriminant analysis is for classifying firms into either a bankrupt or non-bankrupt group. In general, classification involves placing the unit (firm) in the group that it most closely resembles. Unlike regression analysis, the dependent variable in MDA is always the qualitative grouping variable, e.g., sick or health, bankrupt or non-bankrupt. The first step in MDA, therefore, is to establish explicit group classifications. In general, the number of original groups can be two or more, though this study will involve only the binary bankrupt/non-bankrupt classification.

Once the groups have been established, data are collected for the units in the groups. MDA then attempts to derive a linear combination of these characteristics which "best" discriminates between the groups. One advantage of the MDA technique is that it can consider an entire profile of characteristics common to the relevant units or firms, as

well as the interaction of these properties. A univariate study, on the other hand, can only consider the measurements used for group assignments one at a time. Another advantage of MDA is the reduction of the dimensionality of the discrimination space from the number of different dependent variables, to p , the minimum of p and $G-1$, where p is the number of dependent variables and G equals the number of pre-determined groups. The standard approach to multiple discriminant analysis for classification problems is usually formulated by assuming that the two populations, e.g., bankrupt and non-bankrupt firms, are multivariate normal with equal covariance matrices, and that the cost of misclassification is equal. This yields a linear classification rule. As noted above, this research considers only two groups, consisting of bankrupt firms and non-bankrupt firms. Therefore, the analysis is transformed into its simplest form: one dimension. The linear discriminant function:

$$Z = \ln[f_1(\mathbf{x})/f_2(\mathbf{x})] > \ln[(1-q_1)/q_1]$$

$$= \mathbf{x}'\Sigma^{-1}(\mathbf{m}_1 - \mathbf{m}_2) - 1/2(\mathbf{m}_1 + \mathbf{m}_2)'\Sigma^{-1}(\mathbf{m}_1 - \mathbf{m}_2) > \ln[(1-q_1)/q_1]$$

transforms the p individual variable values to a single classification score Z .

where

$\mathbf{x}$ = independent variable

$f_t(\mathbf{x})$ = the probability density function in group t

q_t = the prior probability of membership in group t

$\mathbf{m}_t$ = the p -dimensional vector containing variable means in group t

Σ = the covariance matrix within group t

The classification rule is: Assign $\mathbf{x}$ to bankrupt group if $f_1(\mathbf{x})/f_2(\mathbf{x}) > (1-q_1)/q_1$ and to non-bankrupt group otherwise.

If normality appears to hold but the assumption of equal covariance matrices is seriously violated, then a quadratic rule is appropriate. The optimal rule is: Assign to group 1 if:

$$Z(\mathbf{x}) = \ln[f_1(\mathbf{x})/f_2(\mathbf{x})] > \ln[(1-q_1)/q_1] \quad (2.2.1)$$

$$= 1/2 \ln(\Sigma_2 / \Sigma_1) - 1/2(\mathbf{x}-\mathbf{m}_1)' \Sigma_1^{-1}(\mathbf{x}-\mathbf{m}_1) + 1/2(\mathbf{x}-\mathbf{m}_2)' \Sigma_2^{-1}(\mathbf{x}-\mathbf{m}_2)$$

$$= C_0 - 1/2[\mathbf{x}'(\Sigma_1^{-1} - \Sigma_2^{-1})\mathbf{x} - 2\mathbf{x}'(\Sigma_1^{-1}\mathbf{m}_1 - \Sigma_2^{-1}\mathbf{m}_2)]$$

$$C_0 = \text{constant}$$

Since $\Sigma_1^{-1} - \Sigma_2^{-1}$ does not vanish, one has a quadratic discriminant function, i.e., the function depends on square and cross product terms x_i^2 , x_{ij} , not only on the x_i 's. When group 1 and group 2 are multivariate normal with means $\mathbf{m}_1$ and $\mathbf{m}_2$ and common covariance matrix, we have the important special case, i.e. linear discriminant function. This method was suggested by Welch (1939) as a way to minimize the total probability of misclassification. Other approaches to discriminant function includes Fisher's approach, Bayes theorem approach, unequal costs of misclassification and minimax rule. The details of some of these methods will be discussed in Chapter 3 and section 5.1. As will be shown later, the quadratic rule does not necessarily performs better than the linear rule even when $\Sigma_1 \neq \Sigma_2$. This is most noticeable under non-normality. Huberty (1984) states, "The stability of results yielded by a linear rule is greater than results yielded by a quadratic rule when small samples are used and when the normality condition is not met."

Nevertheless, it is a good practice to check the assumption of homogeneity of covariance matrices for which the Box test can be used .

A classification criterion is ideally evaluated by its performance in the classification of future observations. To test this capability, therefore, an analysis is needed in which the data to be classified are not used in constructing the classification function. One way to test this capabilities is via the *jackknife* procedure of Lachenbruch(1967). Here, each of the n units in the combined sample (bankrupt and non-bankrupt in this study) is classified based on a classification statistic derived from the remaining $(n-1)$ units. Another way to do this is to randomly split the sample and cross validate, though this requires a large sample. When the splitting technique is used, the classification function is computed on one sample called "train sample", and then the error rates are estimated by applying the classification function to the remainder of the split sample called the "test sample". This provided a good check on the external validity of the classification function. More detailed discussion on evaluation of classification criterion can be found in section 5.3 .

The cost of misclassifying must be considered in judging the worth of a classification rule. In this study, misclassifying a bankrupt corporation into non-bankrupt (A) is supposed to be considered more "costly" than misclassifying a non-bankrupt firm into a bankruptcy (B). Therefore, of two procedures having error characters as A and B, with the same overall error rate estimate, B would be considered better since it results in less "costly" misclassifications. Such considerations are outside the scope of this study, and so the investigation of the classification effectiveness of the different methods will

proceed under the assumption that error costs are the same.

2.2.3 Logistic Regression

Logistic regression can be used whenever an individual is to be classified into one of two populations, and can be thought of as the extension of multiple linear regression to a case where the dependent variable is discrete, taking on two possible values. As described earlier, logistic regression represents an alternative method of classification when the multivariate normal model is not justified. The term logistic refers to the fact that the model assumes that the probability of a firm falling into the bankrupt group follows a logistic distribution. Many distribution functions have been proposed for use in the analysis of a dichotomous dependant variable and there are two primary reasons for choosing the logistic distribution. These are: (1) from a mathematical point of view, it is an extremely flexible and easily-used function, and (2) it lends itself to a meaningful interpretation (Hosmer & Lemeshow, 1989). If Y , the dependent or grouping variable, is coded as zero or one then a logistic regression model provides an assumed form for the conditional probability that Y is equal to 1, given $\mathbf{x}$. This will be denoted as $P(Y=1 | \mathbf{x})$ as follows:

$$P(Y=1 | \mathbf{x}) = \pi(\mathbf{x}) = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}},$$

Where $x_1, x_2, \dots, x_p$ are the values taken by a particular units (firm) on each of the p predictor variables, and $\mathbf{x}$ denotes the column vector of these values.

A transformation of $\pi(\mathbf{x})$ is the logit transformation. The transformation is defined, in terms of $\pi(\mathbf{x})$, as follows:

$$\begin{aligned} g(\mathbf{x}) &= \ln[\pi(\mathbf{x})/(1-\pi(\mathbf{x}))] \\ &= \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p, \end{aligned}$$

in which case $\pi(\mathbf{x})$ can be written as

$$\pi(\mathbf{x}) = \frac{e^{g(\mathbf{x})}}{1 + e^{g(\mathbf{x})}},$$

which is the familiar logistic distribution.

It follows that the quantity $1-\pi(\mathbf{x})$ gives the conditional probability that Y is equal to zero given $\mathbf{x}$, $P(Y=0 | \mathbf{x})$. The $g(\mathbf{x})$ then is the logarithm of the odds $P(Y=1 | \mathbf{x})/P(Y=0 | \mathbf{x})$ and from the above formulation, it can be seen that the fundamental assumption in logistic regression analysis is that $\ln(\text{odds})$ is linearly related to the independent variables. Many techniques associated with linear regression analysis therefore can be used with logit analysis, e.g. regression diagnostics (Pregibon, 1981). One of the major advantages of logistic regression is that the $\mathbf{x}$ variables may be discrete or continuous. Design variables or dummy variables can be used in the case of independent variables that are discrete, for example nominal scaled variables such as race, sex, treatment group, etc. This is not a major issue in this study, since none of the independent predictor variables used happen to be discrete.

The maximum likelihood method is typically used to estimate the parameters, $\beta = (\beta_0, \beta_1, \dots, \beta_p)'$, in the logistic regression model, i.e., the maximum likelihood estimators of

these parameters are chosen to be those values which maximize the likelihood function. In this sense, the resulting estimators are those which agree most closely with the observed data.

Efron (1975) states that logit analysis is appropriate when the distributions of the independent variables $\mathbf{x}$ in each of the groups belong to the general exponential family, which indicates a variety of distributions besides the normal. Although we do not have to assume a multivariate normal distribution to use logistic regression, we still need some assumptions, namely: simple random sampling, correct knowledge of group membership (for the "training sample"), and data that are free of miscellaneous errors and outlier.

2.2.4 The Functional Equivalence of Discriminant Analysis and Logistic Regression Analysis

To fully understand the difference between discriminant analysis and logistic regression, one needs to consider the random/non-random status of the grouping variable (Y) and the vector of prediction variables ($\mathbf{x}$). Let $Y = 0, 1$ denote whether an observation belongs to group 0 (bankrupt) or group 1 (non-bankrupt). In standard discriminant analysis, the two states $Y=0$ and $Y=1$ are fixed in advance, and a sample of observation $\mathbf{x}_1$ is taken from group 0, and a corresponding sample $\mathbf{x}_2$ is taken from group 1. Thus the Y variable is fixed, and the $\mathbf{x}$ variables are random. Hence the focus is on the distributions of the $\mathbf{x}$ and on differences in these distributions between the two groups. In some situations, one single simple random sample is taken from the overall population of companies (bankrupt and non-bankrupt), the sample being sorted after the fact into two

groups. In this case, discriminant analysis proceeds as above, since it can be viewed as a conditional analysis of differences in the $\mathbf{x}$, given Y , i.e. with Y fixed. In logistic regression, the standard model is based on one single sample from a combined population. Such a sample, of course, yields a random Y and random $\mathbf{x}$. In contrast to discriminant analysis, the standard logistic regression model proposes a particular distribution for the random variable Y , conditional on $\mathbf{x}$, i.e., a random Y is analyzed as if the $\mathbf{x}$ were originally fixed.

It is natural, therefore, to ask how two such analyses (standard discriminant analysis, with Y fixed and $\mathbf{x}$ random, and logistic regression, with Y random and $\mathbf{x}$ fixed), can be considered to be alternatives. The answer derives from the fact that once classical discriminant analysis has provided estimates of the distributions of $\mathbf{x}$ in both groups (by estimating the parameters of the normal distribution), Bayes' theorem is then involved to convert these probabilities into a statement of the probability that a future observation, with a given $\mathbf{x}$ vector, will fall into group 0 or group 1. Logistic regression starts out with an assumed form for the latter, and so involves making assumptions later in the modelling chain. Instead of involving Bayes' theorem to calculate the probability of distribution of Y conditional on $\mathbf{x}$, as in discriminant analysis, logistic regression directly assumes a form for $P(Y|\mathbf{x})$, and in so doing, removes the need for restrictive assumptions on the distribution of the $\mathbf{x}$.

2.2.5 Sampling Considerations

Irrespective of the discrimination technique used, estimation of the quantities

needed to form a classification function or classification process will require a sample of both kinds of units, in which each unit has been observed for the specified period of time, and values of a set of relevant variables have been obtained on each unit at or up to the start of the observation. Such a sample can be selected in a variety of ways. Following the discussion of the previous section, two of the most common sample selection mechanisms are described here (taken from Afifi and Clark, 1990).

1. Cross-sectional sample: a simple random sample is selected in a random manner and observed for the specified period of time. From this single sample two subsamples result, namely, those who experience the event and those who do not.
2. Case-control sample: two random samples are selected, one for which the event occurred and one for which the event did not occur. Values of the predictive variables must be obtained in a retrospective fashion, i.e., from past records or recollection. As noted in the previous section, the standard approach to logistic regression described in most text assumes a cross-sectional sample selection mechanism. Nevertheless, logistic regression for case control samples (the separate samples case described in the previous section) can be applied using a similar methodology to that used for cross-sectional samples. The only difference is that the parameter in the logistic model corresponding to the constant term (β_0) must be modified to account for possibly different sampling ratios in each of the two groups.

A further complication is encountered in the area of bankruptcy prediction, which derives from the use of matched or paired samples. In a paired design, the complete sample is drawn by first selecting a sample of bankrupt companies, and then matching each of these to one particular non-bankrupt company drawn from a larger database of non-bankrupt firms. The matching can be done in a number of ways, based on a set of matching variables. As an example, a match might consist of two companies belonging to the same size category and having the same SIC code. It appears that all such research ignores the effect of the matching when the discrimination techniques are used for purposes of classifying the likely status of new observations. This issue is regarded as secondary for the purposes of this project, which focuses on a comparison of non-parametric discriminant analysis with the classical methods used in the literature to this point. The problem of how to handle matching is, however, very important and raises unsolved methodological problems that have not been fully formalized and discussed in the literature. One of these problems is the difficulty to classify a single new observation when the classification function is expressed in terms of the relationship of paired bankrupted and non-bankrupt companies. While this research project will not provide a complete solution, the problem will be described in detail , and some suggestions for its solution will be advanced. A detailed exploration of these solutions will be postponed for future research .

CHAPTER 3

LITERATURE REVIEW

The starting point for looking at the analysis of failure is with the work of Beaver(1966). By employing a univariate model to differentiate survivors from failures, he arrived at measures which achieved a moderate level of predictive accuracy. Beaver(1966) defined "failure" as the firm's inability to pay its financial obligations as they mature. The 79 failed firms were matched to 79 non-failed firms by industry and asset size (as close as possible). Beaver compared the financial ratios of failed and non-failed firms in order to investigate which ratios do the best job of predicting failure. The ratios Beaver found useful included cash flow to total debt, net income to total assets, total debt to total assets, working capital to total assets, and the firm's current ratio. Beaver's study deserves credit for his objective testing of commonly used financial ratios, his degree of success in making predictions and his advocacy of the use of cash flows as a theoretical starting point for failure research.

One of the most serious problems in Beaver's study is he tests each ratio independently by a test of means. While the test itself is acceptable, the predictions could have been more accurate if the ratios were used in conjunction with one another. The use of discriminant analysis to create a failure likelihood score from several financial ratios would have addressed this issue. Another problem associated with Beaver's study which is also a common problem with bankruptcy researchers is to use a paired design with an

unpaired analysis.

The following statement by Beaver gave the rationale for using the unpaired analysis with the paired design.

"Although the paired analysis is a legitimate approach to investigating predictive ability, it has one serious drawback. It cannot draw inferences regarding a single observation-only about pairs of observation. The earlier study is restricted to such statements as: Based on an analysis of the ratios, firm A is more (or less) solvent than firm B. However, this is an ambiguous statement because it ignores the obvious question: How solvent is firm B? To be less solvent than firm B may not be so bad, if firm B is an extremely solvent firm. On the other hand, to be more solvent than firm B may be very bad, if firm B is extremely insolvent. In a decision-making situation, the user of ratios wishes to make inferences regarding a single firm--for example, what is the probability of failure for this firm? Such inferences are beyond the scope of the paired analysis."

For this reason, Beaver adopted an unpaired analysis of his paired sample.

Schlesselman (1982) states that matching that is accompanied by an unmatched analysis can reduce the validity of a case-control comparison. Although the primary objective of matching is the elimination of biased comparisons between cases and controls, this objective can be accomplished in general only if matching is accompanied by an analysis that corresponds to the matched design. Unless the analysis properly accounts for

the matching used in the selection phase of a matched case-control study, the estimated risk of bankruptcy associated with financial ratios can be biased as a consequence of matching. Thus, matching is only the first step of a two-step process that should be used to control for confounding: (1) matched design, followed by (2) matched analysis (Schlesselman, 1982). Unfortunately, Schlesselman did not consider the further problem of using a matched analysis to provide a classification tool. As noted by Beaver (1966), this poses difficulties.

Altman(1968) improved on Beaver's work in three basic ways. First, he used discriminant analysis to create a failure likelihood score from several financial ratios(called "Altman's Z-Score" as shown in equation 3.1). Second, Altman controlled for industry type and firm size. Third, he used bankruptcy as the definition for failure--therefore all firms were held to the same standard of "failure". Altman, like Beaver, compared a sample of failed firms to a sample of non-failed firms. Altman's predictive accuracy was better than Beaver's for the period one to two years prior to a possible bankruptcy, but for the period of three to five years prior to a possible bankruptcy Altman showed no great improvement over Beaver. Altman's important contribution was to promote the use of discriminant analysis to create a single bankruptcy classification function from financial ratios. Altman's Z-score was estimated as

$$Z=.012(WC/TA)+0.014(RE/TA)+.033(E/TA)+.006(MVE/BD)+.999(S/TA) \quad (3.1)$$

where Z = Failure classification function
 BD = Book value of Debt
 E = Earnings before interest and taxes
 MVE = Market Value of Equity
 RE = Retained Earning
 S = Sales
 TA = Total Assets
 WC = Working Capital

Altman adopted Fisher's discriminant function in his study. Fisher (1936) has suggested using a linear combination of the observations, and choosing the coefficients so that the ratio of the difference of the means of the linear combination in the two groups to its variance is maximized. In Fisher's approach, let the linear combination be denoted $Z = \lambda'x$. Then the mean of Z is $\lambda'u_1$ in group 1 and $\lambda'u_2$ in group 2; its variance is $\lambda'\Sigma\lambda$ in either population if we assume that the covariance matrices, $\Sigma_1 = \Sigma_2 = \Sigma$. Then we wish to choose λ to maximize

$$\Phi = \frac{(\lambda'u_1 - \lambda'u_2)}{\lambda'\Sigma\lambda}$$

Differentiating Φ with respect to λ we get

$$u_1 - u_2 = \Sigma\lambda \left[(\lambda'u_1 - \lambda'u_2) / \lambda'\Sigma\lambda \right]$$

Since we use λ only to separate the populations, we may multiply λ by any constant we desire. Thus λ is proportional to $\Sigma^{-1}(u_1 - u_2)$. If the parameters are not known,

it is the usual practice to estimate them by $\mathbf{m}_1$, $\mathbf{m}_2$ and $\mathbf{S}$. The assignment rule is to assign an observation to bankrupt group if $Z = (\mathbf{m}_1 - \mathbf{m}_2)' \mathbf{S}^{-1} \mathbf{x}$ is closer to $Z_1 = (\mathbf{m}_1 - \mathbf{m}_2)' \mathbf{S}^{-1} \mathbf{m}_1$ than to Z_2 or to non-bankrupt group otherwise. The midpoint of the interval between Z_1 and Z_2 is $(Z_1 + Z_2)/2 = 1/2 (\mathbf{m}_1 - \mathbf{m}_2)' \mathbf{S}^{-1} (\mathbf{m}_1 + \mathbf{m}_2)$. The difference between Z_1 and Z_2 is $(\mathbf{m}_1 - \mathbf{m}_2)' \mathbf{S}^{-1} (\mathbf{m}_1 - \mathbf{m}_2)$, which is Mahalanobis's D^2 (Lachenbruch, 1975).

The initial sample in Altman's study was composed of sixty-six corporations with thirty-three firms in each of the bankrupt and non-bankrupt groups. Altman applied a matched design, but ignored matching in discriminant analysis.

After Altman(1968), researchers began to extensively use discriminant analysis to create bankruptcy prediction models. Different sets of ratios were employed in these models and while researchers employed different samples covering various time periods they consistently matched firms in the sample by industry.

Altman, Haldeman and Narayanan (1977) improved Altman's model (1968) by using a ZETA model which incorporated inputs with respect to discriminant analysis and utilized a sample of bankrupt firms covering the period 1969-1975. One feature of their study was , not like previous failure models which concentrated either on the broad classification of manufacturers or on specific industries, retailing companies were also included and analyzed on an equal basis with manufacturers by analytical adjustments. Another important feature of their study was that the data had been scrupulously analyzed to include the then most recent changes in financial reporting standards and accepted accounting practices. Some of the these reporting adjustments included capitalization of

lease, reserves, minority interests and other liabilities on the balance sheet, captive finance companies and other non-consolidated subsidiaries, goodwill and intangibles. In order to reduce outlier possibilities and to adhere to statistical assumption, some variables were expressed in logarithmic form. After all of these adjustments, the ZETA model appeared to be quite accurate for up to five years prior to failure with successful classification of well over 90% of the sample one year prior to bankruptcy and 70% accuracy up to five years.

Other major studies employing discriminant analysis include Blum(1974) and Rose and Giroux(1984). Studies which employ a principally cash flow analysis include Wilcox(1973), Casey and Bartczak(1985), Gentry, Newbold and Whitford(1985), and Aziz, Emanuel and Lawson(1988). Other studies add variations to specifically study small business (e.g. Edminister, 1972), and analyze the effect of environmental changes over time (e.g. Ketz, 1978; Norton and Smith, 1979; Mensah, 1983, 1984). Other models which have added to our understanding of the bankruptcy state are logit models (Martin, 1977, Ohlson, 1980), and factor analytic models(West, 1985). The proportional hazard model has recently been applied to the bank failure question (Lane, Looney and Wansley, 1986). Betton (Betton, 1990) applied the proportional hazard model to the problem of corporate bankruptcy prediction.

Zavgren (1985) has argued that some of the above studies are deficient due to the fact that they employ methodologies, such as discriminant analysis, which require sample data to be normally distributed within each group (bankrupt and non-bankrupt). If some

of the variables are not normally distributed, the methods employed may result in selection of an inappropriate set of predictors. To correct for this problem, Zavgren used logit analysis to arrive at her bankruptcy prediction model (as described, logit analysis is not affected by data sets which are not normally distributed). Zavgren argued that the conclusions of most of the earlier studies are difficult to assess because they "play loose" with the assumptions of discriminant analysis. As previously described, discriminant analysis requires that the variables be normally distributed and that the populations have equal variance-covariance matrices. Otherwise, a linear classification rule and the test for distinction between groups become inappropriate. Zavgren (1985) argued none of these authors mentioned above test their data against these assumptions. Confusion over the purposes of discriminant analysis sometimes leads to inappropriate attempts to assess the meaning of individual coefficients. The linear combination of variables unique to a particular classification function renders the importance of individual coefficients difficult to assess. Zavgren's model (a combination of return on investment, financial leverage, short term liquidity, cash position, and capital turnover, receivables turnover and inventory turnover) can thus be considered more robust for her use of logit analysis. It should be noted that Zavgren used a matched design with an unmatched analysis in her paper. For testing the classification power of her procedures, she created a separate matched sample from which she estimated the classification errors by use of her unmatched logistic model.

Lo (1986) has shown that logit analysis is appropriate under a general exponential family assumption on the conditional distribution of variables within each group, and that,

at least for the data he examined, the results of a logit analysis can be quite similar to those achieved with discriminant analysis. Press and Wilson (1978) compare logistic regression and parametric discriminant analysis and conclude that logistic regression is preferable to parametric discriminant analysis in cases for which the variables do not have multivariate normal distributions within classes. However, if the variables have normal within-class distributions, logistic regression is less efficient than parametric discriminant analysis. Efron (1975) shows that with two normal populations having a common covariance matrix, logistic regression is between one half and two thirds as effective as the linear discriminant function to achieve asymptotically the same error rate. Logit and discriminant analysis are distinguished by another characteristic: logit is more robust than discriminant analysis. Efron's (1975) study indicates that if the normality of $f_i(\mathbf{x})$ does obtain, then discriminant analysis is considerably more efficient than logit. Indeed, since the normal discriminant analysis procedure is in fact maximum likelihood estimation when $f_i(\mathbf{x})$ is normally distributed, it is asymptotically efficient in this case. However, the literature review suggests that there have not been any studies which use nonparametric methods of discriminant analysis in bankruptcy prediction. When the distribution within each group is not assumed to have any specific distribution or is assumed to have a distribution different from the multivariate normal distribution, or more generally, from distributions of the exponential family, a nonparametric method should be used to derive classification criteria. Previous research has left a void in the study of bankruptcy prediction when the continuous probability density function is severely non-normal.

It should be remembered that the above comparison will be carried out on what is in reality a matched case-control sample, but in which the matching is ignored for purposes of analysis and comparison of classification power. Thus the classification error rates may be biased by this limitation, though the effect should not differ greatly across the methods. This limitation should not, therefore, bias the comparison of the merits of the three methods unduly.

CHAPTER 4

RESEARCH QUESTIONS AND HYPOTHESES

4.1 Research Questions

The literature leads to the asking of the following research questions:

- (1) How should non-parametric discriminant analysis be applied to predict corporate bankruptcy?
- (2) Which of normal theory discriminant analysis, kernel-method discriminant analysis and logit analysis gives the best prediction of bankruptcy in terms of error count estimate, when applied to a real data sample?
- (3) What is the precise nature of the problems implied by combining matched sample design with an unmatched analysis?

This proposal considers these questions in detail. An outline of the methodology is discussed in Chapter 5.

4.2 Hypotheses

The hypotheses of this study are :

- (1) If the class-conditional probability density distribution is normal, the normal theory discriminant analysis, in terms of error count estimate, will perform best,

followed by logit analysis and the Epanechnikov kernel discriminant analysis.

(2) If the class-conditional probability distribution is not normal or not from the exponential family of distributions then Epanechnikov kernel discriminant analysis will perform better than logit analysis in terms of error count estimate. Normal theory discriminant analysis in this case is invalid.

It should be noted that it is not feasible to test whether or not distribution belongs to the general exponential family. In this research, only test of normality will be conducted. It will be seen that logistic regression is in fact outperformed by the non-parametric technique, from which it can be inferred that the group distributions are not from the exponential family.

CHAPTER 5

METHODOLOGY

5.1 Multiple Discriminant Analysis

In this chapter, further details of the different approaches to discriminant analysis are provided, leading to a discussion of non-parametric techniques. Estimation of error rates, as a means of assessing performance of classification functions is also described.

As outlined in Chapter 2, multiple discriminant analysis computes classification functions for assessing observations into two (or more) groups on the basis of scores on one or more quantitative variables. When the distribution within each group is assumed to be multivariate normal, a parametric method based on multivariate normal distribution theory is used to derive a linear or quadratic classification function. Otherwise, a nonparametric method that does not make any assumptions about the distributions can be used.

The notation below (taken from the SAS/STAT User's Guide) is used to describe the classification methods:

- x** a p -dimensional vector containing the quantitative variables of an observation
- S** the pooled covariance matrix
- t** a subscript to distinguish the groups
- n_t** the number of training set observations in group t

- $\mathbf{m}_t$ the p -dimensional vector containing variable means in group t
- S_t the covariance matrix within group t
- $|S_t|$ the determinant of S_t
- q_t the prior probability of membership in group t
- $p(t|\mathbf{x})$ the posterior probability of an observation $\mathbf{x}$ belonging to group t .

5.1.1 Bayes Theorem

In section 2.2.2 and chapter 3, we had discussed Welsh's approach and Fisher's approach to discriminant function respectively. Another approach that leads to the classification rules is to assign $\mathbf{x}$ to the group with the largest posterior probability, i.e., Bayes theorem approach.

Assuming that the prior probabilities of group membership are known and the group-specific densities at $\mathbf{x}$ can be estimated, discriminant analysis computes $p(t|\mathbf{x})$, the probability of $\mathbf{x}$ belonging to group t , by applying Bayes' theorem:

$$p(t|\mathbf{x}) = q_t f_t(\mathbf{x}) / f(\mathbf{x})$$

where $f_t(\mathbf{x})$ is the group-specific density estimate at $\mathbf{x}$ from group t , and $f(\mathbf{x}) = \sum_t q_t f_t(\mathbf{x})$ is the estimated unconditional density at $\mathbf{x}$.

Discriminant analysis partitions a p -dimensional vector space into regions R_t , where the region R_t is the subspace containing all p -dimensional vectors $\mathbf{y}$ such that $p(t|\mathbf{y})$ is the largest among all groups. An observation is classified as coming from group t if it lies in region R_t . If we assign an observation to bankrupt group when $p(Y=0|\mathbf{x}) > p(Y=1|\mathbf{x})$, this

is equivalent to the rule that minimizes the total probability of misclassification.

5.1.2 Parametric Methods

Assuming that each group has a multivariate normal distribution, discriminant analysis develops a discriminant function or classification criterion using a measure of generalized squared distance. The classification criterion of page 13 (equation 2.2.1) is based on either the individual within-group covariance matrices or on the pooled covariance matrix; it also takes into account the prior probabilities of the classes. The assignment rule of Fisher's discriminant function is equivalent of assigning each observation into the class from which it has the smallest generalized squared distance. The generalized squared distance from $\mathbf{x}$ to group t can be defined as

$$D_t^2(\mathbf{x}) = d_t^2(\mathbf{x}) + g_1(t) + g_2(t)$$

where

$$d_t^2(\mathbf{x}) = (\mathbf{x} - \mathbf{m}_t)' \mathbf{V}_t^{-1} (\mathbf{x} - \mathbf{m}_t), \text{ the squared distance from } \mathbf{x} \text{ to group } t$$

where $\mathbf{V}_t = \mathbf{S}_t$ if the within-group covariance matrices are used, or $\mathbf{V}_t = \mathbf{S}$ if the pooled covariance matrix is used. When the within-group covariance matrices are homogeneous, the pooled covariance matrix is used. Otherwise, the within-group covariance matrices are used.

$$g_1(t) = \log_e S_t \quad \text{if the within-group covariance matrices are used, or}$$

$$g_1(t) = 0 \quad \text{if the pooled covariance matrix is used; and}$$

$$g_2(t) = 2 \log_e(q_t) \quad \text{if the prior probabilities are not all equal. or}$$

$g_2(t)=0$ if the prior probabilities are all equal.

The group-specific density estimate at $\mathbf{x}$ from group t is then given by

$$f_t(\mathbf{x}) = (2\pi)^{-p/2} |\mathbf{V}_t|^{-1/2} \exp(-0.5d_t^2(\mathbf{x}))$$

Discriminant analysis also computes the posterior probability of an observation belonging to each class. Using Bayes' theorem, the posterior probability of $\mathbf{x}$ belonging to group t is

$$p(t|\mathbf{x}) = \frac{q_t f_t(\mathbf{x})}{\sum_u q_u f_u(\mathbf{x})}$$

The posterior probability of $\mathbf{x}$ belonging to group t is then equal to

$$p(t|\mathbf{x}) = \frac{\exp(-0.5D_t^2(\mathbf{x}))}{\sum_u \exp(-0.5D_u^2(\mathbf{x}))}$$

Classification according to the smallest value of $D_t^2(\mathbf{x})$ is equivalent to classifying into group u if setting $t=u$ produces the largest value of $p(t|\mathbf{x})$. (Note that if pooled within-group covariance is used and if prior probabilities are equal, then this is equivalent to placing an observation in the "nearest" group, in terms of Mahalanobis distance). From discussions conducted in the previous sections, we can conclude Fisher's, Welsh's, Bayes theorem and Mahalanobis distance approaches are actually equivalent.

5.1.3 Density Estimation and Nonparametric Discrimination

In view of Bayes' theorem, the computation of $p(t|\mathbf{x})$, the probability of $\mathbf{x}$ belonging

to group t , is a question of estimating the group-specific probability density $f_t(\mathbf{x})$. The probability density function is a fundamental concept in statistics. Consider any random quantity X that has probability density function f . Specifying the function f gives a natural description of the distribution of X , and allows probabilities associated with X to be found from the relation

$$P(a < X < b) = \int_a^b f(x)dx \quad \text{for all } a < b$$

Suppose, now, that we have a set of observed data points assumed to be a sample from an unknown probability density function. Density estimation is the construction of an estimate of the density function from the observed data.

The discussions of the previous section implicitly describe a *parametric* approach to density estimation under the assumption that the data are drawn from one of a known parametric family of distributions, for example, the normal distribution with parameters consisting of the mean μ and variance matrices Σ_1 and Σ_2 , the density f underlying the data is implicitly estimated by finding estimates of μ and σ^2 from the data and substituting these into the formula for the multivariate normal density. Another approach which will be described in Section 5.3 is *nonparametric* in that less rigid assumptions will be made about the distribution of the observed data. Although it will be assumed that the distribution of the observed data has a probability density f , the functional form of f will not be assumed, and f will be estimated from the data. Density estimates of this kind were first proposed by Fix and Hodges (1951) as a way of freeing discriminant analysis from rigid distributional assumptions. Since then, density estimation and related ideas have been

used in a variety of contexts, including discriminant analysis. The nonparametric methods used in this study will be described in detail in next section.

5.2 Non-parametric Discriminant Analysis and Its Kernel Method

Silverman (1986) classified non-parametric pdf estimators into four major types: the histogram, the kernel method, the k-nearest-neighbour(k-NN) method, and the series method. Other non-parametric pdf estimators include the naive estimator, variable kernel method, maximum penalized likelihood estimators. According to Silverman (1986), the naive estimator was introduced by Fix and Hodges(1951) in an unpublished report, while the first published paper to deal explicitly with probability density estimation was by Rosenblatt (1956), who discussed both the naive estimator and the more general kernel estimator.

The Kernel method has been selected as non-parametric method of classification in the study. It will be used to generate a nonparametric density estimate in each group and then to produce a classification criterion. The author's concentration on the kernel method is not intended to imply that the method is the best to use in all circumstances, but there are several reasons for considering the kernel methods first of all. The method is of wide applicability, particularly in the univariate case, and it is certainly worth understanding its behaviour before going on to consider other methods. It is probably the method whose properties are best understood, and discussion of these properties raises

issues which relate to other methods of density estimation.

5.2.1 The Naive Estimator

From the definition of a probability density, if the random variable X has density f , then

$$f(x) = \lim_{h \rightarrow 0} \frac{1}{2h} P(x-h < X < x+h) \quad (5.2.1)$$

For any given h , we can of course estimate $P(x-h < X < x+h)$ by the proportion of the sample falling in the interval $(x-h, x+h)$. Thus a natural estimator $\hat{f}$ of the density is given by choosing a small number h and setting

$$\hat{f}(x) = \frac{1}{2hn} \left[\text{number of } X_1, \dots, X_n \text{ falling in } (x-h, x+h) \right];$$

We shall call this the naive estimator (Rosenblatt, 1956).

To express the estimator more transparently, define the weight function w by

$$w(x) = \begin{cases} (1/2) & \text{if } |x| < 1 \\ 0 & \text{otherwise} \end{cases} \quad (5.2.2)$$

Then it is easy to see that the naive estimator can be written

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n W\left(\frac{x - X_i}{h}\right)$$

It follows from (5.2.2) that the estimate is constructed by placing a 'box' of width $2h$ and height $(2nh)^{-1}$ on each datum x_i , and then summing the heights of the boxes containing the point x , to obtain the estimate. The naive estimator is not wholly satisfactory from the point of view of using density estimates for presentation. "It follows from the definition that $\hat{f}$ is not a continuous function, but has jumps at the points $X_i \pm h$ and has zero derivative everywhere else. This gives the estimates a somewhat ragged character which is not only aesthetically undesirable, but, more seriously, could provide the untrained observer with a misleading impression" (Silverman, 1986).

5.2.2 Definition of the Univariate Kernel Density Estimator

It is easy to generalize the naive estimator to overcome some of the difficulties discussed above. Replace the weight function w by a *kernel function* K which satisfies the condition

$$\int_{-\infty}^{\infty} K(x) dx = 1 \quad (5.2.3)$$

Usually, but not always, K will be a symmetric probability density function, the normal density, for instance, or the weight function w used in the definition of the naive estimator. By analogy with the definition of the naive estimator, the *kernel estimator* K is defined by

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right)$$

where h is the *smoothing parameter*, also called *window width* or *bandwidth* by some authors. Just as the naive estimator can be considered as a sum of 'boxes' centred at the observations, the kernel estimator is a sum of 'bumps' placed at the observations. The kernel function K determines the shape of the bumps while the smoothing parameter h determines their width. Some elementary properties of kernel estimates follow at once from the definition. Provided the kernel K is everywhere non-negative and satisfies the condition (5.2.3) -- in other words is a probability density function -- it follows at once from the definition that $\hat{f}$ will itself be a probability density. Furthermore, $\hat{f}$ will inherit all the continuity and differentiability properties of the kernel K , so that if, for example, K is the normal density function, then $\hat{f}$ will be a smooth curve having derivatives of all orders.

5.2.3 Definition of the Multivariate Kernel Density Estimator

The definition of the kernel estimator is easily generalized to the multivariate case. The multivariate kernel density estimator with kernel K and smoothing parameter is defined by

$$\hat{f}(\mathbf{x}) = \frac{1}{nh^p} \sum_{i=1}^n K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right)$$

The kernel function $K(\mathbf{x})$ is now a function, defined for p -dimensional $\mathbf{x}$, satisfying

$$\int_{R^p} K(\mathbf{x}) d\mathbf{x} = 1$$

Usually, K will be a radially symmetric unimodal probability density function, for example, the standard multivariate normal density function $K(\mathbf{x}) = (2\pi)^{-p/2} \exp(-1/2 \mathbf{x}^T \mathbf{x})$.

Another example is the multivariate Epanechnikov kernel

$$K_e(\mathbf{x}) = \begin{cases} 1/2 V_p^{-1} (p+2) (1 - \mathbf{x}^T \mathbf{x}) & \text{if } \mathbf{x}^T \mathbf{x} < 1 \\ 0 & \text{otherwise} \end{cases} \quad (5.2.4)$$

where V_p is the volume of the unit p -dimensional sphere: $V_1=2$, $V_2=\Pi$, $V_3=4\Pi/3$, etc. The useful kernels for the case $p=2$ are given by

$$K_2(\mathbf{x}) = \begin{cases} 3\pi^{-1} (1 - \mathbf{x}^T \mathbf{x})^2 & \text{if } \mathbf{x}^T \mathbf{x} < 1 \\ 0 & \text{otherwise} \end{cases}$$

and

$$K_3(\mathbf{x}) = \begin{cases} 4\pi^{-1} (1 - \mathbf{x}^T \mathbf{x})^3 & \text{if } \mathbf{x}^T \mathbf{x} < 1 \\ 0 & \text{otherwise} \end{cases}$$

The former is defined as Biweight kernel and the latter is defined as Triweight kernel.

5.2.4 Measures of Discrepancy: Mean Square Error and Mean Integrated Square Error

It is well known that, at least as far as density estimation goes, the choice of smoothing parameter h is critical. On the choice of K , the kernel form, the bulk of the evidence to date suggests that it is not so critical (Silverman, 1986).

Various measures have been studied of the discrepancy of the density estimator $\hat{f}$ from the true density f . The first (Rosenblatt, 1956) and the most widely used way of placing a measure on the *global* accuracy of $\hat{f}$ as an estimator of f is the *mean integrated square error* (abbreviated MISE) defined by

$$\text{MISE}(\hat{f}) = E \int \{\hat{f}(\mathbf{x}) - f(\mathbf{x})\}^2 d\mathbf{x} \quad (5.2.5)$$

Though there are other global measures of discrepancy which may be more appropriate to one's intuitive ideas about what constitutes a globally good estimate, the MISE is by far the most tractable global measure. It is useful to note that, since the integrand is non-negative, the order of integration and expectation in (5.2.5) can be reversed to give the alternative forms

$$\begin{aligned} \text{MISE}(\hat{f}) &= E \int \{\hat{f}(\mathbf{x}) - f(\mathbf{x})\}^2 d\mathbf{x} \\ &= \int \text{MSE}_{\mathbf{x}}(\hat{f}) d\mathbf{x} \\ &= \int \{E\hat{f}(\mathbf{x}) - f(\mathbf{x})\}^2 d\mathbf{x} + \int \text{var}\hat{f}(\mathbf{x}) d\mathbf{x} \end{aligned} \quad (5.2.6)$$

The latter gives the MISE as the sum of the *integrated* square bias and the *integrated* variance. This expression will be used in what follows to examine the issue of choosing a suitable window width h to use with the kernel estimator.

5.2.5 Choosing the Kernel

Approximate expressions can be derived for the bias and the variance of the estimators $\hat{f}$, and these can be used to give some guidance concerning the appropriate choice of kernel and window width. Assume that the kernel K is a radially symmetric

probability density function and that the unknown density f has bounded and continuous second derivatives. Define constants α and β by

$$\alpha = \int t^2 K(t) dt$$

and

$$\beta = \int K(t)^2 dt \quad (5.2.7)$$

Use of the multi-dimensional form of Taylor's theorem then yields the approximations (Silverman, 1986)

$$\text{bias}_h(\mathbf{x}) \approx 1/2 h^2 \alpha V^2 \hat{f}(\mathbf{x}) \quad (5.2.8)$$

$$\text{var } \hat{f}(\mathbf{x}) \approx n^{-1} h^{-p} \beta f(\mathbf{x}) \quad (5.2.9)$$

Combining these gives the approximate mean integrated square error

$$1/4 h^4 \alpha^2 \int \{V^2 f(\mathbf{x})\}^2 d\mathbf{x} + n^{-1} h^{-p} \beta \quad (5.2.10)$$

Hence the approximately optimal window width, in the sense of minimizing mean integrated square error, is given by

$$h_{\text{opt}}^{p+4} = p\beta\alpha^{-2} \{ \int (V^2 f)^2 \}^{-1} n^{-1} \quad (5.2.11)$$

The approximately optimal window width of (5.2.11) converges to zero as n increases, but does so extremely slowly, at the rate $n^{-1/(p+4)}$. Furthermore the appropriate value of h depends on the unknown density being estimated.

The value h_{opt} can be substituted back into (5.2.10) to give the approximate minimum possible mean integrated square error, and hence, to guide the choice of kernel (Silverman, 1986). The Epanechnikov kernel is optimum in the sense of minimizing the

smallest mean integrated square error achievable; however, the other kernels defined in Table 5.2.1 can achieve very similar mean integrated square errors.

5.2.6 Choice of Window Width for a Standard Distribution

The first step beyond choosing the smoothing parameter entirely subjectively is to use the formula (5.2.11) to provide the appropriate value of the window width when f is a standard density, such as the multivariate normal. If ϕ is the unit p -variate normal density, then it can be shown that

$$\int (V^2 \phi)^2 = (2\sqrt{\pi})^p (1/2^p + 1/4^p) \quad (5.2.12)$$

The value given by (5.2.12) can then be substituted back in (5.2.11) to give the optimal window width for the smoothing of normally distributed data with unit variance. The window width will be given by

$$h_{\text{opt}} = A(K)n^{-1/(p+4)} \quad (5.2.13)$$

Where the constant

$$A(K) = \left[p\beta\alpha^{-2} \left\{ \int (V^2 \phi)^2 \right\}^{-1} \right]^{1/(p+4)} \quad (5.2.14)$$

depends on the kernel and is tabulated in Table 5.2.1 taken from Silverman (1986).

Table 5.2.1. The constant $A(K)$, as defined in equation 5.2.14, for various kernels K . Numerical values are given correct to 2 decimal places

Kernel	Dimensionality	$A(K)$
Multivariate normal $K(\mathbf{x}) = (2\pi)^{-p/2} \exp(-1/2 \mathbf{x}^T \mathbf{x})$	2 p	0.96 $\{4/(2p+1)\}^{1/(p+4)}$
Multivariate Epanechnikov $K_e(\mathbf{x}) = \begin{cases} 1/2 V_p^{-1} (p+2) (1 - \mathbf{x}^T \mathbf{x}) & \text{if } \mathbf{x}^T \mathbf{x} < 1 \\ 0 & \text{otherwise} \end{cases}$	2 3 p	1.77 2.78 $8p(p+2)(p+4)(2\sqrt{\pi})^{p-1/(p+4)}$
Biweight $K_2(\mathbf{x}) = \begin{cases} 3\pi^{-1} (1 - \mathbf{x}^T \mathbf{x})^2 & \text{if } \mathbf{x}^T \mathbf{x} < 1 \\ 0 & \text{otherwise} \end{cases}$	2	$(2p+1)V_p$ 2.04
Triweight $K_3(\mathbf{x}) = \begin{cases} 4\pi^{-1} (1 - \mathbf{x}^T \mathbf{x})^3 & \text{if } \mathbf{x}^T \mathbf{x} < 1 \\ 0 & \text{otherwise} \end{cases}$	3	2.29

V_p is the volume of the unit p -dimensional sphere ; e means multivariate Epanechnikov kernel.

5.3 Classification Error-Rate Estimates

A classification criterion can be evaluated by its performance in the classification of observations not used in the estimation. There are two types of error rate estimates that can be used to evaluate the derived classification criterion based on parameters estimated by the training sample:

error-count estimate

posterior probability error-rate estimate

The error-count estimate is calculated by applying the classification criterion derived from the training sample to a test set and then counting the number of misclassified observations. The group-specific error-count estimate is the proportion of misclassified observations in the group in question. When the test set is independent of the training sample, this estimate is unbiased. However, it can have a large variance, especially if the test set is small.

When the same data set is used both to define and to evaluate the classification criterion, the resulting error-count estimate has an optimistic bias and is called an apparent error rate. To reduce the bias, the data can be split into two sets, one set for deriving the discriminant function and the other set for estimating the error rate. Such a split-sample method has the unfortunate effect of reducing the effective sample size. Another way to reduce bias is cross-validation or jackknife (Lachenbruch and Mickey 1968). Cross-validation treats $n-1$ out of n training observations as a training set. It determines the classification functions based on these $n-1$ observations and then applies them to classify the one observation left out. This is done for each of the n training observations. The misclassification rate for each group is the proportion of sample observations in that group that are misclassified. This method achieves a nearly unbiased estimate but with a relatively large variance.

Glick (1978) proposed a new class of estimators, called smoothed error rate

estimators, designed to reduce the resubstitution method's variance as well as its bias. He performed a sampling experiment to compare the bias and variance of a smoothed estimator to those of several other estimators but considered only one member of the class in the univariate normal case. Instead of summing terms that are either zero or one as in the error-count estimator, the smoothed estimator uses a continuum of values between zero and one. The terms then are summed. The resulting posterior probability estimator has a smaller variance than the error-count estimate.

The posterior probability error-rate estimates are smoothed error rate estimates. The posterior probability estimates for each group are based on the posterior probabilities of the observations classified into that same group. The posterior probabilities estimates provide good estimates of the error rate when the posterior probabilities are accurate. However, when a parametric classification criterion (linear or quadratic discriminant function) is derived from a non-normal population, the resulting posterior probability error-rate estimators may not be appropriate. Error count estimate is one of the alternatives in this case. The overall error-rate is estimated through a weighted average of the individual group-specific error-rate estimates, where the prior probabilities are used as the weights.

Hora and Wilcox (1982) compute the posterior probability estimates based on cross-validation. The resulting estimates are intended to have both low variance from using the posterior probability estimate, and low bias from cross-validation. They use Monte Carlo studies on two-group multivariate normal distributions to compare the

cross-validation posterior probability estimates with other aforementioned estimators. They conclude that the cross-validation posterior probability estimator has a lower mean squared error in their simulations.

This study proposes to use the error count estimate of the resubstitution to evaluate the efficiency of the derived classification criterion based on parameters estimated by the training sample in three different methods. The reason for use of error count estimate is for comparison with Altman's results. The error-count estimates are also nature with non-normal population in this case. The cross-validation posterior probability error-rate estimate will also be used on a subset of the data in order to provide a comparison with error count estimate.

5.4 Parameter Estimation for a Matched Analysis

As described by Afifi and Clark (1990) and by Hand (1990), the parameters of the logistic model (β_1 through β_p , but not β_0) for matched case-control data can be estimated using a standard logistic regression program, provided these adjustments are made:

1. For each of the N pairs, compute the difference $x(\text{case}) - x(\text{control})$ for each x variable. These differences will form the new variables for a single sample of size N to be used in the logistic regression equation.
2. Create a new dependent variable Y with a value of 1 for each of the N pairs.

3. Set the constant term in the logistic regression equation equal to zero.

A mathematical description of this procedure is given as described below. In this study, a 1-1 case-control match, there are two subjects within each block. To simplify the notation, let $\mathbf{x}_{1k}$ denote financial ratio vector for bankrupt firm and $\mathbf{x}_{0k}$ the financial ratio vector for survivor in the k block. Using this notation, the conditional likelihood for the k block is

$$l_k(\beta) = \frac{e^{\beta' \mathbf{x}_{1k}}}{e^{\beta' \mathbf{x}_{1k}} + e^{\beta' \mathbf{x}_{0k}}} \quad (5.4.1)$$

Further simplification is obtained when we divide the numerator and denominator of equation (5.4.1) by $e^{\beta' \mathbf{x}_{0k}}$ yield

$$l_k(\beta) = \frac{e^{\beta'(\mathbf{x}_{1k} - \mathbf{x}_{0k})}}{1 + e^{\beta'(\mathbf{x}_{1k} - \mathbf{x}_{0k})}} \quad (5.4.2)$$

The expression on the right-hand side of equation (5.4.2) is identical to a logistic regression model with the constant term set to zero, $\beta_0=0$, and data vector equal to the value of the case minus the value of the control, $\mathbf{x}_k = \mathbf{x}_{1k} - \mathbf{x}_{0k}$. This observation allows us to use standard logistic regression software to compute the conditional maximum likelihood estimates and obtain estimated standard errors of the estimated coefficients. To do this we define the sample size as the number of case-control pairs, use as covariates the differences $\mathbf{x}_k$, set the values of the response variable equal to 1, $y_k=1$, and exclude the constant term from the model. Thus, from a computational point of view, the 1-1 matched design presents no new challenges.

Schlesselman (1982) compares the output that would be obtained from a paired-matched logistic regression analysis to that obtained if the matching was ignored. He explains that if the matching is ignored and the matching variables are associated (correlated) with the exposure variables, then the estimates of the odds ratio of the dependent variables will be biased toward no effect (or an odds ratio of 1). If the matching variables are not associated with the exposure variables, then it is not necessary to use the form of the analysis that accounts for matching. In this case, matching itself is not necessary and achieves nothing. Furthermore, in the latter case, using the analysis that accounts for matching would result in larger estimates of the standard errors of the slope coefficients. As noted, however, Schlesselman does not consider how to use the "matched" analysis in a predictive classification sense. The source of the problem is that the above method fails to estimate β_0 . Without β_0 , the conditional probability of group membership for a new observation can not be estimated directly.

CHAPTER 6

DATA COLLECTION

6.1 The Sample of Bankrupt Firm

The firms to be included in the research sample were selected from the firms on the Wall Street Journal Index (W.S.J.I.), Moody's Industrial Manual, and previous publications. The failed firms were identified as having filed Chapter X or Chapter XI bankruptcy proceedings during the period 1972-1992. Three criteria were used to identify the population boundaries: (i) the period from 1972-1992; (ii) the company must have been traded on some stock exchange or on the over-the-counter market; (iii) the cases represent commercial and industrial failures, including construction, manufacturing and mining and retail and wholesale trades. The reason for choosing the first criterion involving a wide time coverage is to increase the realism of the sample. The second criterion excludes small or privately held corporations. This is crucial, since otherwise the use of firms in Moody's and Standard and Poor's as a source of non-bankrupt firms would be precluded. With the third criterion, railroads, insurance and banking institutions and savings and loan associations are excluded due to their special treatment under US bankruptcy law.

Ideally, all reports for all failed firms satisfying the population constraints should be included in the bankrupt "sample". However, this is clearly impractical. The following

procedure was used to generate a list of failed firms satisfying the inclusion criteria. (1) A primary listing of failed firms was tabulated from the Wall Street Journal Index (W.S.J.I.), Moody's Industrial, Zavgren's paper (1985) and Lo's paper (1986). This list is only for American firms, the Canadian firms are excluded. Similar research concentrating on Canadian firms is planned and will be conducted in the future. Type and date of bankruptcy was recorded. If the name of the company indicated that the firm in question was nonindustrial, then it was excluded. An initial sample of 94 firms was extracted. (2) A secondary listing of firms was tabulated by excluding all firms on the primary listing which had not been traded on one of the stock exchanges (or OTC) during the five-year period prior to the date of bankruptcy. If a company could be traced to one of the exchanges, then the exchange was recorded. This kind of information was derived from stock guides issued by Moody's. Of course, as a practical matter, it was assumed that a stock had not been traded if no evidence could be found to that effect. Nineteen firms were excluded after this step. (3) The remaining firms were traced back on Moody's Industrial in order to find their financial data. Another fourteen firms were excluded since some of required financial data were missing from their financial records. After these three steps, 33 firms which had complete financial records during the five years prior to the bankruptcy filing were left. A breakdown of the bankrupt sample is as follows:

1 year prior to bankruptcy filing: 33 firms

2 years prior to bankruptcy filing: 51 firms

3 years prior to bankruptcy filing: 61 firms

4 years prior to bankruptcy filing: 55 firms

5 years prior to bankruptcy filing: 35 firms

6.2 The Sample of Non-Bankrupt Firms

Most of the bankruptcy studies reported in the literature typically construct sample of non-bankrupt firms by pair-wise matching with individual firm in the bankrupt sample. Typical matching (or control) variables are industry and firm size.

6.2.1 The Matching Debate

Matching raises many issues, one of which, namely, the problem of linking a matched sample to an unmatched analysis, has been discussed earlier. One advantage is that matching may increase homogeneity of within group covariance. Ohlson has argued (as has been previously documented in this paper), that "there are also certain problems related to the matching procedures which have typically been used in MDA. Failed and non-failed firms are matched according to criteria such as size and industry, and these tend to be somewhat arbitrary. It is by no means obvious what is really gained or lost by different matching procedures, including no matching at all." Although majority of the bankruptcy studies use this "matching" methodology, no research has ever provided a quantitative guide to the necessity of matching and on how to choose potential factors. There are very few studies which have compared results from matched data and unmatched data and which have provided recommendations on which is more efficient.

Research in the epidemiology field has shed some light on the issue of matching (Breslow and Day, 1980, Schlesselman, 1982, Rothman, 1986). This line of research shows that under two conditions, a matching factor should be used. First, matching should be used if there is an indirect association between bankruptcy and financial ratios that is due to the potential matching factor. Failure to match or otherwise control for the potential matching factor would result in a spurious association between bankruptcy and financial ratios. The second situation is when financial ratios and the potential matching factors individually alter the risk of bankruptcy, and are also associated. Failure to match or otherwise control for potential matching factor in this instance would result in a biased assessment of the individual effect to bankruptcy. Thompson (1980) has compared the asymptotic variances of the disease-exposure odds ratio based on stratified analyses of unmatched and matched case-control designs. The following major findings emerge from his studies. He states that, in most circumstances, a matched design results in a modest improvement in efficiency, typically resulting in a 5 percent to 15 percent reduction in variance of the estimated disease-exposure odds ratio. The greatest improvements in efficiency due to matching occur when there is a strong association between the study disease and the confounder, the range in the risk of disease across levels of the confounder being twenty-fold or greater.

In this study, industry type and firm size have been chosen as the matching factors in evaluating the relationship between bankruptcy and financial ratios. The rationale for such a choice is consistent with the second of the two situations described above that call

for matching. As noted by Beaver (1966), the same numerical value of a given ratio implies a different probability of failure in different industries, i.e., differences exist among industries that prevent the direct comparison of firms from different industries. In addition, it is a fact that the distributions of financial ratios differ among industries. Evidence in terms of size as a matching variable has been given by Shepperd (1990), Chan et al. (1985) and Buzzell (1975), who have found firm size plays a role in the firm's likelihood of failure, and that size may also be indicative of market share.

A formal argument regarding the benefits, or otherwise, of matching relates to the availability of a proper mechanism that allows for the random selection of bankrupt and non-bankrupt firm from a properly defined population of firms. For reason of cost, the best database on bankrupt and non-bankrupt firms (COMPUSTAT) was not available to this study, so that a "convenience" sample of bankrupt firms had to be constructed, as described earlier. In such a case, pairwise matching on two important firm characteristics is essential in order to induce realistic distributions on the pairwise differences of the unmatched variables.

6.2.2 Selection of Matching Firms

For this study, a survivor firm "matched" a bankrupt firm if the following conditions were met:

1. The firms must be in the same or similar industry as represented by the SIC codes. A survivor's SIC code "matches" if it is within plus or minus 5 of the units digit of the failure's 4 digit SIC code.

2. The size of the case and control firm were compared in the same year which was five years prior to failure.

The mean asset size of bankrupt firms in the sample was \$203 million, with a standard deviation of \$33 million and a range of \$3 million to \$395 million. The mean asset size of the non-bankrupt group of firms was \$327 million, with a standard deviation of \$50 million and a range of \$4 million to \$526 million. Appendix A contains the list of bankrupt firms and non-bankrupt firms, including information on company names, SIC code, bankruptcy date, asset size, etc.

6.3 Financial Ratios Used

The variables employed in this study were similar to the ones employed by Altman (1968). Altman's study, though not as extensive as some others (e.g. Rose and Giroux, 1984), is probably the most frequently cited failure study. Thus, it serves as an excellent starting point for this study. Only a slight deviation from Altman's study was made. Instead of the ratio of earnings before interest and taxes over total assets, net income over total assets was employed. This measure was selected due to its usefulness as a broad measure of solvency and the fact that conforms more closely with regularly tracked industry standard data.

A description of the five financial ratio variables used in the study are given below:

X_1 - Working Capital/Total Assets. The working capital/total assets ratio, frequently

found in studies of corporate problems, is a measure of the net liquid assets of the firm relative to its total capitalization. Working capital is defined as the difference between current assets and current liabilities. Liquidity and size characteristics are explicitly considered. Ordinarily, a firm experiencing consistent operating losses will have shrinking current assets in relation to total assets.

X_2 - Retained Earnings/Total Assets. This measure of cumulative profitability over time was cited earlier as one of the "new" ratios. The age of a firm is implicitly considered in this ratio. For example, a relatively young firm will probably show a low RE/TA ratio because it has not had time to build up its cumulative profits. Therefore, it may be argued that the young firm is somewhat discriminated against in this analysis, and its chance of being classified as bankrupt is relatively higher than another, older firm. But, this is precisely the situation in the real world. The risk of failure is much higher in a firm's earlier years.

X_3 - Net Income/Total Assets. This ratio is calculated by dividing the total assets of a firm into its earnings after interest and tax reductions. In essence, it is a measure of the true productivity of the firm's assets. Since a firm's ultimate existence is based on the earning power of its assets, this ratio appears to be particularly appropriate for studies dealing with corporate failure. Furthermore, insolvency in a bankruptcy sense occurs when the total liabilities exceed a fair valuation of the firm's assets with value determined by the earning power of the assets.

X_4 - Market Value of Equity/Book Value of Total Debt. Equity is measured by the

combined market value of all shares of stock, preferred and common, while debt includes both current and long-term. The measure shows how much the firm's assets can decline in value (measured by market value of equity plus debt) before the liabilities exceed the assets and the firm becomes insolvent.

X_5 - Sales/Total Assets. The capital-turnover ratio is a standard financial ratio illustrating the sales generating ability of the firm's assets. It is one measure of management's capability in dealing with competitive conditions.

CHAPTER 7

DATA ANALYSIS & RESEARCH RESULTS

Three different statistical methods are used to analyze the data in this study, namely, Normal discriminant analysis, Epanechnikov-kernel discriminant analysis and logistic analysis. Evaluation of the overall predictive success of each method is carried out on five versions of the data set. The first data set consists of financial ratios determined for the bankrupt and matching non-bankrupt firms one year prior to the filing of the petition. This will be referred to as data set DS1. DS1 contains 33 paired observations, i.e., 33 bankrupt and 33 non-bankrupt. The second data set (DS2) was determined for the bankrupt and matching non-bankrupt firms from the financial ratios prevailing 2 years prior to the petition. The third through the fifth of the data sets (DS3, DS4, DS5) were similarly determined. To avoid introducing a sampling variance via sampling 33 companies from DS2 through DS5 respectively, the data sets DS2 through DS5 contain data on the paired observation included in DS1, as well as a varying number of paired firms not included in DS1, i.e., using the full size of samples in DS2 through DS5 as the test samples. It would be better if we had enough data to use companies other than the 33 in DS1 as test samples. The purpose of the five data sets was to facilitate the comparison of the performance of each of the three different methods when failure was imminent and when failure was still several years away. Another purpose served by these data sets was comparison of results from this study with some of the

previous results reported in the literature. In this study, the discriminant model will always be estimated using DS1, i.e., using data taken one year prior to bankruptcy. Error rates will be assessed on DS1, and also on DS2 through DS5. The latter will simulate how well the constructed discriminant/classification technique (based on the most recent information prior to bankruptcy) will work on companies whose bankruptcy problem may be further away.

To compare with Altman's findings, only resubstitution is used as chief method to classify bankrupt and non-bankrupt firms with multivariate normal discriminant analysis and Epanechnikov-kernel discriminant analysis. For DS1, error rates are estimated using true resubstitution, and for comparison purpose, cross-validation is also used. Resubstitution can be defined as the sample used to compute the discriminant function would be reused to estimate the error rates. Note that if the sample used to compute the discriminant function is not large, this method gives too optimistic an estimate of the probabilities of misclassification. This optimistic bias can be reduced via cross-validation (no extra data needed). Therefore, an almost unbiased validation test of original sample results by means of cross-validation is also presented in section 7.2.2 and section 7.2.3. Note that even for those firms in DS2 that are included in DS1, the actual financial data on them are different from that used in the DS1 analysis (two years prior, not one). Thus, when DS1-based discriminant/classification methods are used on DS2-DS5, the error rate assessment is a hybrid of resubstitution into the "training sample" (same firm pairs, different data) and substitution into a "test" or "holdout" sample (different firms). This

way of error rates computation i.e., based on pure resubstitution for DS1 and the hybrid resubstitution for DS2-DS5, conforms to the comparison made by Altman in reference. Error count estimate is used as the criteria to compare the efficiency of the three different methods. In section 7.2.2 and section 7.2.3, posterior probability error rate estimate on first year data is also given as reference to error count estimate. The above is summarized in Table 7.1.1.

Table 7.1.1 Design for Data Analysis

Data Set for Error Estimation	Statistical Methods	Classification Methods	Error Rate Estimation
DS1 (original) 33 paired firms	Q. D.A. L. D.A.	Resubstitution Cross-validation	Error Count Posterior Prob.
	Logit Analysis E.. (Equal W.W.) E. (Unequal W.W.)	Resubstitution	Error Count
DS2 DS1 firms- different data, plus 18 additional firms	Q. D.A. L. D.A. Logit Analysis E.. (Equal W.W.) E. (Unequal W.W.)	Hybrid Resubstitution	Error Count
DS3 DS1 firms- different data, plus 28 additional firms	Q. D.A. L. D.A. Logit Analysis E.. (Equal W.W.) E. (Unequal W.W.)	Hybrid Resubstitution	Error Count
DS4 DS1 firms- different data, plus 30 additional firms	Q. D.A. L. D.A. Logit Analysis E.. (Equal W.W.) E. (Unequal W.W.)	Hybrid Resubstitution	Error Count
DS5 DS1 firms- different data, plus 6 additional firms	Q. D.A. L. D.A. Logit Analysis E.. (Equal W.W.) E. (Unequal W.W.)	Hybrid Resubstitution	Error Count

Notation DS data set
 Q.D.A quadratic discriminant analysis
 L.D.A linear discriminant analysis
 E. (Equal W.W.) epanechnikov Kernel with equal window width (h)
 E. (Unequal W.W.) epanechnikov Kernel with unequal window width .

The primary comparison will be based on a matched sample design and an unmatched data analysis. SAS package is used to analyze data in this study. To assess the relative efficiency of the linear vs. quadratic structures when the sample data are not the same as that used to construct the model, i.e. test samples or secondary samples, models based on data from one year prior to bankruptcy are built while using the rest year of data set as holdout test of the original model. This procedure of model selections consistent with Altman's ZETA model (1977).

7.1 Tests Normality and Homogeneity of Covariance

7.1.1 Test of Multivariate Normality

Since the main thrust of this work is to explore methods that work when multivariate normality does not apply, it is first necessary to test for normality in both bankrupt and non-bankrupt groups. A graphical test (Stevens, 1986) is first used to test for multivariate normality of financial ratios in each group. This test is based on the Mahalanobis distances of the subjects in each group, and involves plotting these distances against chi-square percentiles. First, the Mahalanobis distance for each subject j is given by:

$$d_j^2(\mathbf{x}) = (\mathbf{x}_j - \mathbf{m})' \mathbf{S}^{-1} (\mathbf{x}_j - \mathbf{m})$$

where $\mathbf{x}_j$ is the vector of scores for the j th subject, $\mathbf{m}$ is the vector of means for the variables, and $\mathbf{S}$ is the within-group covariance matrix. When the population is

multivariate normal and both N (number of subjects) and $N-p$, where p is the number of financial ratios, are greater than about 25, each of the Mahalanobis distances should behave like a chi-squared variate on one degree of freedom (Johnson & Wichern, 1982). Therefore, we construct for each group a chi-square plot as follows (Stevens, 1986):

1. The distances are ordered from smallest to largest as

$$d_{(1)}^2, d_{(2)}^2, \dots, d_{(N)}^2,$$

2. Then we graph the pairs $(d_{(j)}^2, X_p^2((j-0.5)/N))$, where X_p^2 is the $100(j-0.5)/N$ percentile of the chi-square distribution with p degree of freedom.

The plot for each group should resemble a straight line if the population is multivariate normal. If the multivariate graphical plots show systematic curvature, this is indicative of non-normality, though subjective judgement is required to assess the degree of deviation from normality. Figure 7.1.1 displays the chi-square plot for bankrupt firms taken from data set DS1, i.e., those firms whose financial ratios were determined one year prior to the filing of the petition. Figure 7.1.2 shows the plot on corresponding non-bankrupt firms in data set (DS1). Corresponding plots for bankrupt and non-bankrupt firms whose financial ratios were determined from two through five years prior to the filing of the petition are shown in Appendix B.

Figure 7.1.1 and 7.1.2 both display deviation from a straight line. For both groups, the plots (Figure 7.1.1 and 7.1.2) do deviate from a straight line, but it is difficult to judge the seriousness of deviation. A set of five univariate normality tests have therefore been conducted to determine whether or not the deviations from normality indicated in Figure

7.1.1 and 7.1.2 are statistically significant.

The Shapiro-Wilk statistic, W (available via SAS), was used for the univariate test. The W statistic is the ratio of the best estimator of the variance (based on the square of a linear combination of the order statistics) to the usual corrected sum of squares estimator of the variance. W must be greater than zero and less than or equal to one, with small values of W leading to rejection of the null hypothesis of normality. Note that the distribution of W is highly skewed. Seemingly large values of W (such as 0.90) may be considered small and lead to the rejection of the null hypothesis.

Figure 7.1.3, 7.1.5, 7.1.7, 7.1.9 and 7.1.11 show Shapiro-Wilk test results for all five discriminating variables in bankrupt group respectively, namely, X_1 (working capital/total assets), X_2 (retained earnings/total assets), X_3 (net income/total assets), X_4 (market value equity/book value of total debt), X_5 (sales/total assets). For the bankrupt group, the Shapiro-Wilk test yields very small p-values (< 0.0001) for variables X_2 , X_3 , and X_4 , indicating that these are not normally distributed. Both X_1 ($p=0.075$) and X_5 ($p=0.136$) appear to be normally distributed. Similarly, Figure 7.1.4, 7.1.6, 7.1.8, 7.1.10 and 7.1.12 show Shapiro-Wilk test results for all five discriminating variables in non-bankrupt group respectively. For the non-bankrupt group, the Shapiro-Wilk test yields very small p-values (< 0.0001) for variable X_4 (market value equity/book value of total debt), indicating that X_4 is not normally distributed. Shapiro-Wilk test shows the rest of four variables, namely, X_1 ($p=0.1987$), X_2 ($p=0.1841$), X_3 ($p=0.0894$) and X_5 ($p=0.1726$), are normally distributed.

Since a set of variables are distributed as a multivariate normal if and only if all linear combinations of the set are distributed as univariate normal, the non-normality of the marginal distributions of X_2 , X_3 and X_4 in bankrupt group and the non-normality of X_4 in non-bankrupt group are sufficient to prove that the joint distribution in two separate groups are not multivariate normal. Therefore, it is clear that the data sets are suitable for exploration of non-parametric methods.

For the distribution of the financial ratios in data sets DS2-DS5, the same conclusions were reached as a result of graphical and univariate normality tests as showed in Appendices B.1 to F.10. Appendices B.1 to B.10 display the plots of Mahalanobis distance with Chi-Square on bankrupt and non-bankrupt groups using data sets DS2-DS5. Appendices C.1 to F.10 display the result on test of univariate normality of five variables in bankrupt and non-bankrupt groups using data sets DS2-DS5, including variable group mean, standard deviation, Shapiro-Wilk statistic W , probability value, boxplot and normal probability plot. The bottom and top edges of the box in box plot are located at the sample 25th and 75th percentiles. The center horizontal line is drawn at the sample median and the central plus sign (+) is at the sample mean. The normal probability plot, is a quantile-quantile plot of the data. The empirical quantiles are plotted against the quantiles of a standard normal distribution. Asterisks (*) mark the data values. The plus signs (+) provide a reference straight line that is drawn using the sample mean and standard deviation. If the data are from a normal distribution, the asterisks tend to fall along the reference line.

FIGURE 7.1.1 Plot of Chi-Square With Mahalanobis Distance on Bankrupt Firms Using Data Set DS1

FIGURE 7.1.2 Plot of Chi-Square With Mahalanobis Distance on Non-Bankrupt Firms Using Data Set DS1

Figure 7.1.3 Test of Univariate Normality of X_1 in Bankrupt Group Using DS1

Variable= X_1 Working Capital / Total Assets
 N=33 Mean=0.050726 Std Dev=0.283896
 Shapiro-Wilk statistic (W)=0.9382 Probabiltiy < W=0.0753

Figure 7.1.4 Test of Univariate Normality of X_1 in Non-Bankrupt Group Using DS1

Variable= X_1 Working Capital / Total Assets
 N=33 Mean=0.283053 Std Dev=0.177381
 Shapiro-Wilk statistic (W)=0.95278 Probabiltiy < W=0.1987

Figure 7.1.5 Test of Univariate Normality of X_1 in Bankrupt Group Using DSI

Variable= X_1 Returned Earnings/Total Assets
 N=33 Mean=0.104577 Std Dev=0.139796
 Shapiro-Wilk statistic (W)=0.78932 Probabilitiy < W=0.0001

Figure 7.1.6 Test of Univariate Normality of X_1 in Non-Bankrupt Group Using DSI

Variable= X_1 Returned Earnings/Total Assets
 N=33 Mean=0.353455 Std Dev=0.215449
 Shapiro-Wilk statistic (W)=0.95161 Probabilitiy < W=0.1841

Figure 7.1.7 Test of Univariate Normality of X_1 in Bankrupt Group Using DSI

Variable= X_1 Net Income / Total Assets
 N=33 Mean= -0.09604 Std Dev=0.207242
 Shapiro-Wilk statistic (W)=0.79874 Probabiltiy < W=0.0001

Figure 7.1.8 Test of Univariate Normality of X_1 in Non-Bankrupt Group Using DSI

Variable= X_1 Net Income / Total Assets
 N=33 Mean=0.050746 Std Dev=0.054549
 Shapiro-Wilk statistic (W)=0.94082 Probabiltiy < W=0.0894

Figure 7.1.9 Test of Univariate Normality of X_4 in Bankrupt Group Using DSI

Variable= X_4 Market Value Equity/Book Value of Total Debt
 N=33 Mean=0.113258 Std Dev=0.146633
 Shapiro-Wilk statistic (W)=0.75156 Probability < W=0.0001

Figure 7.1.10 Test of Univariate Normality of X_4 in Non-Bankrupt Group Using DSI

Variable= X_4 Market Value Equity/Book Value of Total Debt
 N=33 Mean=0.187258 Std Dev=0.206049
 Shapiro-Wilk statistic (W)=0.79171 Probability < W=0.0001

Figure 7.1.11 Test of Univariate Normality of X_3 in Bankrupt Group Using DS1

Variable= X_3 Sales / Total Assets
 N=33 Mean=1.395829 Std Dev=0.824081
 Shapiro-Wilk statistic (W)=0.94705 Probabiltiy < W=0.1361

Boxplot

Normal Probability Plot

Figure 7.1.12 Test of Univariate Normality of X_3 in Non-Bankrupt Group Using DS1

Variable= X_3 Sales / Total Assets
 N=33 Mean=1.409197 Std Dev=0.782264
 Shapiro-Wilk statistic (W)=0.95063 Probabiltiy < W=0.1726

Boxplot

Normal Probability Plot

7.1.2 Test of Homogeneity of Covariance

Usually, the Box test, a generalization of the Bartlett univariate homogeneity of variance test, is used to determine whether the group covariance matrices are equal. The test uses the generalized variances, i.e., the determinants of the within group covariance matrices. It is very sensitive to non-normality; thus one may reject with Box test because of a lack of multivariate normality, not because covariance matrices are different. Since the data in the two groups in this study are not distributed according to the multivariate normal, the test results from Box test will not be meaningful and will not be carried out. Though there do exist tests of homogeneity of variance that are robust to non-normality, there is little to be gained from using them here, in view of the different data distribution in the two-groups that is clearly evident from Figure 7.1.1 and 7.1.2. These plots clearly show that pooled estimates of variance are inappropriate. Clearly, separate within-groups estimates of the covariance matrices must be used in the discriminant analysis.

7.2 Normal Discriminant Analysis

The results of the previous section show that normal discriminant analysis, based on the assumption of multivariate normality, is in this case invalid. Nevertheless, normal discriminant analysis will be carried out in order to provide a benchmark from which to compare the performance of the kernel method and of logistic regression analysis, and also to compare with the previous research results of Altman (1968, 1977) and others.

As described earlier, normal discriminant analysis develops a classification criterion based on a measure of generalized square distance, $d_i^2(\mathbf{x}) = (\mathbf{x} - \mathbf{m}_i)' \mathbf{V}_i^{-1} (\mathbf{x} - \mathbf{m}_i)$. The classification criterion is based on either the individual within-group covariance matrices $\mathbf{S}_i$ or on the pooled covariance matrix $\mathbf{S}$. The former results in a quadratic discriminant analysis and the latter results in a linear discriminant analysis. It was shown in the previous section that the within group covariance matrices, $\mathbf{S}_i$, should be used, leading to a quadratic discriminant analysis.

7.2.1 Descriptive Examination of Group Difference

The statistical tests used in this section will, consistent with the discriminant analysis approach, ignore the effect of the pairwise matching. Since the observations within each group can be considered independent, ignoring a possible between group correlation will mean that the tests will be conservative, i.e. they will tend to understate the true difference.

Since the distributions in the two groups are non-normal, the F-test and Wilk's Lambda test are invalid. This does not mean, that these procedures can not be used. A number of authors have examined the performance of these tests under violation of both normality and covariances equality (Mardia, 1971, Hakstian, Roed and Lind, 1979). In order to obtain as clear a picture as possible, the "normal-based" methods were augmented by non-parametric methods.

The F test relates the difference between the average values of the ratios in each group to the variability of values of the ratios within each group. Variable means one

financial statement prior to bankruptcy and the resulting F statistics are presented in Table 7.2.1.1.

Table 7.2.1.1 Variable Means and Test of Significance (Data Set DS1)

Variable	Bankrupt Group Mean (n=33)	Non-Bankrupt Group Mean(n=33)	F-Ratio
X ₁	0.0507	0.2831	2.56*
X ₂	0.1046	0.3536	2.38*
X ₃	-0.0960	0.0575	14.4*
X ₄	0.1133	0.1873	1.97
X ₅	1.3958	1.4092	1.11

* Significant at 0.01 level.

Variables X₁ through X₃ are all significant at the 0.01 level, indicating extremely significant differences in these variables between groups. Variable X₄, X₅ do not show a significant difference between groups and the reason for its inclusion in the variable profile is the volume of sales is proportional to total asset. On a strictly univariate level, all of the ratios indicate higher values for the non-bankrupt firms.

Three nonparametric tests, namely, the Mann-Whitney (Wilcoxon) test, Kolmogorov-Smirnov test and Median test, were used for general information and comparison with the F-test. The nonparametric tests require only that the observations be a random sample and that values can be ordered. These assumptions, especially randomness, are not to be made in this study, but they are less restrictive than those for the two-sample F-test for means. The F-test further requires that the observations be selected from normally distributed populations with equal variances. But, the F-test will

detect true differences between the two populations more often than will the nonparametric test since the F-test uses more information from the data. Therefore, nonparametric tests are generally less powerful than their parametric counterparts when parametric assumptions are valid.

Mann-Whitney test will not work in this study (strictly at least) because it assumes identically shaped distribution, and a difference in mean (or median), namely, it is a test of pure shift. Mann-Whitney ranks all the cases in order of increasing size and computes the test statistic U , the number of times a score from bankrupt group precedes a score from non-bankrupt group. If the samples are from the same population, the distribution of scores from the two groups in the ranked list should be random; an extreme value of U indicates a non-random pattern. The Mann-Whitney test results of variables X_1 to X_5 in data set DS1 are listed in Table 7.2.1.2 below.

Table 7.2.1.2 The Mann-Whitney Test Results of All Five Variables in Data Set DSI

	Mean Rank	Cases	U	W	Corrected Z	2-Tailed P
X_1	Group 1-25.36 Group 2-41.64	33 33	276.0	837.0	-3.443	0.0006*
X_2	Group 1-25.36 Group 2-41.64	33 33	188.0	749.0	-4.605	0.0000*
X_3	Group 1-25.36 Group 2-41.64	33 33	209.0	770.0	-4.303	0.0000*
X_4	Group 1-25.36 Group 2-41.64	33 33	384.5	945.5	-2.052	0.0401
X_5	Group 1-25.36 Group 2-41.64	33 33	276.0	1107.0	-0.019	0.9847

* Significant at 0.01 level.

From Table 7.2.1.2, the observed significance levels for variables X_4 and X_5 are 0.0401 and 0.9847 respectively. Since the significance levels are large, the hypothesis that variable X_4 (Market value of Equity/Book value of total debt) has the same distribution for the bankrupt and non-bankrupt groups and the hypothesis that X_5 (Sales/Total assets) has the same distribution for the bankrupt and non-bankrupt groups are not rejected. Since the significance levels are small for the rest of three variables, namely, X_1 (Working capital/Total assets), X_2 (Retained Earnings/Total assets), X_3 (Net income/total assets), the hypotheses the distributions in bankrupt group of these three variables are the same as their distributions in non-bankrupt group are rejected, namely, X_1 , X_2 and X_3 have

different distributions in bankrupt and non-bankrupt groups.

Similarly, Kolmogorov-Smirnov test computes the observed cumulative distributions for both groups and the maximum positive, negative, and absolute differences. The Kolmogorov-Smirnov Z is then computed along with the two-tailed probability level based on the Smirnov (1948) formula. The one-tailed test can be used to determine whether the values of one group are generally larger than the values of the other group. Table 7.2.1.3 shows K-W test results of variables X_1 to X_5 in data set DS1 in this study.

Table 7.2.1.3 Kolmogorov-Smirnov Test Results of Variables X_1 to X_5 in Data Set DS1

	Absolute	Positive	Negative	K-S Z	2-Tailed P
X_1	0.45455	0.45455	0.00000	1.846	0.002*
X_2	0.54545	0.54545	0.00000	2.216	0.000*
X_3	0.51515	0.51515	0.00000	2.093	0.000*
X_4	0.36364	0.36364	0.00000	1.477	0.025
X_5	0.18182	0.12121	-0.18182	0.739	0.646

* Significant at 0.01 level.

Table 7.2.1.3 shows that variables X_4 and X_5 have the same distributions in bankrupt and non-bankrupt groups while the distributions of variables X_1 , X_2 and X_3 in two groups have significant differences.

The last non-parametric test is Median test. Median test produces a 2x2 contingency table with counts of the number of cases with values greater than the median and less than or equal to the median for the 2 groups. This test may be best, but it is less

powerful. The Median test results of five variables in data set DS1 of this study is present in the following Table 7.2.1.4.

Table 7.2.1.4 The Median Test Results of Variables X_1 to X_5 in Data Set DS1

	Cases	Median	Chi-Square	Significance
X_1	66	0.173	8.7273	0.0031*
X_2	66	0.192	15.5152	0.0001*
X_3	66	0.018	11.8788	0.0006*
X_4	66	0.091	6.0606	0.0138
X_5	66	1.316	0.2424	0.6225

* Significant at 0.01 level.

In Table 7.2.1.4, Median test produces the same results as test results from Mann-Whitney and Kolmogorov-Smirnov test, namely, X_1 , X_2 and X_3 have significantly different distributions in bankrupt and non-bankrupt groups while the rest of three variables have the distributions in bankrupt and non-bankrupt groups from the same population.

A comparison of test results from all four different test methods unravels that four methods produced the same test results which displayed three variables (X_1 , X_2 and X_3) are significant at 0.01 level while variables X_4 and X_5 are not significant at the same level.

Since there is no convenient multivariate generalization to the above test, Wilk's Lambda test was used in this study to test the group differences. Since this test is invalid because the covariance matrices in two group are different. Therefore, this test is used

2 of/de 2

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS
STANDARD REFERENCE MATERIAL 1010a
(ANSI and ISO TEST CHART No. 2)

Table 7.5.1
Summary of Five Year Predictive Accuracy

Prediction Method Interval		Number of Firm Pairs	Corrected Bankruptcy	Classified Non-Bankruptcy	Total Error Rate (Counts)
1	DA (Linear)	33	78.79%	84.85%	18.18%
	DA (Quadratic)	33	72.73%	90.91%	18.18%
	Epanechnikov $h_1 = h_2$	33	93.94%	93.94%	6.06%
	Epanechnikov $h_1 \neq h_2$	33	90.91%	96.97%	6.06%
	Logit	33	75.76%	81.82%	21.21%
2	DA (Linear)	51	82.35%	76.47%	20.59%
	DA (Quadratic)	51	74.51%	82.35%	21.57%
	Epanechnikov $h_1 = h_2$	51	90.20%	82.35%	13.73%
	Epanechnikov $h_1 \neq h_2$	51	88.24%	86.28%	12.74%
	Logit	51	78.43%	70.59%	25.49%
3	DA (Linear)	61	73.77%	68.85%	28.69%
	DA (Quadratic)	61	63.93%	70.49%	32.79%
	Epanechnikov $h_1 = h_2$	61	81.97%	75.41%	21.31%
	Epanechnikov $h_1 \neq h_2$	61	85.25%	75.41%	19.67%
	Logit	61	77.05%	63.93%	21.32%

Number of Observations and Percent Classified (DS1)

		Predicted				Total	
		Bankrupt		Non-Bankrupt			
Observed	Bankrupt	24	72.73%	9	27.27%	33	100%
	Non-Bankrupt	3	9.09%	30	90.91%	33	100%
	Total	27	40.91%	39	59.09%	66	100%
	Priors	0.5000		0.5000			

In Tables 7.2.3.2 and 7.2.3.3, error count estimates and posterior probability error

In the following two tables (table 7.2.2.2 and table 7.2 2.3), error count estimates and posterior probability error rate estimates for bankrupt and non-bankrupt are presented respectively.

Table 7.2.2.2 Error Count Estimates (Data Set DS1)

Estimate	Bankrupt	Non-Bankrupt	Total
Error Rate	0.2727	0.0909	0.1818
Priors	0.5000	0.5000	

Table 7.2.2 2 shows that the model is accurate in classifying 82.82% of the 66 firms in the sample (DS1). Note that the total error rate shown in the table is the sum of the individual error rate (0.2727 and 0.0909), weighted by the assumed values of the priors. Two terms will be used in the following text. The false positive rate is the proportion of firms predicted as bankrupt that were in fact non-bankrupt. The false

negative rate is the proportion of firms predicted as non-bankrupt that were in fact bankrupt. As shown in table 7.2.2.2, the false positive rate proved to be 27.27%, while the false negative rate was much lower at 9.09%.

Table 7.2.2.3 Posterior Probability Error Rate Estimate (Data Set DS1)

Estimate	Bankrupt	Non-Bankrupt	Total
Error Rate	0.1902	0.0472	0.1187
Priors	0.5000	0.5000	

The resubstitution estimate of error rates has been shown to have a large optimistic bias(Hills, 1966; Das Gupta 1974). As noted by Lachenbruch and Michey (1968), one way to reduce these bias is cross-validation (Lachenbruch and Michey 1968). Table 7.2.2.4 through Table 7.2.2.5 show the corresponding empirical results obtained using the cross-validation method.

Table 7.2.2.4 Cross-Validation Summary Using Linear Discriminant Function:
Number of Observations and Percent Classified (DS1)

		Predicted				Total	
		Bankrupt		Non-Bankrupt			
Observed	Bankrupt	22	66.67%	11	33.33%	33	100%
	Non-Bankrupt	3	12.12%	30	87.87%	33	100%
	Total	25	39.39%	41	60.61%	66	100%
	Priors	0.5000		0.5000			

Table 7.2.2.5 Error Count Estimates (Data Set DS1)

Estimate	Bankrupt	Non-Bankrupt	Total
Error Rate	0.3333	0.1212	0.2273
Priors	0.5000	0.5000	

Table 7.2.2.5 shows the model is accurate in classifying 77.27% of the total samples. The false positive rate was 33.33%, when the false negative was at 12.12%. As expected, these are higher than the (biased) estimates produced via resubstitution, as can be seen from a comparison with table 7.2.2.1 and 7.2.2.2.

The results of a simultaneous application of cross-validation and posterior probability error rate estimation is shown in table 7.2.2.6.

Table 7.2.2.6 Posterior Probability Error Rate Estimate (Data Set DS1)

Estimate	Bankrupt	Non-Bankrupt	Total
Error Rate	0.1776	0.057	0.1173
Priors	0.5000	0.5000	

It can be seen that these error rates are virtually identical to those obtained using resubstitution and posterior probability error rate estimation, as shown in Table 7.2.2.3. This is likely to be a factitious result, as the use of posterior probability estimate of error is expected to reduce the variance of the error estimate, but not its optimistic bias.

7.2.3 Empirical Results From Linear Discriminant Function

As stated in section 7.2, when the classification criterion is based on the pooled

covariance matrix S , a linear discriminant function results. Note that the inclusion of the linear discriminant function is for comparison only since the assumptions do not fit this method as demonstrated in section 7.1.1. Table 7.2.3.1 through 7.2.3.6 show results from using resubstitution and cross-validation approaches to assessing classification accuracy, together with error rate estimates computed from error count and posterior probability.

**Table 7.2.3.1 Resubstitution Summary Using Linear Discriminant Function:
Number of Observations and Percent Classified (DS1)**

		Predicted				Total	
		Bankrupt		Non-Bankrupt			
Observed	Bankrupt	26	78.79%	7	21.21%	33	100%
	Non-Bankrupt	5	15.15%	28	84.85%	33	100%
	Total	31	46.97%	35	53.05%	66	100%
	Priors	0.5000		0.5000			

In Tables 7.2.3.2 and 7.2.3.3, error count estimates and posterior probability error rate estimates for bankrupt and non-bankrupt are presented respectively.

Table 7.2.3.2 Error Count Estimates (Data Set DS1)

Estimate	Bankrupt	Non-Bankrupt	Total
Error Rate	0.2121	0.1515	0.1818
Priors	0.5000	0.5000	

Table 7.2.3.2 shows that the model correctly classifies 82.82% of the total samples. The false positive rate is 15.15%, while the false negative error rate was higher at

21.21%.

Table 7.2.3.3 Posterior Probability Error Rate Estimate (Data Set DS1)

Estimate	Bankrupt	Non-Bankrupt	Total
Error Rate	0.2102	0.1410	0.1756
Priors	0.5000	0.5000	

Compared with the results shown from the quadratic method shown in Tables 7.2.2.1 through 7.2.2.3, resubstitution using linear discriminant function resulted in a lower false positive rate, but higher false negative rate, than did the quadratic method though total error count estimate is exactly the same for both (18.18%). The quadratic discriminant function generated a lower posterior probability error rate estimate.

**Table 7.2.3.4 Cross-validation Summary Using Linear Discriminant Function:
Number of Observations and Percent Classified (DS1)**

		Predicted				Total	
		Bankrupt		Non-Bankrupt			
Observed	Bankrupt	26	78.79%	7	21.21%	33	100%
	Non-Bankrupt	3	18.18%	30	81.82%	33	100%
	Total	29	48.48%	37	51.52%	66	100%
	Priors	0.5000		0.5000			

Tables 7.2.3.5 and 7.2.3.6 present error count estimates and posterior probability error rate estimates for bankrupt and non-bankrupt groups using cross-validation and linear

discriminant function.

Table 7.2.3.5 Error Count Estimates (Data Set DS1)

Estimate	Bankrupt	Non-Bankrupt	Total
Error Rate	0.2121	0.1818	0.1970
Priors	0.5000	0.5000	

Table 7.2.3.5 shows the model correctly classifies 80.3% of the total samples. The false positive rate proved to be 18.18%, when the false negative rate was at 21.21%.

Table 7.2.3.6 Posterior Probability Error Rate Estimate (Data Set DS1)

Estimate	Bankrupt	Non-Bankrupt	Total
Error Rate	0.1980	0.1664	0.1822
Priors	0.5000	0.5000	

Compared with the results shown for the quadratic method shown in Table 7.2.2.5 and 7.2.2.6, cross-validation using linear discriminant function resulted in a lower error count rate, but a higher posterior probability error rate estimate, than did the quadratic method.

Tables 7.2.3.3 through 7.2.3.6 indicate resubstitution yields higher estimates of the rate of successful classification than cross-validation. The explanation may be attributed to the known fact that resubstitution exhibits an optimistic bias.

A comparison of results from tables 7.2.2.4 through 7.2.2.6 with results from tables 7.2.3.4 through 7.2.3.6, shows that cross-validation using a linear discriminant function produces lower error count estimate than does cross-validation using quadratic

discriminant function. This is consistent with a statement by Johnson and Wichern (1982) to the effect that , "The quadratic ... rules are appropriate if normality appears to hold but the assumption of equal covariance matrices is seriously violated. However, the assumption of normality seems to be more critical for the quadratic rule than for the linear rule". Huberty (1984) has also stated "The stability of results yielded by a linear rule is greater than results yielded by a quadratic rule when small samples are used and when the normality condition is not met ".

7.3 Results of Epanechnikov Kernel Analysis

As defined in Table 5.2.1, the common kernel methods use uniform, normal, Epanechnikov, biweight and triweight kernels in the density estimation.

Either Mahalanobis distance or Euclidean distance can be used to determine proximity. When a kernel method is used, the Mahalanobis distances are based on either the individual within-group covariance matrices or the pooled covariance matrix.

To summarize, the squared distance between two observation vectors, $\mathbf{x}$ and $\mathbf{y}$, in group t is given by

$$d_t^2(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})' \mathbf{V}_t^{-1} (\mathbf{x} - \mathbf{y})$$

where $\mathbf{V}_t$ has one of the following forms:

$$\mathbf{V}_t = \mathbf{S} \quad \text{the pooled covariance matrix}$$

$$\mathbf{V}_t = \mathbf{S}_t \quad \text{the covariance matrix within group } t$$

The classification of an observation vector $\mathbf{x}$ is based on the estimated group-specific densities from the training set. From these estimated densities, the posterior probabilities of group membership at $\mathbf{x}$ are evaluated. An observation $\mathbf{x}$ is classified into group u if setting $t=u$ produces the largest value of $p(t|\mathbf{x})$. If there is a tie for the largest probability or this largest probability is less than the threshold specified, $\mathbf{x}$ is classified into group OTHER.

The kernel method uses a fixed radius, h , and a specified kernel, K_t , to estimate the group t density at each observation vector $\mathbf{x}$. Let $\mathbf{x}$ be a p -dimensional vector. Then the volume of a p -dimensional unit sphere bounded by $\mathbf{x}'\mathbf{x}=1$ is

$$v_0 = \frac{\pi^{p/2}}{\Gamma(p/2+1)}$$

where Γ represents the gamma function.

Thus, in group t , the volume of a p -dimensional ellipsoid bounded by $\{\mathbf{x} | \mathbf{x}'\mathbf{V}_t^{-1}\mathbf{x}=h^2\}$ is

$$v_h(t) = h^p |\mathbf{V}_t|^{-1/2} v_0 \quad (7.3.1)$$

$v_t(t)$ will be used in calculating the constant $c_t(t)$ in Epanechnikov kernel $K_t(\mathbf{x})$ below.

In this section, non-parametric discriminant analysis using the Epanechnikov kernel approach to density estimation is used. The Epanechnikov approach uses the following as the kernel density in group t .

$$\begin{aligned} K_t(\mathbf{x}) &= c_t(t)(1-\mathbf{x}'\mathbf{V}_t^{-1}\mathbf{x}/h^2) & \text{if } \mathbf{x}'\mathbf{V}_t^{-1}\mathbf{x} < h^2 \\ &= 0 & \text{otherwise} \end{aligned} \quad (7.3.2)$$

where $c_t(t)=(1+p/2)/V_h(t)$, $V_h(t)$ is the value given in equation (7.3.1), p is the dimensions

of the vector of financial ratios, h , is the fixed radius of the kernel and x is the financial ratio vector. In this study, the value of p is 5, h is calculated from the formulation given in Table 5.1.1 by assuming that each group has a multivariate normal distribution with a sample size of 33 (DS1), and by minimizing an approximate mean integrated square error of the estimated density. Note that the assumption of multivariate normality used in calculating h is not critical, but only provides a way of determining the appropriate scale for h . Note that equation 7.3.2 is equivalent of equation 5.2.4. Calculated from table 5.2.1, the value of h (the smoothing parameter) is 1.33. Results for both equal and unequal smoothing parameters will be explained. The use of equal smoothing parameters does not constrain the density estimates to be of equal variance.

Data sets DS2, DS3, DS4 and DS5 are used as test samples for evaluating the efficiency of the derived classification criterion. The parameters of the classification criterion are estimated using DS1, i.e., the data measured one year prior to bankruptcy.

The predictive accuracy of the Epanechnikov kernel method is summarized in Tables 7.3.1 through 7.3.4. Table 7.3.1 below shows the results of resubstitution (for DS1) and hybrid resubstitution (for DS2-DS5) using Epanechnikov kernel density and unequal smoothing parameters.

**Table 7.3.1 Resubstitution Results Using Epanechnikov Kernel Density
with Unequal Smoothing Parameters: Five Years Predictive Accuracy**

Year	Number of Firm Pairs	Correctly Classified Bankruptcy	Correctly Classified Non-Bankruptcy	Total Error Rate (Counts)
1	33	90.91%	96.97%	6.06%
2	51	88.24%	86.28%	12.74%
3	61	85.25%	75.41%	19.67%
4	55	71.43%	76.19%	26.19%
5	35	69.23%	71.80%	29.49%

Table 7.3.2 summarizes the resubstitution results using Epanechnikov kernel density and equal smoothing parameters.

**Table 7.3.2 Resubstitution Results Using Epanechnikov Kernel Density
with Equal Smoothing Parameters: Five Years Predictive Accuracy**

Year	Number of Firm Pairs	Correctly Classified Bankruptcy	Correctly Classified Non-Bankruptcy	Total Error Rate (Counts)
1	33	93.94%	93.94%	6.06%
2	51	90.20%	82.35%	13.73%
3	61	81.97%	75.41%	21.31%
4	55	77.78%	69.84%	26.19%
5	35	71.80%	69.23%	29.49%

The results of Tables 7.3.1 and 7.3.2 indicate that the relative predictive ability of the Epanechnikov model based on data set DS1 decreases when applied successively to data sets DS2 through DS5. i.e., to data sets comprised of measures recorded 2, 3, 4 and 5 years prior to bankruptcy filing. As noted earlier, this can be taken as an indication

of how the efficacy of the method would perform in practice. Impending bankruptcy would be predicted with high probability, with the probability of successful classification decreasing with increasing length of the prediction interval. A comparison of the two tables shows that the model using unequal smoothing parameters exhibits a slightly lower overall error rate in the second and third year prior to filing bankruptcy while the overall error rates in the other years are the same. The difference is too small, however, to provide evidence in favour of using unequal smoothing parameters in general.

7.4 Results Using Logistic Regression Analysis

Central to the consideration of the multiple logistic regression model will be estimation of the coefficients in the model and testing for their significance.

In logistic regression the parameters of the model are estimated using an iterative maximum-likelihood method. The hypotheses test that a coefficient is 0 will be based on the Wald statistic, which has an approximate chi-square distribution in large samples. The sample size in this study is large enough for the chi-squared test to be a good guide. A stepwise model selection was chosen. Variables were entered into and removed from the model in such a way that each forward selection was followed by one or more backward elimination steps. The stepwise selection process terminated if no further variable could be added to the model, or if the variable just entered into the model was the only variable removed in the subsequent backward elimination. The final logistic regression estimation

results for variables in data set DS1 are reported in Table 7.4.1. The final logistic regression estimation results for variables in data sets DS2-DS5 are displayed in Appendix K. Displayed are: maximum likelihood estimates of each parameter; the estimated standard error of each parameter estimate; the Wald chi-squared statistic, computed as the square of the parameter estimate divided by its estimated standard error; the p-value of the Wald statistic referred to a chi-squared distribution with one degree of freedom .

Table 7.4.1 Analysis of Maximum Likelihood Estimates For First Year (DS1)

Variable	Parameter Estimate	Standard Error	Wald Chi-Square	P - Value
Intercept	1.5983	0.5693	7.8813	0.0061
X_2	-6.8186	2.0892	10.652	0.0011
X_3	-13.9374	6.0086	5.3804	0.0204

From the significant variables in Table 7.4.1 (X_2 and X_3), we infer that:

- (i) a larger retained earnings to total assets ratio decreases the probability of bankruptcy;
- (ii) a larger net income to total assets ratio decreases the probability of bankruptcy.

A summary of the classification ability for the final logistic regression models based on data sets DS1-DS5 is given in Table 7.4.2.

Table 7.4.2 Classification Table for Five Years Data Versions

Year	Number of Firm Pairs	Correctly Classified Bankruptcy	Correctly Classified Non-Bankruptcy	Total Error Rate (Counts)
1	33	75.76%	81.82%	21.21%
2	51	78.43%	70.59%	25.49%
3	61	83.60%	73.77%	21.32%
4	55	74.54%	67.73%	28.87%
5	35	77.14%	68.57%	27.15%

Not as expected, Table 7.4.2 indicates that as the length of the predictor interval increases, the predictive ability does not decrease, instead, the third year produces a second best estimation and the performance of fifth year is better than it in fourth year.

7.5 A Summary of the Predictive Accuracy of Three Methods Over Five Year Period

Three different statistical methods have been used to analyze the data in this study, i.e. normal discriminant analysis (linear and quadratic), Epanechnikov kernel discriminant analysis and logit analysis. One of the objectives of this research project have been to find out which of these three methods gives the best prediction of bankruptcy in terms of error count estimate, when applied to a sample of real data which exhibits significant non-normality. The previous results give important evidence of the reliability of the conclusions derived from data obtained one year prior to bankruptcy filing. The overall

effectiveness of the non-parametric discriminant model has also been explored over a range of individuals up to five years prior to bankruptcy, using DS2 through DS5 as test samples. As expected, the relative predictive ability of non-parametric and logistic methods decreases as the length of prediction interval increasing. To properly summarize a comparison of all the methods examined, predictive accuracy from all methods is shown in Table 7.5.1. The results shown are based on error counts using resubstitution for DS1, and hybrid resubstitution for DS2-DS5. The methods listed are normal discriminant analysis (linear and quadratic), logistic regression analysis and the Epanechnikov kernel method with equal and unequal smoothing parameters.

Table 7.5.1
Summary of Five Year Predictive Accuracy

Prediction Interval	Method	Number of Firm Pairs	Corrected Bankruptcy	Classified Non-Bankruptcy	Total Error Rate (Counts)
1	DA (Linear)	33	78.79%	84.85%	18.18%
	DA (Quadratic)	33	72.73%	90.91%	18.18%
	Epanechnikov $h_1 = h_2$	33	93.94%	93.94%	6.06%
	Epanechnikov $h_1 \neq h_2$	33	90.91%	96.97%	6.06%
	Logit	33	75.76%	81.82%	21.21%
2	DA (Linear)	51	82.35%	76.47%	20.59%
	DA (Quadratic)	51	74.51%	82.35%	21.57%
	Epanechnikov $h_1 = h_2$	51	90.20%	82.35%	13.73%
	Epanechnikov $h_1 \neq h_2$	51	88.24%	86.28%	12.74%
	Logit	51	78.43%	70.59%	25.49%
3	DA (Linear)	61	73.77%	68.85%	28.69%
	DA (Quadratic)	61	63.93%	70.49%	32.79%
	Epanechnikov $h_1 = h_2$	61	81.97%	75.41%	21.31%
	Epanechnikov $h_1 \neq h_2$	61	85.25%	75.41%	19.67%
	Logit	61	77.05%	63.93%	21.32%

4	DA (Linear)	55	69.84%	71.43%	29.37%
	DA (Quadratic)	55	55.56%	80.95%	31.75%
	Epanechnikov $h_1 = h_2$	55	77.78%	69.84%	26.19%
	Epanechnikov $h_1 \neq h_2$	55	71.43%	76.19%	26.19%
	Logit	55	69.84%	68.25%	28.87%
5	DA (Linear)	35	66.67%	71.79%	30.77%
	DA (Quadratic)	35	48.72%	76.92%	37.18%
	Epanechnikov $h_1 = h_2$	35	71.80%	69.23%	29.49%
	Epanechnikov $h_1 \neq h_2$	35	69.23%	71.80%	29.49%
	Logit	35	74.36%	56.41%	27.15%

From Table 7.5.1, it can be seen that Epanechnikov model dominates except the last year in which logit model outperforms in a slight edge, approximately consistent with the hypothesis of section 4.2 that the Epanechnikov kernel discriminant analysis will perform better than logit analysis. The logit model performs better than the normal discriminant model for a prediction interval of three years or above. It should be noted that normal discriminant analysis is technically invalid in this case given that the class-conditional probability distribution is not normal. The Epanechnikov method with unequal window width has equal results with equal window width in 3 of 5 years and slightly

lower in other two years, this results in a conclusion that the results from the Epanechnikov method with unequal window width are essentially as good as results from the Epanechnikov method with equal window width. As the length of the prediction interval increases, the predictive ability of all models decreases. The table exhibits an interesting contrast between results for quadratic discriminant analysis and linear discriminant analysis. As noted earlier, classification accuracy is identical for the two discriminant approaches in the first year (data set DS1), with linear showing a slight edge in the bankrupt group and the quadratic in the non-bankrupt group. The most obvious and important differences, however, are in the tests of the bankrupt group. Here, the linear model is clearly superior, with the quadratic misclassifying near fifty percent of bankruptcies five years prior to filing. All of these findings are consistent with results from Altman's ZETA model (1977). Note that if we had estimated the discriminatory models separately for each data set DS1 through DS5, and presented the classification rates obtained by classifying the model estimated for DS3 (say) using the same DS3 data, i.e., by resubstitution, we would have obtained more optimistic results than those showed in table 7.5.1. These results would also have been less realistic, and would not provide any guide as to the usefulness of the classification models for bankruptcy several years in the future.

7.6 A Comparison With Altman's Findings

Table 7.6.1 lists the classification results published in Altman's various research papers, namely, Altman (1968) , results from his 1977 ZETA model, and also results from a model that used variables from the 1968 model together with parameters estimated using the 111-firm ZETA sample (1977). The latter is of interest because it uses the same variables and procedures as were used in this study. Unfortunately, direct comparison of Altman's sample and the sample in this study with ZETA model's sample is impossible since the ZETA 111 sample's data is adjusted along some accounting modifications.

Table 7.6.1

Altman's Classification Results

Year	Method	Number of Firm Pairs	Corrected Bankruptcy	Classified Non-Bankruptcy	Total Error Rates (Counts)
1					
	Altman(1968) (Linear)	66	93.94%	96.97%	4.55%
	Altman's(1977) (Linear)	111	96.2%	89.7%	7.2%
	Altman's(1977) (Quadratic)	111	94.3%	91.4%	7.2%
	1968 Variables With ZETA Parameters	111	92.5%	84.5%	11.5%
2					
	Altman(1968) (Linear)	66	71.9%	93.9%	17.1%
	Altman's(1977) (Linear)	111	84.9%	93.1%	11.0%
	Altman's(1977) (Quadratic)	111	77.4%	91.9%	15.3%

	1968 Variables With ZETA Parameters	111	83.0%	86.2%	15.4%
3	Altman(1968) (Linear)	66	48.3%	n.a.	
	Altman's(1977) (Linear)	111	74.5%	91.4%	16.5%
	Altman's(1977) (Quadratic)	111	62.7%	92.1%	21.1%
	1968 Variables With ZETA Parameters	111	72.7%	89.7%	18.8%
4	Altman(1968) (Linear)	66	28.6%	n.a.	
	Altman's(1977) (Linear)	111	68.1%	89.5%	20.2%
	Altman's(1977) (Quadratic)	111	57.4%	87.8%	26.0%
	1968 Variables With ZETA Parameters	111	57.5%	83.0%	29.75%
5	Altman(1968) (Linear)	66	36.0%	n.a.	
	Altman's(1977) (Linear)	111	69.8%	82.1%	23.2%
	Altman's(1977) (Quadratic)	111	46.5%	87.5%	30.3%
	1968 Variables With ZETA Parameters	111	44.2%	82.1%	36.85%

To make the comparison of tables 7.5.1 and 7.6.1 clearer, a table is given that ranks the models in terms of the predictive ability over the 5 years. Note that direct

comparisons of models are not useful, as the data sets on which the different studies have been conducted are different.

Table 7.6.2 Ranking the Predictive Ability of Models

Rank	Year	1	2	3	4	5
1		Altman (1968)	Altman (L 1977)	Altman (L 1977)	Altman (L 1977)	Altman (L 1977)
2		Epanech. ($h_1 \neq h_2$ & $h_1 = h_2$)	Epanech ($h_1 \neq h_2$)	1968 Var ZETA para	Altman (Q 1977)	Logit
3		Altman (Q&L1977)	Epanech ($h_1 = h_2$)	Epanech ($h_1 \neq h_2$)	Epanech. ($h_1 \neq h_2$ & $h_1 = h_2$)	Epanech. ($h_1 \neq h_2$ & $h_1 = h_2$)
4		1968 Var ZETA para	Altman (Q 1977)	Altman (Q 1977)	Logit	Altman (Q 1977)
5		DA (Q&L)	1968 Var ZETA para	Epanech ($h_1 = h_2$)	DA (L)	DA (L)

Legend:

Q: Quadratic discriminant analysis

L: Linear discriminant analysis

var: Variables

para: Parameters

DA: Discriminant Analysis

$h_1 = h_2$: Equal window width

$h_1 \neq h_2$: Unequal window width

Logit: Logit analysis

Epanech: Epanechnikov kernel method

Table 7.6.2 shows that in the first year, Epanechnikov models perform better than Altman's ZETA linear model (1977). As time is leading, Altman's ZETA linear model performs better. To give a general ranking in terms of time range, table 7.6.2 implicitly shows Epanechnikov models (both pooled within-group covariance and within-group covariance) perform better than Altman's quadratic discriminant model (1977) except in the 4th year and worse than Altman's linear discriminant model (1977) except for first year. The predictive ability of Logit model relatively increases as the lead time increases. Logit model achieves an impressive accuracy in the fifth year, ranked only after Altman's ZETA linear model.

Even though Altman's model (1968) started off (one year prior to bankruptcy) with very low error rates, these rates grow to be very similar to those obtained from the study for the normal discriminant analysis, both linear and quadratic. Except for first year, Altman's (1977) linear model generated quite lower error rates, even lower than the error rates from Epanechnikov method. But, As noted earlier, direct comparison with Altman's ZETA models (1977) is not useful since the ZETA 111 sample's data is adjusted along some accounting modification. Another possible explanation is that the Epanechnikov method is evaluated using a hybrid resubstitution, i.e., a test sample that includes some companies not included in the estimating data set DS1. Thus the results of Table 7.5.1 contain a lower "optimistic bias".

CHAPTER 8

CONCLUSIONS

This chapter reviews the findings of this study in light of certain practical limitations. The chapter is divided into four sections. The first section presents a summary of the overall findings. The second section discusses the value of the study. While the third section focuses on limitations associated with the study, the fourth section maps a direction for future research.

8.1 Summary of the Study

This study examined the accuracy of the predictive ability of three different statistical methodologies (normal discriminant analysis, Epanechnikov nonparametric discriminant analysis and logit analysis) in the context of bankruptcy prediction. The objectives were to examine the performance of models associated with these three methodologies when the predictor variables exhibited significant lack of multivariate normality.

In order to examine the attributes of bankruptcy problems, literature in the area of bankruptcy prediction was reviewed. The attributes of three commonly used bankruptcy predictive models - univariate ratio analysis, normal discriminant analysis, and logit analysis were carefully described. For the empirical part of the study, 33 bankrupt firms

were selected and financial data on these firms was obtained over a five year period prior to bankruptcy filing. Another 33 non-bankrupt firms matched on industry and total assets in the fifth year prior to bankruptcy were extracted from Moody's Industrial. In addition, different numbers of firms have also been added to the list of bankrupted firms from second year to fifth year. Since Altman's model (1968) has become a standard of comparison for subsequent bankrupt classification studies, the five variables similar to the ones employed by Altman (1968) were adopted in this study. The analysis was based on matched design with an unmatched analysis, in keeping with the majority of studies in this field.

8.1.1 Comparison Among Models Associated with Three Methodologies

To answer the first research question, namely, which of normal theory discriminant analysis, kernel method discriminant analysis and logit analysis gives the best prediction of bankruptcy in terms of error count estimate, models based on these three methods were applied to the same data sets. Graphic tests combined with univariate normality tests indicated that the data sets were not according to a multivariate normal distribution. Of the three methods, models based on non-parametric kernel discriminant analysis yielded the lowest error rates in all, but last year in which logit model outperforms non-parametric kernel discriminant analysis in a slightly edge. This finding is almost consistent with our hypothesis. But in the initial two years, normal discriminant models turned out to be superior to the logit model, while from the third year prior to bankruptcy, the logit model outperformed normal discriminant models. That normal discriminant analysis should

outperform logit analysis both one and two years prior to bankruptcy is surprising, given that the assumption behind normal discriminant analysis are violated. Further explanation of this result is warranted.

8.1.2 Ranking Predictive Ability of All Models

The comparison on the performance of models were extend to models with Altman's previous research. Altman's ZETA model with 111 ZETA sample performs best. Although this is only a general comparison, it still poses a need to conduct a kernel analysis based on Altman's ZETA sampling. It is likely that the predictive ability of models constructed in this study will increase after the data samples are modified by Altman's sampling method which was described in literature review. Another interesting finding is that Linear discriminant model performs better than quadratic discriminant models and Epanechnikov kernel method with unequal window width yields slightly better results than the method using equal window width over 5 years. As time is leading, logit model outperforms normal discriminant model, even performs better than Epanechnikov kernel model and Altman's ZETA quadratic model in the last year. Although normal discriminant analysis is invalid when financial variables are not multivariate normal distributed, normal discriminant model still produces a quite useful prediction . This is one of the reason why researchers still use normal discriminant model when its pre-requisite conditions are violated. A question regards to this is raised here, "how much deviation exists as normal discriminant model is applied when normal distribution and equal within-group covariance are not met?".

8.1.3 Issues Related to Matched Design and Matched Analysis

Most research in the area of bankruptcy prediction is based on matched samples. It appears that all such research ignores the effect of the matching when the discriminating techniques are used for the purpose of classifying the likely status of new observations. In this study the question on when a matched design should be used and when a matched analysis needed to be conducted as well as the differences between case-control matching and random matching have been clearly identified and discussed. From a computational point of view, the 1-1 matched design presents no new challenge to logit analysis, as described in section 5.4. The problem is classifying a single new observation based on a logit model estimated from a 1-1 matched design. As can be seen from the description of a matched logit analysis in section 5.4, the constant term in the linear function can not be evaluated. Thus log odds can not be evaluated, whereas, posterior probabilities can not. Since posterior probabilities are required for classifying new observation, further work is required before the results of Hosmer and Lemeshow (1989) on matched analysis can be applied to classification. Once this problem is solved, it may be possible to extend the methodology to the case of normal discriminant analysis and non-parametric kernel based discriminant analysis.

8.2 Value of the Research

This research has explored the use of a nonparametric method to predict corporate

bankruptcy. This research has several values. First, no other study has used the method of nonparametric discriminant analysis on bankruptcy data and it thus comprises a generalization of the corporate bankruptcy prediction problem to the case of non-normal data. More importantly, this study will have value in that its results can be generalized to other fields in finance. Multivariate discriminant analysis has been applied to such financial problems as consumer credit evaluation, investment classification and liquidation and merger. The study will shed some light on research in the above areas. Second, this study compares three different statistical methods in the analysis of corporate bankruptcy with matched case-control design. It is the first paper to clearly described the matching problems associated with bankruptcy prediction problems and to elaborate the methodological problem associated with matching that remains to be solved. Third, this study offers a chance to make comparison of the efficiency of three different statistical methods in bankruptcy prediction.

8.3 Limitations of the Study

It is difficult in the present state of knowledge to give a fair assessment of the usefulness of the nonparametric discrimination techniques. Theoretical properties of these techniques display asymptotic optimality in a certain sense. Loosely speaking, the idea of this theoretical work is that, if large samples are available, consistent estimates of the underlying densities can be obtained and therefore, under very mild conditions, the

eventual behaviour of the nonparametric rules will be as good as if the population densities were actually known. The practical difficulty of this argument is that the sample sizes actually required may be extremely large, especially if the data are of high dimensionality. Second, since the conclusions of this study are based on one real-life sample, there are limitations introduced in the generalization of the results to a real life application. Third, it is not easy to compare the results of this study with those reported by others. First, most studies have not considered data from the seventies and eighties. Second, studies have differed in their selection of predictors. Third, in the primary research, the analysis does not account for the matching used in the selection phase of a case-control study. This could result in a bias in the estimated relative risk of bankruptcy based on financial ratios, but since this bias is present in each of the three methods studied in this thesis, the simplified methodology should not greatly affect the comparison of classical discriminant analysis, logistic regression and non-parametric discrimination. Finally, because computerized data bases were not available due to financial limitation, the matched non-bankrupt firms were not selected as large a pool as has been used in some of the previous studies. This probably accounts for the different distributions observed in the two groups. Better matching would tend to produce more similar (but still non-normal) distribution, and would provide a more controlled examination of the effects of non-normality.

8.4 Directions for Future Studies

The following issues need to be paid attention to in the future.

Scope, time and financial constraints did not allow for the application of the same statistical techniques to a separate analysis of Canadian data. It is technically possible to incorporate both Canadian and American data and to make a comparison of the determinants of corporate bankruptcy in the United States and Canada using the methods described. If differences can be found, an explanation of the attributes of these differences would be of great interest.

As noted earlier, the ZETA sampling of training sample and test sample would greatly increase the accuracy of predictive ability of our models. If this were combined with the selection of matched non-bankrupt firms from a large database, the accuracy, reliability and validity of models constructed in this study might be improved.

Finally, the methodological problems with a matched design and unmatched analysis need to be addressed in a future study. A generic method should be devised for classifying a simple new observation when a matched sample has been analyzed taking into account the matching.

REFERENCES

- Afifi, A.A., and Clark V., "Computer-Aided Multivariate Analysis", Van Nostrand Reinhold Company, New York, 1990.
- Altman, Edward I., 1968, "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy", *Journal of Finance*, 3, 589-609.
- Altman, Edward I., 1971, "Corporate Bankruptcy in America", D.C. Heath and Company, pg. 2-4.
- Altman, Edward I., Robert G. Haldeman and P. Narayanan, "Zeta Analysis, A New Model to Identify Bankruptcy Risk of Corporations", *Journal of Banking and Finance*, 1 (June 1977), 29-54
- Aziz. Abdul, David C. Emanuel and Gerald H. Lawson, "Bankruptcy Prediction-An Investigation of Cash Flow Based Models," *Journal of Management Studies*, 25 (September 1988), 419-37.
- Bathory, A. (1984) "Predicting Corporate Collapse: Credit Analysis in the Determination and Forecasting of Insolvent Companies", *Financial Times Business Information*, London.
- Beaver, William, 1966, "Financial Ratios as Predictions of Failure", *Journal of Accounting Research*, (Empirical Research in Accounting, Selected Studies, Supplement 1966), 71-111.
- Betton, Sandra, 1990, "Bankruptcy Prediction: A Proportional Hazard Approach", ASAC 1990 Conference.
- Blum, Marc, "Failing Company Discriminant Analysis", *Journal of Accounting Research*, 12 (Spring 1974), 1-25
- Breiman, L., Meisel W. and Purcell, E. (1977), "Variable Kernel Estimates of Multivariate Densities", *Technometrics*, 19, 135-144.
- Breslow, N.E. and Day, N.E. (1980). *Statistical Methods in Cancer Research, Vol. 1, The Analysis of Case-Control Studies*. International Agency on Cancer, Lyon, France.
- Casey, C. and M. Bartczak, "Using Operating Cash Flow Data to Predict Financial

Distress: Some Extensions" , Journal of Accounting Research, 23 (Spring 1985), 384-401.

Cencov, N.N. (1962), "Evaluation of An Unknown Distribution Density From Observations", Soviet Mathematics, 3, 1559-1562.

Draper, N.R. and Hunter W.G., 1969, "Transformations: Some Examples Revisited," Technometrics, 11, 23-40.

Economic Report of the President, Transmitted to the U.S. Congress, February 1986.

Edminster, R. O., "An Empirical Test of Financial Ratio Analysis for Small Business Failure Prediction", Journal of Financial and Quantitative Analysis (March 1972), 1477-98.

Efron, B. (1975), "The Efficiency of Logistic Regression Compared to Normal Discriminant Analysis", Journal of the American Statistical Association, 70, 892-898.

Eisenbeis, R. A., "Pitfalls in the Application of Discriminant Analysis in Business and Economics", The Journal of Finance (June 1977), 975-900.

Fix, E. and Hodges, J.L., (1951), "Discriminatory Analysis, Nonparametric Estimation: Consistency Properties". Report No. 4, Project no. 21-49-004, USAF School of Aviation Medicine, Randolph Field, Texas.

Gentry, J.A., P. Newbold and D. Whitford, "Classifying Bankrupt Firms with Fund Flow Components", Journal of Accounting research (Spring 1985), 112-34.

Glick, N. (1978), "Additive Estimators for Probabilities of Correct Classification", Pattern Recognition, 10, 211-222.

Good, I.J. and Gaskins, R.A. (1971), "Nonparametric Roughness Penalties for Probability Densities", Biometrika, 58, 255-277.

Grimm, H, 1980, "Transformation of Variables versus Nonparametrics" 32. Nonparametric Statistical Inference, Budapest (Hungary) 1980.

Hand, D.J. 1981, "Discrimination and Classification", John Wiley & Sons, pg. 16.

- Heinz, Kres, "Statistical Tables for Multivariate Analysis", Springer-Verlag, 1983
- Hora, S.C. and Wilcox, J.B. (1982), "Estimation of Error Rates in Several-Population Discriminant Analysis", *Journal of Marketing Research*, XIX, 57-61.
- Hosmer, D. W. and Lemeshow, S., "Applied Logistic Regression", John Wiley & Sons, Inc. (1989)
- Huberty, C.J., 1984, "Issues in the Use of and Interpretation of Discriminant Analysis", *Psychological Bulletin*, 95
- Ketz, E. J., "The Effect of General Price-Level Adjustments on the Predictive Ability of Financial Ratios", *Journal of Accounting Research (Empirical Research in Accounting, Selected Studies, Supplement, 1978)*, 273-84.
- Lachenbruch, Peter A., "Discriminant Analysis", Hafner Press, 1975.
- Lachenbruch, P.A. and Mickey, M.A. (1968), "Estimation of Error Rates in Discriminant Analysis", *Technometrics*, 10, 1-10.
- Lane, William, Stephen Looney and James W. Wansley, "An Application of the Cox Proportional Hazards Model to Bank Failure", *Journal of Banking and Finance*, 10 (1986), 511-531.
- Lo, Andrew W., 1986, "Logit Versus Discriminant Analysis, A Specification Test and Application to Corporate Bankruptcies", *Journal of Econometrics*, 31, 151-78.
- Loftsgaarden, D.O. and Quesenberry, C.P. (1965), "A Nonparametric Estimate of A Multivariate Density Function", *Annals of Math. Statistics.*, 36, 1040-1951.
- Martin, Daniel, "Early Warning of Bank Failure", *Journal of Banking and Finance*, 1(1977), 249-276.
- McFadden, D., 1976, "A Comment on Discriminant Analysis 'Versus' Logit Analysis", *Annals of Economic and Social Measurement* 5, 511-523
- Mensah, Yaw. M., "The Differential Bankruptcy Predictive Ability of Specific Price-Level Adjustments: Some Empirical Evidence", *the Accounting Review*, 58 (April 1983), 228-46.
- Modigliani, F. and M.H. Miller, "The Cost of Capital, Corporation Finance, and the

Theory of Investment", *The American Economic Review*, (June 1958), 453-487.

Morse, D. and Wayne Shaw. "Investing in Bankrupt Firms", *Journal of Finance*, 43(1988), 1193-1206.

Norton, C. L. and R. E. Smith, "A Comparison of General Price Level & Historical Cost Financial Statements in the Prediction of Bankruptcy", *Accounting Review*, 54 (January 1979), 72-87.

Ohlson, James A. (1980), "Financial Ratios and the Probabilistic Prediction of Bankruptcy", *Journal of Accounting Research* Vol.18 No.1.

Pinches, G.E., K.A. Mingo and J.K. Caruthers (1973), "The Stability of Financial Patterns in Industrial Organizations", *Journal of Finance* (May 1973), 389-396

_____ and A.A. Eubank (1975), "The Hierarchical Classification of Financial Ratios", *Journal of Business Research* (October 1975), 295-310.

Pregibon, D. (1981), "Logistic Regression Diagnostics", *Annals of Statistics*, 9, 705-724.

Press, S.J. and Wilson, S. (1978), "Choosing Between Logistic Regression and Discriminant Analysis", *Journal of the American Statistical Association*, 73, 699-705.

Rose, Peter S. and Gary A. Giroux, "Predicting Corporate Bankruptcy: An Analytical and Empirical Evaluation", *Review of Business and Economic Research* (Spring 1984), 1-12

Rosenblatt, M., "Remarks on Some Nonparametric Estimates of A Density Function", *Annals of Mathematical Statistics*, 27, (1956), 832-837.

Rothman, K.J., 1986, *Modern Epidemiology*, Little Brown, Boston.

Santomero, Anthony M. and Joseph D. Vinso, "Estimating the Probability of Failure for Commercial Banks and the Banking System", *Journal of Banking and Finance*, 1 (1977), 185-205.

SAS/STAT User's Guide, Release 6.03 Edition, SAS Institute Inc.

Schlesselman, J.J. (1982), "Case-Control Studies". Oxford University Press, New York.

- Scott, A.J., and Wild, C.J., 1986, "Fitting Models Under Case-control or Choice Based Sampling", *Journal of the Royal Statistical Association, Series B*, 48, 170-182.
- Scott, James, 1976, "A Theory of Optimal Capital Structure", *Journal of Economic*, Spring.
- Scott, James, 1981, "The Probability of Bankruptcy", *Journal of Banking and Finance* 5, 317-344.
- Sheppard, J. Paul, 1990, "Bankruptcy Prediction: the Effects of Firm Financial Condition, Size and Industry Environment", ASAC 1990 Conference.
- Silverman, B. W., "Density Estimation for Statistics and Data Analysis", Chapman and Hall.
- Smirnov, N. V., 1948, "Table for Estimating the Goodness of Fit of Empirical Distributions", *Annals of Mathematical Statistics*, 19, 279-281.
- Smith, V. (1982), "Default Risk, Scale, and the Home-made Leverage Theorem" , *American Economic review*, (February, 1972), 66-76.
- Snapinn, S. M. and Knoke, J.D. (1985), "An Evaluation of Smoothed Classification Error-Rate Estimators", *Technometrics*, Vol 27, No. 2.
- Stevens James, "Applied Multivariate Statistic for the Social Science", Lawrence Erlbaum Associates, Publishers.
- Stiglitz, J.E. (1972), "Some Aspects of the Pure Theory of Corporate Finance: Bankruptcies and Take-Overs", *The Bell Journal of Economics and Management Science* (Autumn, 1972), 458-482.
- Tinsley, P.A. (1970), "Capital Structure, Precautionary Balances, and Valuation of the Firm: the Problem of Financial and Quantitative Analysis", (March, 1970), 33-62.
- Wertz, W. (1978), "Statistical Density Estimation: A Survey", Göttingen: Vandenhoeck and Ruprecht.
- West, Robert Craig, "A Factor-Analytic Approach to Bank Condition", *Journal of Banking and Finance*, 9 (1985), 253-266.
- Weston, J. Fred and Eugene Brigham, 1966, "Managerial Finance", 2nd Edition. New

York: Holt, Rinehart & Winston.

Wilcox, J. W., "A Prediction of Business Failure Using Accounting Data", *Journal of Accounting Research* (Empirical research in Accounting. Selected Studies, Supplement to Volume 2, 1973), 463-79.

Wittle, P., (1958), "On the Smoothing of Probability Density Functions", *Journal of Royal Statistic Society B*, 20, 334-343.

Zavgren, Christine V., "Assessing the Vulnerability of Failure of American Industrial Firms: A Logistic Analysis", *Journal of Business and Accounting*, 12 (Spring 1985), 19-45.

Appendix A The List of Bankrupt and Non-Bankrupt Firms

TYPE	IND_NO	COMPANY_NO	COMP_NAME	BKTCY_DATE	YEAR	COUNTRY
1	1382		Mesa Petroleum Company	1986	1985	us
2	1382		Texas Oil & Gas Corp			us
1	1382		Freeport-Memoran Energy Part	1989	1988	us
2	1382		The Penn Central Corporation			us
1	1382		Petroleum Investments, Ltd.	1990	1989	us
2	1382		Hershey Oil Corp			us
1	1383		Verna Corp	1986	1985	us
2	1383		Damson Oil Corp			us
1	1384		Belden & Blake Energe Co.	1991	1990	us
2	1384		Barnwell Industries, Inc			us
1	1389		Mesa Limited Partnership	1991	1990	us
2	1389		Pennzoil Company			us
1	2200	266057	Duplan	1976	1974	us
1	2200	524030	Lehigh Valley Railrd	1977	1976	us
2	2200	12347	Albany Intl. Corp.			us
2	2200	88665	Bibb Company			us
1	2270	817460	Sequoyah Industries	1974	1973	us
2	2270	574803	Masland(O.H.) & Sons			us
1	2300	101241	Botany Industries	1973	1971	us
2	2300	95293	Blue Bell			us
1	2300	269461	Eagle Clothes	1978	1976	us
2	2300	471016	Jantzen Inc.			us
1	2300		Garland Corp	1980	1979	us
2	2300		Movie Star			us
1	2300		CS Group	1982	1980	us
2	2300		Wolf (Howard B)			us
1	2450	959078	Western Orbis	1976	1975	us
1	2450	202651	Commodore	1976	1975	us
2	2450	731588	Poloron Products Inc.			us
2	2450	206039	Conchemco Inc.			us
1	2510		Brody (B) Seating Co.	1981	1978	us
2	2510		Jensen Industries			us
1	2521		GF Corp	1990	1989	us
2	2521		La-Z-Boy Chair Co.			us
1	2600		Saxon Industries	1982	1980	us
2	2600		Southwest Forest Industries			us
1	2650	703181	Paterson Parchment	1977	1973	us
2	2650	828658	Simkins Industries			us
1	2732		Sorg Incorporated	1989	1987	us
2	2732		Harland (John H.) Co			us
1	3069	763121	Richardson Co.	1975	1974	us
2	3069	24591	American Biltrite Inc.			us
1	3140	26367	American Girl Fashions, Inc	1975	1973	us
2	3140	498552	Kleinert's Inc.			us
1	3140	358604	Frier Industries	1978	1974	us
2	3140	864261	Suave Shoe Corp			us
1	3250	233048	DCA Development	1973	1971	us
2	3250	826750	Sikes corp.			us
1	3270	577438	Maule Industries	1977	1974	us
2	3270	882451	Texas Industries Inc.			us
1	3310	880447	Tennessee Forging Steel Corp.	1977	1976	us
2	3310		MMA Co.			us
1	3310	11608	Alan Wood Steel	1978	1975	us
2	3310	549866	Lukens Steel Co.			us
1	3316		Wheeling-Pittsburgh Steel Corp	1990	1989	us
2	3316		Morrison Knudsen Corporation			us
1	3341	829880	Sitkin Smelting	1978	1977	us
2	3341	743396	Proler International Corp			us
1	3350	755400	Reading Industries, Inc.	1974	1973	us

APPENDIX B.7 PLOT OF MAHALANOBIS DISTANCE WITH CHI-SQUARE ON BANKRUPT GROUP USING DATA SET DS5

121

APPENDIX B.8 PLOT OF MAHALANOBIS DISTANCE WITH CHI-SQUARE ON NON-BANKRUPT GROUP USING DATA SET DS5

TYPE	IND_NO	COMPANY_NO	COMP_NAME	BKTCY_DATE	YEAR	COUNTRY
1	4210		Cooper-Jarrett, Inc.	1982	1980	us
2	4210		Eazor Express, Inc.			us
1	4400		Pacific Far East Line, Inc.	1978	1976	us
2	4400		Overseas Shipholding Group			us
1	4700		Shulman Transport Enterprise	1978	1977	us
2	4700		Dereco Inc.			us
1	4832		SFN Companies, Inc	1986	1985	us
2	4832		Affiliated Publications, Inc			us
1	4911		Myers (L.E.) Co and Intl. Ltd	1982	1981	us
2	4911		Salem Corp			us
1	5099		Fireco Sales	1982	1979	us
2	5099		Ronco Teleproducts, Inc.			us
1	5112		Marathon Office Supply Inc	1989	1988	us
2	5112		Servicemaster Limited Partner			us
1	8059		Anta Corp	1984	1983	us
2	8059		Berger Brunswick Corp			us
1	8091		Western Health Plans, Inc	1989	1988	us
2	8091		Angelica Corp			us

Label:

Type

1 "Bankrupt Firm"

2 "Non-Bankrupt Firm"

IND_NO

"Company SIC Code"

COMPANY_NO

"Company Number"

COMP_NAME

"The Name of the Firm"

BKTCY_DATE

"The Year of Bankruptcy Filing"

YEAR

"The Last Period which Financial Data Was Published"

COUNTRY

"The Country Origin of the Firm"

APPENDIX B.1 PLOT OF MAHALANOBIS DISTANCE WITH CHI-SQUARE ON BANKRUPT GROUP USING DATA SET DS2

118

APPENDIX B.2 PLOT OF MAHALANOBIS DISTANCE WITH CHI-SQUARE ON NON-BANKRUPT GROUP USING DATA SET DS2

APPENDIX B.3 PLOT OF MAHALANOBIS DISTANCE WITH CHI-SQUARE ON BANKRUPT GROUP USING DATA SET DS3

119

APPENDIX B.4 PLOT OF MAHALANOBIS DISTANCE WITH CHI-SQUARE ON NON-BANKRUPT GROUP USING DATA SET DS3

APPENDIX B.5 PLOT OF MAHALANOBIS DISTANCE WITH CHI-SQUARE ON BANKRUPT GROUP USING DATA SET DS4

APPENDIX B.6 PLOT OF MAHALANOBIS DISTANCE WITH CHI-SQUARE ON NON-BANKRUPT GROUP USING DATA SET DS4

APPENDIX B.7 PLOT OF MAHALANOBIS DISTANCE WITH CHI-SQUARE ON BANKRUPT GROUP USING DATA SET DS5

121

APPENDIX B.8 PLOT OF MAHALANOBIS DISTANCE WITH CHI-SQUARE ON NON-BANKRUPT GROUP USING DATA SET DS5

APPENDIX C.1

Test of Univariate Normality of X_1 in Bankrupt Group Using DS2

Variable= X_1 Working Capital / Total Assets
 N=51 Mean=0.135466 Std Dev=0.244217
 Shapiro-Wilk statistic (W)=0.94574 Probability < W=0.0346

APPENDIX C.2

Test of Univariate Normality of X_1 in Non-Bankrupt Group Using DS2

Variable= X_1 Working Capital / Total Assets
 N=51 Mean=0.285999 Std Dev=0.214439
 Shapiro-Wilk statistic (W)=0.96982 Probability < W=0.3631

APPENDIX C.3

Test of Univariate Normality of X_1 in Bankrupt Group Using DS2

Variable= X_1 Returned Earnings/Total Assets
 N=51 Mean=0.109727 Std Dev=0.160615
 Shapiro-Wilk statistic (W)=0.84294 Probability < W=0.0001

Boxplot

Normal Probability Plot

APPENDIX C.4

Test of Univariate Normality of X_1 in Non-Bankrupt Group Using DS2

Variable= X_1 Returned Earnings/Total Assets
 N=51 Mean=0.316482 Std Dev=0.20094
 Shapiro-Wilk statistic (W)=0.95435 Probability < W=0.0856

Boxplot

Normal Probability Plot

APPENDIX C.5

Test of Univariate Normality of X_1 in Bankrupt Group Using DS2

Variable= X_1 Net Income / Total Assets
 N=51 Mean=-0.5867 Std Dev=0.195673
 Shapiro-Wilk statistic (W)=0.89500 Probability < W=0.0001

APPENDIX C.6

Test of Univariate Normality of X_1 in Non-Bankrupt Group Using DS2

Variable= X_1 Net Income / Total Assets
 N=51 Mean=0.035864 Std Dev=0.08737
 Shapiro-Wilk statistic (W)=0.88154 Probability < W=0.0001

APPENDIX C.7

Test of Univariate Normality of X_1 in Bankrupt Group Using DS2

Variable= X_1 Market Value Equity/Book Value of Total Debt
 N=51 Mean=0.095529 Std Dev=0.116464
 Shapiro-Wilk statistic (W)=0.75183 Probability < W=0.0001

APPENDIX C.8

Test of Univariate Normality of X_1 in Non-Bankrupt Group Using DS2

Variable= X_1 Market Value Equity/Book Value of Total Debt
 N=51 Mean=0.188637 Std Dev=0.209428
 Shapiro-Wilk statistic (W)=0.78248 Probability < W=0.0001

APPENDIX C.9

Test of Univariate Normality of X_i in Bankrupt Group Using DS2

Variable= X_i Sales / Total Assets
 N=51 Mean=1.380611 Std Dev=0.78537
 Shapiro-Wilk statistic (W)=0.95476 Probability < W=0.0893

APPENDIX C.10

Test of Univariate Normality of X_i in Non-Bankrupt Group Using DS2

Variable= X_i Sales / Total Assets
 N=51 Mean=1.41342 Std Dev=0.746021
 Shapiro-Wilk statistic (W)=0.93629 Probability < W=0.0124

Boxplot

APPENDIX D.1

Test of Univariate Normality of X_i in Bankrupt Group Using DS3Variable= X_i

Working Capital / Total Assets

N=61

Mean=0.190494

Std Dev=0.197515

Shapiro-Wilk statistic (W)=0.97132 Probability < W=0.3381

Boxplot

Normal Probability Plot

APPENDIX D.2

Test of Univariate Normality of X_i in Non-Bankrupt Group Using DS3Variable= X_i

Working Capital / Total Assets

N=61

Mean=0.315668

Std Dev=0.199681

Shapiro-Wilk statistic (W)=0.94832 Probability < W=0.0238

Boxplot

Normal Probability Plot

APPENDIX D.3

Test of Univariate Normality of X_1 in Bankrupt Group Using DS3

Variable= X_1 Returned Earnings/Total Assets
 N=61 Mean=0.114439 Std Dev=0.18219
 Shapiro-Wilk statistic (W)=0.89756 Probability < W=0.0001

APPENDIX D.4

Test of Univariate Normality of X_1 in Non-Bankrupt Group Using DS3

Variable= X_1 Returned Earnings/Total Assets
 N=61 Mean=0.301461 Std Dev=0.199828
 Shapiro-Wilk statistic (W)=0.97445 Probability < W=0.4468

APPENDIX D.5

Test of Univariate Normality of X_1 in Bankrupt Group Using DS3

Variable= X_1 Net Income / Total Assets
 N=61 Mean= -0.04077 Std Dev=0.129237
 Shapiro-Wilk statistic (W)=0.85843 Probability < W=0.0001

Box plot

Normal Probability Plot

APPENDIX D.6

Test of Univariate Normality of X_1 in Non-Bankrupt Group Using DS3

Variable= X_1 Net Income / Total Assets
 N=61 Mean=0.051394 Std Dev=0.101881
 Shapiro-Wilk statistic (W)=0.86089 Probability < W=0.0001

Box plot

Normal Probability Plot

APPENDIX D.7

Test of Univariate Normality of X_1 in Bankrupt Group Using DS3Variable= X_1

Market Value Equity/Book Value of Total Debt

N=61

Mean=0.107346

Std Dev=0.119104

Shapiro-Wilk statistic (W)=0.79595 Probability < W=0.0001

APPENDIX D.8

Test of Univariate Normality of X_1 in Non-Bankrupt Group Using DS3Variable= X_1

Market Value Equity/Book Value of Total Debt

N=61

Mean=0.189174

Std Dev=0.196629

Shapiro-Wilk statistic (W)=0.81457 Probability < W=0.0001

APPENDIX D.9

Test of Univariate Normality of X_1 in Bankrupt Group Using DS3

Variable= X_1 Sales / Total Assets
 N=61 Mean=1.3275 Std Dev=0.73539
 Shapiro-Wilk statistic (W)=0.95575 Probability < W=0.0606

Box plot

Normal Probability Plot

APPENDIX D.10

Test of Univariate Normality of X_1 in Non-Bankrupt Group Using DS3

Variable= X_1 Sales / Total Assets
 N=61 Mean=1.344044 Std Dev=0.658901
 Shapiro-Wilk statistic (W)=0.97089 Probability < W=0.3247

Box plot

Normal Probability Plot

APPENDIX E.1

Test of Univariate Normality of X_1 in Bankrupt Group Using DS4

Variable= X_1 Working Capital / Total Assets
 N=55 Mean=0.220698 Std Dev=0.175711
 Shapiro-Wilk statistic (W)=0.95503 Probability < W=0.0749

APPENDIX E.2

Test of Univariate Normality of X_1 in Non-Bankrupt Group Using DS4

Variable= X_1 Working Capital / Total Assets
 N=55 Mean=0.320023 Std Dev=0.97075
 Shapiro-Wilk statistic (W)=0.96611 Probability < W=0.2366

APPENDIX E.3

Test of Univariate Normality of X_1 in Bankrupt Group Using DS4

Variable= X_1 Returned Earnings/Total Assets
 N=55 Mean=0.142545 Std Dev=0.228435
 Shapiro-Wilk statistic (W)=0.82231 Probability < W=0.0001

Box plot

APPENDIX E.4

Test of Univariate Normality of X_1 in Non-Bankrupt Group Using DS4

Variable= X_1 Returned Earnings/Total Assets
 N=55 Mean=0.31583 Std Dev=0.18315
 Shapiro-Wilk statistic (W)=0.96189 Probability < W=0.1557

Box plot

APPENDIX E.5

Test of Univariate Normality of X_j in Bankrupt Group Using DS4

Variable= X_j Net Income Total Assets
 N=55 Mean= -0.00121 Std Dev=0.110097
 Shapiro-Wilk statistic (W)=0.88623 Probability < W=0.0001

APPENDIX E.6

Test of Univariate Normality of X_j in Non-Bankrupt Group Using DS4

Variable= X_j Net Income / Total Assets
 N=55 Mean=0.059708 Std Dev=0.073633
 Shapiro-Wilk statistic (W)=0.87547 Probability < W=0.0894

APPENDIX E.7

Test of Univariate Normality of X_1 in Bankrupt Group Using DS4Variable= X_1

Market Value Equity/Book Value of Total Debt

N=55

Mean=0.109442

Std Dev=0.132038

Shapiro-Wilk statistic (W)=0.75558 Probability < W=0.0001

Normal Probability Plot

APPENDIX E.8

Test of Univariate Normality of X_1 in Non-Bankrupt Group Using DS4Variable= X_1

Market Value Equity/Book Value of Total Debt

N=55

Mean=0.165656

Std Dev=0.196557

Shapiro-Wilk statistic (W)=0.76156 Probability < W=0.0001

Normal Probability Plot

APPENDIX E.9

Test of Univariate Normality of X_1 in Bankrupt Group Using DS4

Variable= X_1 Sales / Total Assets
 N=55 Mean=1.343138 Std Dev=0.687697
 Shapiro-Wilk statistic (W)=0.95773 Probability < W=0.1005

Boxplot

Normal Probability Plot

APPENDIX E.10

Test of Univariate Normality of X_1 in Non-Bankrupt Group Using DS4

Variable= X_1 Sales / Total Assets
 N=55 Mean=1.4336 Std Dev=0.773194
 Shapiro-Wilk statistic (W)=0.94702 Probability < W=0.0324

Boxplot

Normal Probability Plot

APPENDIX F.1

Test of Univariate Normality of X_i in Bankrupt Group Using DS5

Variable= X_i Working Capital / Total Assets
 N=35 Mean=0.215793 Std Dev=0.210153
 Shapiro-Wilk statistic (W)=0.93209 Probability < W=0.0404

Pos plot

APPENDIX F.2

Test of Univariate Normality of X_i in Non-Bankrupt Group Using DS5

Variable= X_i Working Capital / Total Assets
 N=35 Mean=0.344935 Std Dev=0.215778
 Shapiro-Wilk statistic (W)=0.97124 Probability < W=0.5597

Pos plot

APPENDIX F.3

Test of Univariate Normality of X_1 in Bankrupt Group Using DS5

Variable= X_1 Returned Earnings/Total Assets
 N=35 Mean=0.132811 Std Dev=0.167407
 Shapiro-Wilk statistic (W)=0.89438 Probability < W=0.0026

Boxplot

Normal Probability Plot

APPENDIX F.4

Test of Univariate Normality of X_1 in Non-Bankrupt Group Using DS5

Variable= X_1 Returned Earnings/Total Assets
 N=35 Mean=0.328962 Std Dev=0.206161
 Shapiro-Wilk statistic (W)=0.96294 Probability < W=0.3481

Boxplot

Normal Probability Plot

APPENDIX F.5

Test of Univariate Normality of X_1 in Bankrupt Group Using DS5

Variable= X_1 Net Income / Total Assets
 N=35 Mean= 0.001792 Std Dev=0.183923
 Shapiro-Wilk statistic (W)=0.7816 Probabilitiy < W=0.0001

Box plot

Normal Probability Plot

APPENDIX F.6

Test of Univariate Normality of X_1 in Non-Bankrupt Group Using DS5

Variable= X_1 Net Income / Total Assets
 N=35 Mean=0.043763 Std Dev=0.136286
 Shapiro-Wilk statistic (W)=0.57208 Probabilitiy < W=0.0001

Box plot

Normal Probability Plot

APPENDIX F.7

Test of Univariate Normality of X_i in Bankrupt Group Using DS5

Variable= X_i Market Value Equity/Book Value of Total Debt
 N=35 Mean=0.105167 Std Dev=0.130255
 Shapiro-Wilk statistic (W)=0.72730 Probability < W=0.0001

Box plot

Normal Probability Plot

APPENDIX F.8

Test of Univariate Normality of X_i in Non-Bankrupt Group Using DS5

Variable= X_i Market Value Equity/Book Value of Total Debt
 N=35 Mean=0.139041 Std Dev=0.161638
 Shapiro-Wilk statistic (W)=0.76905 Probability < W=0.0001

Box plot

Normal Probability Plot

APPENDIX F.9

Test of Univariate Normality of X_1 in Bankrupt Group Using DS5

Variable= X_1 Sales / Total Assets
 N=35 Mean=1.346412 Std Dev=0.747958
 Shapiro-Wilk statistic (W)=0.91761 Probability < W=0.0139

Boxplot

APPENDIX F.10

Test of Univariate Normality of X_1 in Non-Bankrupt Group Using DS5

Variable= X_1 Sales / Total Assets
 N=35 Mean=1.294175 Std Dev=0.678776
 Shapiro-Wilk statistic (W)=0.97254 Probability < W=0.5975

Boxplot

APPENDIX G

The Results of Five Tests on Variables X_1 to X_5 in Data Sets DS2

Appendix G.1 Variable Means and Test of Significance (Data Set DS2)

Variable	Bankrupt Group Mean (n=33)	Non-Bankrupt Group Mean(n=33)	F-Ratio
X_1	0.1355	0.2860	10.94*
X_2	0.1097	0.3165	32.95*
X_3	-0.0587	0.0359	9.92*
X_4	0.0955	0.1886	7.70*
X_5	1.3806	1.4134	0.047

* Significant at 0.01 level.

Appendix G.2 The Mann-Whitney Test Results of All Five Variables in Data Set DS2

	Mean Rank	Cases	U	W	Corrected Z	2-Tailed P
X_1	Group 1-42.49 Group 2-60.51	51 51	841.0	2167.0	-3.075	0.0021*
X_2	Group 1-36.10 Group 2-66.90	51 51	515.0	1841.0	-5.292	0.0000*
X_3	Group 1-40.63 Group 2-62.37	51 51	746.0	2072.0	-3.711	0.0002*
X_4	Group 1-25.36 Group 2-41.64	51 51	867.5	2193.5	-2.898	0.0038*
X_5	Group 1-51.20 Group 2-51.80	33 33	1285.0	2611.0	-0.104	0.9174

* Significant at 0.01 level.

Appendix G.3 Kolmogorov-Smirnov Test Results of Variables X_1 to X_5 in Data Set DS2

	Absolute	Positive	Negative	K-S Z	2-Tailed P
X_1	0.29412	0.29412	0.00000	1.485	0.024
X_2	0.56863	0.56863	0.00000	2.871	0.000*
X_3	0.41176	0.41175	-0.0392	2.079	0.000*
X_4	0.31373	0.31373	0.00000	1.584	0.013
X_5	0.09804	0.09804	-0.09804	0.495	0.967

* Significant at 0.01 level.

Appendix G.4 The Median Test Results of Variables X_1 to X_5 in Data Set DS2

	Cases	Median	Chi-Square	Significance
X_1	102	0.237	3.9216	0.0477
X_2	102	0.173	30.7451	0.0000*
X_3	102	0.034	10.0392	0.0015*
X_4	102	0.089	7.6863	0.0056*
X_5	102	1.328	0.0000	1.0000

* Significant at 0.01 level.

Appendix G.5 Multivariate Statistics and Exact F Statistics (Data Set DS2)

S=1 M=1.5 N=47

Statistics	Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	0.702741	8.1216	5	96	0.0001

APPENDIX H

The Results of Five Tests on Variables X_1 to X_5 in Data Sets DS3

Appendix H.1 Variable Means and Test of Significance (Data Set DS3)

Variable	Bankrupt Group Mean (n=33)	Non-Bankrupt Group Mean(n=33)	F-Ratio
X_1	0.1905	0.3157	12.12*
X_2	0.1144	0.3017	29.18*
X_3	-0.0477	0.0514	19.13*
X_4	0.1074	0.1892	7.729*
X_5	1.3275	1.3440	0.017

* Significant at 0.01 level.

Appendix H.2 The Mann-Whitney Test Results of All Five Variables in Data Set DS3

	Mean Rank	Cases	U	W	Corrected Z	2-Tailed P
X_1	Group 1-50.97 Group 2-72.03	61 61	1218.0	3109.0	-3.290	0.0010*
X_2	Group 1-44.58 Group 2-78.42	61 61	828.5	2719.5	-5.305	0.0000*
X_3	Group 1-45.69 Group 2-77.31	61 61	896.0	2787.0	-4.939	0.0000*
X_4	Group 1-53.10 Group 2-69.90	61 61	1348.0	3239.0	-2.624	0.0087*
X_5	Group 1-60.93 Group 2-62.07	61 61	1826.0	3717.0	-0.177	0.8598

* Significant at 0.01 level.

Appendix H.3 Kolmogorov-Smirnov Test Results of Variables X_1 to X_5 in Data Set DS3

	Absolute	Positive	Negative	K-S Z	2-Tailed P
X_1	0.32787	0.32787	0.00000	1.811	0.003*
X_2	0.55738	0.55738	-0.01639	3.078	0.000*
X_3	0.44262	0.44262	0.00000	2.444	0.000*
X_4	0.22951	0.22951	0.00000	1.268	0.080
X_5	0.11475	0.11475	-0.09836	0.634	0.817

* Significant at 0.01 level.

Appendix H.4 The Median Test Results of Variables X_1 to X_5 in Data Set DS3

	Cases	Median	Chi-Square	Significance
X_1	122	0.289	3.2787	0.0702
X_2	122	0.173	29.5082	0.0000*
X_3	122	0.030	18.8852	0.0000*
X_4	122	0.088	3.2787	0.0702
X_5	122	1.274	0.1311	0.7172

* Significant at 0.01 level.

Appendix H.5 Multivariate Statistics and Exact F Statistics (Data Set DS3)

S=1 M=1.5 N=57

Statistics	Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	0.73876	8.2039	5	116	0.0001

APPENDIX I

The Results of Five Tests on Variables X_1 to X_5 in Data Sets DS4

Appendix I.1 Variable Means and Test of Significance (Data Set DS4)

Variable	Bankrupt Group Mean (n=33)	Non-Bankrupt Group Mean(n=33)	F-Ratio
X_1	0.2210	0.3202	7.783*
X_2	0.1426	0.3158	19.27*
X_3	-0.0012	0.0597	11.64*
X_4	0.1094	0.1657	3.099
X_5	1.3431	1.4336	0.420

* Significant at 0.01 level.

Appendix I.2 The Mann-Whitney Test Results of All Five Variables in Data Set DS4

	Mean Rank	Cases	U	W	Corrected Z	2-Tailed P
X_1	Group 1-48.00 Group 2-63.00	55 55	1100.0	2640.0	-2.466	0.0137
X_2	Group 1-40.42 Group 2-70.58	55 55	683.0	2223.0	-4.972	0.0000*
X_3	Group 1-42.25 Group 2-68.75	55 55	784.0	2324.0	-4.355	0.0000*
X_4	Group 1-50.36 Group 2-60.64	55 55	1230.0	2770.0	-1.689	0.0912
X_5	Group 1-53.70 Group 2-57.30	55 55	1413.5	2953.5	-0.592	0.5539

* Significant at 0.01 level.

Appendix 1.3 Kolmogorov-Smirnov Test Results of Variables X_1 to X_5 in Data Set DS4

	Absolute	Positive	Negative	K-S Z	2-Tailed P
X_1	0.27273	0.27273	0.00000	1.430	0.033
X_2	0.52727	0.52727	-0.05455	2.765	0.000*
X_3	0.47273	0.47273	-0.01818	2.479	0.000*
X_4	0.20000	0.20000	0.00000	1.049	0.221
X_5	0.12727	0.12727	-0.05455	0.667	0.765

* Significant at 0.01 level.

Appendix 1.4 The Median Test Results of Variables X_1 to X_5 in Data Set DS4

	Cases	Median	Chi-Square	Significance
X_1	110	0.260	0.5818	0.4456
X_2	110	0.204	24.5818	0.0000*
X_3	110	0.042	14.5455	0.0001*
X_4	110	0.073	1.3091	0.2526
X_5	110	1.308	0.5818	0.4456

* Significant at 0.01 level.

Appendix 1.5 Multivariate Statistics and Exact F Statistics (Data Set DS4)

S=1 M=1.5 N=51

Statistics	W-Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	0.80060	5.1804	5	104	0.0003

APPENDIX J

The Results of Five Tests on Variables X_1 to X_5 in Data Sets DS5

Appendix J.1 Variable Means and Test of Significance (Data Set DS5)

Variable	Bankrupt Group Mean (n=33)	Non-Bankrupt Group Mean(n=33)	F-Ratio
X_1	0.2158	0.3449	6.434
X_2	0.1328	0.3290	19.09*
X_3	-0.0018	0.0438	1.177
X_4	0.1052	0.1390	0.932
X_5	1.3464	1.2941	0.094

* Significant at 0.01 level.

Appendix J.2 The Mann-Whitney Test Results of All Five Variables in Data Set DS5

	Mean Rank	Cases	U	W	Corrected Z	2-Tailed P
X_1	Group 1-30.17 Group 2-40.83	35 35	426.0	1056.0	-2.191	0.0285
X_2	Group 1-25.63 Group 2-45.37	35 35	267.0	897.0	-4.075	0.0000*
X_3	Group 1-30.51 Group 2-40.49	35 35	438.0	1068.0	-2.050	0.0404
X_4	Group 1-33.70 Group 2-37.30	35 35	549.5	1179.5	-0.740	0.4592
X_5	Group 1-35.91 Group 2-35.09	35 35	598.0	1257.0	-0.170	0.8648

* Significant at 0.01 level.

Appendix J.3 Kolmogorov-Smirnov Test Results of Variables X_1 to X_5 in Data Set DS5

	Absolute	Positive	Negative	K-S Z	2-Tailed P
X_1	0.28571	0.28571	0.00000	1.195	0.115
X_2	0.48571	0.48571	0.00000	2.032	0.001*
X_3	0.31429	0.31429	-0.05714	1.315	0.063
X_4	0.20000	0.20000	-0.5714	0.837	0.486
X_5	0.14286	0.14286	-0.08571	0.598	0.867

* Significant at 0.01 level.

Appendix J.4 The Median Test Results of Variables X_1 to X_5 in Data Set DS5

	Cases	Median	Chi-Square	Significance
X_1	70	0.296	2.0571	0.1515
X_2	70	0.187	11.2000	0.0008*
X_3	70	0.039	3.6571	0.0558
X_4	70	0.077	0.0000	1.0000
X_5	70	1.281	0.0000	1.0000

* Significant at 0.01 level.

Appendix J.5 Multivariate Statistics and Exact F Statistics (Data Set DS5)

S=1 M=1.5 N=31

Statistics	Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	0.7584	4.0784	5	64	0.0028

APPENDIX K

Analysis of Maximum Likelihood Estimates and Classification Tables For Data Sets DS2-DS5

Appendix K.1 Analysis of Maximum Likelihood Estimates For Data Set DS2

Variable	Parameter Estimate	Standard Error	Wald Chi-Square	P - Value
Intercept	1.2588	0.3457	13.2580	0.0003
X ₁	-6.2837	1.4106	19.8440	0.0001

Note: Variables met the 0.05 significance level entered into the model

Appendix K.2 Classification Table (DS2)

		Predicted		
		Bankrupt	No-Bankrupt	Total
Observed	Bankrupt	41	10	51
	Non-Bankrupt	17	34	51
Total		58	44	102

Sensitivity=80.4% Specificity=66.7% Correct=73.5% False Positive Rate=29.3%
False Negative Rate=22.7%

Appendix K.3 Analysis of Maximum Likelihood Estimates For Data Set DS3

Variable	Parameter Estimate	Standard Error	Wald Chi-Square	P - Value
Intercept	0.9203	0.3232	8.1060	0.0044
X ₂	-4.3070	1.3030	10.9259	0.0009
X ₃	-5.2024	2.3717	4.8155	0.2883

Note: Variables met the 0.05 significance level entered into the model

Appendix K.4 Classification Table (DS3)

		Predicted		
		Bankrupt	No-Bankrupt	Total
Observed	Bankrupt	51	10	61
	Non-Bankrupt	16	45	61
	Total	67	55	122

Sensitivity=83.6% Specificity=73.8% Correct=78.7% False Positive Rate=23.9%
False Negative Rate=18.2%

Appendix K.5 Analysis of Maximum Likelihood Estimates For Data Set DS4

Variable	Parameter Estimate	Standard Error	Wald Chi-Square	P - Value
Intercept	0.9849	0.3246	9.2047	0.0024
X ₂	-3.4416	1.1150	9.5281	0.0020
X ₃	-6.4817	3.0264	4.5870	0.0322

Note: Variables met the 0.05 significance level entered into the model

Appendix K.6 Classification Table (DS4)

		Predicted		
		Bankrupt	No-Bankrupt	Total
Observed	Bankrupt	41	14	55
	Non-Bankrupt	18	37	55
	Total	59	51	110

Sensitivity=74.5% Specificity=67.3% Correct=70.9% False Positive Rate=30.5%
False Negative Rate=27.5%

Appendix K.7 Analysis of Maximum Likelihood Estimates For Data Set DS5

Variable	Parameter Estimate	Standard Error	Wald Chi-Square	P - Value
Intercept	1.2488	0.4297	8.4457	0.0037
X ₂	-5.6490	1.6070	12.3565	0.0004

Note: Variables met the 0.05 significance level entered into the model

Appendix K.8 Classification Table (DS5)

		Predicted		
		Bankrupt	No-Bankrupt	Total
Observed	Bankrupt	27	8	35
	Non-Bankrupt	11	24	35
	Total	38	32	102

Sensitivity=77.1% Specificity=68.6% Correct=72.9% False Positive Rate=28.9%
False Negative Rate=25.0%

END

20-07-94

FIN