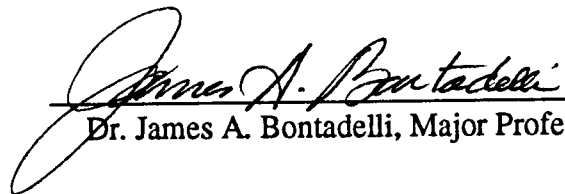
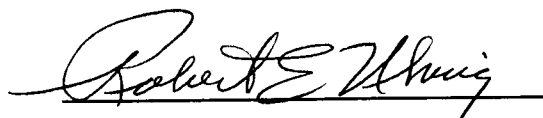
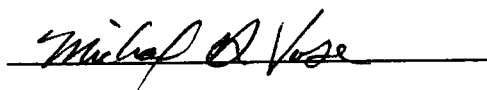


To the Graduate Council:

I am submitting herewith a dissertation written by Michael William Howard entitled "Fuzzy Expert System for Predicting Financial Performance of Small, Technology-Based Companies." I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Engineering Science.


Dr. James A. Bontadelli, Major Professor

We have read this dissertation
and recommend its acceptance:

Accepted for the Council:


Associate Vice Chancellor and
Dean of The Graduate School

*Fuzzy Expert System for Predicting
Financial Performance of
Small, Technology-Based Companies*

A Dissertation Presented for the
Doctor of Philosophy Degree
The University of Tennessee, Knoxville

Michael William Howard
August 1996

Copyright © Michael William Howard, 1996

All rights reserved

Dedication

This dissertation is dedicated to the people that supported me in my educational endeavors, especially my parents John and Eleanor Howard.

A very special dedication is given to my lovely wife Karen who had the confidence in me to support my dream of a doctoral education, and to our son Brian for his patience and sacrifice throughout this journey.

Acknowledgments

In the course of this dissertation research, several individuals have been influential in its completion. I extend a sincere appreciation to my committee members: Dr. Elden L. DePorter, for assistance in optimization analysis; Dr. Robert E. Uhrig, for support in fuzzy logic and neural networks; and Dr. Michael D. Vose, for assistance with genetic algorithms. Their advice, suggestions, and guidance were instrumental in the completion of this research. I especially thank my committee chairperson and mentor, Dr. James A. Bontadelli for his continued guidance and direction. Dr. Bontadelli spent several hours constructively evaluating my research and providing insight into the information content, especially the engineering economic aspects of this research. My hopes are that I may be as respected by my peers, valued by my students, and significant to the profession as my committee members.

I also want to acknowledge Mr. Ryan Carroll who assisted in development of the genetic algorithm and fuzzy logic software used in this research. His expertise and support were essential in bringing this undertaking to fruition.

I am grateful for the encouragement from staff members of the Tennessee Center for Research and Development. Their assistance and support are very much appreciated.

Special acknowledgment, deepest appreciation and admiration are owed to my wife Karen, for her patience, support and constant encouragement throughout this journey, and without which this would not have been possible. A special thanks to our son, Brian, for without him my higher education would not be as important.

Abstract

This research discusses the development of an expert system methodology for identifying the fuzzy rules and membership functions used to predict the financial performance of small, technology-based companies. The training data consists of four years of financial information on 250 small, publicly traded, technology-based companies. A genetic algorithm determines the expert system fuzzy rules and membership functions using the training data. A fuzzy implication operator then determines the resultant output that represents the predicted financial performance. Additional companies are used to validate the fuzzy expert system model.

The training data consists of 114 financial and 31 economic independent variables for each company. Multiple linear regression analysis was used to determine the significant variables in predicting a company's financial performance as measured by the economic value added (EVA). The significant independent variables are then used by a genetic algorithm to derive the "best fit" fuzzy rules and membership functions to develop the expert system model. This research found that none of the 31 external economic variables significantly contributed to predicting a company's EVA.

The number of fuzzy rules identified using a genetic algorithm strongly corresponds to the number of data clusters found using k-mean cluster analysis. Additional information available in the training data for learning new rules decrease significantly beyond a specific number of data clusters. As a result, this research

suggests that cluster analysis may be used to identify the number of fuzzy rules potentially available in a training data set.

The fuzzy expert system model with 20 or more rules outperformed the “best fit” ten variable multiple linear regression model in predicting a company’s EVA. After reducing the number of independent variables to include only the five most significant independent variables that contribute to improving the regression model’s coefficient of determination, the fuzzy expert system model produced additional prediction improvement. This improvement is partially due to the reduced search space resulting in a higher probability that the genetic algorithm will identify a “better fit” fuzzy rule and membership function combination.

This research found that a fuzzy expert system model with fewer significant independent variables was less sensitive to system noise and more responsive to general trends. Therefore, the fuzzy expert system that included independent variables which contribute measurably to increasing the regression model’s coefficient of determination, significantly outperformed the multiple linear regression model, with the same variables, in predicting a company’s future financial performance.

Table of Contents

Chapter	Page
Chapter 1 - Introduction.....	1
A. Rationale for the Research	5
B. Summary of the Research Problem	7
C. Research Questions Investigated	7
D. Limitations of the Research.....	8
E. Organization of the Dissertation	10
Chapter 2 - Review of the Literature.....	12
A. Historical Overview of Theory and The Research Literature.....	12
B. Contributions This Study Will Make to the Research Field	28
C. Summary.....	29
Chapter 3 - Research Procedures.....	31
A. Research Methodology	31
B. Specific Research Procedures.....	31
C. Summary.....	39
Chapter 4 - Research Knowledge Domain	41
A. Introduction.....	41
B. Economic Value Added.....	42
C. Financial Summary	46
D. Independent Financial Variables	48
E. Summary.....	56
Chapter 5 - Independent Variable Selection	58
A. Introduction.....	58
B. Multiple Linear Regression Analysis	59
C. Experimental Results - Independent Variable Selection.....	68
D. Summary	93

Chapter 6 - Cluster Analysis	96
A. Introduction.....	96
B. Experimental Results - Cluster Analysis.....	103
C. Summary.....	107
Chapter 7 - Genetic Algorithm and Fuzzy Expert System.....	109
A. Introduction.....	109
B. Example - A Genetic Algorithm	112
C. Fuzzy Set Theory	118
D. Fuzzy Logic.....	123
E. Fuzzy Expert System.....	127
F. Developing a Fuzzy Expert System.....	130
G. Experimental Results - Fuzzy Expert System.....	142
H. Summary	154
Chapter 8 - Validation.....	156
A. Introduction.....	156
B. Validation Results	157
Chapter 9 - Information Content.....	161
A. Introduction.....	161
B. Information Content Analysis	161
Chapter 10 - Conclusions	168
A. Summary	168
B. Future Research	174
References	178
Vita.....	187

List of Tables

Table	Page
Table 2-1. Genetic Algorithm Applications.	19
Table 3-1. Example of Typical Internal and External Company Attributes.	34
Table 5-1. ANOVA and Parameter Estimates (CPEVA).	71
Table 5-2. ANOVA and Parameter Estimates (CPEVA, CPEVA ²).	72
Table 5-3. ANOVA and Parameter Estimates (SHE).	74
Table 5-4. ANOVA and Parameter Estimates (SHE, SHE ²).	75
Table 5-5. ANOVA and Parameter Estimates (CPEVA, SHE).	75
Table 5-6. ANOVA and Parameter Estimates (CPEVA, cTLSHER).	76
Table 5-7. ANOVA and Parameter Estimates (CPEVA, SHE, cTLSHER).	76
Table 5-8. Independent Variables.	78
Table 5-9. Regression Model A Independent Variables.	80
Table 5-10. ANOVA and Parameter Estimates (Model A).	81
Table 5-11. Stepwise Regression Control with Step History (Model A).	83
Table 5-12. Model A with Cross Product and Quadratic Terms.	85
Table 5-13. ANOVA and Parameter Estimates (Model A1).	86
Table 5-14. Regression Model B Variables.	87
Table 5-15. Stepwise Regression Control with Step History (Model B).	87
Table 5-16. ANOVA and Parameter Estimates (Model B).	89
Table 5-17. ANOVA and Parameter Estimates (Model B1).	90
Table 5-18. ANOVA and Parameter Estimates (Balance Sheet Variables).	91
Table 5-19. ANOVA and Parameter Estimates (Income Statement Variables).	91
Table 5-20. ANOVA and Parameter Estimates (Income Statement Variables with CPEVA).	92
Table 5-21. ANOVA and Parameter Estimates (Balance Sheet and Income Statement Variables).	93

Table 7-1. Model A1 - 30 Rules.....	144
Table 7-2. Model B1 - 50 Rules.....	147

List of Figures

Figure	Page
Figure 3-1. Block Diagram of GA-based Learning System.	37
Figure 3-2. Block Diagram of the Fuzzy Expert System.	38
Figure 5-1. Simple Linear Regression.....	60
Figure 5-2. Regression Variability.....	63
Figure 5-3. Leverage Plot of Next-Period EVA vs. Predicted Next-Period EVA.	72
Figure 5-4. Regression Plot of NPEVA vs. CPEVA.....	73
Figure 5-5. Leverage Plot (Model A).....	81
Figure 5-6. Pareto Chart of Regressor Effects (Model A).....	84
Figure 5-7. Pareto Chart of Regression Effects (Model B).	88
Figure 6-1. Cluster Analysis.....	97
Figure 6-2. Cluster Data Set	99
Figure 6-3. Cluster Data Set - Initial Assignment.	100
Figure 6-4. Cluster Center - First Iteration.	101
Figure 6-5. Cluster Data Set - Second Assignment.....	101
Figure 6-6. Cluster Center - Second Iteration.....	102
Figure 6-7. Cluster Data Set - Third Assignment.....	102
Figure 6-8. Cluster Center - Third Iteration.....	103
Figure 6-9. R^2 vs. Number of Clusters for Model A.....	105
Figure 6-10. R^2 vs. Number of Clusters for Model A1.....	105
Figure 6-11. R^2 vs. Number of Clusters for Model B.....	106
Figure 6-12. R^2 vs. Number of Clusters for Model B1.....	106
Figure 7-1. Example of a Genetic Algorithm Minimum Point Search.	109
Figure 7-2. Normal and Convex Membership Function for Fuzzy Number A.....	120
Figure 7-3. Normal and Convex Membership Function for Fuzzy Number B.....	121
Figure 7-4. Fuzzy Rule Combination.....	124

Figure 7-5. Fuzzy Associative Memory for Project Investment.....	126
Figure 7-6. Block Diagram of a Fuzzy Logic Control System.....	126
Figure 7-7. Fuzzy Associative Memory for Electrical Energy Consumption.....	131
Figure 7-8. Fuzzy Membership Function for Time.....	132
Figure 7-9. Block Diagram of the Fuzzy Logic Module.....	135
Figure 7-10. Chromosome Rank Selection.	137
Figure 7-11. NMSE vs. Number of Rules for Model A.....	143
Figure 7-12. NMSE vs. Number of Rules for Model A1.....	143
Figure 7-13. Model A1 Variable Membership Functions.....	145
Figure 7-14. NMSE vs. Number of Rules for Model B.....	146
Figure 7-15. NMSE vs. Number of Rules for Model B1.....	146
Figure 7-16. NMSE vs. Number of Rules for Models A and A1.....	148
Figure 7-17. NMSE vs. Number of Rules for Models B and B1.	148
Figure 7-18. NMSE vs. Number of Rules for Models A and B.	149
Figure 7-19. MSE vs. Number of Rules for Models A1 and B1.	149
Figure 7-20. NMSE vs. Number of Rules.....	150
Figure 8-1. Model A and A1 Validation Results.....	158
Figure 8-2. Model B and B1 Validation Results.	158
Figure 9-1. NPEVA Output Membership Function (Maximum 0.7).....	164
Figure 9-2. NPEVA Output Membership Function (Maximum 1.0).....	165
Figure 9-3. Surface Profile - NPEVA versus R & D and S, G & A Expenses.....	166
Figure 9-4. Surface Profile - NPEVA versus SHE and TL.....	166
Figure 9-5. Surface Profile - NPEVA versus GP and SHE.	167

Nomenclature

r	Company generated rate of return
c^*	Rate of return required by investors
R^2	Coefficient of regression
y	Dependent variable (Next-period EVA)
x_i	Independent variable i
β_k	Regression coefficient k
ε	Regression error term
σ^2	Variance
n	Number of observations in data set
k	Number of variables
b_j	Least Square Estimator of regression coefficient β
p	Total number of model parameters
R	Correlation matrix
$r_{k-1,k}$	Simple correlation between x_{k-1} and x_k
σ_ε^2	Error variance
C_p	Mallows' criteria
C_k	Cluster k
k_c	Centroid of C_k cluster
v_i	Chromosome population I
F	Population fitness
P_i	Chromosome probability selection
q_i	Chromosome cumulative probability
a_i	Chromosome element i
l	Schema length or number of elements in genetic string
k	Number of alphabets in chromosome
n	Number of chromosomes in population

H	Specific schema
$o(H)$	Schema order
$\delta(H)$	Schema length or distance between first and last specific string position
$m(H, t)$	Number of schema H in population of chromosomes $A(t)$ at time t
f_i	Fitness value for chromosome I
p_c	Chromosomes crossover probability
p_m	Chromosomes mutation probability
$\mu_A(x)$	Membership function A
α -cut	Alpha cut for fuzzy membership function
y_i	Dependent variable observation i

List of Abbreviations

ANOVA	Analysis of Variance
ATCF	After Tax Cash Flow
ATR	Asset Turnover Ratio
BESF	Break-Even Sales Factor
CA	Current Assets
CL	Current Liabilities
cBEM	Change in Break-Even Margin
cNSWCR	Change in Net Sales to Working Capital Ratio
COG	Cost of Goods
CPEVA	Current-period EVA
cPTIICR	Change in Pre-Tax Income to Invested Capital Ratio
cSHER	Change in Shareholder's Equity Ratio
cTLSHER	Change in Total Liabilities to Shareholder's Equity Ratio
DC	Do Not Care
EKA	Explicit Knowledge Acquisition
EPS	Earnings per Share
EVA	Economic Value Added
FLC	Fuzzy Logic Controller
GA	Genetic Algorithm
GBML	Genetic Based Machine Learning
GP	Gross Profit
H	High
IE	Interest Expenses
IKA	Implicit Knowledge Acquisition
L	Low
LTLSHER	Long-term Liabilities to Shareholder's Equity Ratio
M	Medium

MF	Membership Function
MSE	Mean Square Error
NI	Net Income
NITAR	Net Income to Total Assets Ratio
NISHER	Net Income to Shareholder's Equity Ratio
NPV	Net Present Value
NSA	Net Sales
NSCAR	Net Sales to Current Assets Ratio
NSPER	Net Sales to Plant and Equipment Ratio
NSWCR	Net Sales to Working Capital Ratio
NMSE	Normalized Mean Square Error
NPMSE	Normalized Predicted Mean Square Error
OPATCF	Operating Profit After Tax Cash Flow
PTICR	Pre-Tax Income to Invested Capital Ratio
PTISHER	Pre-Tax Income to Shareholder's Equity Ratio
PTITAR	Pre-tax Income to Total Assets Ratio
QR	Quick Ratio
R ²	Coefficient of Determination
RA	Regression Analysis
RDE	Research and Development Expenses
RDENSR	Research and Development Expenses to Net Sales Ratio
RPG	Rule Projection Graph
RPM	Rule Projection Matrix
SCR	Sales to Cash Ratio
SEC	Securities and Exchange Commission
SGAE	Sales, General and Administrative Expenses
SGAENSR	Sales, General and Administrative Expenses to Net Sales Ratio
SHE	Shareholder's Equity
SIC	Standard Industrial Classification

SS _E	Sum Square Error
TA	Total Assets
TL	Total Liabilities
TLTAR	Total Liabilities to Total Assets Ratio
TLSHE	Total Liabilities and Shareholder's Equity
TLSHER	Total Liabilities to Shareholder's Equity Ratio
VIF	Variance Inflation Factor
VH	Very High
VL	Very Low

Chapter 1

Introduction

Entrepreneurial self-determination releases vast amounts of human energy in the form of "early-stage" companies based on a wide array of technologies. These technology-based emerging companies build the cornerstones of the U. S. economic growth with approximately 80 percent of all new jobs created by small businesses. In creating new jobs, these small businesses are estimated to be 10 to 50 times more productive in engineering design/development and overall organizational productivity than their large business counterparts.

Recent statistics indicate that more than 80 percent of all technology-based businesses fail within their first five years of operation. With such a large number of technology-based business failures, how does a company increase its chances of survival?

For these "early-stage" companies to be successful, several factors must converge at once. The company must have a management team experienced in business development that thoroughly understands the market and technology being pursued by the company. There must be a market need for the company's product, and the company must have sufficient financing to fuel its growth. In addition, the

company must have a clearly developed business plan that defines exactly where the firm wants to go and how it plans to get there.

The existing research into the failure of small technology companies, as well as the writer's experience, indicate that the major reasons for business failure are a lack of management dedication and experience, lack of proper start-up financing, the inability to focus on a clear set of goals and objectives, and an inadequate understanding of the company's product, technology, and market. Billions of dollars are spent yearly by would-be entrepreneurs (both outside and inside an existing organization) to create businesses without adequate business planning, only to see their financial inputs and dreams destroyed. New business development is one of the most demanding, yet challenging tasks any individual or organization can undertake.

This dissertation topic, the development of a fuzzy expert system for predicting financial performance of small, technology-based companies, is part of the intense research interest occurring in the artificial intelligence community. In 1994, the U. S. Commerce Department [1] released a report on artificial intelligence titled *Critical Technology Assessment of the U. S. Artificial Intelligence Sector*. In this report, the Commerce Department estimates that the total U. S. research funding for artificial intelligence is approximately \$200 million per year with the federal government providing approximately 80 percent of this total. The report also estimates that the global market for artificial intelligence is \$900 million per year with knowledge-based systems having the most commercial success with a \$350 million per year market

share. Another study, conducted in 1992 by Arthur D. Little, indicated that the market for artificial intelligence in the financial services industry is estimated to be more than \$300 million per year and growing at 44 percent per year [2]. In yet another recent survey, nearly 33 percent of financial brokerage firms, 50 percent of banks, and 80 percent of insurance companies indicated they are considering investing in artificial intelligence applications [3].

A promising new direction in the implementation of artificial intelligence application is in the design of a knowledge-based learning system that involves the hybrid use of fuzzy logic and a genetic algorithm to intensify the abilities of the system to learn and adapt through genetic evolution. This dissertation develops an approach to creating a knowledge-based learning system by incorporating a genetic algorithm learning mechanism into a new hybrid fuzzy logic reasoning method to achieve a significant learning capability. The genetic algorithm learning mechanism provides for the selection of the fuzzy rules and associated membership functions.

Specifically, the development in this dissertation involves a knowledge-based learning methodology that results in a model which identifies the rules to predict the financial performance of small, technology-based companies. The rules are determined using a genetic algorithm. In addition, the fuzzy membership functions for each rule's antecedent and consequent variable are also determined using a genetic algorithm. The rules are implemented using a fuzzy implication operator to calculate a resultant

output that represents the predicted financial performance of the small, technology-based company.

The knowledge-base, or training data, comes from 250 small, publicly traded, technology-based companies, and is used for determining the fuzzy rules and associated membership functions. An additional 91 small, publicly traded, technology-based companies is used for validating the model. Finally, the normalized mean square error resulting from applying the fuzzy rules and membership functions to the training and validation data is compared to the normalized mean square error using a multiple linear regression model for the same independent variables. The results are discussed along with the information content provided by each model.

This research has two objectives. The first is to develop a *knowledge-based learning system* that can be used to improve the probability of success for small, technology-based businesses by identifying the financial rules used by these companies and their resultant output. The second objective is to develop a *robust adaptive learning technique* as the key element in determining the fuzzy expert system rules and membership functions. The accomplishment of these two objectives will be beneficial in this application, as well as in expanding artificial intelligence and knowledge-based systems in solving other engineering and business applications.

A. Rationale for the Research

This dissertation develops a new approach for creating the rules of a expert system that can be used to increase the probability of success for small, technology-based companies. The rules are developed using a genetic algorithm and then evaluated using a fuzzy implication operator.

In traditional system model design, the initial step is to obtain a mathematical model of the system. This model represents the *a priori* information about the system. Many real-world systems, however, do not have a well-known structure. Rather, they have complex and nonlinear characteristics as well as processes that are not clearly understood. A useful approach to overcome this difficulty in system design is to use a fuzzy rule-based expert system. The fuzzy rule-based expert system is expressed as a set of condition-action rules (called *fuzzy rules*) which describe conditions about the process state using linguistic terms such as low, medium, high, and very high. The recommended action or output of the fuzzy rule-based expert system also uses linguistic terms such as “increase moderately” or “only slightly.” An example of a fuzzy rule is the following:

“IF Error is Very Small and Temperature is Large,
THEN Slightly Decrease Throttle Pressure”

Recently, a number of successful studies used fuzzy expert systems where the design of the fuzzy system is based on human expert experience and/or knowledge (see Chapter 2). However, one of the main difficulties in developing these fuzzy expert systems is the acquisition of the area (domain) of knowledge. Typically,

knowledge is acquired by an engineer or analyst conducting a series of interviews with an expert in the domain of interest. As is often the case, a domain expert cannot tell linguistically what kind of action is taken or what inference to use in a particular situation. Furthermore, there is no reason to believe that an expert's rule-base is optimal. Despite the many new tools and methodologies available, the process of determining the rules from a domain expert are very difficult, time-consuming, expensive, and prone to error. Another problem with expert systems is that as the model approaches the limits of the knowledge domain, the performance of the system degrades significantly.

To overcome these problems, a system must be able to continuously acquire new knowledge and easily adapt to changing conditions. That is; the system must *learn*. For these reasons, it is appealing to consider the development of a learning rule-based system with evolutionary search procedures using a genetic algorithm. One of the most significant reasons for using a genetic algorithm is that it requires little or no *a priori* domain knowledge. Even though initial results from research in this area are encouraging, most of the research has been demonstrated using simple control problems such as the cart-pole balancing example (see Chapter 2).

A genetic algorithm is a randomized, parallel, and global search method based on natural selection and genetics. This research suggests that a genetic algorithm can be used to identify the selection of fuzzy rules and membership functions for the development of an intelligent fuzzy knowledge-based expert system.

B. Summary of the Research Problem

As discussed in the previous section, several problems currently exist with traditional system model design and the development of expert systems. The research problem addressed in this dissertation is the development and evaluation of a robust adaptive learning technique using historical information about a system to determine the best fit fuzzy expert rules and membership functions. This method will overcome many problems currently existing with traditional system model design and expert system development by *not* requiring a mathematical model of the system or a domain expert.

In addition, the methods developed in this dissertation identify the fuzzy rules, (i.e., guidelines) for predicting the financial performance of small, technology-based companies. The fuzzy rules will assist managers in better understanding the very complex financial interactions involved in creating successful technology-based companies. The financial expert system can also be used to assist managers in reviewing financial scenarios (as a tool in budgetary planning) and for developing different financial alternatives for large capital equipment purchases.

C. Research Questions Investigated

The four research questions addressed in this dissertation are summarized below:

1. What are the dominant financial attributes that define superior financial performance of small, publicly owned, technology-based companies? For example, are net sales, return-on-investment, current ratio, research and development expenses, and gross margin significant factors in explaining and predicting the future financial performance of these companies?
2. Once the dominant financial attributes are identified, what combination and quantity of attributes create the fuzzy rules to explain and predict financial performance? For example, a typical fuzzy rule may be, "If net sales are HIGH, and return-on-investment is MEDIUM, and research and development expenses are VERY HIGH, and cost of money is LOW, and . . . , then the financial performance of the company next-period is likely to be VERY GOOD."
3. Knowing the fuzzy rules, what information content is available to provide guidelines to improve chances of survival for a small, publicly owned, technology-based company?
4. Finally, how does the normalized mean square error of the fuzzy rule set and membership function model compare to that for a multiple linear regression model?

D. Limitations of the Research

The resulting answers to these four questions are limited to the knowledge-domain used in the research; that is, small, technology-based companies. The fuzzy rules and associated membership functions are learned (developed) using publicly

available financial information from the Securities and Exchange Commission (SEC). For publicly owned firms, Congress passed the Securities Act of 1933 that prescribed the methods of accounting used in preparing financial statements for stockholders of publicly held corporations. Congress delegated its authority to the SEC, an agency of the federal government, for legal authority in prescribing standard accounting principles. As such, publicly owned firms are required to submit yearly detailed audited financial statements. Specifically, the financial information is available from Disclosure Inc., a Bethesda, Maryland company that compiles information on more than 11,000 companies directly from the SEC. The publicly owned firms used as the knowledge domain in this research are limited to those firms with a market value of less than \$120 million, and includes Standard Industrial Classification (SIC) Group Code 35 (Industrial and Commercial Machinery and Computer Equipment), Group Code 36 (Electronic and Other Electrical Equipment and Components Except Computer Equipment), and Group Code 38 (Measuring, Analyzing, and Controlling Instruments).

The information used in this research is all audited financial information collected by the SEC for existing publicly owned firms. The reasons for using this information in the research are as follows:

1. The financial information is easily available on a large number of small, technology-based companies.

2. The data used is based on audited financial information that represents financial statements prepared in accordance with “generally accepted accounting principles.” Such “principles” have evolved over time or were decreed acceptable by the Financial Accounting Standards Board (FASB). Thus, the audited financial information follows standard principles in describing the financial health of the firm. The data used in this research is compiled from annual reports that contain the following financial statements:
- Statement of financial position
 - Statement of net income
 - Statement of changes in financial position
 - Various supporting statements or schedules
 - Notes to the financial statements
 - Opinion of the independent certified public accountant
3. These financial statements are an objective method of judging the health of a company. In essence, the financial statements reflect the capabilities of the management team and its ability to consistently make successful strategic and tactical decisions.

E. Organization of the Dissertation

The dissertation is organized into ten chapters starting with Chapter 1, “Introduction.” Chapter 2, “Review of the Literature,” presents a historical overview

of artificial intelligence and its evolution into classical expert systems. The practical limitations and associated problems of classical expert systems are reviewed along with a discussion of a relatively new approach in developing expert systems referred to as a learning system or machine learning. Research in learning systems are reviewed, followed by a discussion of the contribution that genetic algorithms have played in the development of these systems. Next, the review examines fuzzy sets and the development of fuzzy logic control. Finally, the literature review concludes with the research done on combining a genetic algorithm and fuzzy logic to create a new hybrid learning system. Also included in Chapter 2 is a discussion of the contributions this study makes to the research field.

Chapter 3, "Research Procedures," outlines the ten specific steps performed in this research along with a detailed description and reason for each step. Chapter 4 through Chapter 7 discusses in greater detail the development of the learning system as it applies to the financial prediction of small, technology-based companies along with the experimental results of each step. Chapter 8, "Validation," highlights the comparisons of the fuzzy expert systems developed in this dissertation with the classical multiple linear regression model. In Chapter 9, "Information Content," the information provided by the output of each model is discussed, as well as its usefulness in making more informed decisions. Chapter 10, "Conclusion," presents a summary of the dissertation research results along with recommendations for future research based on the conclusions reached.

Chapter 2

Review of the Literature

A. Historical Overview of Theory and The Research Literature

The research discussed in this dissertation expands traditional expert systems by combining the evolutionary search of a genetic algorithm with fuzzy logic to create a hybrid evolutionary learning system model for predicting the financial performance of small, technology-based companies. Because multiple techniques are being integrated to create a new learning method, the literature review will begin by presenting a historical overview of artificial intelligence followed by a similar overview of genetic algorithms and fuzzy logic.

Artificial intelligence (AI) is the name for a group of technology fields that enable computers to understand and manipulate information. Expert systems are a specific branch or subset of artificial intelligence. The engineering goal of AI and expert systems is to solve real-world problems.

Research in AI began in the 1950s. Then rapid advances in computer technology opened an avenue for the present surge in AI research and application activities. In 1956, Dr. Marvin Minsky, Dr. Claude Shannon, and other researchers coined the term artificial intelligence for the theme of a conference held at Dartmouth College. AI efforts grew steadily over the following decades with many of the early advances being focused on neural networks. As often occurs with most new technical

disciplines, the first couple of decades of work in AI were characterized mainly by theoretical research with little practical applied use. However, since approximately 1985, several practical uses of AI tools in different engineering applications have occurred.

In the 1970s, the AI field began to subdivide into four major areas: (1) knowledge presentation, (2) understanding natural language, (3) computer vision, and (4) speech input. Knowledge presentation was the start of the pace-setting expert system technology, particularly in medicine and engineering [4]. Expert system technology is a specific branch of AI that makes use of specialized information and knowledge to solve problems at the level of a human expert.

By the 1980s, specialized AI purpose hardware became available which eventually lead to customized very large scale integrated (VLSI) circuits for specific AI applications. During the 1990s, the use of AI tools grew as the computing power and capacity of personal computers became great enough to handle many practical AI applications.

The research in this dissertation is based on an extension of the expert system technology launched in the late 1960s. Expert systems have experienced significant growth and popularity since their commercial introduction in the early 1980s. Today, expert systems are used in many fields including medicine, engineering, and business.

Dr. Edward Feigenbaum, an original developer of expert systems technology, defines an expert system as “an intelligent computer program that uses knowledge and

inference procedures to solve problems that are difficult enough to require significant human expertise for their solutions.” [5] Thus, an expert system is a computer-based system that emulates the decision-making ability of a human expert.

The knowledge required to solve a specific problem in an expert system is called the system’s *knowledge-domain* and may be developed from generally available sources including technical books and magazines, or obtained from knowledgeable persons. The expertise contained within the expert system is called the system’s *knowledge-base*. The expert system’s inference engine draws conclusions from the knowledge-base.

Classical types of expert systems were developed using expertise available from knowledgeable persons based on IF . . . THEN type rules, called rule-based production systems. For example, “IF it is cold and raining outside, THEN put on a coat.” The process of extracting rules from experts is referred to as explicit knowledge acquisition (EKA). However, formalizing the knowledge of experts into rules is not simple, especially when the knowledge has never been systematically explored and developed. There may be inconsistencies, ambiguities, and duplications with the expert’s knowledge that may not be readily apparent. Rule-based production systems also tend to be very expensive to develop as well as time consuming, mainly because experts frequently find it difficult to explain how they make their decisions [6]. Consequently, a practical limitation of many expert systems is the lack of causal knowledge [7]. That is, the rule set developed by the knowledgeable person and

incorporated into the expert system's knowledge-base does not represent a complete understanding of the underlying causes and effects in the system; especially the nonlinear effects associated with describing a system.

To eliminate the problem of extracting rules from knowledge-based experts, researchers started in the 1960s to develop expert systems that allowed them to learn rules by example through rule induction. Thus, the expert system would derive its rules from data of successfully solved cases provided by the expert system developer. This relatively new approach in developing the expert system rule-base is referred to as a learning system, and is sometimes referred to as machine learning. This new approach of extracting rules from historical data is called implicit knowledge acquisition (IKA). Thus, the required knowledge from the expert system is implicitly embedded in the examples. For this reason, expert systems that are based on the IKA approach tend to be less expensive than their counterparts [8].

Learning systems have the potential to exceed the performance of knowledge-based experts and the potential to discover new relationships among concepts and hypotheses by analyzing the records of hundreds of cases. Thus, the process of learning automatically holds the promise of incorporating knowledge into the expert system without the need of a knowledge-based expert.

Research in learning systems may be distinguished by three major periods. The first period was characterized by neural modeling and decision-theoretic techniques in the late 1950s and early 1960s. During this period, most of the research was

theoretical in nature and lead to the development of neural model-based machines based on perceptron-type learning systems [9]. Related research lead to the development of evolutionary processes that through random mutation and “natural” selection might create a system capable of identifying optimal rules [10, 11, 12]. A second major paradigm, which started in the early 1960s, is based on work done by psychologists and early AI researchers on models of human learning. This paradigm utilized logic or graph structure representations. Influential developers during this period included E. B. Hunt and P. H. Winston [13]. The third paradigm represents the most recent period starting in the mid-1970s. During this period, researchers began concentrating on developing a wider spectrum of new learning methods.

There is currently very little scientific evidence comparing learning system models with classical expert systems using EKA-derived rules in terms of system accuracy; however, several learning system models have been successfully implemented in a number of real-work applications.

The early researchers in evolutionary optimization include W. W. Bledsoe [14] and H. J. Bremermann [15]. Bledsoe and Bremermann suggested the use of binary string coding for evolutionary optimization. The theoretical foundation for using evolutionary optimization as a learning system was published by John Holland in 1962 and 1965 [16, 17]. Holland laid the foundation for genetic algorithms as an evolutionary optimization technique at the University of Michigan in the 1960s while researching artificial intelligence. Holland, who noted that learning can occur not only

by adaptation in a single generation but also by evolutionary adaptation over many generations, was inspired by Darwin's theory of evolution in which only the fittest survive from one generation to the next [18]. As a result of his research in the 1960s and early 1970s, he published his first comprehensive series of papers on genetic algorithms in 1975 [19].

After Holland's paper titled "Adaptation in Natural and Artificial Systems" was published in 1975, several other researchers also published papers on genetic algorithms, and by the early 1980s genetic algorithms were showing broad promise in solving a variety of optimization problems. Leaders in the field began holding regular conferences every other year. In addition, articles appeared in a variety of scientific journals discussing genetic algorithm techniques. In 1989, David Goldberg, a former student of Holland's, published a book that discussed the scientific basis for a genetic algorithm and cited 73 successful applications [20]. Goldberg also provided genetic algorithm software called GENESIS and OOGS to colleagues studying genetic algorithms. For the first time, the book and software made genetic algorithm techniques available to thousands of researchers interested in exploring in more detail the unique features of this approach. In 1991, Lawrence Davis, who also studied under Holland, published a handbook on genetic algorithms [21]. The combination of these two books, the genetic algorithm software that had become available, and the proliferation of research articles, resulted in genetic algorithms becoming recognized as an important technique in optimization analysis.

Some recent examples of applying a genetic algorithm are shown in Table 2-1 [22]. In 1978, researchers reported the first practical implementation involving a genetic algorithm in a learning system [23]. This was followed by a variety of other learning system applications using a genetic algorithm. The first system called the Cognitive System Level One (CS-1) was utilized to learn seven- and 13-node maze tasks and processed a serial population of individual rules or classifiers as the basic unit of genetic manipulation. The CS-1 system showed for the first time that a GA-based learning system could significantly outperform a random walk [24]. Since 1978, a number of researchers extended these ideas in different ways to solve many learning system applications. In 1980, S. F. Smith of the University of Pittsburgh reported the first use of a genetic algorithm in evaluating a population of rule combinations in parallel [25]. Smith's Learning System Level One (LS-1) processed string rules with genetic operators and genetic manipulation. In the LS-1 system, rule sets mate, cross, and mutate to create new and better combination of rules. In addition, LS-1 contains an inference engine to blend rules together to arrive at a combined result. Smith's research expanded the practical use of genetic algorithms in developing learning systems. Since the development of the CS-1 and the LS-1 learning system, several other systems were developed. David Goldberg outlines many examples of genetic-based learning systems in his classical book, *Genetic Algorithms in Search, Optimization, and Machine Learning* [26].

Table 2-1. Genetic Algorithm Applications [27].

Subject Area	Description
Biology	Simulation of the evolution of single-celled organism populations Genetic Based Machine Learning (GBML) system for medical diagnosis
Business	GBML system for describing consumer preferences
Computer Science, Engineering and Operations Research	Adaptive document clustering Recursive adaptive filter design Steady-state and transient optimization of a gas pipeline Job shop scheduling VLSI circuit layout/compaction On-off, steady-state optimization of an oil pump-pipeline system Keyboard configuration design Structural optimization (plane truss) Aircraft landing strut weight optimization Classroom scheduling
Image Processing/Pattern Recognition	Selection of detectors for binary pattern recognition Image registration to minimize image differences Search for image feature detectors
Social Science	Calibration of population migration model Simulation of the evolution of behavioral norms

Research results presented at the International Conferences on Genetic Algorithms, coupled with special Workshops on Genetic Algorithms for Machine Learning, and articles appearing in regular issues of *Machine Learning*, document a diverse interest in applying genetic algorithms to learning systems. Typically, the articles limit themselves to focus on a number of central issues in applying a genetic algorithm to learning systems such as string representation, crossover, and mutation operators. A recent book edited by John Grefenstette includes several articles on using a genetic algorithm for concept learning from examples, reinforcement learning for control, and theoretical analysis of a genetic algorithm [28]. Another recent book by Hojjat Adeli and Shih-Lin Hung [29] explores using neural network computing to improve several learning algorithms including perceptron learning, backpropagation learning, gradient learning algorithms, hybrid genetic/neural networks, and fuzzy neural network learning models.

Emerging in parallel with the development of genetic algorithms as a learning mechanism was fuzzy set theory. In the mid-1960s at the University of California at Berkeley, Lotfi A. Zadeh developed a model for the quantification of information using fuzzy sets. Fuzzy sets emerged from the notion of possibility theory as first described in the article by Zadeh titled "Fuzzy Sets" [30]. In classical set theory there is a strict concept of membership to a set, that is, an element either belongs or does not belong to the set. In 1965, Zadeh introduced the concept of fuzzy sets where a more flexible concept of membership or degree of membership is possible. For

example, an element's degree of membership to a set is indicated by numbers between '0' and '1', not just the numbers '0' and '1'. The mapping of a crisp number to the element's degree of membership is performed by a membership function.

The evolution of fuzzy sets leads to the development of possibility theory and a technique for linguistic information. Fuzzy numbers, a method of representing linguistic information, can calculate, for example, the consequent of a set of fuzzy rules. A definition and example of fuzzy numbers and mathematical techniques for combining them are discussed in Chapter 4.

Zadeh's development of fuzzy set theory in the 1960s leads to the use of fuzzy logic control which has emerged as one of the most active areas of research in applying fuzzy set theory. The major advantage of fuzzy logic control is that it provides a way of converting linguistic control knowledge into an automatic control strategy. In the mid-1970s, Mamdani developed the pioneering research on fuzzy logic control [31, 32, 33]. Since this time, hundreds of articles involving the application of fuzzy logic control have been published [34].

Fuzzy logic control is accomplished by using a number of parallel fuzzy rules linked together by a fuzzy inference process. The intrinsic benefits of using fuzzy logic control results from the fuzzy controller being able to handle a higher degree of uncertainty, unstructured nonlinearity, and the lack of a formal mathematical model of the system. The benefits of fuzzy logic control are extensively discussed by Earl Cox in *Fuzzy Logic for Business and Industry* [35]. Cox summarizes the three key benefits

of fuzzy logic as a modeling technique as reduced application development costs, execution costs, and maintenance costs. In addition, Cox believes that parallel processing and the ability to deal with imprecise information give fuzzy logic control significantly higher computational capabilities. As a result, fuzzy logic has gained considerable attention because of the benefits outlined above and its capability of emulating humans' approximate reasoning ability.

As summarized above, fuzzy logic control technology has been successfully applied to a number of control problems. However, fuzzy logic control still requires extensive tuning of the applications rule set and the control variable membership functions before the application can be successfully used. The considerable amount of research that has been done in this area has resulted in better techniques for tuning the fuzzy controller (rule set and membership functions). This research is generally classified into two categories; those approaches based on tuning the membership function and those based on tuning the rule set [36, 37, 38, 39, 40, 41]. Because of the large number of degrees of freedom in the design of a fuzzy controller, it is important to develop new methodologies that provide the most appropriate level of control for the application.

This research focuses on expanding the field of genetic algorithms and fuzzy logic by combining the evolutionary search of a genetic algorithm with fuzzy logic into a new application area. In addition, this research should further expand evolutionary learning system models as well as methods for tuning a fuzzy rule set and membership

functions. Existing research in combining a genetic algorithm and fuzzy logic is primarily limited to classical control theory applications modeled from well-behaved differentiable mathematical equations and a small set of independent variables. A typical example used in many existing research efforts includes the development of optimal fuzzy control of an inverted cart and pendulum. Among the present researchers, C. L. Karr [42, 43, 44, 45] pioneered the application of combining a genetic algorithm and fuzzy logic to control an air-injected hydrocyclone and a liquid level controller. Karr performed his early work in the late 1980s while working on his dissertation under Goldberg at the University of Alabama. In another application, Karr used a genetic algorithm to develop the membership functions for the fuzzy logic controller of a computer simulated cart-pole balancing system [46]. In these two examples, Karr demonstrated that the GA-designed fuzzy logic controller outperformed a human-designed fuzzy logic controller. Later, Karr expanded his earlier work and developed a controller for the pH value in an acid-base experiment [47]. Typically, Karr's work incorporates a small set of independent variables, a mathematical objective function, and binary encoding of the GA to identify the fuzzy membership functions and the fuzzy rule set.

In addition to research by Karr, Philips Thrift [48] outlines some general procedures for developing a GA-designed fuzzy logic controller for centering a cart with mass m moving on a one-dimensional track using three independent variables, a mathematical objective function, and integer encoding (0, 1, 2, 3, 4, and 5) of the GA

to identify the fuzzy rules given fixed membership functions. Similarly, Michael Lee [49] used a genetic algorithm to find the rules and membership functions with a TSK fuzzy model and rule-consequent weights, a small set of independent variables, a mathematical objective function, and binary encoding of the GA. In [50, 51], Lee demonstrates dynamically controlled parameter settings of GAs in calculating fuzzy rule sets. Also, Wen-Ruey Hwang's [52] research work used a small set of independent variables, a mathematical objective function, and binary encoding of the GA to identify the fuzzy rules as an alternative means of tuning a digital proportional, integral, and derivative (PID) controller.

In addition to Hwang, Chir-Ho Chang's [53] research also combined a genetic algorithm and fuzzy logic. Chang used three variables and three term membership functions to create a 3 by 3 by 3 matrix using GA binary encoding. In this application, a GA is also used to tune the rules and membership functions of a PID controller. Likewise, Daihee Park [54] used a GA to determine the fuzzy relation matrix for the two variables used in the steady-state control of a DC motor. The GA in this application used binary encoding techniques for determining the membership functions; also, various fuzzy implication operators were investigated. Park's research results showed that the optimal fuzzy controller was obtained when both the fuzzy membership functions and the resulting 6 by 5 fuzzy relation matrix was searched using a genetic algorithm.

Cooper considered the application of a genetic algorithm as a general methodology for automatically generating fuzzy logic controllers [55, 56]. In contrast to his earlier work where every input-output combination was enumerated, he developed an encoding scheme which maintained only those rules necessary for controlling the target system and demonstrated the process on the cart-pole type problem with four and six independent variables and one dependent variable. In a more recent article [57], Cooper demonstrates the use of a genetic algorithm to evolve the composition and fuzzy rule base for controlling a boat rudder.

In a recent article, Jindrich Liska and Stephen Melsheimer [58] presented a technique for constructing a fuzzy logic controller for a two input and one output system used to control a gas furnace. Their method used 296 observation points for training the fuzzy rules and membership functions. They limited the fuzzy logic controller to 10 rules and six membership functions per variable. In the Liska and Melsheimer model, the GA is further fine-tuned once it converges toward a near optimum point by using a conjugate gradient method. The fuzzy logic controller using GA-based tuning was also compared to a neural network based model. The results showed the GA approach compared well in terms of both prediction accuracy and in the tuning of parameters.

A review of the November 1995 *Engineering Index* [59], which provides abstracted information from approximately 2,600 engineering and technology journals and selected government reports and books, indicates that since 1990 there was more

than 6,000 articles published on fuzzy logic and more than 3,800 articles on genetic algorithms. However, since 1990, less than 50 articles were published on combining genetic algorithms with fuzzy logic to create a robust learning system.

This dissertation develops a fuzzy expert system for predicting the financial performance of small, technology-based companies by developing an integrated approach for identifying the significant attributes (independent variables), number of fuzzy rules, and the corresponding rule set and membership functions using a genetic algorithm. The objective function is to identify the rule set and associated fuzzy membership functions for predicting the economic value added (EVA^{*}) resulting from the continued operation of a small, technology-based company. *The EVA per time period of a firm is the after tax net operating profit minus the cost of capital, including both debt and equity capital.* In essence, the EVA measures the wealth increment a company creates each time period. Generating continuous positive EVA increments is the key to creating wealth for investors.

There is no one magic formula that can always be used to measure the long-term impact of a business on its ability to create wealth for its shareholders. However, EVA, developed by Stern Stewart & Co. [60] in New York, is one of the most popular tools financial executives use today [61]. For example, AT&T has made EVA its central financial measure since 1992. They report that "EVA makes financial

* EVA is a registered trademark of Stern Stewart & Co.

performance relevant to a whole new audience, from plant managers to payables clerks [62].”

AT&T's Chief Financial Officer is quoted as saying, “We calculated our EVA back to 1984 and found an almost perfect correlation with stock price. The key reason for this is that such measures show investors what they really care about, the net cash return on their capital, rather than some other type of performance viewed through what are often regarded as the often distorting lens of accounting rules” [63].

In addition to AT&T, Coca-Cola uses EVA to guide management decisions including whether to build new plants, purchase a competitor, or sell a division. The strongest selling point of the EVA economic measure is the improved ability to closely track the increases or declines in market value. For example, Coca-Cola's EVA increased 983%, shadowing its stock price which increased 900% since 1983. Eli Lilly also found a correlation between EVA and stock price [64].

Other companies that use EVA as a measure of wealth creation includes Wal-Mart, Whirlpool, Georgia-Pacific, Polaroid, Sprint, Teledyne, and Quaker Oats [65]. Investors are also using EVA as an economic performance measure. “The first thing we look at when we pick companies is, are they motivated by EVA? We prefer it to measures like EPS or return on equity,” says Eugene Vesell, managing director of Oppenheimer Capital, a \$30 billion mutual fund [66].

No references were found in the literature concerning the development of a robust learning system for predicting the financial performance of small, technology-based companies.

B. Contributions This Study Will Make to the Research Field

As can be seen from the previous discussion, most of the early work in combining a genetic algorithm and fuzzy logic is confined to classical control theory applications modeled for well-behaved systems, a small set of independent variables, and binary encoding of the genetic string.

This research will contribute to the further development and practical use of knowledge-based systems by developing an integrated approach for identifying the significant attributes (independent variables) involved, the number of fuzzy rules, and the corresponding membership functions. The learning system application developed in this dissertation is a non traditional example of fuzzy logic control which applies a genetic algorithm and fuzzy logic to a non-differentiable equation application incorporating a large number of independent variables. In addition, this dissertation will present a technique that selects significant independent variables from a large list of potential variables and selects the number of fuzzy rules used in the fuzzy expert system model from the clustering of these variables.

The study will also expand the research field to a new application area; identifying the significant fuzzy rules that help predict superior financial performance

for small, technology-based companies. Once the significant rules are known and their interactions evaluated, this research will also help small companies focus resources on variables they have the most control over to accomplish a desired financial outcome. This technique may also help to provide an alternative approach for benchmarking industry segments and individual companies.

C. Summary

This chapter briefly discussed the historical evolution of expert systems along with recent developments leading to hybrid expert systems that combine a genetic algorithm and fuzzy logic control. John Holland laid the foundation for genetic algorithms in the late 1960s and early 1970s while he was a student at the University of Michigan. As a result of Holland's work, several of his former students expanded his research which led to two influential books written by David Goldberg and Lawrence Davis.

While Holland continued to expand genetic algorithms as a learning system, Lotfi A. Zadeh of the University of California at Berkeley developed the theory of fuzzy sets. Zadeh's research led to the growth of fuzzy logic control. C. L. Karr recognized the limitations of fuzzy logic control, especially in tuning the membership function and corresponding rule set, while a student of Goldberg's at the University of Alabama. As a result, he was one of the early researchers in applying a genetic algorithm to search for optimal fuzzy membership functions for fuzzy logic control

applications. Later, several other researchers, including Karr, applied a genetic algorithm as a learning technique to identify not just fuzzy membership functions but also fuzzy rules.

This research objective is to expand the earlier work using a genetic algorithm as a learning technique by solving a non traditional fuzzy logic control problem which incorporates a large number of variables and a training set based on historical data. It is also the objective of this research to better explain the independent variables and fuzzy rules involved in predicting financial performance for small, technology-based companies as measured by a firm's EVA.

The EVA concept, as developed by Stern Stewart and Co., focuses management control on the financial performance of a company as measured by the actual wealth increment added each operating period. Thus, in this dissertation the fuzzy expert system is used to predict a small, technology-based company's wealth increment added the next operating period or the companies next-period EVA.

Chapter 3

Research Procedures

A. Research Methodology

The analytical research methodology in this study makes use of historical financial data from 341 small, publicly owned, technology-based companies. From this data, multiple linear regression analysis techniques are used to identify the most significant variables that help to explain a firm's financial performance as measured by the economic value added (EVA) each operating period. These significant variables are then used to develop an expert system model based on the evolutionary learning technique of a genetic algorithm, which determines the best fit fuzzy rules and membership functions. Outlined below are the primary steps performed in this research.

B. Specific Research Procedures

The following ten research steps are used in this study to develop a robust learning expert system model to predict the financial performance of small, technology-based companies.

Step 1. Determine the dependent variable used to measure a company's financial performance (see Chapter 4).

Profitability as determined by the accounting system has traditionally been used to measure a firm's financial performance. However, using a firm's profit per time period as the dominant criterion for determining actual change in the economic wealth of a firm has shown only limited value. As a result, this research suggests that a better criterion for measuring a firm's periodic financial performance should be economic value added (EVA). As previously stated, a firm's EVA is calculated as the after-tax net operating profit per time period, minus the cost of capital for the time period, and includes both debt and equity capital in the calculation. In essence, the EVA measures the wealth a company creates each time period. Generating a positive EVA period after period is the key to creating wealth for a company and its owners.

Step 2. Select and identify the knowledge domain for the expert system (see Chapter 4).

The knowledge domain for this research study is limited to publicly owned, small technology-based companies. Other knowledge-domains, that is, industry segments, could easily be identified using the procedure outlined in this dissertation. Publicly traded firms are defined as being actively traded on either the New York Stock Exchange, American Stock Exchange, or the NASDAQ. Small companies are defined as having a market value of less than \$120 million and an "Unqualified"

Auditor's Report. Technology-based companies are defined as having one or more of the following SIC Codes:

- Major Group No. 35: Industrial and Commercial Machinery and Computer Equipment
- Major Group No. 36: Electronic and Other Electrical Equipment and Components Except Computer Equipment
- Major Group No. 38: Measuring, Analyzing, and Controlling Instruments

Step 3. Identify from the total knowledge domain a random sample of companies with two to four years of historical financial data for training the expert system (see Chapter 4).

The financial information on 341 companies meeting the criteria outlined in Step 2 above is randomly selected for the study from more than 1,000 potential companies in the Disclosure/SEC database from Disclosure, Inc., located in Bethesda, Maryland. The Disclosure/SEC database contains information on more than 10,000 publicly traded companies and is compiled from documents filed with the U. S. Securities and Exchange Commission (SEC). The database is updated monthly and includes more than 250 fields of information per company. Of the 341 companies selected, 250 are used for the training data set and 91 are used for the validation (test) data set.

Step 4. Identify the internal and external company financial attributes to be used in the analysis (see Chapter 4).

Table 3-1 below shows a list of some of the typical internal and external company financial attributes analyzed. More than 145 different financial attributes, including year-to-year changes in each attribute, are evaluated in this research study. A survey of the existing literature and personal conversations with several business evaluation experts also helped to identify some of the specific attributes. The objective is to identify a list of potential financial attributes that may explain the future financial performance of small, technology-based companies.

Table 3-1. Example of Typical Internal and External Company Attributes.

Revenue	Net Income	Assets
Liabilities	Shareholders Equity	Gross Margin
Breakeven Margin	Working Capital	Interest Expense
Business Cycle Indicators	Gross Domestic Production	R&D Expense
Profit Growth	Return on Gross Profit	Equity Ratio
Turnover Ratios	Z-Score Bankruptcy Model	Operating Profit
Prime Rate	Cost of Capital	Leading Indicators

Step 5. Use a multiple linear regression analysis model, dependent variable being next-period EVA, to identify those individual regression coefficients for financial attributes that most significantly influence the dependent variable (see Chapter 5).

The regressor variables that most significantly influence the dependent variable, that is, maximize R^2 , are identified in this step using multiple linear regression analysis and the data for the 250 firms in the training set. The method used to identify the “best fit” multiple linear regression model is one that, for a given number of variables, produces the minimum error sum of squares, or equivalently, the maximum R^2 .

Two selection techniques are used to help identify the best fit multiple linear regression model. First, the financial variables are selected based on an examination of the selected model’s variance inflation factor, t-statistics on individual regression coefficients, and the model’s coefficient of determination (R^2). Second, the model variables are selected using an exhaustive search technique to find the model with the maximum R^2 . The exhaustive search technique begins by selecting one- and two-variable models and examines all possible pairwise interchanges with the variables not currently in the model. If any interchanges increase the R^2 , then that interchange resulting in the largest R^2 is retained. The process is repeated until the pair that results in the maximum R^2 is identified. At this point, after selecting a third variable, the interchange process is repeated. In both methods, examining the model’s residuals help to identify the best fit model. The final subset of variables for the model is selected based on the convergence of results from these two techniques.

There are several advantages to using multiple linear regression to identify significant variables. First, the method is an established and proven statistical technique. Second, the method is used as a quantitative technique for drawing

inferences about relationships between dependent and independent variables. A third advantage is its use as a technique to detect the degree of importance of *each* independent variable in explaining the variation in response to the dependent variable [67].

Step 6. Use statistical clustering to identify the number of data clusters in the training set. The number of fuzzy rules corresponds to the number of data clusters (see Chapter 6).

The nearest centroid sorting algorithm is used to identify the number of data clusters. Using this method, a set of cluster seeds is initially selected from each observation. The seeds are then replaced by the means of the temporary clusters using Euclidean distances. This iterative process continues until no additional changes occur in the clusters [68, 69]. The number of clusters in the training set acts as a guide in selecting the optimal number of fuzzy rules. For example, each cluster represents one fuzzy rule.

Step 7. Knowing the most significant factors from Step 5, and the number of fuzzy rules from Step 6, use a genetic algorithm[†] to search for the fuzzy rule set and fuzzy membership functions (see Chapter 7).

[†] Genetic algorithm software was written by Ryan Carroll, graduate student in the Department of Computer Science, The University of Tennessee at Knoxville.

As previously discussed, the classical approach to identifying the fuzzy rule set and membership functions uses an “expert” or “experts” to define the rule set based on their knowledge of the subject area. However, this dissertation suggests using a GA-based search technique to identify the “best fit” fuzzy rule set and fuzzy membership functions. Thus, use a genetic algorithm to search for the fuzzy rules and membership functions that best classify the training set.

Figure 3-1 shows a block diagram of the learning system. The genetic algorithm uses the training data (raw data in Figure 3-1) to learn the fuzzy rule set and membership functions.

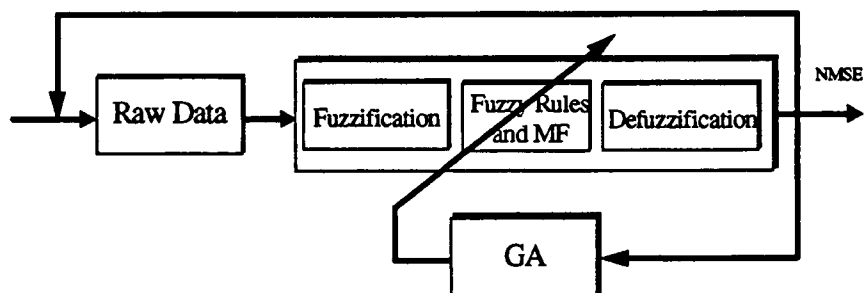


Figure 3-1. Block Diagram of GA-based Learning System.

The objective of the GA is to minimize the normalized mean square error (NMSE) between the known next-period EVA values in the training set and the estimated next-period EVA values using the fuzzy rule set and membership functions (MF in Figure 3-1). The genetic algorithm continues to evolve, or learn new rules,

until obtaining a minimum NMSE or reaching a preset generation. This research evaluates and tests four different multiple linear regression models and their respective fuzzy models.

Step 8. Develop the expert system using fuzzy rule sets and fuzzy membership functions obtained from the GA-based learning system (see Chapter 7).

Figure 3-2 shows the development of the fuzzy expert system using the information obtained in the previous research step. For example, to calculate the estimated next-period EVA, enter a company's current-period financial attributes into the EVA model containing the fuzzy rules and membership functions; using the fuzzy implication operator, the estimated next-period EVA is then calculated.

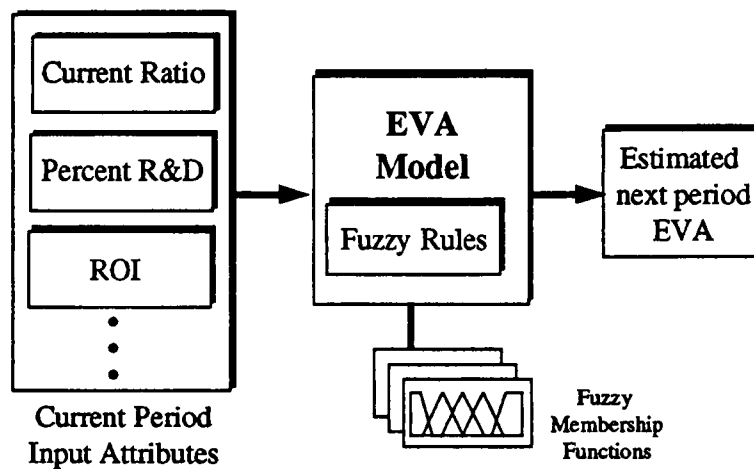


Figure 3-2. Block Diagram of the Fuzzy Expert System.

Step 9. Test the expert system by comparing results with a separate set of 91 companies (see Chapter 8).

To test the expert system rules and membership functions, compare them to the known EVA values for 91 companies (referred to as the validation or test set). In addition, compare the results of the four different expert system models with their respective multiple linear regression models.

Step 10. Evaluate the information content provided by the expert system and the multiple linear regression models (see Chapter 9).

A model is only as useful as the information content it provides for the application involved. For this reason, it is important to evaluate the information content provided by each of the models as well as their advantages and disadvantages.

C. Summary

The completion of these ten research steps discussed in this chapter leads to a better understanding of the significant independent variables and rule set used in predicting the financial performance of small, technology-based companies.

In summary, this research study uses an adaptive search technique; that is, a genetic algorithm, as the key element in the design and implementation of a robust concept learning system for financial classification of small, technology-based companies. The GA-based fuzzy rules and membership functions are evaluated using

fuzzy implication operators. The specific research includes ten steps starting with the identification of the dependent variable, EVA, and concluding with the validation of the expert system model developed. The knowledge domain for developing the learning systems consists of 250 small, technology-based, publicly owned firms, and the validation of the research approach uses 91 additional firms of the same type.

The following six chapters discuss in more detail the ten research steps outlined in this chapter. Chapter 4 begins by discussing the predicted value (dependent variable), EVA, and its relationship to market value. Next, Chapter 5 reviews regression analysis and discusses a method for identifying the significant variables used to predict a firm's next-period EVA based on the training data set. After selecting the significant variables, Chapter 6 discusses cluster analysis to determine the potential number of fuzzy rules. This research uses the k-mean method to determine the number of clusters or fuzzy rules.

After determining the significant independent variables and the number of fuzzy rules, Chapter 7 discusses how a genetic algorithm is used to evolve the best fit fuzzy rules and membership functions. Finally, the chapter concludes with a discussion of fuzzy sets and fuzzy logic and its use in developing a fuzzy expert system. Chapter 8 discusses the comparison of the multiple linear regression analysis model with the fuzzy expert system model using a separate set of 91 companies. Finally, Chapter 9 evaluates the information content provided by the multiple linear regression and the fuzzy expert system model.

Chapter 4

Research Knowledge Domain

A. Introduction

The expert system developed in this dissertation is based on a knowledge domain composed of the annual financial statements submitted to the Securities and Exchange Commission (SEC) from 1989 to 1993 for 250 small, publicly owned, technology-based companies. The 250 companies are referred to in this study as the training data set. The expert system's fuzzy rules and membership functions are derived from the training data set using a genetic algorithm, and are then validated using the annual financial statements submitted to the SEC from 1989 to 1993 for an additional 91 small, publicly owned, technology-based companies. The 91 companies are referred to in this study as the validation or test data set. Both the training and the validation data sets represent companies randomly selected from a much larger population of publicly owned, technology-based companies whose annual financial statements are compiled by Disclosure Inc. as outlined in Chapter 3 (Step 3).

This chapter includes an introduction to the Economic Value Added (EVA) method of economic performance measurement, and a summary of some important financial statistics of the combined training and validation data sets. These financial statistics are then compared with a statistical summary developed by Dun and

Bradstreet to verify that the training and validation data sets used in the study are a reasonable representation of a much larger population of technology-based companies.

B. Economic Value Added

All companies are in the business of acquiring and then allocating scarce capital resources to projects that will generate after-tax cash flows (ATCFs) with positive present equivalent values. Because of the relationship between the financial objectives of a firm and its market value, it is essential that the present values of the ATCFs generated from capital invested in projects by the company add to its economic wealth.

What determines the market value of a company, as depicted by the price per share of common stock, is therefore a function of its total after-tax cash flow. Specifically, the present value of the total ATCF of a firm prior to the allocation of any financing costs for a period, or the present value of the operating profit after-tax cash flow (OPATCF), is the incremental value that the owners can expect to achieve during the time period. The market value of a firm is then the present value of the incremental values achieved plus the original capital required to generate the cash flows. Simply, the incremental after-tax cash flow for a time period that contributes to increasing the market value of a company is that portion remaining after subtracting the cost of capital for the period. Unlike other economic analysis procedures, financial capital in this research includes debt capital as well as the equity capital invested by a

firm's shareholders. Therefore, a firm's market value equals the total capital invested in the firm plus the present value of the periodic incremental after tax cash flows generated by the firm resulting from the invested capital.

Bennett Stewart of Stern Stewart and Company [70] refers to the operating profit after-tax cash flow, less the cost of capital employed to produce the cash flow, as the firm's Economic Value Added (EVA). For example, EVA will increase if a positive cash flow can be generated without requiring any additional capital. However, if the company needs new capital, EVA will only be increased if the return on the invested capital is greater than the cost of capital. Discounting the periodic EVA over the life of the firm, and adding the amount of capital employed, is a measure of the firm's market value. Therefore, if the EVA for a time period is positive, wealth is added to the firm because operating after-tax cash flow exceeds the cost of the capital employed. If EVA is negative, the cost of capital employed exceeded the OPATCF. Therefore, the firm should consider alternative operating strategies including identifying projects with a higher rate of return or eliminating projects with a lower rate of return. In essence, a negative EVA results in a decrease in the wealth of the firm.

Mathematically, EVA (for any time period) is based on the difference between the rate of return, r , that a company earns within its business and the firm's required return, c^* , times the capital employed in the company. The current weighted average cost of capital, c^* , for a firm is the average return available to investors and lenders in

other investment opportunities with comparable levels of risk. Again, capital is defined to include both the debt and equity capital invested in the company.

For example, if a firm's first-year return on investment is 20 percent and the investor requires a 15 percent return on capital of \$100, the EVA would be \$5. Considering a one-year period, the market value would be \$105. However, if the firm's return is 10 percent, the market value is decreased to \$95, again considering a one-year period. Here, the EVA would be a negative \$5. It is obvious that a firm should only invest in projects that earn a return in excess of the cost of capital to improve their market value. The rate of return generated by the firm, r , is equal to the ATCF divided by the amount of capital. The ATCF can also be viewed as the amount of available cash after taxes to all financial providers of capital to the firm. Therefore, $EVA = (r - c^*)(\text{capital}) = \text{operating profits} - \text{cost of capital}$ [71].

Thus, EVA is the residual income, or operating profits, less a charge for the use of capital. That is, EVA is the difference between the profits a company derives from the use of capital, and the charge for capital incurred by the investors [72]. From examining the EVA calculation, a firm has three options for generating positive EVA as stated below [73]:

1. Increase after-tax cash flow without requiring additional capital.
2. Increase capital so long as the additional capital is invested in projects that generate a return greater than the cost of capital.

3. Free capital and return the capital to investors so long as the earnings lost is more than offset by a savings in the cost of capital.

Researchers have conducted several studies to verify the relationship between the market value of a firm and its' EVA. Remember, a firm's market value is a measure of the value a company has created for its shareholders. Thus, the market value of a firm is the stock market's view of the firm's ability to generate a positive EVA, and hence productively utilize scarce capital resources. A statistical analysis conducted in 1992 [74] compared the market value of 1,000 companies and their EVA. The study performed a regression analysis of the companies market value compared with some common performance indicators such as return on equity, cash flow growth, earnings-per-share growth, asset growth, dividend growth, and sales growth. In the study, no indicators had a coefficient of regression (R^2) of greater than 0.35 except EVA which had a correlation of 0.50. Consequently, EVA was able to explain as much as 50 percent of the variation in the market value of a firm, whereas a common performance indicator like return on equity (used by many financial analysts) was only able to explain 35 percent of the variation in market value, and only nine percent of the variation in market value was explained by sales growth. Several other studies found similar comparisons [75].

This study uses EVA to measure a firm's financial performance. The next two sections summarize some of the important financial attributes or variables used to predict EVA in this research study.

C. Financial Summary

A review of the financial data from the 341 technology-based companies (includes the training and validation data sets) indicates that the average EVA was a negative \$2.3 million per year, with a standard deviation of \$4.6 million. The annual EVA ranged from \$24.8 million to a negative \$30.5 million, with a median value of negative \$923,000. Of the 341 companies, 25 percent had a positive annual EVA.

The average assets of the 341 companies were \$57.4 million. The total assets ranged from \$743.5 million to \$0, with a median of \$30.8 million. Seventy-five percent of the companies had total assets greater than \$12 million.

The average amount of shareholder's equity for the 341 companies was \$22.6 million, with a standard deviation of \$27.1 million and ranged from \$224.8 million to negative \$158.5 million. The median shareholder's equity was \$13.8 million. Seventy-five percent of all companies had shareholder equity amount greater than \$5 million.

The average liabilities of the 341 companies were \$34.7 million. The total liabilities ranged from \$751.9 million to \$0. The median value was \$14 million, and seventy-five percent of the companies had total liabilities greater than \$4.5 million.

The average debt to equity ratio of the 341 companies was 1.148. According to a 1993 study by Dun and Bradstreet [76], the average debt to equity ratio of 5,292 technology-based companies were 0.923.

The average annual sales of the 341 companies were \$72.5 million. The annual sales ranged from \$1.1 billion to \$0, with a median of \$40.6 million. Seventy-five percent of the companies had total sales of greater than \$17 million.

The average gross margin of the 341 companies in the training and validation data set was 33.9 percent, with a standard deviation of 0.4 percent. The gross margin ranged from 100 percent to 1.5 percent, with a median value of 31.8 percent. Seventy-five percent of the 341 companies in this study had a gross margin greater than 23.7 percent. According to a 1993 survey by Dun and Bradstreet [77], the average gross margin of 5,292 technology-based companies were 38 percent.

The average annual amount of research and development expenses for the 341 companies reviewed in this research was \$1.8 million, with a standard deviation of \$103,000. The annual research and development expenses ranged from \$38.0 million to \$0 with a median value of \$94,000. The average annual percentage of research and development expenses to net annual sales was 3.4 percent, with a standard deviation of 4.8 percent. The annual research and development expenses to net annual sales ranged from 38 percent to 0 percent, with a median value of 0.5 percent.

The average amount of annual sales, general and administrative expenses of the 341 companies in this study was \$15.6 million. The total annual sales, general

and administrative expenses ranged from \$134 million to \$0, with a median value of \$9.3 million. The average annual percentage of total sales, general and administrative expenses to net annual sales was 25.6 percent, with a standard deviation of 13.9 percent. The total annual sales, general and administrative expenses to net annual sales ranged from 147 percent to 0 percent, with a median value of 23 percent. Seventy-five percent of all companies had total annual sales, general and administrative expenses to net annual sales of greater than 7 percent.

The average annual net income for the 341 companies was \$398,000, with a standard deviation of \$7.6 million. The net annual income ranged from \$60.1 million to a negative \$66.8 million, with a median value of \$748,000. The average net annual income to net annual sales was 1.4 percent, with a standard deviation of 14.6 percent. The net annual income to net annual sales ranged from 43.6 percent to a negative 96 percent, with a median value of 2 percent. Twenty-five percent of the 341 companies in this study had total net annual income to net annual sales of greater than 5 percent. According to a 1993 survey by Dun and Bradstreet [78], the average net annual income to net annual sales of 5,292 technology-based companies were 4.37 percent.

D. Independent Financial Variables

More than 145 specific financial attributes, including year-to-year changes in each attribute, were evaluated in this research in the selection of independent variables for the regression models. Those familiar with financial terminology associated with

financial Income and Balance Sheet statements, may wish to proceed to Chapter 5, “Independent Variable Selection.” Otherwise, several of the variables found in this research to be significant and thus used in predicting a firm’s next-period EVA are briefly discussed below.

The Income Statement, often referred to as a firm’s Profit and Loss Statement, represents the results of earnings activity during a stated period, usually one year. In contrast, the Balance Sheet represents the firm’s assets, liabilities, and equities at a specific time. Described below are several attributes used in this study to describe a firm’s Income Statement. Also included is the abbreviation used in this dissertation for describing each attribute.

- | | |
|-----|---|
| NSA | <u>Net SA</u> les. Net sales, often referred to as net revenues, is a measure of the inflows from selling goods and providing services to customers during a stated financial period. |
| COG | <u>C</u> ost <u>Q</u> f <u>G</u> oods. Cost of goods is the combined cost of three major cost elements; direct materials, direct labor, and indirect manufacturing costs. Indirect manufacturing costs includes all costs other than direct material and direct labor that are associated with the manufacturing process. Other terms describing indirect manufacturing costs include factory overhead, factory burden, and manufacturing overhead. |

GP	<u>G</u> ross <u>P</u> rofit. Gross profit is the excess sales after cost of goods expense. The gross profit is the contributing margin available for non-manufacturing expenses such as research and development expenses; sales, general and administrative expenses; and interest expenses.
OPE	<u>O</u> perating <u>E</u> xpenses. Costs incurred in sales, general and administrative, advertising, research and development, and interest. In general, operating expenses also includes depreciation.
SGAE	<u>S</u> ales, <u>G</u> eneral and <u>A</u> ddministrative <u>E</u> xpenses. Expenses associated with selling and marketing a firm's products and services. Expense also includes the costs associated with the general and administrative operations of the firm.
RDE	<u>R</u> esearch and <u>D</u> evelopment <u>E</u> xpenses. Expenses associated with the research and development of the firm's products and services.
IE	<u>I</u> nterest <u>E</u> xpenses. Expenses associated with the payment for the use of a firm's debt capital.
PTI	<u>P</u> re- <u>T</u> ax <u>I</u> ncome. The difference between gross profit and operating expenses.
NI	<u>N</u> et <u>I</u> ncome. Net income is the cumulative earnings or profits after taxes.

The Balance Sheet is a listing of the firm's assets, liabilities, and shareholder's equity at a specified time. An asset is an item that has the ability or potential to provide future services or benefits to a firm. A liability is a creditor's claims on the resources of a firm. Shareholder's equity is the owner's claim on the firm's assets. Below is a description of several attributes used in this study in describing a firm's Balance Sheet. Also included is the abbreviation used in this dissertation for describing each attribute.

CA	<u>C</u> urrent <u>A</u> ssets. Cash and other assets that are expected to be turned into cash, sold, or exchanged within the normal operating cycle of the firm, usually one year. Current assets include cash, marketable securities, receivables, inventory, and current prepayments.
TA	<u>T</u> otal <u>A</u> ssets. Total assets include all current assets plus additional items such as plant and equipment, long-term securities, patents, goodwill, land, and accumulated depreciation.
CL	<u>C</u> urrent <u>L</u> iabilitys. A debt or other obligation that must be discharged within a short period of time, usually one year.
LTL	<u>L</u> ong <u>T</u> erm <u>L</u> iabilitys. Loans and notes payables usually over a period of more than one year.
TL	<u>T</u> otal <u>L</u> iabilitys. Total liabilities is the sum of current and long-term liabilities.

SHE Shareholder's Equity. Shareholder's equity is the difference between total assets and total liabilities. Shareholder's equity is also the sum of capital stock and retained earnings.

WC Working Capital. A measure of the margin of protection for current creditors and reflects the ability to finance current operations.

Working capital is the difference between current assets and current liabilities.

Ratio analysis is a common financial method used to compare two or more items. Just as the ratio MPG (Miles Per Gallon) measures the miles an automobile can travel per gallon of gasoline, financial ratios also measure the amount of one item when compared to another item. The resulting ratios can then be conveniently compared to other firms and used as a benchmark in making financial decisions.

Used by financial analysts for more than 200 years, ratio analysis is a key component in the banking industry. Robert Morris Associates was among the early pioneers in ratio analysis. In 1928, two individuals associated with Robert Morris wrote a classic book on financial ratio analysis titled, *Ratio Analysis of Financial Statements* [79]. This book was the first comprehensive text on ratios useful in measuring a firm's financial strength. *101 Business Ratios* [80] is another useful text on ratio analysis.

The usefulness of ratio information is limited by the understanding and meaning of the terms. Some common important uses for business financial ratios include: measuring a firm's profitability and return on investment, identifying trends, diagnosing problems, comparing one company and one time period to another, and measuring managerial efficiency. Listed below is a short discussion of several of the ratios found to be significant in this research and used in predicting a firm's next-period EVA. Also included are the abbreviations used to describe each ratio in this research.

- LTL~~S~~HER** Long-Term Liabilities to SHareholder's Equity Ratio. This ratio is a measure of what is owed by the firm to what is owned. When the ratio is greater than 1.0, capitalization by lenders is greater than that of shareholders.
- QR** Quick Ratio. Quick ratio is also known as the "acid test" and is cash plus marketable securities and receivables divided by current liabilities. The ratio is a measure of a firm's liquidity. A ratio of 1.0 or greater is strongly desirable.
- ATR** Asset Turnover Ratio. This ratio is a measure of a firm's net sales divided by total assets. ATR measures a firm's productive use of their assets. A higher ratio is favorable.
- TL~~T~~AR** Total Liabilities to Total Assets Ratio. This ratio is a measure of what is owed by the firm to the value of the assets used by the business. It

measures how successful the business is in using debt to build the business.

- TLSSHER** Total Liabilities to Shareholder's Equity Ratio. Measures the relationship between capital provided by creditors and that provided by investors. A high ratio indicates a greater risk being assumed by creditors. A firm with a lower ratio usually has greater freedom in borrowing capital.
- BEM** Break-Even Margin. Percentage of a firm's total expenses including the cost of goods; research and development expenses; sales, general and administrative expenses; and interest expenses divided by net sales. The larger the break-even margin, the greater a firm's expenses, and therefore, the less profitable the firm. The smaller the break-even margin ratio, the more favorable is the firm.
- BESF** Break-Even Sales Factor. Percentage of net income divided by a firm's total expenses including the cost of goods; research and development expenses; sales, general and administrative expenses; and interest expenses. The larger the break-even sales factor, the more profitable is the firm. The break-even sales factor shows by what percentage sales must change for break-even operations to occur.
- NSWCR** Net Sales to Working Capital Ratio. Working capital is the differences between a firm's current assets (cash, inventory, notes, and

accounts receivable) minus current liabilities (notes and accounts payable, accrued expenses, and taxes). This ratio is calculated as net sales divided by working capital.

- PTICR** Pre-Tax Income to Invested Capital Ratio. This ratio is a measure of a firm's pre-tax income divided by shareholder's equity and long-term liabilities.
- NSCAR** Net Sales to Current Assets Ratio. This ratio is a measure of a firm's net sales divided by total current assets. The ratio measures a firm's productive use of their current assets. A higher ratio is favorable.
- NSPER** Net Sales to Plant and Equipment Ratio. Measures the production used by fixed assets in generating sales.
- RDENSR** Research and Development Expenses to Net Sales Ratio. This ratio measures the relationship between research and development expenses and net sales. Generally, in a young, technology-based company, this ratio will be high due to the need for research and development expenses to develop competitive products and services.
- PTISHER** Pre-Tax Income to SHareholder's Equity Ratio. Expresses the rate of return before tax on capital provided by investors.
- PTTAR** Pre-Tax Income to Total Assets Ratio. Pre-tax income to total assets ratio expresses the pre-tax return on total assets. This measures the effectiveness of the firm in employing the resources available to it.

- SCR Sales to Cash Ratio. This ratio is a measure of a firm's net sales divided by total cash assets. This ratio measures a firm's productive use of cash. A higher ratio is favorable.
- SGAENSR Sales, General and Addministrative Expenses to Net Sales Ratio. This ratio measures the relationship between sales, general and administrative expenses to net sales. Sales, general and administrative expenses are those costs associated with selling and marketing a firm's products and services. SGAENSR also includes expenses associated with the general and administrative operations of the firm.
- SHER Shareholder's Equity Ratio. This ratio measures the relationship between shareholder's equity and long-term liabilities.

E. Summary

The expert system discussed in this dissertation is developed using yearly financial statements from 1989 to 1993 for 250 small, publicly owned, technology-based companies. The 250 companies are referred to as the training data set. The expert system's fuzzy rules and membership functions are derived from the training data using a genetic algorithm and validated using financial statements from 1989 to 1993 of an additional 91 small, publicly owned, technology-based companies. The 91 companies are referred to as the validation data set.

Both the training and the validation data set are randomly selected from a much larger population of publicly owned, technology-based companies whose yearly financial statements are compiled by Disclosure Inc.

This chapter included an introduction to Economic Value Added (EVA) as a measure of operating and financial performance, and a summary of some important financial statistics of the combined training and validation data set. These financial statistics are compared with the Dun and Bradstreet statistical summary, and are found to be a reasonable comparison to a much larger population of technology-based companies.

Chapter 5, "Independent Variable Selection" begins with an introduction to multiple linear regression analysis and concludes with the identification of the significant independent financial variables useful in predicting a firm's next-period EVA.

Chapter 5

Independent Variable Selection

A. Introduction

Variable selection, often referred to as variable filtering, filters out irrelevant attributes or independent variables before the induction process occurs. The selection of significant independent variables for inclusion in a multiple linear regression equation is a familiar problem in research. Fortunately, theories on hypothesis testing provide guidelines in specifying which variables are important. Often, however, several variables are available for evaluating a particular attribute and little basis exists for a specific choice.

In this research, more than 145 different independent financial variables are reviewed to determine their significance in predicting the dependent variable (a firm's next-period EVA). Of the 145 variables, 114 are considered internal variables; that is, variables that are associated with a firm's financial statements as discussed in the previous section. The remaining 31 variables are considered external economic variables, including general business cycle indicators such as the prime interest rate, new orders by manufacturers, the average work week of production workers, contracts and orders for plant and equipment, inventories in the manufacturing sector, an index of stock prices, the S&P 500 composite index, indexes of industrial production, the composite index of leading economic indicators, and the gross

domestic product. In addition, the period-to-period changes in the 145 variables are also evaluated.

This chapter begins with an introduction to multiple linear regression analysis and concludes with the selection of four best fit multiple linear regression models for predicting a firm's next-period EVA.

B. Multiple Linear Regression Analysis

A linear regression model that involves more than one variable is referred to as a multiple linear regression model. For example, suppose we choose the following relationship

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \varepsilon_i \quad (\text{B-1})$$

where y_i represents a firm's next-period EVA and x_{ki} represents the independent variables, or attributes, that influence y_i the dependent variable. The parameters β_k are the regression coefficients. The error term, ε_i , represents the difference between the known y_i and the predicted $\hat{y}_i$ value. The regression analysis model assumes that each observation, y_i , can be described by equation B-1 where ε_i is a random error with mean zero and variance σ^2 . A description of a simple linear regression model involving only one independent variable is shown in Figure 5-1.

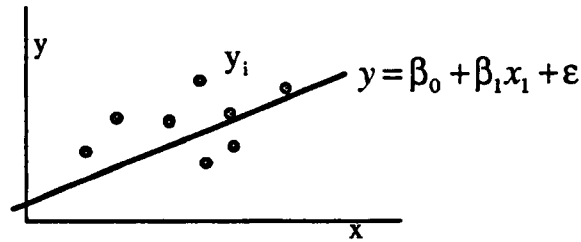


Figure 5-1. Simple Linear Regression.

The term “linear” is used to refer to the regression equation because it is a linear function of the unknown regression coefficients, β_k . The method of least squares is used to determine the values of the regression coefficients that minimize the error term, ϵ_i , in equation B-1 for an application.

Using matrix notation, the formula $y = X\beta + \epsilon$ can be used to model the observed data based on equation B-1. It is simpler to solve for the least square estimators when the observed data are presented in matrix notation. A general set of observed data is shown below where n is the number of observations and k is the number of variables.

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdot & \cdot & \cdot & x_{k1} \\ 1 & & & & & & \\ \cdot & & & & & & \\ \cdot & & & & & & \\ \cdot & & & & & & \\ 1 & x_{1n} & \cdot & \cdot & \cdot & \cdot & x_{kn} \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \cdot \\ \beta_k \end{bmatrix} \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \cdot \\ \cdot \\ \cdot \\ \epsilon_n \end{bmatrix}$$

The minimum error, ϵ , is found by using the least square estimates, $\mathbf{b} = b_0, b_1, \dots, b_k$, for the regression coefficient vector β . The error sum squares, SS_E , is calculated as shown in equation B-2 where y_i is the observed value and $\hat{y}_i$ is the value predicted by the regression model. Thus, the SS_E is the sum of squares of the residuals, that is, the amount of variation in the dependent variable, y , not explained by the multiple linear regression model. The objective of the least square estimator vector, $\mathbf{b}$, is to minimize the SS_E .

$$\text{Sum Square Error} = SS_E = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = (\mathbf{y} - \mathbf{Xb})^T (\mathbf{y} - \mathbf{Xb}) \quad (\text{B-2})$$

and equation B-2 can be simplified as follows:

$$SS_E = \mathbf{y}^T \mathbf{y} - \mathbf{b}^T \mathbf{X}^T \mathbf{y} - \mathbf{y}^T \mathbf{Xb} + \mathbf{b}^T \mathbf{X}^T \mathbf{Xb} = \mathbf{y}^T \mathbf{y} - 2\mathbf{b}^T \mathbf{X}^T \mathbf{y} + \mathbf{b}^T \mathbf{X}^T \mathbf{Xb}$$

The minimum SS_E is found by taking the partial derivative of SS_E with respect to $\mathbf{b}$.

$$\frac{\partial SS_E}{\partial \mathbf{b}} = -2\mathbf{X}^T \mathbf{y} + 2(\mathbf{X}^T \mathbf{X})\mathbf{b} = 0$$

$$\text{thus, } \mathbf{X}^T \mathbf{y} = (\mathbf{X}^T \mathbf{X})\mathbf{b}$$

$$\text{and, } \mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

where $\mathbf{b}$ is the least squares estimator of the vector β .

$$\mathbf{b} \text{ is a minimum because } \frac{\partial^2 SS_E}{\partial \mathbf{b}^2} > 0.$$

Consequently, the best fit multiple linear regression model that results in minimizing the error, ϵ , is $\hat{\mathbf{y}} = \mathbf{Xb}$.

The mean square error (MSE) of the model is defined as the SS_E divided by its degrees of freedom. The normalized mean square error (NMSE) is defined as the MSE divided by the σ_y^2 . The NMSE and the MSE are frequently used to determine how well the observed data fits the model.

The $\mathbf{X}^T \mathbf{X}$ matrix is a $(p \times p)$ symmetric matrix where p is the total number of model parameters. The diagonal elements of $\mathbf{X}^T \mathbf{X}$ are the sums of squares of the elements in the columns of $\mathbf{X}$, and the off-diagonal elements are the sums of cross-products of the elements in the columns of $\mathbf{X}$. The $\mathbf{X}^T \mathbf{X}$, when centered and scaled, is called the correlation matrix. The correlation matrix, $\mathbf{R}$, takes the form

$$\mathbf{R} = \begin{bmatrix} 1 & r_{12} & r_{13} & \dots & r_{1k} \\ & 1 & r_{23} & \dots & r_{2k} \\ & & & \ddots & \\ & & & & r_{k-1,k} \\ & & & & 1 \end{bmatrix}$$

where $r_{k-1,k}$ is the simple linear correlation between x_{k-1} and x_k . When $r_{k-1,k}$ is close to ± 1 there is a strong correlation between x_{k-1} and x_k .

In most multiple linear regression problems, some of the variables being modeled as independent variables may actually be intercorrelated (referred to as multicollinearity). Multicollinearity can have serious effects on the estimates of the regression coefficients and the general applicability of the selected model. In addition, multicollinearity also inflates the variance of the regression coefficients as will be seen

in the following derivation. For this reason, detection of multicollinearity in selecting the regression model variables is essential.

Figure 5-2 below and equation B-3 show that the total variability of an observation, y_i , about the mean of all the observations, $\bar{y}$, can be partitioned into three sums of squares [81].

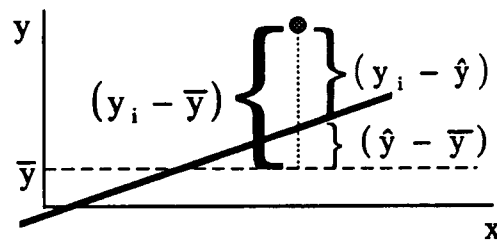


Figure 5-2. Regression Variability.

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2 = SS_T = SS_R + SS_E \quad (B-3)$$

that is,

$$\left(\begin{array}{c} \text{Total Variability} \\ \text{In Response, } SS_T \end{array} \right) = \left(\begin{array}{c} \text{Variability Explained} \\ \text{by Model, } SS_R \end{array} \right) + \left(\begin{array}{c} \text{Variability} \\ \text{not explained} \\ \text{by model, } SS_E \end{array} \right)$$

$$R^2 = \frac{\text{Variability Explained by Model}}{\text{Total Variability In Response}} = \text{Coefficient of Determination} = \frac{SS_R}{SS_T}$$

R^2 , the Coefficient of Determination, is often used to judge the adequacy of a regression model. When $R^2 = 1$, SS_E and $NMSE = 0$, thus the regression equation is a perfect fit to the observed data. The R^2 value is one criterion used in selecting the appropriate regression model.

In addition to its use in calculating R^2 , the SS_T value is used in selecting the appropriate regression model. Regression models assume that $E(\epsilon) = 0$, and that the variance is a function of the error variance, thus $V(\epsilon) = V(y) = \sigma^2$, or $\sigma_y^2 \equiv \sigma_\epsilon^2$.

$$\text{Therefore, } \sigma_y^2 \equiv \hat{S}_\epsilon^2 = \frac{(\mathbf{y} - \mathbf{Xb})^T (\mathbf{y} - \mathbf{Xb})}{n - p} = \sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{n - p} = \frac{SS_E}{n - p} = \text{MSE},$$

where p is the number of parameters in the model and n is the number of observed data points [82].

In addition to R^2 and the model variance, the estimated variance of the regression coefficients is also used in selecting the appropriate regression model. The variance estimates of the regression coefficients, $\mathbf{b}$, are required for testing hypotheses about the regression coefficients. The determination of the variance for the regression coefficients is shown below.

$$\sigma_b^2 = \text{VAR}(\hat{\mathbf{b}}) = \text{VAR}\left[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}\right] = \left[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T\right] \text{VAR}(\mathbf{y})$$

$$\text{Note: } \text{VAR}(\mathbf{A}\mathbf{y}) = \mathbf{A} \text{VAR}(\mathbf{y}) \mathbf{A}^T, \text{ Let } \mathbf{A} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$$

$$\text{and } \left[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T\right]^T = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1}. \text{ VAR}(\mathbf{y}) = \sigma_\epsilon^2 \mathbf{I}$$

$$\text{Then, } \sigma_b^2 = \left[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T\right] [\sigma_\epsilon^2 \mathbf{I}] \left[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T\right]^T$$

Note: $(AB)^T = B^T A^T$, $(A^T)^{-1} = (A^{-1})^T$ and $(A^T)^T = A$

Then, $\left[(X^T X)^{-1} X^T\right]^T = X \left[(X^T X)^{-1}\right]^T = X \left[(X^T X)^T\right]^{-1} = X(X^T X)^{-1}$

and,

$$\sigma_b^2 = \left[(X^T X)^{-1} X^T\right] [\sigma_\epsilon^2 I] \left[X(X^T X)^{-1}\right] = \sigma_\epsilon^2 (X^T X)^{-1} (X^T X) (X^T X)^{-1} = \sigma_\epsilon^2 (X^T X)^{-1}$$

$$\text{Thus, } \sigma_b^2 = \sigma_\epsilon^2 (X^T X)^{-1}. \quad (\text{B-4})$$

This implies that the variance of the vector b coefficients appears on the diagonal elements of $(X^T X)^{-1}$ apart from the error variance σ_ϵ^2 . The covariance appears on the off-diagonal elements.

In multiple linear regression, certain tests of hypotheses about individual regression coefficients are useful in measuring the adequacy of the model. The test of the significance of the regression coefficients determines if there is a linear relationship between the dependent variable and one or more of the independent variables. The appropriate hypotheses are:

$$H_0: \beta_i = 0$$

$$H_A: \beta_i \neq 0$$

Rejection of $H_0: \beta_i = 0$ implies that the independent variable x_i contributes significantly to the model and should not be deleted from the model. To test the hypotheses, the σ_b^2 must be known or estimated from the data set. Since the true σ_b^2

is unknown, the t-distribution is used to test the hypotheses. The criterion for rejecting the hypothesis is:

$$|t_0| > t_{\frac{\alpha}{2}, n-p} \quad \text{where } |t_0| = \left| \frac{b_i}{\sigma_b} \right|$$

and, $\sigma_b = \sqrt{\sigma_\epsilon^2 \text{diag}_i(\mathbf{X}^T \mathbf{X})^{-1}} = \sigma_\epsilon \sqrt{\text{diag}_i(\mathbf{X}^T \mathbf{X})^{-1}}$ from equation B-4.

$$\text{Thus, if } \left\{ |t_0| = \left| \frac{b_i}{\sigma_\epsilon \sqrt{\text{diag}_i(\mathbf{X}^T \mathbf{X})^{-1}}} \right| > t_{\frac{\alpha}{2}, n-p} \right\} \text{ then reject } H_0. \quad (\text{B-5})$$

As seen in the derivation of equation B-5, the factor that inflates the σ_b value is a function of the $(\mathbf{X}^T \mathbf{X})^{-1}$ diagonal coefficients. These coefficients, $\text{diag}_i(\mathbf{X}^T \mathbf{X})^{-1}$ are called the Variance Inflation Factors (VIF).

Underfitting and overfitting the data is another important aspect to consider in model selection. Underfitting is defined as ignoring important model terms needed to fit the data set, and overfitting is defined as including terms in the model that have minimum marginal effects. Overfitting results in a model that performs well on the training data set but may perform poorly on the validation data set. Thus, the overfit model is including (fitting) noise in the training data set that may not be present in the validation data set. One criterion used to assist in determining the model fit is the Mallows's criterion, C_p . Equation B-6 shows the C_p calculation [83].

$$C_p = p + \frac{(\text{MSE of reduced model} - \text{MSE of full model})(n - p)}{\text{MSE of full model}} \quad (\text{B-6})$$

When C_p is equal to p , the number of parameters in the model, a proper compromise is reached between underfitting and overfitting the data.

The selection of the independent variables for the multiple linear regression models in this research is based on standard statistical techniques. This includes the evaluation of the R^2 value, the R matrix, the VIF values, hypotheses testing of the regression coefficients, C_p analysis, and analysis of the model residual values.

In summary, the strategy for selecting the best fit regression equation uses the method of least squares, and is outlined in the steps below:

- Step 1. Accomplish initial exploratory data analysis.
- Step 2. Develop one or more tentative regression models. In most cases, several of the independent variables will be intercorrelated or multicollinear which can inflate the variance of the regression coefficients. For this reason, an initial step in identifying tentative regression models is to start with a full-model that incorporates all possible independent variables and identify all variables that tend to be intercorrelated. This usually eliminates several of the independent variables.
- Step 3. Revise the regression model and/or develop a new model. Based on the remaining independent variables, revise the model and identify those variables that are more significant in predicting the dependent variable.
- Step 4. Identify the most suitable model and test the model with the validation data set.

Step 5. Make inferences on the basis of the regression model.

After selecting the significant independent financial variables, the next step in developing the expert system is to identify the number of data clusters given the selected variables. Chapter 6 discusses the method used to identify the number of data clusters.

C. Experimental Results - Independent Variable Selection

Stepwise regression, forward selection, backward elimination, and MaxR procedures are used in the variable selection process to help identify the most significant independent variables. However, it is well known that just using these variable selection techniques does not always yield the best set of independent variables for describing the relationship to a dependent variable. As a result, several iterative approaches are necessary along with a thorough evaluation of the data set to arrive at a "best fit" multiple linear regression model.

The independent variables determined using stepwise, forward, backward, and MaxR procedures are compared and modified as necessary to arrive at a final set of significant variables. Surprisingly, in this research, none of the 31 external economic variables were significant at a p-value of less than 7 percent in predicting a firm's next-period EVA. In addition, since the data set is compiled from financial statements from 1989 to 1993, a check of autocorrelation is performed to determine

if the training set data is time series dependent. The formal statistical test for detecting autocorrelation, the Durbin-Watson test, is performed along with other hypothesis testing techniques on the data set. None of the independent variables identified appeared to exhibit significant autocorrelation.

In addition to autocorrelation, it is also important to detect the existence of multicollinearity. Multicollinearity is concerned with the linear dependencies of multiple variables. When several variables are linearly dependent with each other, it becomes more difficult to sort out the impact of each regressor on the response. In addition, multicollinearity can severely inhibit the ability to estimate regression coefficient values and their interpretation in understanding the problem. In summary, when multicollinearity is present, the variances of the regression coefficients are inflated. This inflation of variance is easily detected by calculating the eigenvalues of the correlation matrix. Not surprisingly, several of the regressor variables were found to be collinear with other variables and as a result are not included in the final multiple linear regression equation.

The selection of the independent variables that are significant in explaining the dependent variable, a firm's next-period EVA, is determined as discussed in Section B of this chapter. This research developed several multiple linear regression models. In all models, however, the most significant independent variable is the firm's current-period EVA (CPEVA).

When including only the firm's current-period EVA as the single regressor variable for predicting a firm's next-period EVA (NPEVA), the single regressor variable explains 52.7 percent of the variation in a firm's next-period EVA. For example, the single linear regression equation, $NPEVA = - \$955.78 + 0.628(CPEVA)$, explains 52.7 percent of the variation in the predicted response, NPEVA. Analysis of the residuals indicate they are normally distributed with a mean of zero.

The CPEVA regression coefficient, 0.628, is the unit dollar increase in the firm's next-period EVA per dollar increase in the firm's current-period EVA. The analysis of variance (ANOVA) table for the single regression equation appears in Table 5-1. From Table 5-1, the F-statistic is significant at less than 0.0001 indicating that there is, indeed, a statistically significant linear trend between a firm's next-period EVA and the current-period EVA. Also shown in Table 5-1 is the parameter estimate which indicates that the regression coefficient for the independent variable, current-period EVA, is statistically significant in explaining the next-period's EVA. Figure 5-3 is the leverage plot of the linear regression model when only including the single regressor. The leverage plot shows the relationship between the actual NPEVA and the predicted NPEVA. In Figure 5-3, the distance from the observation to the mean NPEVA is the actual residual without the linear fit. The distance from the observation to the regression line is the residual or error associated with the linear model. The leverage plot also shows the confidence curves for the linear fit. Clearly,

Figure 5-3 and Table 5-1 indicate that the CPEVA is a significant variable in predicting a firm's next-period EVA.

The t-statistic, equation B-5, is used for testing whether the individual regression coefficient, term CPEVA in Table 5-1, is zero (the null hypothesis, H_0). The t-statistic is summarized in the column labeled t-Ratio in Table 5-1. The next column, Prob>|t| gives the p -value for the two tailed t-statistic test. For example, the p -value of 0.0000 for CPEVA states that if you reject the null hypothesis, H_0 : (CPEVA = 0), there is a 0.0000 chance of erroneous rejection therefore indicating that the CPEVA is a significant variable in predicting a firm's next-period EVA.

Table 5-1. ANOVA and Parameter Estimates (CPEVA).

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob>F
Model	1	1.13e+10	1e+10	1112.26	0.0001
Error	998	1.01e+10	10131233		
C Total	999	2.14e+10			

Term	Estimate	Std Error	t-Ratio	Prob> t
Intercept	-955.7791	109.63	-8.72	0.0000
CPEVA	0.6281179	0.01883	33.35	0.0000

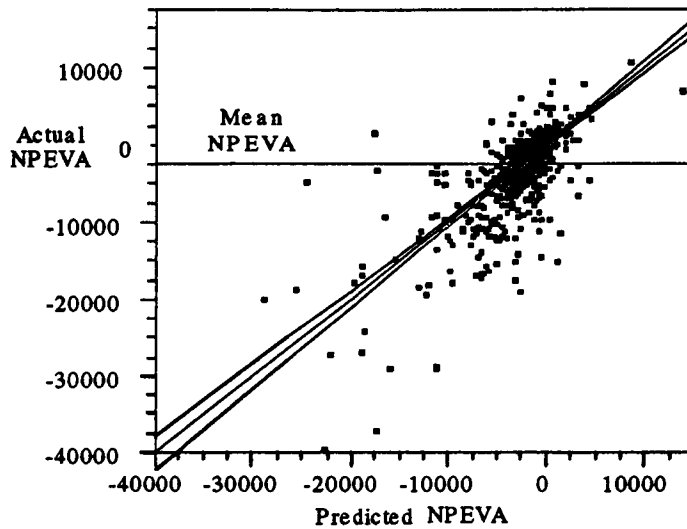


Figure 5-3. Leverage Plot of Next-period EVA vs. Predicted Next-period EVA.

When the second-order effect of CPEVA is included in the model, the R^2 value increases from 52.7 percent to 53.05 percent. Therefore, adding the second-order effect shows very little reduction in the MSE. The entries in Table 5-2 show the ANOVA results and the parameter estimates when both the first- and second-order effects of CPEVA are included.

Table 5-2. ANOVA and Parameter Estimates (CPEVA, CPEVA²).

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob>F
Model	2	1.13e+10	6e+9	563.2320	0.0000
Error	997	1e+10	10068235		
C Total	999	2.14e+10			

Term	Estimate	Std Error	t-Ratio	Prob> t
Intercept	-919.0085	110.14	-8.34	0.0000
CPEVA	0.688711	0.0293	23.49	0.0000
CPEVA ²	0.000003	0	2.69	0.0072

Figure 5-4 shows the regression plot of the first- and second-order CPEVA terms and the 95 percent confidence intervals on mean response (center line in Figure 5-4).

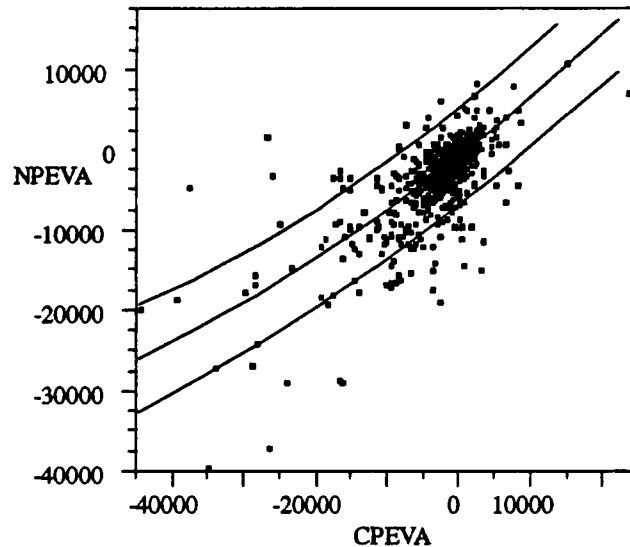


Figure 5-4. Regression Plot of NPEVA vs. CPEVA.

The total shareholder's equity (SHE) is also found to be a significant variable in predicting EVA for a firm during the next-period. A regression model, which has only the single independent variable SHE, is able to explain 40.3 percent of the variation in a firm's next-period EVA. This compares to 52.7 percent when only the single independent variable CPEVA was used in the model. For example, the single linear regression equation, $NPEVA = \$33.01 - 0.112(SHE)$, explains 40.3 percent of the variation in the response. The SHE coefficient, -0.112, is the unit dollar decrease in the next-period EVA per dollar increase in shareholder's equity. An ANOVA table

and parameter estimates for this single variable regression equation appears in Table 5-3. As seen in Table 5-3, the F-statistic is significant at less than 0.0001 value indicating that there is a statistically significant linear trend between the firm's next-period EVA and shareholder's equity. The parameter estimate also indicates that the regression coefficient for the independent variable, shareholder's equity, is statistically significant in predicting the next-period EVA.

Table 5-3. ANOVA and Parameter Estimates (SHE).

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob>F
Model	1	8615520309	9e+9	673.6353	0.0001
Error	998	1.28e+10	12789592		
C Total	999	2.14e+10			

Term	Estimate	Std Error	t-Ratio	Prob> t
Intercept	33.015705	147.008	0.22	0.8223
SHE	-0.111885	0.00431	-25.95	0.0000

When the second-order effect of SHE is included in the model, the R^2 value increases from 40.3 percent to 43.6 percent. Again, by adding the second-order effect of the variable, there is very little reduction in the MSE value. The entries in Table 5-4 show the results of the ANOVA and the parameter estimates, when both the first- and second-order effects of SHE are included in the model.

Table 5-4. ANOVA and Parameter Estimates (SHE, SHE²).

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob>F
Model	2	9319090624	5e+9	385.1904	0.0000
Error	997	1.21e+10	12096733		
C Total	999	2.14e+10			

Term	Estimate	Std Error	t-Ratio	Prob> t
Intercept	-540.3972	161.536	-3.35	0.0009
SHE	-0.056771	0.00835	-6.80	0.0000
SHE ²	-5.4e-7	0	-7.63	0.0000

When including the current-period EVA and the shareholder's equity, the R² value increases to 64.6 percent. See Table 5.5 for ANOVA and the parameter estimates when both independent variables are included in the model.

Table 5.5. ANOVA and Parameter Estimates (CPEVA, SHE).

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob>F
Model	2	1.38e+10	7e+9	908.8603	0.0000
Error	997	7572827624	7595614		
C Total	999	2.14e+10			

Term	Estimate	Std Error	t-Ratio	Prob> t
Intercept	179.24423	113.429	1.58	0.1144
CPEVA	0.4775921	0.01827	26.14	0.0000
SHE	-0.068031	0.00372	-18.28	0.0000

When including the current-period EVA and the change in total liabilities to shareholder's equity ratio (cTLSHER), the R² value is 54.2 percent. See Table 5-6 for the ANOVA results and the parameter estimates when both independent variables are included in the model.

Table 5-6. ANOVA and Parameter Estimates (CPEVA, cTLSHER).

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob>F
Model	2	1.15881e10	5.794e+9	589.9683	0.000
Error	997	9791463249	9820926		
C Total	999	2.13795e10			

Term	Estimate	Std Error	t-Ratio	Prob> t
Intercept	-958.4919	107.9395	-8.88	0.0000
CPEVA	0.64302	0.018726	34.34	0.0000
cTLSHER	0.4681257	0.082073	5.70	0.0000

When CPEVA, SHE, and cTLSHER are all included as independent variables in the model, the R^2 value is 65.4 percent. Table 5-7 shows the ANOVA results and the parameter estimates when the model includes all three independent variables. An analysis of the residuals indicate they are normally distributed with a mean of 0.000 and a standard deviation of 2,753.5.

Table 5-7. ANOVA and Parameter Estimates (CPEVA, SHE, cTLSHER).

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob>F
Model	3	1.39749e10	4.6583e9	626.5921	0.000
Error	996	7404614942	7434352		
C Total	999	2.13795e10			

Term	Estimate	Std Error	t-Ratio	Prob> t
Intercept	148.30945	112.4064	1.32	0.1873
CPEVA	0.4922981	0.018336	26.85	0.0000
SHE	-0.066295	0.0037	-17.92	0.0000
cTLSHER	0.3413291	0.071757	4.76	0.0000

The multiple linear regression equation determined using these three most significant independent variables is:

$$\text{NPEVA} = 148.311 + 0.492(\text{CPEVA}) - 0.066(\text{SHE}) + 0.341(\text{cTLSHER}). \quad (\text{C-1})$$

The coefficients of the regression equation C-1 indicate that as the current-period EVA of a firm increase by one dollar unit, the firm's next-period EVA increases by only 0.49 dollars. However, as the shareholder's equity in the firm increases by one dollar unit, the next-period EVA decreases by 0.066 dollars. In addition, as the ratio of the period-to-period change in total liabilities with respect to total shareholder's equity increases in the firm, the next-period EVA will increase 0.341 dollars. The above three term multiple linear regression model explains 65.4 percent of the variation in the next-period EVA achieved by a firm.

The relationship between shareholder's equity and total liabilities, and a firm's EVA is obvious from the EVA calculation. Recall from Chapter 4, "Research Knowledge Domain," EVA is defined as the net operating profit after-tax cash flow minus the cost of capital, including both debt and equity capital. Thus, EVA is the after-tax residual income, or operating profit, less a charge for the use of total capital. Regression equation C-1 indicates that as the amount of shareholder's equity increases, the firm's next-period EVA decreases. Likewise, as the ratio of total liabilities to shareholder's equity increases, the firm's next-period EVA will also increase. Since the cost of equity capital to a firm is greater than the cost of debt

capital, the use of equity capital above a selected amount in the capital structure will tend to reduce the next-period EVA to a greater extent than the use of debt capital.

Several other independent variables were also evaluated in the research effort, and include those shown in Table 5-8. However, no combination of one or two independent variables, other than those previously discussed, produce an R^2 value greater than 33 percent. However, it is interesting that none of the external variables reviewed such as the prime rate, the gross domestic product, and the composite index of leading economic indicators is a significant factor in predicting the next-period EVA of a firm. As a result, none of the multiple linear regression models discussed in this research incorporate external financial variables.

Table 5-8. Independent Variables.

Independent Variable	R^2	b_0	$b_1/p\text{-value}$	$b_2/p\text{-value}$
NI	5.1%	-244.5	0.133/0.000	N/A
NSA + NSA ²	17.8%	-558.5	-0.03/0.000	-2.1e-8/0.000
GP + GP ²	19.5%	9.75	0.155/0.000	8e-7/0.000
RDE + RDE ²	11.8%	-1817.4	-0.261/0.000	-6e-6/0.050
SGAE + SGAE ²	21.6%	34.2	-0.206/0.000	-1.2e-6/0.000
TA + TA ²	32.6%	264.8	-0.061/0.000	-8e-8/0.000
CL + CL ²	20.2%	-844.7	-0.104/0.000	-2e-7/0.000
TL + TL ²	23.9%	-631.7	-0.0689/0.000	-1e-7/0.000
RDENSR	0.00%	-	-	-
TLTAR	0.00%	-	-	-
TLSHER	0.00%	-	-	-
NITAR	0.00%	-	-	-
NISHER	0.00%	-	-	-

The training data set spans 250 companies over a four-year period. Because the data is acquired over a period of four successive years, it is necessary to

determine if autocorrelation is present in the data set. For example, are the independent variables influenced by time? Several types of analyses are performed including residual analysis, the Durbin-Watson test, and development of separate regression equations for each time period. In summary, none of the significant independent variables appear to be influenced by time, and autocorrelation was not statistically detected.

Two multiple linear regression models referred to as Model A and Model B and two subset multiple linear regression models referred to as Model A1 and Model B1 are used as independent variables for determining the fuzzy rules and membership functions using a genetic algorithm. Model A includes the ten most significant independent variables identified using the analytical process discussed in Section B, and Model B includes the ten most significant independent variables identified when the single most important independent variable, a firm's current-period EVA (CPEVA) is excluded from the selection list. Multiple linear regression Model A and Model B each have a subset model referred to as Model A1 and Model B1 which only includes those variables in the parent model that significantly contributes to increasing the R^2 value.

Multiple linear regression Model A includes the independent variables shown in Table 5-9.

Table 5-9. Regression Model A Independent Variables.

Variable	Description
CPEVA	Current-Period (Year) EVA
NSA	Net Sales
NSCAR	Net Sales to Current Assets Ratio
PTISHER	Pre-Tax Income to Shareholder's Equity Ratio
SGAENSR	Sales, General and Administrative Expenses to Net Sales Ratio
SHE	Shareholder's Equity
TL	Total Liabilities
cBEM	Change in Break-Even Margin
cSHER	Change in Shareholder's Equity Ratio
cTLSHER	Change in Total Liabilities to Shareholder's Equity Ratio

Using the independent variables shown in Table 5-9 for the regression model, the R^2 value is 67.1 percent and the NMSE is 0.33. The ANOVA results and the parameter estimates for Model A are shown in Table 5-10.

The leverage plot shown in Figure 5-5 is a plot of the predicted next-period EVA values (Model A) versus the actual next-period EVA values using the training data set. Figure 5-5 also shows the 95 percent confidence interval curves for the linear fit. Table 5-10 shows the predicted values calculated for the regression coefficients. As Table 5-10 also indicates, the ten independent variables used in Model A are all significant at a p-value of less than 7 percent.

Table 5-10. ANOVA and Parameter Estimates (Model A).

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob>F
Model	10	1.43368e10	1.4337e9	201.3284	0.0000
Error	989	7042757820	7121090		
C Total	999	2.13795e10			

Term	Estimate	Std Error	t-Ratio	Prob> t
Intercept	17.500583	309.8259	0.06	0.9550
CPEVA	0.4836368	0.018593	26.01	0.0000
SHE	-0.073509	0.004884	-15.05	0.0000
cTLSHER	0.3578067	0.070637	5.07	0.0000
NSCAR	227.05608	98.19421	2.31	0.0210
cSHER	-2.870695	0.829975	-3.46	0.0006
TL	-0.007378	0.002066	-3.57	0.0004
NSA	0.00483	0.001791	2.70	0.0071
PTISHER	466.94297	240.7637	1.94	0.0527
cBEM	27.847036	14.56180	1.91	0.0561
SGAENSR	-1181.262	649.6988	-1.82	0.0693

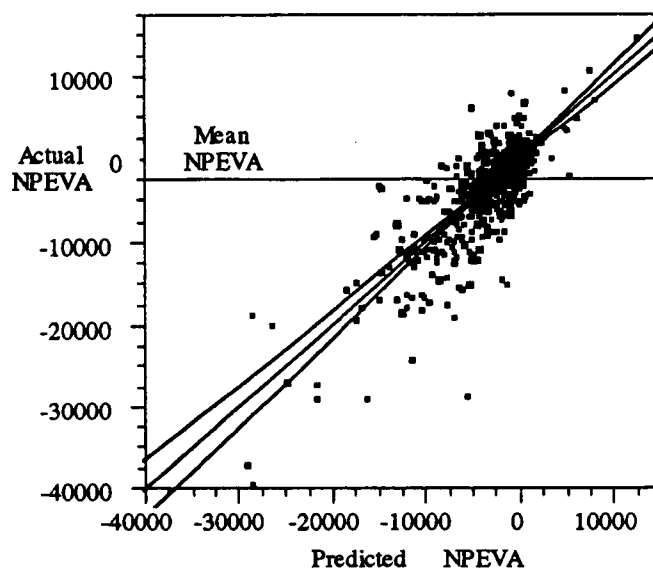


Figure 5-5. Leverage Plot (Model A).

As discussed earlier, stepwise regression, forward selection, backward elimination, and MaxR procedures help to identify the set of most significant independent variables. Table 5-11, Stepwise Regression Control with Step History, shows the successive entry of each regression variable, the related significance probability, the contribution to maximizing the R^2 value, and the C_p value. For example, Table 5-11 again indicates that the most significant independent variable in contributing to maximizing the R^2 value is the current-period EVA of the firm. Including the CPEVA variable as the single regression variable, produces an R^2 value of 52.7 percent. The next most significant variable to be added to the regression equation is SHE. When CPEVA and SHE are both included in the regression equation, the R^2 value increases to 64.6 percent. This is an increase of almost 12 percent. The remaining eight independent variables are significant but contribute much less to increasing the R^2 value.

One problem with the selection of independent variables for the model is the impact of underspecifying (i.e., ignoring important model terms) or overspecifying (i.e., including model terms that have minimal marginal effect). As previously discussed, one criterion used to assist in determining the fit of the model to the training data set is the Mallows' criteria, C_p .

When a regression model is underspecified, the variation that would be accounted for by the ignored variables is deposited in the residual sum of squares and thus inflates the error sum of squares, SS_E . However, a model that is overspecified

produces variances that are larger than those for a simpler version of the regression model. A proper compromise is a model with a value of $C_p = p$. Table 5-11 shows the C_p values as each variable is added to the model. Clearly, for a model which only includes CPEVA, C_p indicates that the model is significantly underspecified. As additional independent variables are added to the model, the C_p value approaches p , the number of regression terms in the model. When ten regression variables are included in the model, $p = 11$, and the C_p value is equal to p , the preferred model is identified. Table 5-11 shows the coefficient values of the preferred model.

Table 5-11. Stepwise Regression Control with Step History (Model A).

Step	Parameter	Action	Sig Prob	Seq SS	R ² Value	C _p	p
1	CPEVA	Entered	0.0000	1.127e10	0.5271	423.86	2
2	SHE	Entered	0.0000	2.5381e9	0.6458	69.437	3
3	cTLSHER	Entered	0.0000	1.6821e8	0.6537	47.815	4
4	NSCAR	Entered	0.0001	1.0856e8	0.6587	34.569	5
5	cSHER	Entered	0.0008	82559353	0.6626	24.976	6
6	TL	Entered	0.0300	34129704	0.6642	22.183	7
7	NSA	Entered	0.0047	57631452	0.6669	16.090	8
8	PTISHER	Entered	0.0375	31041010	0.6683	13.731	9
9	cBEM	Entered	0.0648	24390176	0.6695	12.306	10
10	SGAENSR	Entered	0.0693	23540460	0.6706	11.000	11

Figure 5-6 is a Pareto chart of the absolute values of the scaled parameter estimates for the Model A variables, arranged from high to low. The scaled estimates shown in the figure are the regression parameters multiplied by the standard deviation of the regression variable. This shows the estimated relative influence of each

variable and is useful in judging its influence on the predicted value, the next-period EVA of a firm. The figure shows a cumulative curve with almost 70 percent of the effect on next-period EVA being attributed to the current-period EVA and the amount of shareholder's equity. The other 30 percent contribution to next-period EVA is composed of the remaining eight regression terms in the model. Again, the Pareto chart confirms that the current-period EVA and the amount of shareholder's equity are the most dominant variables in predicting the next-period EVA of a firm.

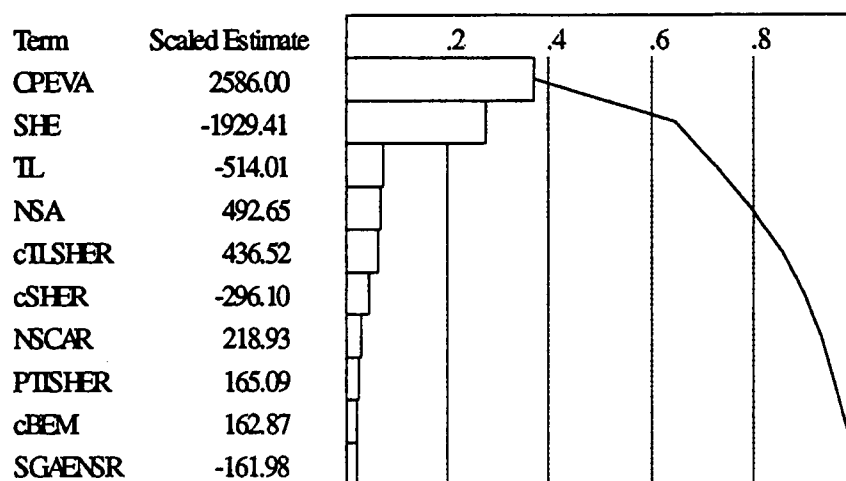


Figure 5-6. Pareto Chart of Regressor Effects (Model A).

It is interesting to note that when all of the quadratic and crossproduct terms of the selected 10 variables are included in the model (for a total of 65 Model A regression variables), the R^2 value increases only 6 percent (to almost 73 percent),

and the NMSE decreases from 0.33 to 0.28. Table 5-12 shows that most of the increase in the R^2 value is due to the 45 crossproduct terms.

Table 5-12. Model A with Cross Product and Quadratic Terms.

Source	DF	Sum of Squares	R^2 Value	F-Ratio	Prob>F
Linear	10	14336775437	0.6706	235.0	0.0000
Quadratic	10	351385006	0.0164	5.759	0.0000
Crossproduct	45	992763351	0.0464	3.616	0.0000
Total Regress	65	15680923794	0.7335	39.540	0.0000

Model A1 is a subset of Model A but only includes the five most significant independent variables which includes: a firm's current-period EVA, the amount of shareholder's equity, the change in total liabilities to shareholder's equity ratio, the net sales to current asset ratio, and the change in shareholder's equity ratio. The difference in the R^2 value for Model A and Model A1 is less than 1 percent. When including all the linear, quadratic, and crossproduct terms for Model A1, the R^2 value is increased from 66 percent to 69 percent and the NMSE is decreased from 0.34 to 0.32. Table 5-13 shows the ANOVA results and the parameter estimates for Model A1.

The second regression model, Model B, is developed from the list of 145 possible independent variables, but excludes from this list the most significant variable, the current-period EVA as a potential selection. Independent variables shareholder's equity; sales, general and administrative expenses to net sales ratio;

Table 5-13. ANOVA and Parameter Estimates (Model A1).

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob>F
Model	5	1.4166e+10	2.8332e9	390.4087	0.0000
Error	994	7213490621	7257033		
C Total	999	2.13795e10			

Term	Estimate	Std Error	t-Ratio	Prob> t
Intercept	-612.9345	222.9812	-2.75	0.0061
CPEVA	0.4919986	0.018374	26.78	0.0000
SHE	-0.066439	0.003659	-18.16	0.0000
cTLSHER	0.3379993	0.070914	4.77	0.0000
NSCAR	362.22426	89.16585	4.06	0.0001
cSHER	-2.820561	0.836242	-3.37	0.0008

change in break-even margin; change in shareholder's equity; and net sales to current assets ratio are common among both Models A and B. The purpose for developing Model B is to determine the effect that eliminating a firm's current-period EVA has on predicting its next-period EVA. The R^2 value for Model B is 66.9 percent and includes ten regression variables. The NMSE for Model B is 0.34. Table 5-14 shows regression Model B with all regression variables.

As shown in Table 5-15, the most significant independent variable is the amount of shareholder's equity. Recall that in Model A; the shareholder's equity is the second most significant independent variable (after current-period EVA) in predicting the next-period EVA of a firm.

Table 5-14. Regression Model B Variables.

Variable	Description
SHE	Shareholder's Equity
GP	Gross Profit
RDE	Research and Development Expenses
SGAE	Sales, General and Administrative Expenses
IE	Interest Expenses
SGAENSR	Sales, General and Administrative Expenses to Net Sales Ratio
cBEM	Change in Break-Even Margin
cSHER	Change in Shareholder's Equity Ratio
cNSWCR	Change in Net Sales to Working Capital Ratio
NSCAR	Net Sales to Current Assets Ratio

Table 5-15. Stepwise Regression Control with Step History (Model B).

Step	Parameter	Action	Sig Prob	Seq SS	R ² Value	Cp	p
1	SHE	Entered	0.0000	8.6155e9	0.4030	789.25	2
2	RDE	Entered	0.0000	5.0836e8	0.4268	720.14	3
3	NSCAR	Entered	0.0000	3.1186e8	0.4413	678.53	4
4	SGAENSR	Entered	0.0000	2.1753e8	0.4515	650.1	5
5	GP	Entered	0.0000	2.2907e8	0.4622	620.06	6
6	SGAE	Entered	0.0000	1.9606e9	0.5539	347.83	7
7	SGAENSR	Removed	0.6222	2332865	0.5538	346.16	6
8	IE	Entered	0.0000	2.1572e9	0.6547	46.443	7
9	cNSWCR	Entered	0.0000	1.4328e8	0.6614	28.404	8
10	cSHER	Entered	0.0005	86787140	0.6655	18.265	9
11	cBEM	Entered	0.0106	47006911	0.6677	13.691	10
12	SGAENSR	Entered	0.0306	33536810	0.6693	11.000	11

Figure 5-7 shows the Pareto chart for Model B. As shown, a firm's gross profit accounts for almost 40 percent of the cumulative influence on a firm's next-period EVA followed by sales, general and administrative expenses, which accounts for approximately 25 percent. Together, gross profit and sales, general and administrative expenses account for almost 65 percent of the relative influence on a firm's next-period EVA. Thirty percent is influenced by a firm's shareholder's equity, interest expenses, and research and development expenses. Table 5-16 shows the ANOVA and parameter estimates for Model B.

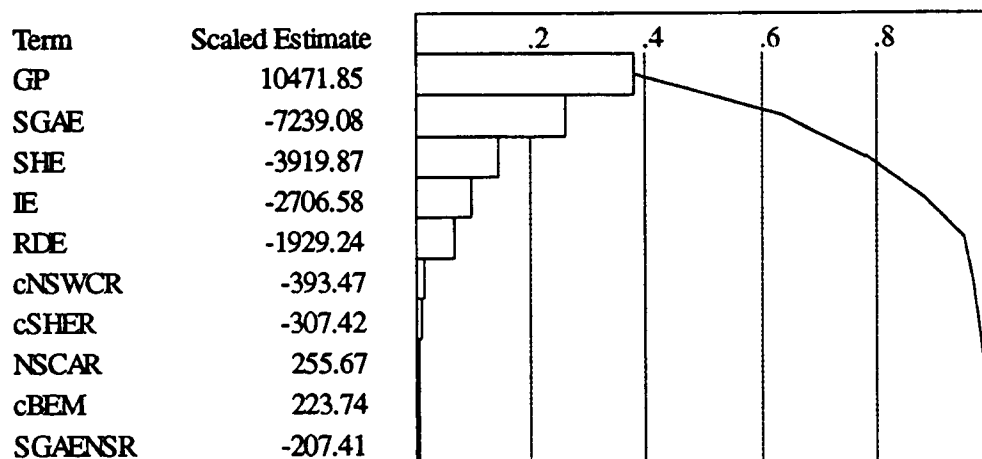


Figure 5-7. Pareto Chart of Regression Effects (Model B).

Table 5-16. ANOVA and Parameter Estimates (Model B).

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob>F
Model	10	1.43085e10	1.4308e9	200.1261	0.0000
Error	989	7071074599	7149722		
C Total	999	2.13795e10			

Term	Estimate	Std Error	t-Ratio	Prob> t
Intercept	-33.42202	319.9495	-0.10	0.9168
SHE	-0.149344	0.004846	-30.82	0.0000
GP	0.3936913	0.015684	25.10	0.0000
RDE	-0.475555	0.026005	-18.29	0.0000
SGAE	-0.371779	0.01767	-21.04	0.0000
IE	-0.502158	0.027814	-18.05	0.0000
SGAENSR	-1512.57	698.3917	-2.17	0.0306
cBEM	38.254074	14.63065	2.61	0.0091
cSHER	-2.980381	0.838908	-3.55	0.0004
cNSWCR	-0.153281	0.033564	-4.57	0.0000
NSCAR	265.15728	93.52795	2.84	0.0047

Model B1 is a subset of Model B but only includes the six most significant independent variables: the amount of shareholder's equity; research and development expenses; the net sales to current assets ratio; sales, general and administrative expenses; gross profit; and interest expenses. The independent variables selected for Model B1 produce an R^2 value of 65 percent and a NMSE of 0.35. The difference in the R^2 value for Model B and Model B1 is less than 1.5 percent. Table 5-17 shows the ANOVA results and the parameter estimates for Model B1.

Table 5-17. ANOVA and Parameter Estimates (Model B1).

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob>F
Model	6	1.39979e10	2.333e+9	313.8370	0.0000
Error	993	7381681428	7433717		
C Total	999	2.13795e10			

Term	Estimate	Std Error	t-Ratio	Prob> t
Intercept	-484.5224	227.2167	-2.13	0.0332
SHE	-0.145572	0.004719	-30.84	0.0000
RDE	-0.450925	0.025945	-17.38	0.0000
NSCAR	282.76689	92.45533	3.06	0.0023
SGAE	-0.365066	0.016771	-21.77	0.0000
GP	0.3815322	0.015502	24.61	0.0000
IE	-0.47426	0.02784	-17.03	0.0000

Several other linear regression models were also considered in this research effort, including models based only on balance sheet variables, income statement variables, and a combination of balance sheet and income statement variables.

The regression model developed based only on independent variables associated with a firm's balance sheet produced an R^2 value of 42.2 percent. Table 5-18 shows the ANOVA results and the parameter estimates. In this model, a significant amount of collinearity existed with other regression variables, especially total liabilities to shareholder's equity. As a result, a regression model based only on balance sheet variables is not desirable.

The regression model developed based only on independent variables associated with the income statement of a firm produced an R^2 value of 34.2 percent. Table 5-19 shows the ANOVA results and the parameter estimates. Again, a regression model based only on income statement variables is *not* desirable.

Table 5-18. ANOVA and Parameter Estimates (Balance Sheet Variables).

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob>F
Model	5	9028851785	1.8058e9	145.3309	0.000
Error	994	1.23507e10	12425233		
C Total	999	2.13795e10			

Term	Estimate	Std Error	t-Ratio	Prob> t
Intercept	120.43822	148.4181	0.81	0.4173
CA	0.0169389	0.008063	2.10	0.0359
TA	-0.244379	0.112968	-2.16	0.0308
CL	-0.029348	0.00672	-4.37	0.0000
TL	0.2443725	0.113264	2.16	0.0312
SHE	0.1282548	0.113117	1.13	0.2571

Table 5-19. ANOVA and Parameter Estimates (Income Statement Variables).

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob>F
Model	7	7307818562	1.044e+9	73.5960	0.000
Error	992	1.40717e10	14185196		
C Total	999	2.13795e10			

Term	Estimate	Std Error	t-Ratio	Prob> t
Intercept	-804.4961	156.5268	-5.14	0.0000
NSA	-1.433646	0.823073	-1.74	0.0818
COG	1.4140918	0.822641	1.72	0.0859
GP	1.7246515	0.824008	2.09	0.0366
RDE	-0.502772	0.044174	-11.38	0.0000
SGAE	-0.36036	0.033396	-10.79	0.0000
IE	-0.121044	0.04144	-2.92	0.0036
NI	-0.050193	0.023522	-2.13	0.0331

A regression model developed based only on independent variables associated with the income statement of a firm, but also including the current-period EVA as a variable, produced an R^2 value of 61.8 percent. Table 5-20 shows the ANOVA results and the parameter estimates.

Finally, the regression model developed based only on independent variables associated with the balance sheet and the income statement of a firm, produced an R^2 value of 65.5 percent. Table 5-21 shows the ANOVA results and the parameter estimates.

Table 5-20. ANOVA and Parameter Estimates (Income Statement Variables with CPEVA).

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob>F
Model	8	1.32063e10	1.6508e9	200.1569	0.000
Error	991	8173236883	8247464		
C Total	999	2.13795e10			

Term	Estimate	Std Error	t-Ratio	Prob> t
Intercept	14.359178	123.2175	0.12	0.9073
CPEVA	0.837904	0.031332	26.74	0.0000
NSA	-0.482741	0.628604	-0.77	0.4427
COG	0.4812216	0.628237	0.77	0.4439
GP	0.2130002	0.630848	0.34	0.7357
RDE	0.2560655	0.044042	5.81	0.0000
SGAE	0.2764775	0.034864	7.93	0.0000
IE	0.274401	0.034887	7.87	0.0000
NI	-0.019716	0.017972	-1.10	0.2729

Table 5-21. ANOVA and Parameter Estimates (Balance Sheet and Income Statement Variables).

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob>F
Model	12	1.39942e10	1.1662e9	155.8526	0.000
Error	987	7385332261	7482606		
C Total	999	2.13795e10			

Term	Estimate	Std Error	t-Ratio	Prob> t
Intercept	130.16829	118.8398	1.10	0.2736
CA	-0.006468	0.007722	-0.84	0.4024
TA	-0.042953	0.088137	-0.49	0.6261
CL	0.002783	0.005981	0.47	0.6418
TL	0.0393507	0.088587	0.44	0.6570
SHE	-0.104175	0.088348	-1.18	0.2386
NSA	-0.222176	0.604399	-0.37	0.7133
COG	0.2261724	0.603963	0.37	0.7081
GP	0.6153009	0.604904	1.02	0.3093
RDE	-0.459608	0.032524	-14.13	0.0000
SGAE	-0.372819	0.024651	-15.12	0.0000
IE	-0.463331	0.052866	-8.76	0.0000
NI	-0.006222	0.018036	-0.34	0.7302

Cluster analysis (see Chapter 6) was performed on the significant independent variables selected for Models A and B to estimate the potential number of fuzzy rules capable of being learned from the training data set. Cluster analysis was also performed on Models A1 and B1.

D. Summary

Multiple linear regression analysis was used, as discussed in this chapter, to determine the most significant independent financial variables useful in predicting the

next-period EVA of a firm. The current-period EVA was found to be the most significant independent variable and explains 52.7 percent of the variation in the predicted next-period EVA of a firm. The next two most significant variables are the amount of shareholder's equity and the change in total liabilities to shareholder's equity. These two variables plus the current-period EVA explains 65.4 percent of the variation in the next-period EVA amount. A formal statistical test for detecting autocorrelation was performed but none was detected in the variables selected.

Other variables found to be significant in predicting the next-period EVA of a firm included net sales; net sales to current assets ratio; pre-tax income to shareholder's equity ratio; sales, general and administrative expenses to net sales ratio; total liabilities; gross profit; research and development expenses; and change in break-even margin.

Two multiple linear regression models, Model A and Model B, and two subset multiple linear regression models, Model A1 and Model B1, are used as the independent variables for determining the number of clusters and for determining the fuzzy rules and membership functions using a genetic algorithm. Model A includes the ten most significant independent variables identified including a firm's current-period EVA. Model B includes the ten most significant independent variables identified when excluding from the selection list the single most important independent variable, a firm's current-period EVA. Multiple linear regression Model A and B have a subset model called Model A1 and Model B1 which only includes those variables in the

parent model that significantly contributes to increasing R^2 . None of the external financial variables were found significant in predicting a firm's next-period EVA.

Chapter 6

Cluster Analysis

A. Introduction

Cluster analysis is a statistical technique that is used to identify groups of data within the set of observed data. In this research, the number of groups of data (clusters) corresponds to the number of potential fuzzy rules available in the training data set. By assuming that the training data came from a homogeneous population, the number of fuzzy rules that could be learned from the population may be significantly limited, and potentially just involve a single rule. However, most data sets are nonhomogeneous and will consist of several different groups. Each group of data contains its own uniqueness. To understand this better, consider the cluster analysis illustration in Figure 6-1. This figure is a plot of data points on which two variables are measured; a firm's next-period EVA (NPEVA) and the corresponding amount of shareholder's equity (SHE). The figure shows what appears to be approximately 11 data clusters thus corresponding to 11 potential fuzzy rules required in describing the set of data. For example, the rule describing the cluster C_A data set may be, "IF shareholder's equity is high, then next-period EVA is likely to be low."

A number of methods have been proposed for grouping data sets into clusters. However, most of the available techniques tend to be exploratory and descriptive in nature [84]. As a result, small perturbations in a data set might lead to a different

number of clusters. Thus, the optimal number of clusters for any given application may have to be estimated subjectively. Fortunately, in this application, the number of clusters and therefore the number of rules, do not need to be exact but only requires an estimate within a specific range.

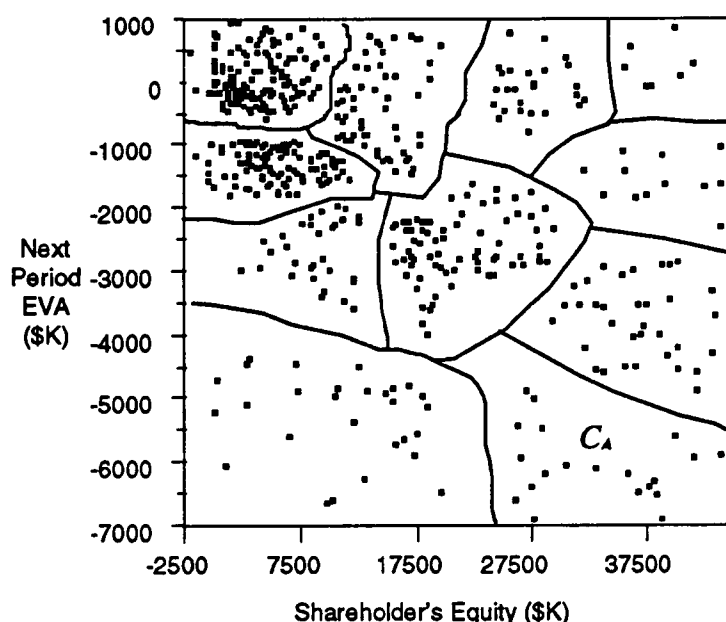


Figure 6-1. Cluster Analysis.

The cluster process used in this research is based on the k-mean clustering algorithms. The k-mean method originally proposed by J. MacQueen [85] in 1967 is probably the most widely applied nonhierarchical clustering technique [86]. The k-mean clustering algorithm minimizes the Euclidean distance between each observation point and its cluster. To group a set of N observations into C_k clusters, the algorithm

begins by randomly selecting the k_c center for each C_k cluster. An individual observation is assigned to the C_k cluster if the observation point, i , is nearer to the center of C_k than to all the other cluster centers. Based on the assignment of observation points to each cluster, new cluster centers are determined by using the centers of gravity of each cluster. These new centers create a new partition for assigning each observation and again determine new cluster centers. This process continues until two succeeding iterations lead to the same partition or after meeting a conveniently chosen criterion.

The quality of the fitted clusters is performed similar to linear regression analysis. That is, the coefficient of determination, often referred to symbolically as R^2 , is used as a measure of the cluster fit. The R^2 is defined as the sum of squares of the Euclidean distances between the data points of a cluster and its centroid, divided by the total sum of squares. The sum of squares of the Euclidean distances is a measure of the tightness of the cluster. Therefore, if too few clusters are chosen, the error sum of squares will be large resulting in a low R^2 value. The mean square error (MSE) is the error sum of squares divided by the degrees of freedom, and is a measure of how well the clusters fit the data set. Obviously, the smaller the MSE, the better the observations fit the clusters. However, the MSE can be forced to be artificially low by overfitting the data. By choosing a large number of clusters that may approach the number of individual observations, the MSE approaches zero. Typically, as the

number of clusters increase, the MSE tends to reach a point where additional clusters have minimal effect on its value.

The following example will demonstrate the k-mean cluster method. In this example, 20 observations are identified as shown in Figure 6-2.

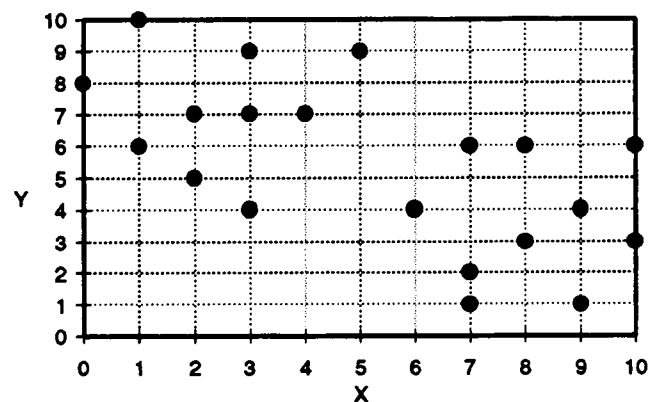


Figure 6-2. Cluster Data Set

The first step in using the k-mean clustering algorithm is to initially estimate the number of clusters and randomly select the centers for each cluster. In the example above, two clusters are selected with centers at (2, 1) and (7, 9). Each individual observation in Figure 6-2 is assigned to a cluster center based on minimizing the Euclidean distances between the cluster center and each observation point. Figure 6-3 shows the initial cluster assignment for each observation. The initial assignment of observation points to the two randomly selected cluster centers has an MSE of 28.

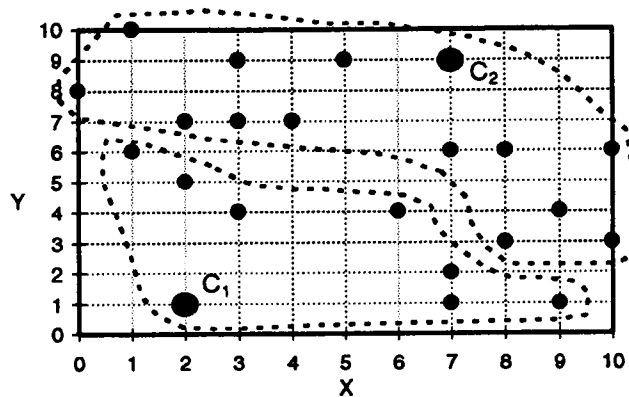


Figure 6-3. Cluster Data Set - Initial Assignment.

Based on the initial assignment of the 20 observation points to each cluster shown in Figure 6-3, new cluster centers are identified by calculating the centers of gravity for each cluster. The new cluster centers are shown in Figure 6-4. The next step is to assign each individual observation in Figure 6-2 to the new cluster centers based on minimizing the Euclidean distances between the new cluster centers and each observation point. Figure 6-5 shows the assignment of each observation to the new clusters. The MSE for the second assignment of observation points to the new cluster centers is 14. This is a 50 percent reduction in the MSE from the initial cluster assignment. The second and third iterations have an MSE of 8 and 6, respectively. This is shown in Figures 6-6 through 6-8. The MSE may be further reduced (or the R^2 value increased) by initially selecting more than two clusters. However, as the number of clusters increase, the MSE approaches a minimum (or the R^2 value approaches a maximum) where additional clusters will have little effect on further reducing MSE.

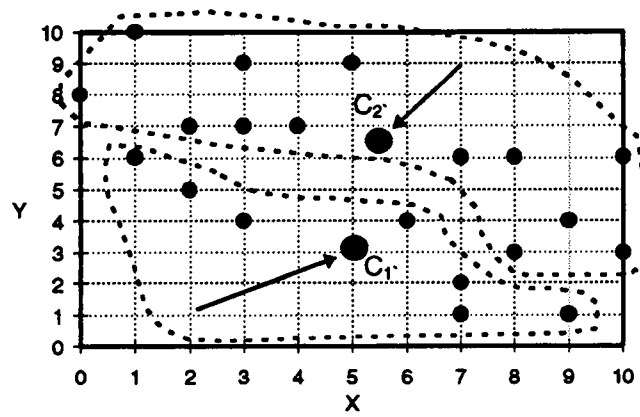


Figure 6-4. Cluster Center - First Iteration.

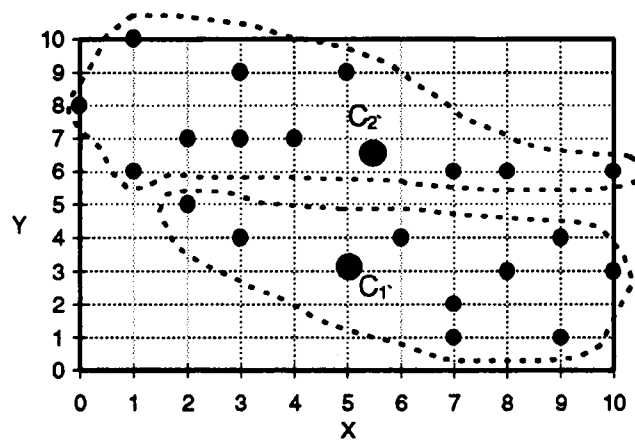


Figure 6-5. Cluster Data Set - Second Assignment.

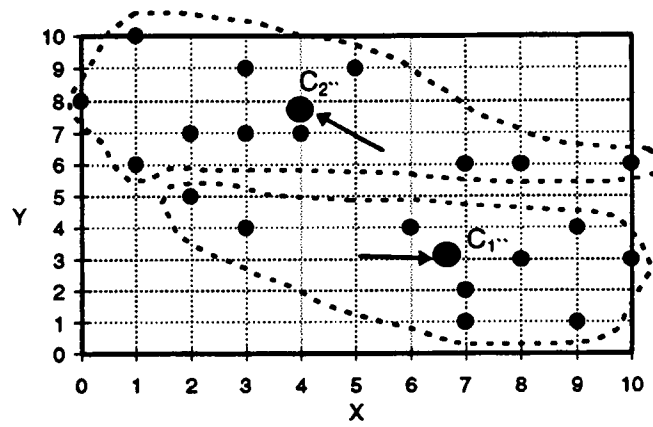


Figure 6-6. Cluster Center - Second Iteration.

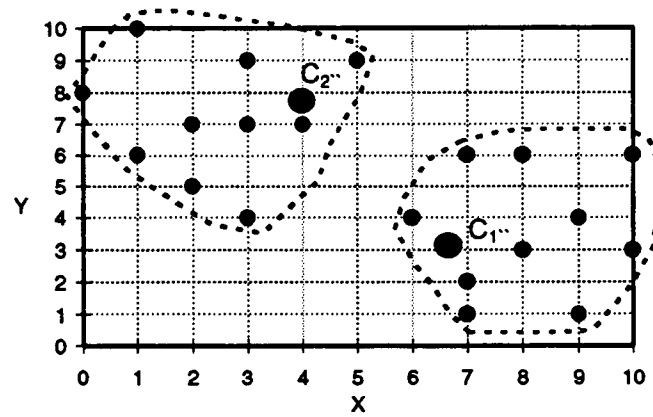


Figure 6-7. Cluster Data Set - Third Assignment.

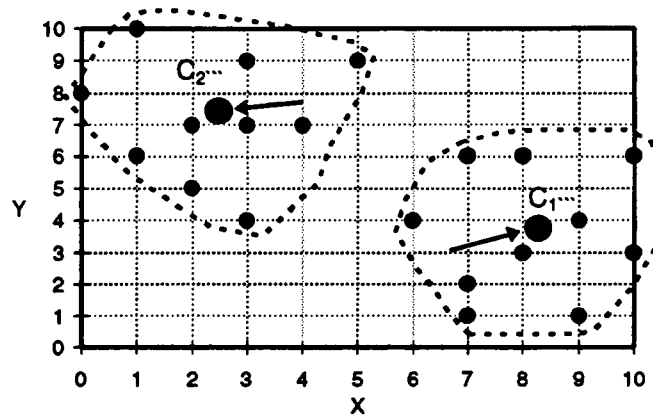


Figure 6-8. Cluster Center - Third Iteration.

After identifying the significant independent variables for the four multiple linear regression models, and determining the number of data clusters in the training data set, a genetic algorithm is used to search for the best fit fuzzy rules and membership functions. This topic is discussed in Chapter 7. Before proceeding, however, the next section discusses the experimental results of applying cluster analysis to the training data set for Models A, A1, B and B1 variables discussed in Chapter 5.

B. Experimental Results - Cluster Analysis

As previously discussed, the purpose of cluster analysis is to find groups of data points within the observed data set. In this research, the groups of data (clusters) estimate the number of potential fuzzy rules available in the training data set.

In Figure 6-9, the R^2 value is plotted versus the number of clusters for the data related to the variables in Model A. This figure also shows that the R^2 value for more than 20 clusters approaches a constant value above 80 percent. This would indicate that beyond 20 clusters, that is 20 fuzzy rules, there is little evidence to suggest that additional rules would significantly contribute to improving the fuzzy expert system model.

With the data from the variables in Model A1, (Figure 6-10), the R^2 value for more than 30 clusters approaches a constant value. Figure 6-11 shows the R^2 value versus the number of clusters for Model B. For Model B, the number of clusters tend to approach a constant beyond 30 clusters. Again, this would indicate that beyond 30 clusters, or 30 fuzzy rules, there is little evidence to suggest that additional rules would significantly contribute to the expert system model. Finally, Figure 6-12 shows the R^2 value versus the number of clusters for the data related to Model B1. For Model B1, the R^2 value tends to approach a constant beyond 50 data clusters.

From Figures 6-9 and 6-10, and Figures 6-11 and 6-12, the more independent variables per model, the greater R^2 value is for the same number of clusters. For example, the Model A variables with 20 clusters produce an R^2 of 0.80; however, Model A1 variables require almost 50 clusters to produce the same R^2 value. In addition, more independent variables per model also result in the R^2 reaching a constant value with fewer clusters. Consequently, the more significant independent variables included in the model, (that is, Models A and B), results in fewer clusters and

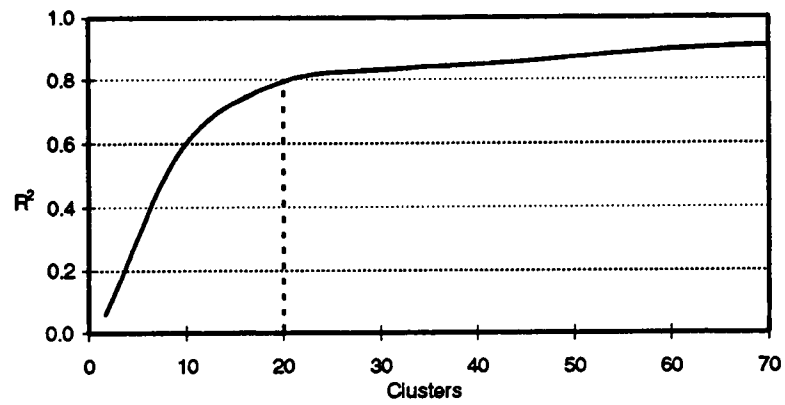


Figure 6-9. R^2 vs. Number of Clusters for Model A.

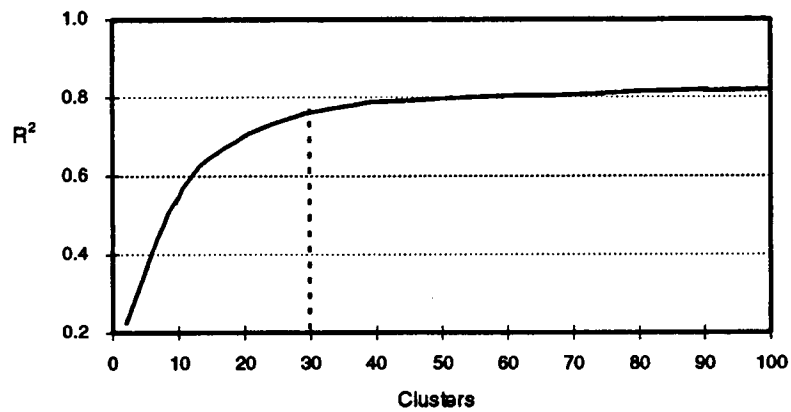


Figure 6-10. R^2 vs. Number of Clusters for Model A1.

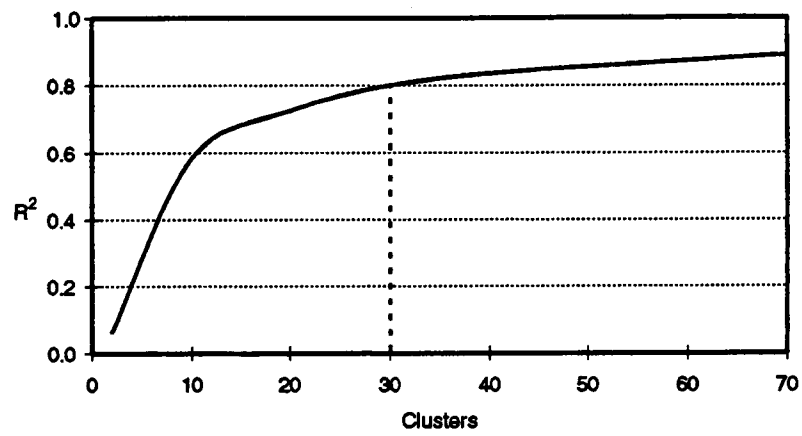


Figure 6-11. R^2 vs. Number of Clusters for Model B.

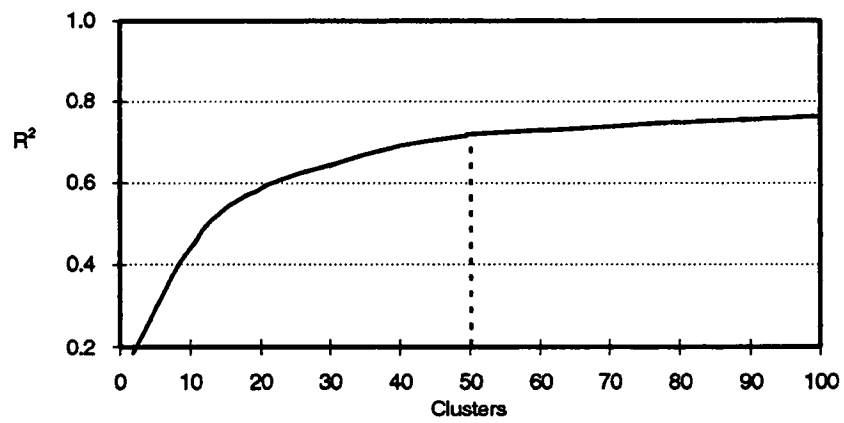


Figure 6-12. R^2 vs. Number of Clusters for Model B1.

a higher R^2 than does fewer significant independent variables. Thus, more significant independent variables results in better fit clusters, that is, the distance between the cluster's centroid and the cluster's observations are minimized.

Performing cluster analysis on the training data set helps to identify the number of fuzzy rules potentially available. Using this information, the substring used for encoding the fuzzy rules in the genetic algorithm (Chapter 7) can be reduced to correspond to the number of clusters identified using cluster analysis. Without cluster analysis, the number of potential rules for several independent variables can be extremely large resulting in unreasonably large genetic string encoding requirements. The computation time required by the genetic algorithm to search for the best fit fuzzy rules and membership functions increases significantly as a function of the number of fuzzy rules contained in the genetic string. Therefore, identifying the number of clusters contained in the training data set prior to using the genetic algorithm to search for the best fit fuzzy rules and membership functions can significantly reduce the time required to develop a fuzzy expert system. Cluster analysis, by identifying the optimal number of fuzzy rules available in the training data set, also helps to prevent overfitting the model with an excessive number of fuzzy rules.

C. Summary

Cluster analysis identifies groups within the observation data. In this research, groups of data clusters correspond to the number of potential rules available in the

training set. The cluster process used in this research is based on the k-mean clustering algorithms. The k-mean clustering algorithm minimizes the Euclidean distance between each observation point and its cluster.

The number of clusters is used by the genetic algorithm as a guide in determining the optimal number of fuzzy rules as discussed in Chapter 7.

Chapter 7

Genetic Algorithm and Fuzzy Expert System

A. Introduction

A genetic algorithm is a search technique based on Darwin's theory of evolution by natural selection. A genetic algorithm uses the concept of survival of the fittest among a set of search points to form a search procedure that crossbreeds and mutates among the points that is biased toward allowing only the fittest to survive into subsequent generations.

As an introduction, consider Figure 7-1 that represents the MSE versus a combination of potential IF . . . THEN rules. In this example, a genetic algorithm is used to search for the best fit (smallest MSE) combination of rules. This example is similar to that used in this research. However, a genetic algorithm may also be used to find the best fit solution to any multivariable function.

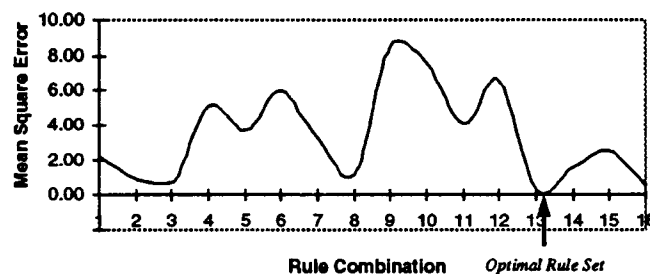


Figure 7-1. Example of a Genetic Algorithm Minimum Point Search.

A genetic algorithm is considered best suited to finding the solution to problems with discontinuous, multimodal, and noisy search spaces. For a number of real-world problems, traditional nonlinear search techniques are unsuitable and, as a result, are limited in their application.

In the current literature, four general classes of procedures used in solving nonlinear optimization problems are identified. They include multivariable search methods, multivariable elimination procedures, random searches, and stochastic approximation procedures. A genetic algorithm is different from the traditional optimization methods in four ways [87, 88]:

1. Genetic algorithm works with a coding of the variable set, not the parameters themselves.
2. Genetic algorithm searches from a population of points, not from a single point.
3. Genetic algorithm uses the value of the objective function, not derivatives or other auxiliary knowledge.
4. Genetic algorithm uses probabilistic transition rules, not deterministic rules.

Classical genetic algorithm codes the variable set of data as a finite-length string of zeros and ones, or as a binary set of numbers. These coded strings are called genetic strings or chromosomes. However, more recent research, especially by Davis

[89], shows that floating-point encoding may also be useful in many applications. The binary set of coded numbers is then mapped to a specific value for each variable in the interval to be studied. The genetic algorithm search starts with a population of binary strings and thereafter generates successive populations of strings until it reaches an optimal population.

Many search techniques require considerable more auxiliary information to work properly. By contrast, a genetic algorithm does not need auxiliary information. A genetic algorithm only requires the value of the objective function associated with individual strings. Once the objective function is calculated for a population of binary strings, probabilistic transition rules guide the search. A genetic algorithm uses randomized steps in the search procedure as a tool to guide the search toward optimal regions of the search space [90].

Thus, taken together, these four differences, direct use of a coding of the variable set, search from a population of search points, blindness to auxiliary information, and randomized operators, contribute to the robustness and resulting advantage of a genetic algorithm over other traditional nonlinear optimization techniques [91].

The genetic algorithm technique is summarized below in six steps [92]:

1. Initialize a population of chromosomes, i.e. binary string inputs.
2. Evaluate each chromosome in the population.

3. Create chromosomes by mating current chromosomes; apply mutation and recombination as the parent chromosome mate.
4. Delete members of the population to make room for the new chromosomes.
5. Evaluate the new chromosomes and insert them into the population.
6. If time is up, stop and return the best chromosome; if not, go back to 3.

B. Example - A Genetic Algorithm

The genetic algorithm technique summarized above will be discussed in more detail using an example. The function to be evaluated is defined as:

$$f(X_1, X_2) = 2X_1^2 - X_2^2 + 2X_2.$$

The problem is to find values for X_1 and X_2 in the range of $[0, 6]$ which maximizes the function f , i.e., to find X_0 such that $f(X_0) \geq f(X_1, X_2)$, for all $X_1, X_2 \in [0, 6]$. The selected population size of search points is five, and the precision is five discrete steps per unit number. The probability of crossover is selected to be 25 percent and the probability of mutation is 1 percent.

The domain of the variables X_1 and X_2 has length six, thus the precision requirement implies that the range $[0, 6]$ be divided into at least six times five equal size ranges or 30 discrete steps. This means the binary vector (chromosome) requires five bits:

$$16 = 2^4 \leq 30 \leq 2^5 = 32.$$

To map from a binary string into a real number X , perform the following transformation:

$$X = 0.0 + (\text{Base 10 of binary discrete unit}) * (6.0/30.0).$$

For example, if the binary unit is [10011], the base 10 representation of the binary number is 19.0 and the real number X is $(0.0 + 19.0 * 6.0/30.0)$ or 3.80.

Initially, randomize the first population of the chromosomes. In the problem above, the selected initial population is shown below where v_i represents chromosome i :

First Population

	X_1	X_2	X_1	X_2	$f(X_1, X_2)$
$v_1 =$	10011	10000	3.80	3.20	25.04
$v_2 =$	11100	00100	5.60	0.80	63.68
$v_3 =$	10100	11001	4.00	5.00	17.00
$v_4 =$	11010	00101	5.20	1.00	55.08
$v_5 =$	11101	10100	5.80	4.00	59.28

Clearly, the chromosome v_2 is the best of the five chromosomes since its evaluation returns the highest value of the function, and chromosome v_3 has the smallest value.

The total fitness of the population is defined as

$$F = \sum_{i=1 \text{ to } 5} (v_i) = 220.08.$$

The probability of a selection p_i for each chromosome v_i ($i = 1, \dots, 5$) is:

$$\begin{aligned} p_1 &= f(v_1) / F = 0.114 \\ p_2 &= f(v_2) / F = 0.289 \\ p_3 &= f(v_3) / F = 0.077 \\ p_4 &= f(v_4) / F = 0.250 \\ p_5 &= f(v_5) / F = 0.269 \end{aligned}$$

The cumulative probabilities q_i for each chromosome v_i are:

$$\begin{aligned}q_1 &= 0.114 \\q_2 &= 0.403 \\q_3 &= 0.480 \\q_4 &= 0.731 \\q_5 &= 1.000\end{aligned}$$

Next, select five random values between $[0, 1]$ for the new population.

Assume the random sequence is 0.1168, 0.3530, 0.1768, 0.5078, and 0.7436. The first random number 0.1168 is greater than q_1 but less than q_2 , and chromosome v_2 is selected for the new population. Similarly, the other four chromosomes are selected to generate the following new set of five chromosomes:

	X_1	X_2
$v'_1 =$	11100	00100
$v'_2 =$	11100	00100
$v'_3 =$	11100	00100
$v'_4 =$	11010	00101
$v'_5 =$	11101	10100

The first recombination operator (crossover) is now applied to the chromosomes in the new population. This action is called crossover. The crossover probability being used is $p_c = 0.25$. Therefore, on the average, 25 percent of the chromosomes will undergo a crossover action. The next step is to again select five random values between $[0, 1]$ for the crossover action. Assume the random sequence is 0.1673, 0.9224, 0.7546, 0.0754, and 0.2995. Choose all elements in the random sequence that are less than or equal to 0.25; chromosomes v'_1 and v'_4 are selected for

the crossover action. The crossover point in v'_1 and v'_4 is selected from a random number between $[1, 10]$; say 4. Thus, the crossover point is four, meaning the crossover point is four bits from the left most bit in the chromosome. For example, the first four bits of v'_1 and v'_4 are switched to form v''_1 and v''_4 . The new population after crossover now becomes:

	X_1	X_2
$v''_1 =$	11010	00100
$v''_2 =$	11100	00100
$v''_3 =$	11100	00100
$v''_4 =$	11100	00101
$v''_5 =$	11101	10100

The next recombination operator (mutation) is performed on a bit-by-bit basis. The probability of mutation being used is $p_m = 0.01$. Therefore, on the average 1 percent of the bits are expected to mutate. Since there are five times 10 or 50 bits, 1 bit is randomly selected for mutation. After randomly selecting the 17th bit (the binary value of the 17th bit changes from 0 to 1), the final population is:

	X_1	X_2
$v'''_1 =$	11010	00100
$v'''_2 =$	11100	01100
$v'''_3 =$	11100	00100
$v'''_4 =$	11100	00101
$v'''_5 =$	11101	10100

Mutation is sometimes needed to keep the algorithm from getting “stuck” at a less-than-best solution. At this point, one iteration or one generation in the genetic

procedure has been completed. The new generation of chromosomes is evaluated and the results are shown below:

Generation 1

	X ₁	X ₂	X ₁	X ₂	f(X ₁ ,X ₂)
v ₁ =	11010	00100	5.20	0.80	55.04
v ₂ =	11100	01100	5.60	2.40	61.76
v ₃ =	11100	00100	5.60	0.80	63.68
v ₄ =	11100	00101	5.60	1.00	63.72
v ₅ =	11101	10100	5.80	4.00	59.28

Chromosome v₄ is now the best of the five chromosomes, since its evaluation returns the highest value, and chromosome v₁ is the smallest. The total fitness of the new population F is now 303.48 for an improvement of 83.40 or 39 percent.

Now we are ready to run the selection process again, applying the genetic recombination operators, and evaluating the next generation. Below is the output for

Generation 2.

Generation 2

	X ₁	X ₂	X ₁	X ₂	f(X ₁ ,X ₂)
v ₁ =	11100	00100	5.60	0.80	63.68
v ₂ =	11110	00100	6.00	0.80	72.96
v ₃ =	11100	01100	5.60	2.40	61.76
v ₄ =	11100	00100	5.60	0.80	63.68
v ₅ =	11010	00101	5.20	1.00	<u>55.08</u>
F =					317.16

The total fitness of the second generation is now 317.16 for an improvement of 13.68 or 5 percent. Using the same procedural steps, a third generation of the selection process is accomplished. The results are shown below:

Generation 3

	X ₁	X ₂	X ₁	X ₂	f(X ₁ ,X ₂)
v ₁ =	11100	00100	5.60	0.80	63.68
v ₂ =	11110	00100	6.00	0.80	72.96
v ₃ =	11100	01100	5.60	2.40	61.76
v ₄ =	11101	01100	5.80	2.40	66.32
v ₅ =	11100	00100	5.60	0.80	<u>63.68</u>
F =					328.40

The total fitness of the third generation is now 328.40 for an improvement of 11.24 or 4 percent.

The maximum value selected from each population is summarized below:

Generation	X ₁	X ₂	f(X ₁ ,X ₂)	F
Initial Population	5.60	0.80	63.68	220.08
First Generation	5.60	1.00	63.72	303.48
Second Generation	6.00	0.80	72.96	317.16
Third Generation	6.00	0.80	72.96	328.40
Global Maximum	6.00	1.00	73.00	

As seen in the summary table above, the fitness value of the population improves during each generation [93].

In summary, a genetic algorithm is a search technique based on the mechanics of natural selection and survival of the fittest. As the example showed, a genetic algorithm combines survival of the fittest among a given set of search points to form a search algorithm that crossbreeds and mutates and is biased toward allowing only the fittest to survive subsequent generations. The output result is the set of search points that includes the best solution found to a nonlinear objective function.

A genetic algorithm is best suited to finding the solution to problems that have discontinuous, multimodal, and noisy search spaces. Traditional nonlinear solution techniques depend on the restrictive requirements of continuity and the existence of derivatives. An early developer of genetic algorithms, John Holland, said genetic algorithms provide a far greater range of potential solutions to nonlinear problems than conventional search programs [94].

Before continuing to discuss genetic algorithms as it relates to this research, a brief introduction of fuzzy set theory is needed.

C. Fuzzy Set Theory

Most of the phenomena encountered in everyday life are imprecise. That is; there is a certain amount of uncertainty, impression or fuzziness in the description of our encounters. For example, a person is described as tall but no one can identify a single number at which a person is tall and another person is short. This kind of imprecision is a common phenomena whether in engineering, medicine, or business.

To mathematically describe the fuzziness associated with imprecision, Lotfi A. Zadeh developed a model for the quantification of information using fuzzy sets and their associated fuzzy numbers [95]. Fuzzy numbers are fuzzy sets used with applications where an explicit representation of the ambiguity of the projected data is required. For example, fuzzy numbers take into account the “almost,” “about,” and

“likelihood,” of numerical data. Therefore, fuzzy numbers provide a convenient method to represent knowledge about certain types of events.

A fuzzy number is defined mathematically by assigning to each possible element in the union of discourse a value representing its grade of membership in the fuzzy set. This grade then represents the degree to which the element is similar or compatible to the concept represented by the fuzzy set. The fuzzy set may illustrate a set of elements that represent the total sales in a particular year. For example, let A equal the set of all possible values of total sales in a particular year, thus:

$$A = \{a_1, a_2, a_3, \dots a_n\}.$$

The elements of set A are further defined by the set's membership grade. A higher membership grade means the element is more likely to occur.

A fuzzy number is defined by its characteristic function (called the membership function) that takes on values in the interval $[0, 1]$, instead of in the binary set of $\{0, 1\}$ used for crisp numbers. For example, let X be a referential set of all real numbers that denotes a universal set. An ordinary subset, A , of this real number set is defined by its characteristic membership function and takes the form:

$$\forall x \in X: \quad \mu_A(x) \in [0, 1],$$

where $[0, 1]$ denotes the interval of real numbers from 0 to 1 which is the range of the membership grade for the fuzzy number [96]. The membership function is read as “a membership function” μ_A which maps every element of the universal set to the interval $[0, 1]$.

Fuzzy numbers also must be normal and convex. That is, the largest value of $\mu_A(x)$ must be equal to 1, and the fuzzy number membership set must "not go up-and-down" more than once. Figure 7-2 shows an example of a fuzzy number.

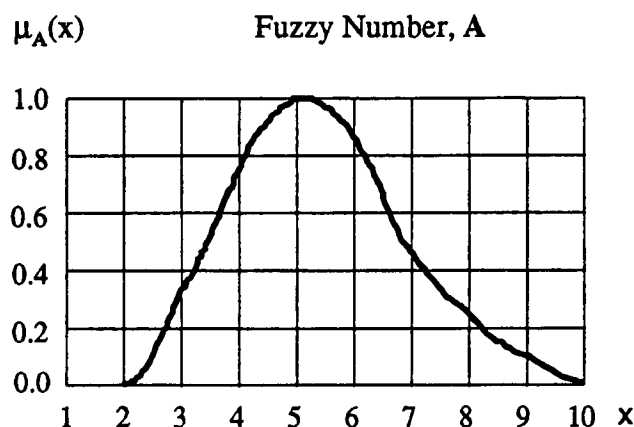


Figure 7-2. Normal and Convex Membership Function for Fuzzy Number A.

The fuzzy number A shown graphically in Figure 7-2 may be interpreted as representing the degree that the number X is possible. For example, when X is equal to 2 or 10, the degree of membership is 0, thus representing that the possibility that X is equal to 2 or 10 is zero or very impossible. However, when X is equal to 5, the degree or membership is 1.0, meaning that 5 is extremely possible. When X is equal to all other numbers between 2 and 10, the membership function represents the degree of possibility for the number X. A membership function may also be expressed by an α -cut, or a set of α -cuts. For example, for $\alpha = 0.4$ in Figure 7-2, the α -cut is 3.2 to 7.2.

The α -cut thus represents the domain of the membership function at a specific α value [97].

Modifiers of fuzzy numbers include linguistic terms such as *extremely*, *very*, *somewhat*, *quite*, *more or less*, *slightly*, *most*, *several*, *few*, *almost impossible*, *quite possible*, *quite true*, and *mostly true*.

A convenient way to represent a fuzzy number is by defining the lowest value, the highest value, and the most likely value or best value of the uncertain number.

Figure 7-3 shows how these three points produce a triangular fuzzy number.

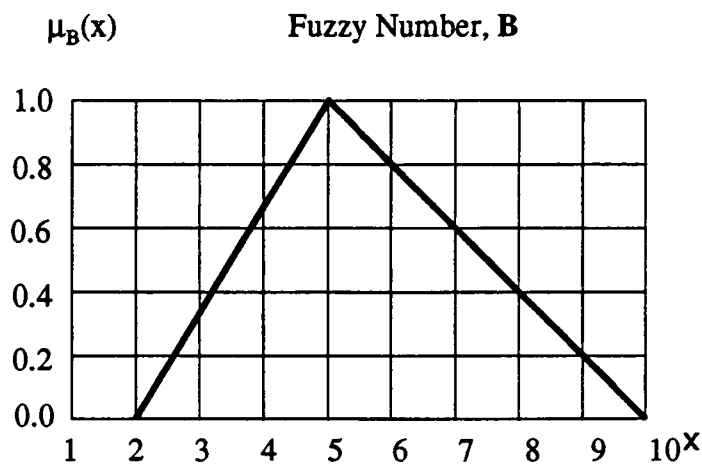


Figure 7-3. Normal and Convex Membership Function for Fuzzy Number B.

Fuzzy numbers are an extension of interval arithmetic and form the basis for fuzzy mathematics. Thus, fuzzy mathematics is a basic tool for dealing with fuzzy

numbers. In addition, the familiar operations of arithmetic, addition, subtraction, multiplication, and division can be defined for fuzzy numbers as well [98].

Fuzzy mathematics includes methods to add, subtract, multiply, and divide fuzzy numbers. When performing a mathematical operation on one or more fuzzy numbers, a new fuzzy number is produced. This resulting fuzzy number is uniquely defined by its membership function. For example, adding fuzzy numbers A and B can be represented by the new fuzzy number C. The new fuzzy number C can be defined by its membership function, $\mu_C(x) = \mu_A(x) + \mu_B(x)$. The addition of the membership functions may be defined as the addition of its α -cuts as follows [99]:

$$A + B = [a_1(\alpha), a_2(\alpha)] + [b_1(\alpha), b_2(\alpha)],$$

where (α) is a specific α -cut, and $a_1(\alpha)$, $b_1(\alpha)$ and $a_2(\alpha)$, $b_2(\alpha)$ represent the left- and right-hand X values of the membership function.

If we add the previously defined fuzzy numbers A and B at $\alpha=0.4$, C is defined as follows:

$$A = [3.2^{(0.4)}, 7.2^{(0.4)}], \text{ and } B = [3.1^{(0.4)}, 8.0^{(0.4)}], \text{ then} \\ C = [3.2^{(0.4)}+3.1^{(0.4)}, 7.2^{(0.4)}+8.0^{(0.4)}], \text{ or } [6.3^{(0.4)}, 15.2^{(0.4)}]$$

Subtraction is performed in a similar manner. For example, if we subtract fuzzy number B from A at $\alpha=0.4$, C is calculated as follows:

$$C = [3.2^{(0.4)}-3.1^{(0.4)}, 7.2^{(0.4)}-8.0^{(0.4)}], \text{ or } [0.1^{(0.4)}, -0.8^{(0.4)}]$$

Division of two fuzzy numbers is computed by:

$$A / B = [a_1(\alpha)/b_2(\alpha), a_2(\alpha)/b_1(\alpha)],$$

and multiplication is computed by [100, 101]:

$$A \times B = [a_1(\alpha) \times b_1(\alpha), a_2(\alpha) \times b_2(\alpha)].$$

D. Fuzzy Logic

The expansion of fuzzy sets and fuzzy numbers lead to the concept of fuzzy logic. Fuzzy logic is a method of mapping a fuzzy input set to a fuzzy output set. For example, a fuzzy input set may be the fuzzy set called temperature and the fuzzy output set may be the fuzzy set called pressure. Fuzzy logic is used to map the fuzzy set “temperature” to the fuzzy set “pressure.” One method to map two or more fuzzy sets uses a fuzzy rule with fuzzy set qualifiers such as high, medium, and low. For example, “IF input temperature 1 is high, and input temperature 2 is low, THEN pressure is medium.” However, more complex systems are typically characterized by incorporating several fuzzy rules. Multiple fuzzy rules are combined using an inference procedure as discussed below.

Fuzzy mathematics combines fuzzy rule sets. For example, assume we take two fuzzy rules, R1 and R2, along with the antecedent input variable and consequent output variable membership functions as shown in Figure 7-4. *Calculate the combined output by taking the minimum of the input value per linguistic value that corresponds to a specific rule set and the maximum of the consequent variable membership function.* As an example, for R1 in Figure 7-4, the minimum input value for the three linguistic values is 0.3, (0.3 VL), and the maximum consequent variable

membership function is L (low). This method of combining rules is called the Zedah AND operator. Several other implication operators are frequently used depending on the class of problems being investigated [102]. However, this research uses the Zedah AND operator exclusively. The Zedah AND operator applies all fuzzy rules in parallel and with equal weighting.

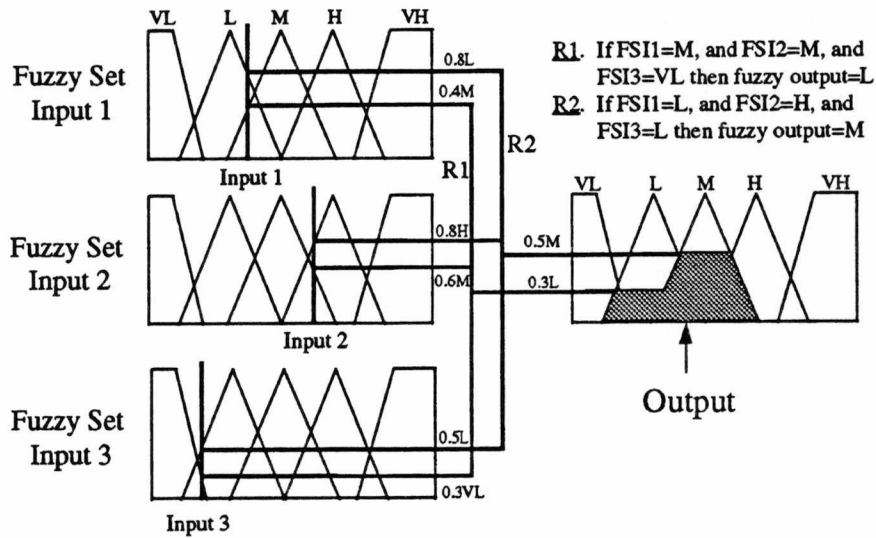


Figure 7-4. Fuzzy Rule Combination.

The necessity for alternative implication operators is motivated by the strict minimum and maximum compensation of the Zedah AND operator. For example, applying the Zadeh AND operator yields the following truth value:

$$\mu_{consequent} = \mu_{antecedent-1}[x] \cdot \mu_{antecedent-2}[y] \cdot \mu_{antecedent-3}[z]$$

$$\mu_{consequent} = \min(0.56, 0.86, 0.21) = 0.21$$

Clearly, from the result above, the minimum of any truth value controls the entire expression. For this reason, other implication operators have recently been developed including several types of functional compensatory operators. Other types of implication operators are not reviewed in this dissertation.

Several methods are also widely used to calculate or defuzzify the output membership function including center of area method, center of maxima method, and mean of maxima method [103]. Thus, in Figure 7-4 and throughout this research, the final output value is calculated as the centroid of the combined consequent membership function.

The fuzzy rules follow commonsense behavior of the system and are written in terms of the membership function linguistic labels (see Figure 7-5). For example, a typical fuzzy rule from Figure 7-5 is "If Net Sales are LOW and Net Profits are MEDIUM, then Project Investment should be MEDIUM." The illustration in Figure 7-5 refers to a fuzzy associative memory where antecedent 1 is net sales, antecedent 2 is net profits and the consequent is project investment.

In this research, the fuzzy associative memory and the fuzzy membership functions are both acquired using the genetic algorithm's search technique and based on known, historical information.

		Net Sales (Antecedent 1)				
		Very Low	Low	Medium	High	Very High
Net Profit (Antecedent 2)	Very Low	Very Low	Very Low	Very Low	Very Low	Very Low
	Low	Low	Low	Low	Very Low	Very Low
	Medium	Medium	Medium	Medium	Low	Low
	High	High	High	High	Medium	Medium
	Very High	Medium	Medium	High	High	Very High

Figure 7-5. Fuzzy Associative Memory for Project Investment.

In summary, the fuzzy logic system model discussed earlier consists of four modules: a fuzzy rule base, a fuzzy inference engine, and the fuzzification and defuzzification modules. The interconnections among these modules are shown in Figure 7-6.

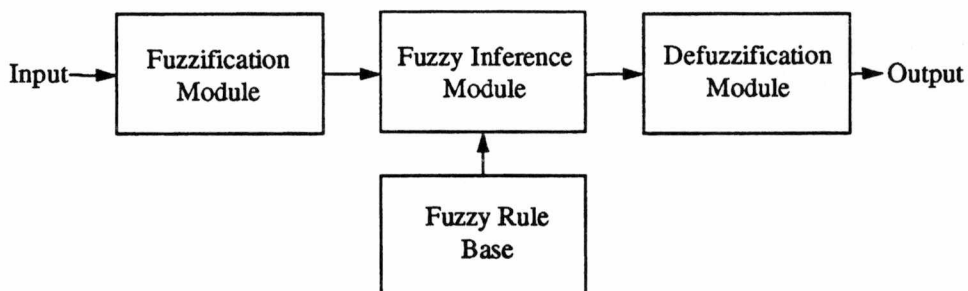


Figure 7-6. Block Diagram of a Fuzzy Logic Control System.

E. Fuzzy Expert System

How does one determine the “best fit” fuzzy rules and membership functions for a specific application? The construction of the rule-base and membership functions is the crucial and most difficult aspect of developing a fuzzy logic control system. It is also a major criticism concerning fuzzy logic control models because there are no universally accepted method for systematically developing the rules and membership functions of a fuzzy logic controller. The classical approach used in developing a fuzzy logic control system model is to use an “expert” or “experts” to define the rule set based on their knowledge of the subject area. This research, however, uses a genetic algorithm and historical information of the system as the method to determine the fuzzy rules and membership functions.

The genetic algorithm is used to determine the best fit fuzzy rules and membership functions using a set of historical information for learning the rules and membership functions. The genetic algorithm used in this research incorporates a hybrid genetic string with the following substrings.

(SS1) Substring 1 = Antecedent membership function (Floating Point)

(SS2) Substring 2 = Consequent membership function (Floating Point)

(SS3) Substring 3 = Antecedent rule value (Integer)

(SS4) Substring 4 = Consequent rule value (Integer)

The genetic string is a combination of the substrings shown above:

Genetic String = [Float_SS1, SS2] [Integer_SS3, SS4]

Example: For 10 input variables and 1 output variable, 5 membership functions per variable and 50 rules, the genetic string consists of a floating point [11 x 5] plus an integer [50 x 11] array.

The genetic algorithm steps involved in identifying the “best fit” fuzzy rule set and membership functions for this research are listed below.

- Step 1: Generate a random set of n possible genetic strings (Each genetic string consists of the four substrings) Note: Each genetic string represents all variable membership functions and rule combinations.
- Step 2: Calculate the EVA amount using fuzzy logic for each genetic string and compare with known historical EVA data.
- Step 3: Determine genetic fitness value by calculating the normalized mean square error of the calculated EVA using fuzzy logic compared to the known historical value.
- Step 4: Produce a next generation of genetic strings and go to Step 2. Stop when normalized mean square error is less than an established preset value.
- Step 5: The optimal rule set is the last generated set of genetic strings.

It is interesting to note that with just 10 attributes and 5 linguistic values per attribute, the combination of potential rules is more than 9.7 million [(linguistic value per attribute)^{number of attributes}]. Given the 9.7 million rules, the genetic algorithm

searches for the combination of rules that best fits the knowledge base to generate the “expert” rule base. Using this technique, no expert is required to generate the rule set.

The reproduction operation of the genetic algorithm used in this research is performed using a ranking selection procedure. The previous example discussed in this chapter used proportional selection. In ranking selection, the population of chromosomes are ranked according to their fitness, and the probability of selecting a chromosome for reproduction is given by a specific function. The concept of elitism is also used in choosing the best fit chromosomes. That is, if no increasingly fit chromosomes are found from one generation to the next, the elitist policy carries forward the most fit individual from the previous generation into the next generation.

The crossover operation in the genetic algorithm is performed differently on the floating point substring than the binary substring. For the binary substring, the crossover operation is performed similar to what was described in the previous example. However, for the floating point substring, the crossover is performed by taking the randomly weighted mean of two substrings to create the new substring.

The mutation operation for the binary substring is also performed similar to that described in the previous example. For the floating point substring, however, the mutation operation is performed by adding or subtracting a random number to the substring, within some preset range, to create the new substring.

The fuzzy implication operator used in this research is based on the Zadeh AND operation. The defuzzification method is based on the centroid of the combined fuzzy output membership functions.

The fuzzy membership functions are restricted to the Gaussian form. One advantage of using the Gaussian membership functions is that they have non-zero values at every point in the domain. This is in contrast to the triangular membership function commonly used that has zero values outside the range of the triangle.

F. Developing a Fuzzy Expert System

The following example will briefly discuss the development of a fuzzy expert system using a hybrid genetic string similar to that used in this research.

Suppose a company is interested in identifying the rules that can better predict the electrical energy consumption for a large aluminum processing facility based on two attributes: the time of day and outside temperature change. For example, a typical rule used by the plant supervisor to predict the electrical energy consumption might be "IF it is very early morning and the temperature is dropping, THEN the electrical energy consumption is very high."

Because the electrical energy consumption of the aluminum manufacturing facility has been closely monitored, extensive data is available for a range of times and temperature conditions. The genetic algorithm uses this historical information to

identify the best fit fuzzy rules and membership functions to predict future electrical consumption.

With two attribute variables and one consequent variable (five linguistic terms per variable), there is a total combination of 25 possible rules (See Figure 7-7 below). For example, Rule R11 is “IF the time is very early and the temperature is dropping fast, THEN the electrical energy consumption is likely to be <consequent C11>.”

		Time (Antecedent 1)				
		Very Early	Early	Normal	Late	Very Late
Temperature (Antecedent 2)	Dropping Fast	C11	C12	C13	C14	C15
	Dropping	C21	C22	C23	C24	C25
	Static	C31	C32	C33	C34	C35
	Rising	C41	C42	C43	C44	C45
	Rising Fast	C51	C52	C53	C54	C55

Figure 7-7. Fuzzy Associative Memory for Electrical Energy Consumption.

The fuzzy expert system uses the evolutionary search capabilities of the genetic algorithm, and historical information on electrical energy consumption, to identify the rules that are best able to predict electrical consumption, i.e., C11 in Figure 7-7 in the above example.

In addition to learning the best fit fuzzy rules, the genetic algorithm also searches for the best fit fuzzy membership functions that define the linguistic terms like temperature is dropping fast and the time is very early. For example, the fuzzy membership functions for time can be represented by the Gaussian membership functions shown in Figure 7-8.

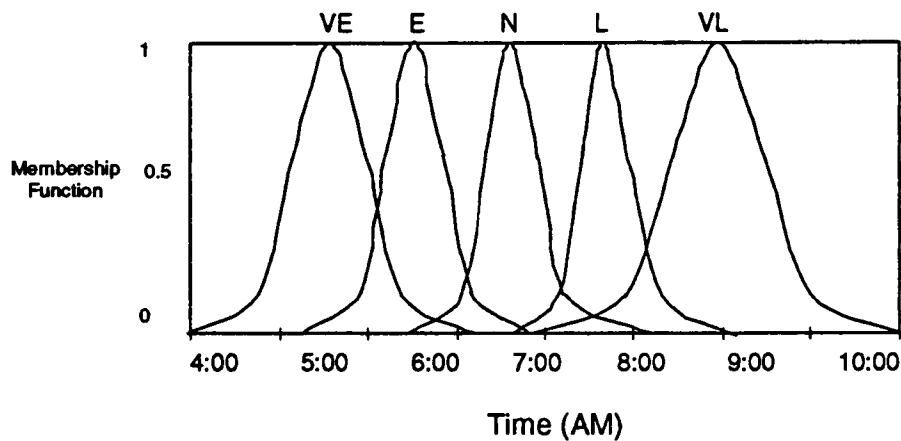


Figure 7-8. Fuzzy Membership Function for Time.

The first step in developing a genetic algorithm as a learning mechanism to determine the best fit fuzzy rules and membership functions is to define the genetic encoding technique. In this research, the total genetic string consists of a hybrid of floating point and integer coding strings. The genetic string actually consists of four substrings as shown below:

(SS1) Substring 1 = Antecedent membership function (Floating Point)

(SS2) Substring 2 = Consequent membership function (Floating Point)

(SS3) Substring 3 = Antecedent rule value (Integer)

(SS4) Substring 4 = Consequent rule value (Integer)

For example, the fuzzy membership functions shown in Figure 7-8 can be described as Gaussian functions with two unknown values: the mean, μ , and the standard deviation, σ . Therefore, Substring 1 (antecedent membership function) may be represented by a floating point number with values for μ and σ . The fuzzy membership function represented by Substring 1 is $\mu_{ai}^m \sigma_{ai}^m$ and represents the mean and standard deviation for antecedent i and membership function m :

$$\text{Substring 1: } [\mu_{a1}^1 \sigma_{a1}^1, \mu_{a1}^2 \sigma_{a1}^2 \dots \mu_{a2}^5 \sigma_{a2}^5]$$

The consequent membership function can be similarly represented as Substring 2 where $\mu_c^m \sigma_c^m$ represents the mean and standard deviation for single consequence c and the membership function m :

$$\text{Substring 2: } [\mu_c^1 \sigma_c^1, \mu_c^2 \sigma_c^2 \dots \mu_c^5 \sigma_c^5]$$

The fuzzy rules may be represented by integer values as shown below where Rab represents the linguistic values for antecedent a and antecedent b :

$$\text{Substring 3: } [R11, R12, R13 \dots R55]$$

The consequent rule value may be similarly written as:

$$\text{Substring 4: } [C11, C12, C13 \dots C55]$$

Finally, the combined genetic hybrid string is shown below:

$$\text{Genetic String: } [\mu_{a1}^1 \sigma_{a1}^1, \mu_{a1}^2 \sigma_{a1}^2 \dots \mu_{a2}^5 \sigma_{a2}^5, \mu_c^1 \sigma_c^1, \mu_c^2 \sigma_c^2 \dots \mu_c^5 \sigma_c^5, R11 \dots R55, C11 \dots C55]$$

It is important to note that in the electrical energy consumption example, there is a maximum possible combination of 25 rules, therefore, the genetic string includes 25 rule combinations and 25 consequence values. However, for most applications and demonstrated in this research, the number of possible rule combinations is very large and requires cluster analysis to determine the number of potential rules available given the training data set. Otherwise, encoding all possible combinations of potential rules in a genetic string would become impracticable. For example, assume cluster analysis was performed on the historical electrical energy consumption data and 10 clusters were identified. Therefore, 10 rules and consequence values would be encoded in the genetic string. Using the basic example above, a hypothetical hybrid genetic string with 10 encoded rules and membership functions may appear as shown in F-1:

$$v_1: [4.01 \ 0.96 \ 5.82 \ 0.81 \ 7.30 \ 1.21 \ 8.36 \ 0.76 \ 9.82 \ 1.31 \ -6.21 \ 1.25 \ -1.60 \ 2.30 \\ 0.02 \ 1.24 \ 2.96 \ 1.22 \ 4.96 \ 2.30 \ 224551423331215143343242351124] \quad (F-1)$$

Outlined in the following discussions are the five genetic algorithm steps involved (and previously identified) in determining the “best fit” fuzzy rule set for this example and the research discussed in this dissertation.

Step 1: Generate a random set of n possible genetic strings similar to F-1 (Each genetic string consists of four substrings).

Step 2: Calculate the electrical energy consumption using fuzzy logic (discussed in Section D) for each genetic string, and compare the results with the historical electrical energy consumption data, i.e., the training data set. For example, each genetic string in the population is decoded to form the fuzzy rules and membership functions. These fuzzy rules and membership functions are then used by the fuzzy logic module shown in Figure 7-9 to calculate the predicted electrical energy consumption.

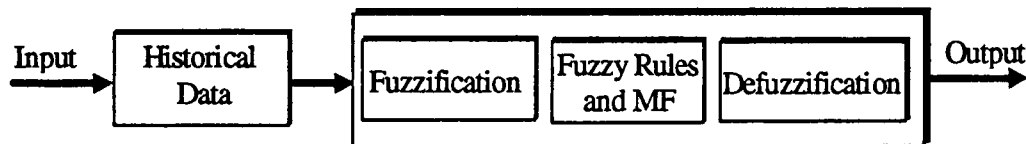


Figure 7-9. Block Diagram of the Fuzzy Logic Module.

The output error is the difference between the predicted value using fuzzy logic and the historical value. The objective of the genetic algorithm is to search for the best fit fuzzy rules and membership functions that minimizes the output error.

Step 2 determines the genetic fitness value by calculating the differences between the values using fuzzy logic and the historical value. Once the fitness values are calculated, the first genetic algorithm operation called reproduction is performed on the chromosomes. Several types of reproduction operations

are frequently used by a genetic algorithm such as steady-state reproduction and ranking reproduction. This research uses ranking reproduction as previously discussed and illustrated below.

Example:

Suppose the initial population is randomly selected for five chromosomes similar to F-1. The following fitness values are then calculated, which may represent in the previous example the difference between the predicted electrical energy consumption and the actual electrical energy consumption as determined using the historical information (training data set):

First Population

<u>chromosome</u>	<u>fitness</u>
$v_1 =$	63
$v_2 =$	21
$v_3 =$	10
$v_4 =$	86
$v_5 =$	43

The average fitness value for the randomly selected first population is 44.6.

Using the ranking selection procedure, the probability of selecting a specific chromosome for reproduction is give by $P\{select(v_i)\} = \frac{2}{v} - \frac{2}{v(v+1)}q$ where

v is the number of chromosomes and q is the chromosome fitness rank. The five chromosomes are ranked by increasing value of their fitness measure which is the difference between the fuzzy logic value and the related historical

value (v_3 , v_2 , v_5 , v_1 , and v_4), and are shown in Figure 7-10 based on the above equation to calculate the selection probability.

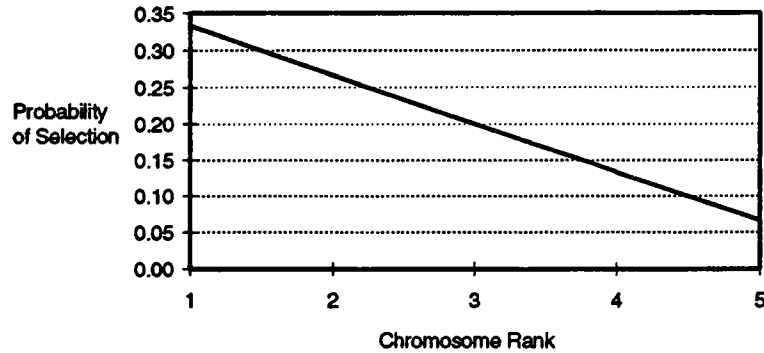


Figure 7-10. Chromosome Rank Selection.

From Figure 7-10, the best fit chromosome (that is, the one with the least output error), v_3 (rank 1), has a 33 percent chance of being selected for reproduction, and the least fit chromosome, v_4 (rank 5), has a 7 percent chance of reproduction.

Next, five random values are selected between $[0, 1]$ for the new population.

Assume the random sequences are 0.2584, 0.6873, 0.1245, 0.3867, and 0.8947. The first random number 0.2584 is less than q_2 but greater than q_3 ; therefore, select q_2 or chromosome v_2 for the new population. Similarly, select the other chromosomes to generate the following new population:

First Population Reproduction

<u>chromosome</u>	<u>fitness</u>
$v'_2 =$	21
$v'_3 =$	10
$v'_1 =$	63
$v'_3 =$	10
$v'_3 =$	10

The average fitness value for the first reproduction is 22.8. Following reproduction, the genetic algorithm recombination operators, crossover and mutation, are performed as discussed in Step 3.

Step 3: Perform crossover and mutation operations on the chromosomes in Step 2 to generate a new population of five chromosomes.

As previously discussed, the crossover operation is performed differently on a floating point substring then on a binary substring. For the integer substring, the crossover operation is performed similar to the procedure described earlier in the example in Section B. However, for the floating point substring, the crossover is performed by taking the randomly weighted mean of two substrings to create the new substring.

Example (Floating Point Crossover):

$v_1 =$ [4.01]
 $v_2 =$ [5.92]

Crossover is computed using a randomly weighted mean calculated using

$$\frac{\{(1+\tau w)b+(1-\tau w)a\}}{2} \text{ where } b > a \text{ and represents the floating point}$$

chromosome value, τ is a random number between $[-1,1]$, and w is a parameter specifying the degree that the crossover may vary from the mean of the two chromosomes. Thus, if $w = 0$, the crossover will be the mean of the parent chromosome values and if $w = 1$, the crossover will produce a random number between the parent chromosome values. In the example above, if $w = 0.5$ and $\tau = 0.18$, the new crossover value is 5.05.

For integer values, crossover is performed on a randomly selected number of integer values by copying directly from the two parents.

Example (Integer Crossover):

$$v_1 = [1 \ 3 \ 2 \ 4 \ 2]$$

$$v_2 = [2 \ 5 \ 3 \ 2 \ 3]$$

For example, using the five random numbers generated in Step 1 (0.2584, 0.6873, 0.1245, 0.3867, 0.8947) and a crossover cutoff point of 0.5, would result in the first integer value 1 being copied to the new chromosome.

However, the second random value is greater than the crossover cutoff point 0.5 resulting in the integer value 3 not being copied to the new chromosome.

Similarly, the other integer values create the final chromosome being

$$[1 \ 5 \ 2 \ 4 \ 3].$$

The mutation operation for the integer substring is also performed similar to the procedure described earlier in the example in Section B. For example, a random number is selected for every integer value. If the random number is less than a preset value, the integer value is replaced with a random integer within the range of 1 to 5. For the floating point substring, the mutation operation is performed by adding or subtracting to the substring, a random number, within some preset range to create the new substring.

Both crossover and mutation are performed at the knowledge level, that is, the maximum and minimum value resulting from the crossover and mutation operation is limited to within the preset range for the floating point and integer values. For example, the integer range is limited to integer values between 1 and 5.

Step 4: Produce the next generation of genetic strings and go to Step 2. Stop when the error is less than a preset value.

The crossover and mutation operations discussed in Step 3 produces a new set of chromosomes using the parent chromosomes from Step 2. The new parent chromosomes are then evaluated to determine their fitness values. If the average fitness value of the new chromosome is less than a preset value or the number of generations has reached a preset value, go to Step 5; otherwise, go back to Step 2 and perform a new generation of the algorithm.

Step 5: The best fit rule set and membership functions are the last generated set of genetic strings.

The genetic algorithm and fuzzy logic software used in this research was implemented on several networked SPARCStation 1, 5 and 10 hardware platforms located at the University of Tennessee Computer Science Department in Knoxville, Tennessee. The SPARCStations are developed by Sun Microsystems, Inc. located in Mountain View, California. The genetic algorithm and fuzzy logic software used in this research is parallelized using Parallel Virtual Machine (PVM). PVM is a software system that allows a heterogeneous network of computer systems, i.e., SPARCStations, to appear as a single large parallel computer. By using PVM, the large computational problem inherent in a genetic algorithm can be solved by using the aggregate power of several computers appearing as a single computer [104].

The time required to execute the genetic algorithm and fuzzy logic software using PVM is a function of the number of parallel processors. For example, Model A variables (10 independent variables) with 20 rules using two SPARCstation 10s, 24 SPARCstation 5s, and 31 SPARCstation 1s require an average execution time per generation of 27 seconds. The experiments performed in this research was conducted using a minimum of 2,000 generations requiring 15 hours per complete run.

Analysis of the software indicates that 69 percent of the computation time is involved with the fuzzification and applying the rules and 25 percent of the time is

involved with the defuzzification process. The remaining 6 percent is involved in the computer interface and general overhead.

G. Experimental Results - Fuzzy Expert System

The genetic algorithm searches for the “best fit” fuzzy rules and membership functions based on the independent variables previously identified using multiple linear regression analysis. For each of the four linear regression models, Models A, A1, B, and B1 from Chapter 5, the genetic algorithm is used to identify the fuzzy rules and membership functions for set of 3, 10, 20, 30, 40, 50, and 60 rules. The genetic algorithm performs a minimum of 2,000 generations or continues until the change in the NMSE for the last 200 generations approaches zero.

Figures 7-11, 7-12, 7-14, and 7-15 respectively show the NMSE versus the number of rules used for Models A, A1, B, and B1. In each figure, the ‘R’ value designates the NMSE for the multiple linear regression model. Tables 7-1 and 7-2 show the best fit fuzzy rules for Models A1 and B1. Figure 7-13 shows the fuzzy membership functions for Model A1.

Figures 7-16 and 7-17 respectively show the combination of Models A and A1, and Models B and B1. In addition, Figures 7-18 and 7-19 respectively show the combination of Models A and B, and Models A1 and B1. Figure 7-20 shows the NMSE versus all models for 3, 10, 20, 30, 40, 50 and 60 rules. The NMSE for the

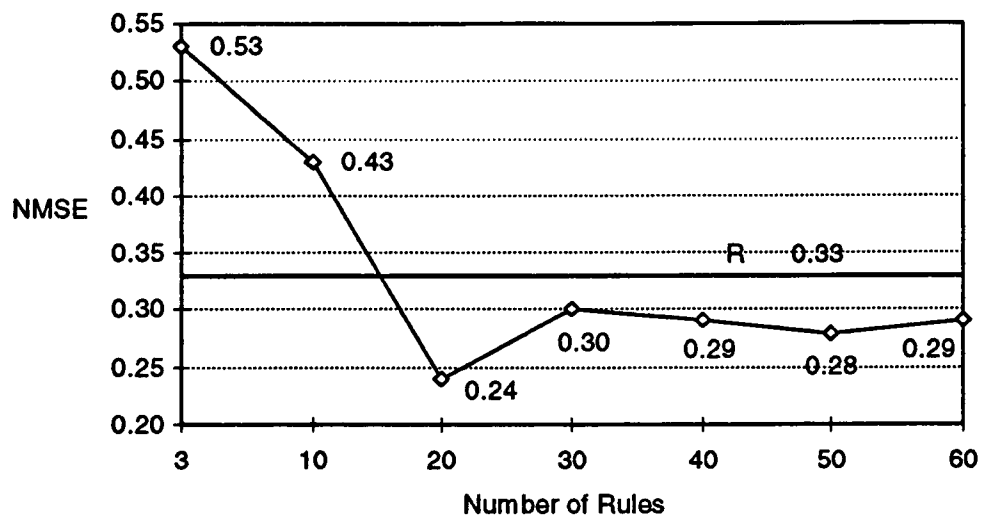


Figure 7-11. NMSE vs. Number of Rules for Model A.

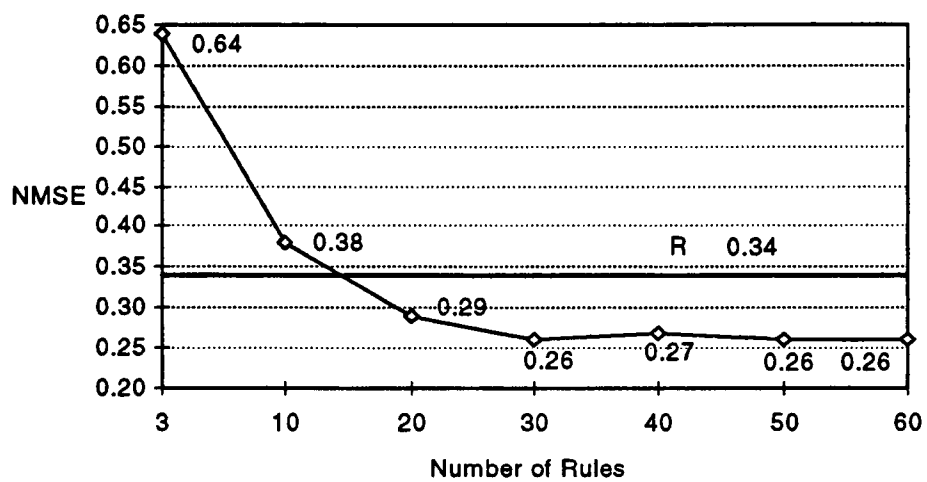
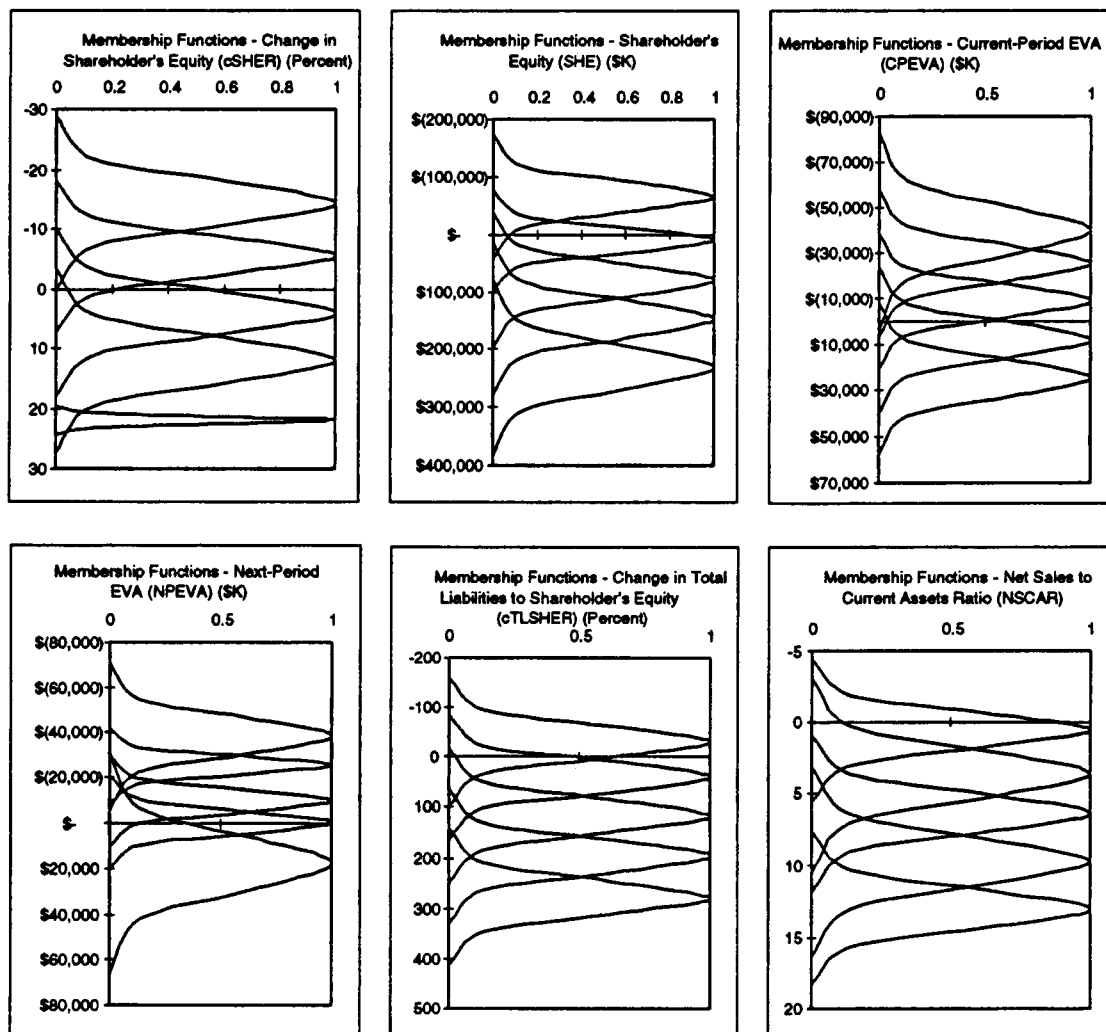


Figure 7-12. NMSE vs. Number of Rules for Model A1.

Table 7-1. Model A1 - 30 Rules.

Rule	IF	CPEVA	SHE	cSHER	NSCAR	cTLSHER	THEN	NPEVA
1		L	L	H	DC	VL		H
2		M	M	VL	DC	L		VL
3		M	L	VL	M	H		L
4		H	L	H	M	DC		VH
5		H	M	M	M	L		M
6		H	VL	VH	H	M		VL
7		H	VL	H	H	L		VH
8		H	DC	VH	VL	DC		VL
9		VL	H	L	L	DC		L
10		DC	DC	L	M	L		H
11		M	M	L	DC	VL		M
12		VH	VL	L	VL	DC		VH
13		VL	VH	L	L	DC		VL
14		VH	M	M	M	DC		VH
15		VL	L	L	M	VL		L
16		VH	H	L	DC	L		VH
17		VL	DC	H	VL	L		VL
18		VH	VH	L	M	M		VL
19		H	DC	DC	DC	VL		H
20		L	H	L	DC	M		VH
21		VL	H	VL	M	VH		H
22		DC	M	DC	VL	VH		H
23		VH	L	H	M	M		L
24		L	L	L	H	DC		VL
25		L	VL	H	VL	VL		VL
26		L	M	DC	L	VH		H
27		H	H	M	VL	VL		L
28		H	VL	M	L	VH		L
29		VH	L	L	L	M		M
30		VL	VL	VL	L	H		VH



The membership functions arranged from top to bottom are VL, L, M, H, and VH.

Figure 7-13. Model A1 Variable Membership Functions.

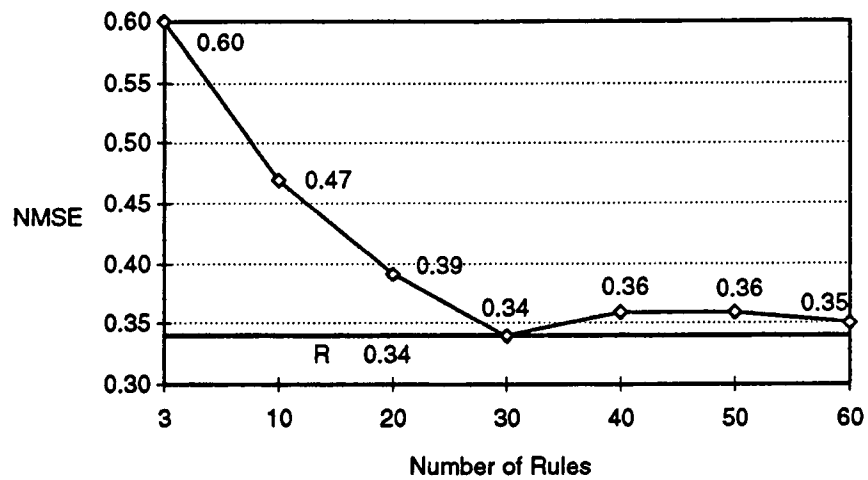


Figure 7-14. NMSE vs. Number of Rules for Model B.

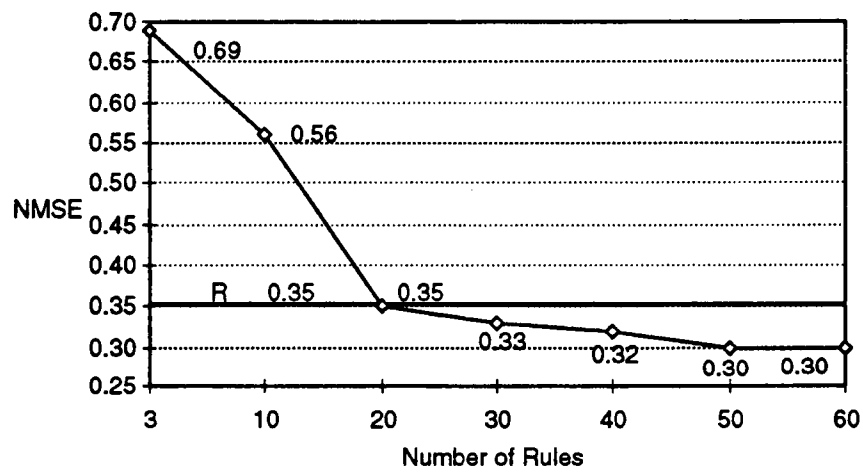


Figure 7-15. NMSE vs. Number of Rules for Model B1.

Table 7-2. Model B1 - 50 Rules.

Rule	IF	NSCAR	SHE	GP	RDE	SGAE	IE	THEN	NPEVA
1		M	H	L	H	VH	H		M
2		H	VH	VH	VL	H	M		M
3		VL	M	VH	VL	DC	VH		L
4		DC	M	DC	M	VL	VH		H
5		L	DC	H	L	L	DC		VH
6		VL	M	DC	DC	DC	VH		M
7		M	M	L	H	M	VH		VH
8		DC	H	H	DC	L	H		H
9		H	VL	VH	M	H	L		VH
10		M	VH	VL	VH	H	DC		M
11		DC	DC	L	L	VH	DC		VL
12		DC	VL	VH	VL	DC	VH		VH
13		DC	M	L	L	VL	DC		H
14		VH	VH	M	H	H	M		VH
15		VL	DC	VL	L	M	VH		L
16		VL	M	VL	VH	M	H		H
17		VH	M	M	VH	M	L		L
18		VL	L	L	M	L	VL		M
19		M	DC	L	M	VH	VH		VH
20		M	VH	DC	H	VL	VH		L
21		L	H	VL	VL	DC	M		H
22		L	L	H	H	H	H		VH
23		L	L	M	VH	DC	DC		L
24		M	VL	H	M	H	H		M
25		VL	DC	M	M	H	VL		H
26		M	M	DC	VL	L	VH		M
27		DC	L	VL	VL	VL	L		VH
28		H	M	VH	L	L	VH		L
29		DC	DC	L	VL	DC	M		L
30		L	L	M	VL	L	H		VH
31		VL	VH	M	VH	L	M		L
32		L	H	VL	DC	L	H		VL
33		DC	M	H	DC	VL	DC		L
34		H	DC	VH	M	VH	H		L
35		DC	L	DC	VH	DC	DC		M
36		M	M	VH	M	VH	M		H
37		DC	VL	DC	VH	H	DC		H
38		VL	L	L	M	L	VH		M
39		M	H	L	M	VL	DC		VL
40		L	L	DC	VL	DC	DC		H
41		VL	VL	VH	L	L	L		VL
42		DC	VL	DC	VL	DC	DC		H
43		H	VH	L	L	DC	VL		H
44		VL	VH	H	H	DC	M		VL
45		VH	L	VL	VH	DC	VH		VH
46		L	DC	H	VL	M	VL		H
47		H	DC	M	VL	M	DC		M
48		DC	M	L	VL	M	DC		M
49		VH	M	H	VL	DC	DC		VH
50		VL	M	H	VL	M	L		L

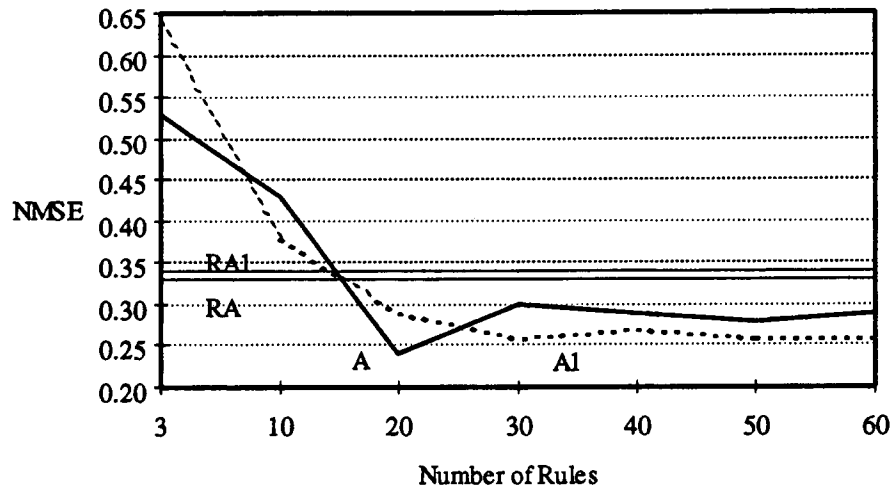


Figure 7-16. NMSE vs. Number of Rules for Models A and A1.

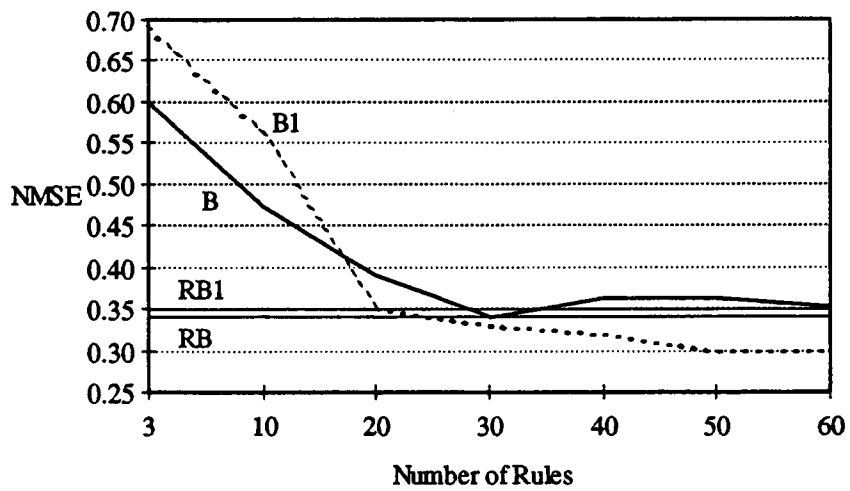


Figure 7-17. NMSE vs. Number of Rules for Models B and B1.

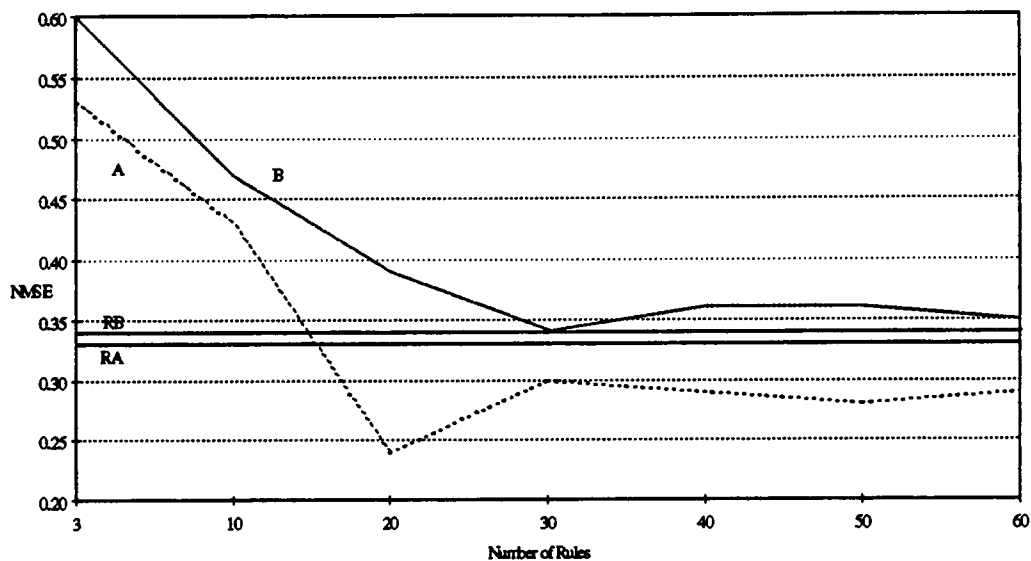


Figure 7-18. NMSE vs. Number of Rules for Models A and B.

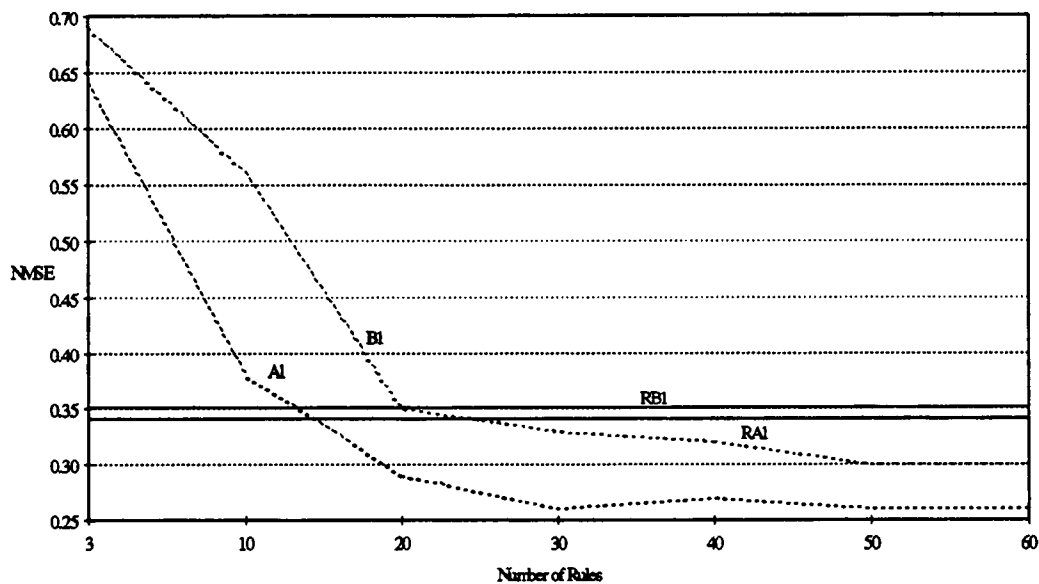
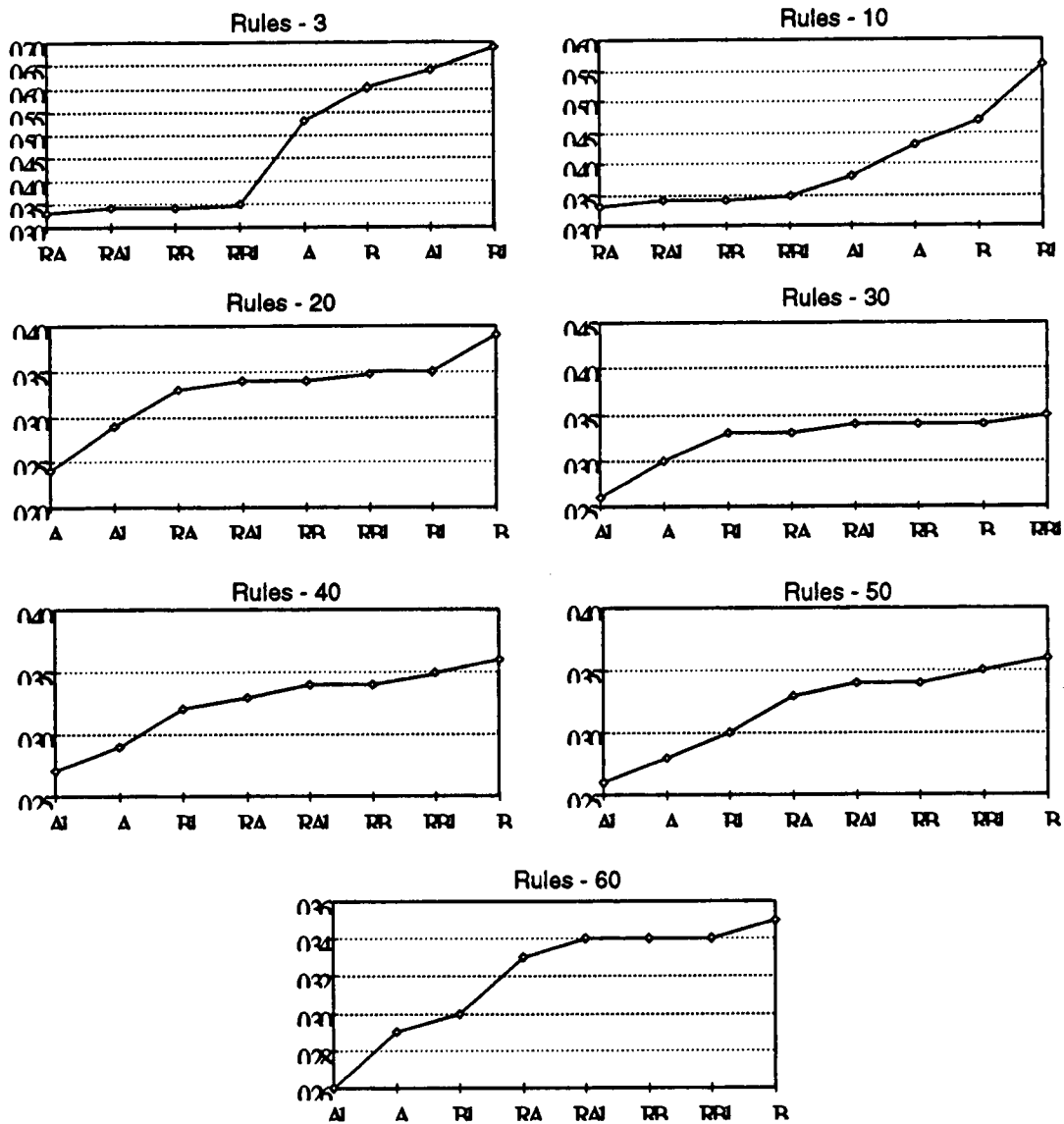


Figure 7-19. MSE vs. Number of Rules for Models A1 and B1.



Note: 'R' preceding the model name indicates a multiple linear regression model.

Figure 7-20. NMSE vs. Number of Rules.

multiple linear regression model is shown in Figures 7-16 through 7-19, and is designated by 'R' followed by the name of the model.

Several conclusions can be inferred from the experimental results shown in Figure 7-11 through Figure 7-19, and Table 7-1 and Table 7-2.

As shown in Figure 7-20, the fuzzy expert system model produces a smaller NMSE than does a multiple linear regression model for 20 or more rules. Below 20 rules, the multiple linear regression model has a smaller NMSE. The fuzzy expert system model also outperforms the multiple linear regression model for 20 or more rules even when the multiple linear regression model includes all quadratic and crossproduct terms. When CPEVA is included as an independent variable, there is a significant reduction in the NMSE using the fuzzy expert system model versus the multiple linear regression model from approximately 23 to 28 percent.

When the number of independent variables in Models A and B are reduced to include only those independent variables that result in a minimum increase of one percent improvement in R^2 , thus Models A1 and B1, the NMSE using the multiple linear regression equation slightly increases. However, when using the reduced independent variable set, Models A1 and B1, the fuzzy expert system results are a significant reduction in the NMSE, especially as compared to Models A and B shown in Figures 7-12, 7-15, 7-16, 7-17, 7-19 and 7-20.

Why does the NMSE increase for the fuzzy expert system model yet decreases for the multiple linear regression model when additional independent (significant)

variables are added? For Models A and B, the total number of potential rules and membership functions is over 362,797,056 whereas with Models A1 and B1, the total number of potential rules and membership functions is only 46,656 and 279,936 respectively. Thus, the total search space for the genetic algorithm for Model A1 is 7,776 times smaller than for Model A, and Model B1 is 1,296 times smaller than Model B. The genetic algorithm is using the same number of chromosomes for each model to search for the “best fit” fuzzy rules and membership function, however, for Models A1 and B1, the potential search space is significantly smaller resulting in a higher probability of identifying a better fit fuzzy rule and membership combination. However, in Figure 7-11, the minimum NMSE for Model A with 20 rules is 0.24 which is less than that found for the other models. The explanation is that the genetic string for Model A with 20 rules consists of a floating point [11 x 5] plus a binary [20 x 11] array for a total of 275 elements. However, Model A1 with 40 rules has the same number of genetic string elements as Model A with 20 rules. Therefore, the smaller the search space, the greater the probability that the genetic algorithm will find a better set of fuzzy rules and membership functions give a specific number of evaluations. In the research discussed in this dissertation, the genetic algorithm constrains the search space by the number of rules and the number of independent variables.

Another conclusion inferred from the experimental results is that beyond a specific number of rules, adding more rules do not contribute to reducing the NMSE.

For example, in Model A, 20 rules results in the minimum NMSE. This also correlates to the number of clusters at which the change in the R^2 value in Figure 6-9 approaches a minimum. This same analogy is also true for Model B which has a minimum number of 30 clusters.

Why does the NMSE not improve with additional rules? As shown using cluster analysis in Models A and B, the additional information that is available in the training data set for learning new rules decreases significantly beyond approximately 20 to 30 rules. Beyond 20 to 30 rules, additional clusters are represented by only a few, and in some cases just one single observation, therefore, corresponding to a single rule for a single observation. Thus, a single rule's contribution to reducing the NMSE for a single observation is small considering the 1,000 observations contained in the training data set. Another reason why the NMSE does not improve with additional rules is the same explanation discussed earlier; more rules results in the genetic algorithm searching over a larger search space to identify the "best fit" fuzzy rules and membership functions.

When the number of independent variables are reduced in the model, however, the number of rules that result in a minimum NMSE increase. That is, with fewer variables used in classifying the data set, more clusters or rules are required for learning the "best fit" fuzzy rules and membership functions. For example, with Model A, the minimum NMSE results in 20 rules, however with Model A1, the minimum NMSE results in 30 rules. For Model B, the minimum NMSE occurs with 30 rules,

and with Model B1, the minimum NMSE occurs with 50 rules. From Figures 6-9, 6-10, 6-11 and 6-12, the change in the R^2 value versus the number of clusters approaches zero beyond the number of rules that correspond to the minimum NMSE using the fuzzy expert system model. Thus, the more independent variables, the more specific each individual cluster resulting in a better fit of the data, and the fewer number of clusters required for describing the training data set. This can also be seen by observing that the R^2 value for 50 clusters is greater for Models A and B than it is for Models A1 and B1.

Given the four models, the minimum NMSE results in Model A with 20 rules followed by Model A1 with 30 rules. The fuzzy rules for Model A1 is shown in Table 7-1 and the fuzzy membership functions are shown in Figure 7-13. Notice that in Figure 7-13, the fuzzy membership functions are not uniformly spaced but have been evolved by the genetic algorithm to minimize the NMSE when applied to the training data set.

The difference in the fuzzy expert system for Model A with 20 rules versus the multiple linear regression equation for Model A is a reduction in the NMSE of approximately 28 percent.

H. Summary

A genetic algorithm is a search technique based on Darwin's theory of evolution by natural selection. In this research, a genetic algorithm uses historical

information, which is called the training data set, to search for the “best fit” fuzzy rules and membership functions that minimizes the error between the predicted value using the fuzzy rules and membership functions and the training data set.

Four sets of “best fit” fuzzy rules and membership functions, using 3 to 60 rule combinations, were found using a genetic algorithm. These best fit fuzzy rules and membership functions are compared to a multiple linear regression model. In summary, the results of this comparison are as follows:

- The fuzzy expert system model produces a smaller NMSE than does a multiple linear regression model for 20 or more rules.
- When CPEVA is included as an independent variable, there is a significant reduction in the NMSE using the fuzzy expert system model versus the multiple linear regression model from approximately 23 to 28 percent.
- When the number of independent variables in Models A and B is reduced to include only those independent variables that result in a minimum increase of one percent improvement in R^2 (Models A1 and B1), the NMSE using the fuzzy expert system decreases.
- The smaller the search space, the greater the probability that the genetic algorithm will find a better set of fuzzy rules and membership functions given a specific number of evaluations.
- Beyond a specific number of rules, adding more rules do not contribute to reducing the NMSE.

Chapter 8

Validation

A. Introduction

As previously discussed, the validation data set includes 91 companies chosen at random from a population of small, publicly owned, technology-based companies. A statistical hypothesis test of the training and validation data set using a Student's t-test at the 0.05 confidence level was performed for the mean values of current-period EVA, next-period EVA, shareholder's equity, and change in total debt to shareholder's equity ratio. There was no statistical evidence of any significant difference between the mean values for the training and validation data sets among these pairs.

For each of the four fuzzy expert system models discussed in Chapter 7, the validation data set is tested using the fuzzy rules and membership functions corresponding to the minimum NMSE for each model. The four multiple linear regression models are also tested using the validation data set. Figure 8-1 and 8-2 show the results. In each figure, the NMSE for a specific model is shown for both the validation data set, designated by a small 'v' following the name of the model, and the training data set designated by a small 't' following the name of the model.

B. Validation Results

Several conclusions can be inferred from analyzing the validation data set using the multiple linear regression and the fuzzy expert system models. The comparative results are depicted in Figures 8-1 and 8-2.

The fuzzy expert system with the Model A variables does not reduce the NMSE, as compared to the multiple linear regression model, when using the validation data set. For example, the fuzzy expert system using the Model A variables has a NMSE of 0.24 with the training data set but only 0.37 with the validation data. Likewise, the fuzzy expert system with the Model B variables does not reduce the NMSE using the validation data. The same is also true for Models A and B when applying the multiple linear regression model to the validation data. However, for Models A1 and B1, which have the reduced number of variables, there is a significant reduction in the NMSE when applying the fuzzy expert system model to the validation data set in comparison to the multiple linear regression model.

Why do Models A and B, which have more independent variables than Models A1 and B1, perform worse when applied to the validation data set? This phenomenon may be explained by excessive learning in the development of Models A and B when applied to the training data set. As demonstrated from the multiple linear regression analysis, there is considerable noise associated with the training data set. This was verified by the large MSE shown in Tables 5-10 and 5-16. For the regression model, the total error sum of squares equals 33 percent of the total sum of squares.

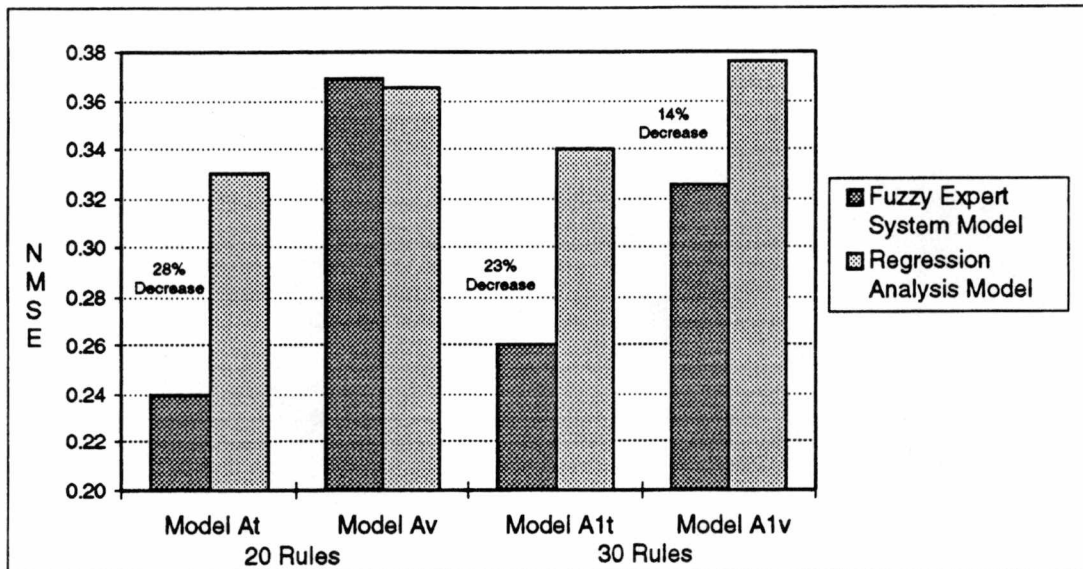


Figure 8-1. Model A and A1 Validation Results.

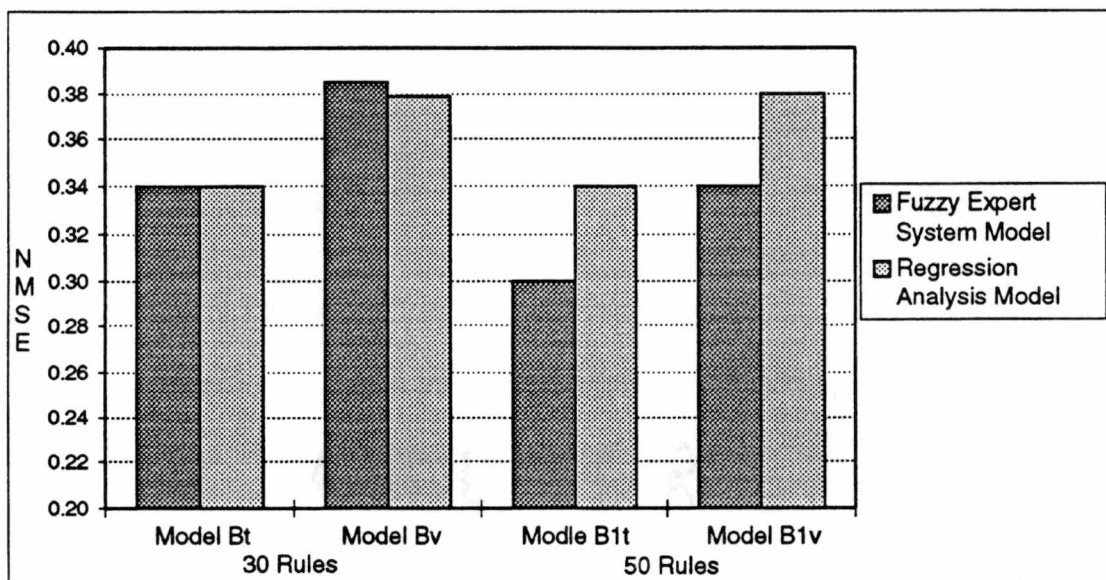


Figure 8-2. Model B and B1 Validation Results.

Thus, approximately 33 percent of the variation in the data can be explained by the error term in the regression equation. This error term can also be viewed as unexplained noise in the observation data. The genetic algorithm does not distinguish between the actual signal and random noise in the training data. The genetic algorithm attempts to find the “best fit” fuzzy rules and membership functions given the data. As a result, the “best fit” fuzzy rules and membership functions learned in developing Models A and B includes results that minimize the NMSE given the *combined signal and noise*. However, the noise associated with the training data set may not be the same noise as that found in the validation data set. As a result, a model with fewer attributes or independent variables will tend to be less sensitive to random noise and more responsive to general trends in the observation data.

In conclusion, the fuzzy expert system with independent variables that only include terms that contribute measurably to increasing the R^2 value in the multiple linear regression equation, outperforms the multiple linear regression model when applied to both the training and validation data set. In addition, the best multiple linear regression and fuzzy expert system model, i.e., the model with the smallest NMSE, has the most significant independent variables.

However, an important criteria for a good learning expert system model for use in this application is one that provides an accurate prediction of the financial performance of small technology based companies, (within the limits of the data) as

well as provides insight and understanding into the predictive structure of the data.

The next section discusses the information content provided by both types of models.

Chapter 9

Information Content

A. Introduction

The usefulness of a model is not based just on its ability to accurately predict a future event, but also on its ability to provide the user with additional information content about a specific application. The information content provided by the multiple linear regression model and the fuzzy expert system model in this research is different, and each provides its own unique perspective into the analysis of the problem. No research in model development would be complete without an analysis of the information content provided by the model. For this reason, a comparative discussion of the information content provided by the multiple linear regression model and the fuzzy expert system model is provided in this chapter.

B. Information Content Analysis

The information content provided using a multiple linear regression analysis model includes the significant contribution made by each of the independent variables or attributes given in the observation data. In addition, the regression coefficients indicate the direction and magnitude the independent variable contributes to the predicted variable. For example, in Model A1, the multiple linear regression model

(equation) is: $NPEVA = -\$612.93 + 0.492(CPEVA) - 0.066(SHE) + 0.338(cTLSHER) + 362.2(NSCAR) - 2.82(cSHER)$. The Model A1 multiple linear regression equation indicates that a one unit increase in the current-period EVA will result in a 0.492 unit increase in the next-period EVA. Likewise, a one unit increase in shareholder's equity will decrease the next-period EVA by 0.066 units.

Advanced multiple linear and non linear regression models that include crossproduct and quadratic terms become more difficult to interpret. In most cases, these multiple linear regression equations have no relationship to the physical structure of a problem other than providing additional regression coefficient terms to reduce the MSE and increase the R^2 value.

The fuzzy expert rules and membership functions that were shown in Table 7-1 and Figure 7-13 respectively, provide additional information, plus a better fit to the training and validation data set than does the multiple linear regression model. For example, consider Rule 7 in Table 7-1: IF CPEVA is VH, and cSHER is H, and NSCAR is H, and cTLSHER is L, THEN NPEVA is VH. Rule 7 is unique, that is, none of the other 29 rules in Table 7-1 are identical to Rule 7. In fact, all the rules are unique. Therefore, based on the Zedah AND implication operator used in this research, if a firm had independent variables equal to the value stated in Rule 7 to a degree of 1.0, the predicted NPEVA would be VH to a degree of 1.0. However, in most cases, the final NPEVA result will be a combination of several rules applied in parallel. The parallel nature of the fuzzy implication operation also makes it more

difficult to develop an intuitive feeling for the rules. Nevertheless, the rules do provide additional insight into the problem. For example, consider the information in Table 7-2 which shows the fuzzy rules for Model B1. It can be concluded from this information that when a firm incurs a minimum amount of high research and development costs, the next-period EVA is likely to be medium. This would indicate there is a positive relationship between the amount of research and development expenses and a firm's next-period EVA.

In addition to fuzzy rules, additional information content is provided by observing an independent variable's membership function. For example, in Figure 7-13, a very high shareholder's equity ranging from \$200 million to \$260 million to a degree of 0.5 is shown. Likewise, a very high shareholder's equity is \$232 million to a degree of 1.0. This information can be very useful when describing and comparing the financial statements of different companies. A fuzzy membership function can be used to provide quantitative meaning when describing financial terms like very high profits, low gross margins, and high research and development expenses.

Additional information content is also provided by observing the output membership function for a specific set of input variables. For example, consider the management of a firm which is developing the next year's financial budget. As part of the planning process, the management team estimates the values for each significant independent variable and then uses the fuzzy expert system to generate a predicted NPEVA. Figures 9-1 and 9-2 show the output fuzzy membership functions that might

be generated given a set of estimated crisp input values generated by the management team. The output membership function shown in Figure 9-1 indicates there is a potential that the NPEVA value may range from a low of negative \$70 million to a high of \$70 million with the most likely value being the centroid or \$8 million. With a wide range in potential output values, none with a membership value above 0.7, considerable variation in NPEVA can be expected. Contrast this with Figure 9-2. In Figure 9-2, the potential NPEVA ranges from a negative \$30 million to \$70 million with the most likely value of \$20 million corresponding to a membership function value of 1.0. Consequently, Figure 9-2 shows that the estimated input values used in generating the output membership functions is more likely to occur than those shown in Figure 9-1.

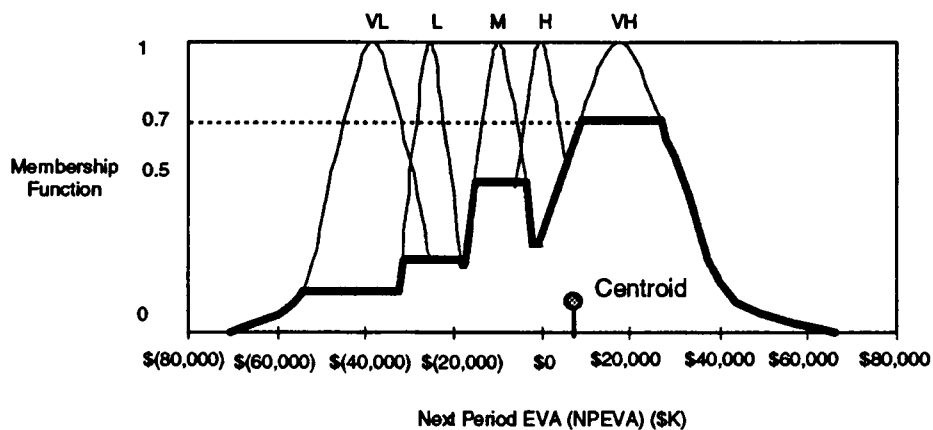


Figure 9-1. NPEVA Output Membership Function (Maximum 0.7).

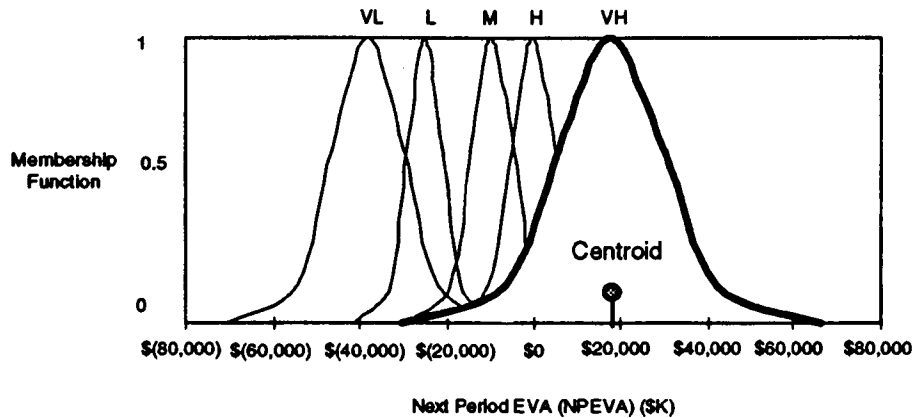


Figure 9-2. NPEVA Output Membership Function (Maximum 1.0).

Figure 9-3 is a three dimensional membership function surface view that represents the mapping of R & D expenses and SG&A expenses to the NPEVA value. For example, Figure 9-3 shows the surface value of the fuzzy expert system for Model B1 with medium (as defined by their membership function) NSCAR, SHE, IE, and GP. The lighter intensity pixels represent a higher NPEVA value. Interpreting Figure 9-3 indicates that as the SG&A expenses decrease, and R&D expenses increase, the NPEVA is likely to increase. However, as SG&A expenses increase, and R&D expenses decrease, the NPEVA is very low. Membership surface views can be used to obtain an intuitive feeling of the relationship between two independent variables and the response. In addition, surface views can be used to explore the nonlinear relationships between independent variables. The surface map can also be used to test the intuitiveness of the fuzzy rules, and as a pictorial tool in explaining the relationships between independent variables and the predicted variable. Figures 9-4

and 9-5 show additional membership function surface views for other significant independent variables.

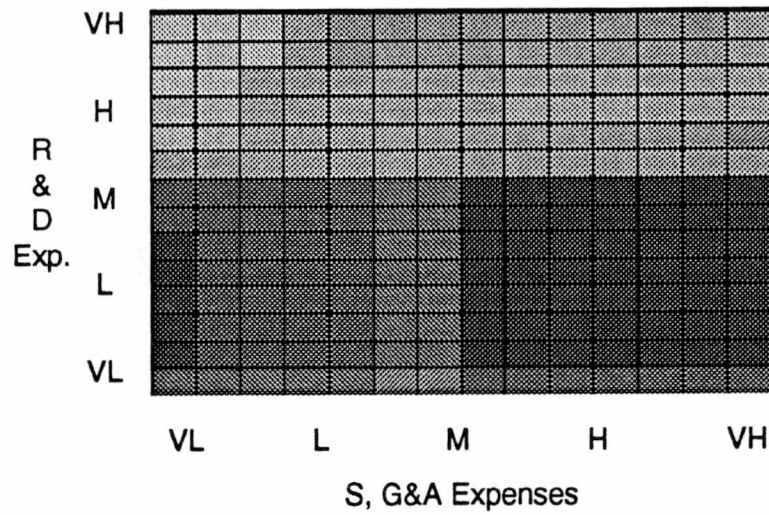


Figure 9-3. Surface Profile - NPEVA versus R &D and S, G & A Expenses.

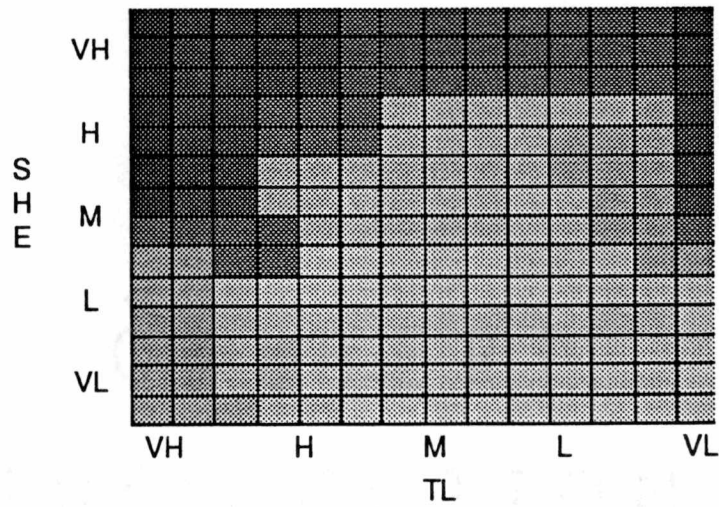


Figure 9-4. Surface Profile - NPEVA versus SHE and TL.

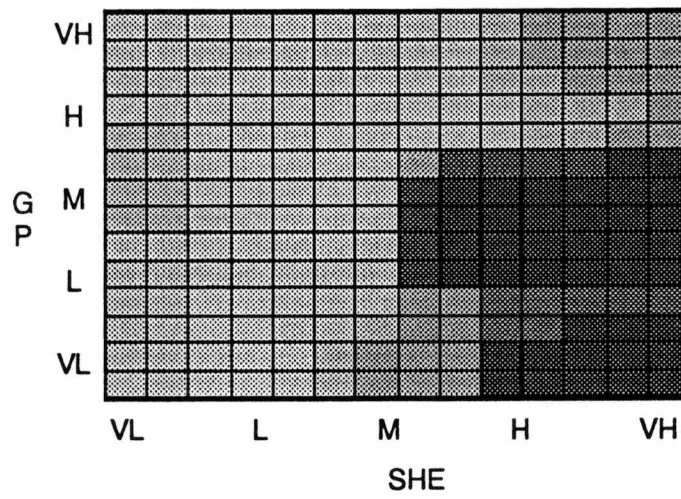


Figure 9-5. Surface Profile - NPEVA versus GP and SHE.

Chapter 10

Conclusions

A. Summary

How do you make a financial decision about a firm? Do you look at profitability of the firm, geographic location, product quality, age of the company, or competitive advantage? There are many factors to consider in making a good financial decision, but are they the *right* ones?

Most people make informed financial decisions through personal experience and knowledge. But the number of factors to consider is large. Any problem that takes into consideration only 10 attributes (for example company age, location, productivity, etc.) mathematically has a combination of several million potential rules. How do you know what is the best combination of rules? For most people, the selection process is based on past experience and knowledge. Yet, there is a way to find the best fit rules based on historical performance using technology recently developed and used in this research. To put it simply, using a computer with a genetic algorithm and fuzzy mathematics, can help to determine the best rules for successful financial decision making.

This dissertation discusses the development of a knowledge-based learning expert system methodology that can be used to identify the rules which predict the financial performance of small, technology-based companies. A genetic algorithm is

used to determine the best rules. In addition, the genetic algorithm determines the fuzzy membership functions for antecedent and consequent variable related to each rule. A fuzzy implication operator then determines the resultant output that represents the predicted financial performance of the small, technology-based company.

The knowledge-base used in this research consists of 250 small, publicly traded, technology-based companies. The fuzzy rules and associated membership functions are developed from this historical data base. An additional 91 small, publicly traded, technology-based companies is used to validate the model.

More than 145 different independent variables for each company in the knowledge-base (training data set) are evaluated to determine significance of each variable in predicting the financial performance of a firm as measured by the next-period EVA, the dependent variable used. Of the 145 different independent variables, 114 variables are internal financial variables, that is, variables that are associated with a firm's financial statements. The remaining 31 variables are external economic variables including general business cycle indicators such as the prime interest rate, new orders in the manufacturing sector, the S&P 500 composite index, indexes of industrial production, the composite index of leading economic indicators, and the gross domestic product. In addition, the period-to-period changes in the 145 variables are also evaluated.

To assist in determining the significance of a variable in predicting a firm's next-period EVA, several multiple linear regression models are developed and

evaluated using statistical analysis techniques. Surprisingly, none of the external economic variables significantly contributed in predicting a firm's next-period EVA.

Four "best fit" multiple linear regression models are developed. Using the independent variables identified in the four models, cluster analysis is used to determine the number of potential fuzzy rules for each model. The independent variables in the multiple linear regression models are then used by the genetic algorithm to derive the "best fit" fuzzy rules and membership functions. Finally, the 91 company validation data set is used in validating the four multiple linear regression models and their corresponding fuzzy rules and membership functions.

The fuzzy expert system model produces a smaller NMSE than does the multiple linear regression model for 20 or more rules. Below 20 rules, the multiple linear regression model has a smaller NMSE. After reducing the number of independent variables in the multiple linear regression model to include only those variables that contribute to improving the R^2 value, the fuzzy expert system generated a significant reduction in the NMSE. The smaller NMSE is partially due to the reduced search space involved resulting in a higher probability of identifying a "better fit" fuzzy rule and membership function combination.

The NMSE for the fuzzy expert system does not necessarily improve with additional rules. The additional information that is available in a training data set for learning new rules decreases significantly beyond a specific number of data clusters. Beyond this, additional data clusters are represented by only a few, and in some cases

just one single observation, therefore, corresponding to a single rule for a single observation. Thus, a single rule's contribution to reducing the NMSE for a single observation is small considering the observations contained in the training data set.

The independent variables identified as the "best fit" based on multiple linear regression analysis also correspond to the "best fit" variables used by the fuzzy expert system. These independent variables also correspond to the "best fit" multiple linear regression and the fuzzy expert system model for the validation data set.

The optimal number of fuzzy rules for a specific set of independent variables strongly corresponds to the number at which the R^2 value becomes constant. As a result, cluster analysis may be used to assist in identifying the number of fuzzy rules potentially available in a knowledge-base. Using cluster analysis to identify the number of fuzzy rules can significantly reduce the time required to develop a fuzzy expert system.

The reduction in the NMSE, when using the fuzzy expert system with the training data set, does not necessarily contribute to a reduction in the NMSE when applied to the validation data set. However, for the reduced Models A1 and B1, there is a significant reduction in the NMSE when applying the fuzzy expert system model to the validation data. Excessive learning applied to the training data could explain this phenomenon.

As demonstrated from the multiple linear regression analysis, there is considerable noise associated with the training data. The genetic algorithm attempts

to search for the “best fit” fuzzy rules and membership functions given the training data. The genetic algorithm does not distinguish between the actual signal and random noise and attempts to find the “best fit” fuzzy rules and membership functions given the data. As a result, the “best fit” fuzzy rules and membership functions learned include results that minimize the NMSE given the combined signal and noise. However, the noise associated with the training data set may not be the same noise as that found in the validation data set. As a result, a model with fewer attributes or independent variables will tend to be less sensitive to random noise and more responsive to general trends in the observation data. Therefore, the fuzzy expert system with independent variables that only include terms that contribute measurably to increasing the R^2 value in the multiple linear regression equation, outperformed the multiple linear regression model when applied to both the training and validation data set.

The information content provided by the multiple linear regression and the fuzzy expert system model provide its own unique perspective into the analysis of a problem. By using a multiple linear regression model, the information content provided includes the significant contribution made by a set of independent variables or attributes given the observation data. In addition, the regression coefficients indicate the direction and magnitude the independent variable contributes to the predicted variable.

Advanced multiple linear and nonlinear regression equations that include crossproduct and quadratic terms become more difficult to interpret. In most cases, these multiple linear regression equations have no relationship to the physical structure of a problem other than providing additional regression coefficient terms to reduce the MSE and increase the R^2 .

The fuzzy expert rules and membership functions provide additional information, plus a better fit to the training and validation data set than does a multiple linear regression model. The fuzzy rules provide an additional insight into the application and can be used to better intuitively understand the problem.

In addition to fuzzy rules, additional information content is also provided by observing an independent variable's membership function. This information can be very useful when describing companies financial statements. A fuzzy membership function can be used to provide a quantitative meaning in describing financial terms like very high profits, low gross margins, and high research and development expenses.

Additional information content is also provided by observing a two- or three-dimensional output membership function for a specific set of input variables. Membership surface views can be used to create an intuitive feeling of the relationship between two independent variables and the response. In addition, surface views explore the nonlinear relationships between independent variables. The surface views can also be used to test the intuitiveness of the fuzzy rules and as a pictorial tool in explaining the relationships between independent variables and the predicted variable.

Other business and engineering applications that can take advantage of the type of fuzzy expert system model discussed in this dissertation includes:

- Competitive Analysis
- Risk Capital Venture Investing
- Optimal Site Selection
- Production Planning and Facility Layout
- Medical Diagnostic
- Mergers and Acquisition Planning

B. Future Research

This research shows that the combined use of multiple linear regression analysis, cluster analysis, and a genetic algorithm can determine the “best fit” fuzzy rules and membership functions for a given set of training data. This combination creates a technique that can be used to develop a learning expert system model based on historical information.

However, there are several additional areas of research that need exploring before the technique discussed in this dissertation can be used for more advanced applications. Some suggestions for future research include:

1. This research only included financial attributes as independent variables for predicting a future EVA of a firm. However, several other non financial attributes such as management experience and companies competitive advantages (discussed in Chapter 1, “Introduction”) are also important in determining the financial success of a firm. If a reasonable and cost-effective method is developed

for quantifying some of these non financial attributes, they should be used in the development of the model.

2. The training and validation data sets included only small, technology-based, publicly owned firms. Based on preliminary research, the fuzzy rules should also apply to small, technology-based privately owned firms, however, the magnitude of the fuzzy membership functions will likely change. Additional research is needed to verify this fact. Furthermore, research could determine if the fuzzy rules and membership functions developed in this research are also applicable to non technology-based companies.
3. This research uses multiple linear regression analysis to determine the number and actual set of significant independent variables for the models. However, investigating other techniques such as using quadratic and higher order terms and crossproduct terms could also help to determine the significant independent variables.
4. This research uses the k-mean cluster analysis technique to determine the number of data clusters as a guide to the number of fuzzy rules. Other clustering techniques have recently been developed including the use of neural networks to cluster data. Additional analysis should be performed to investigate alternative clustering approaches.
5. The genetic algorithm encoding technique used in this research includes a hybrid of both floating and binary encoding. Several other encoding techniques are

available and should be explored. In addition, the genetic algorithm searches for the “best fit” fuzzy rules and membership functions given a constraint on the number of rules. Other types of search strategies could also be used, for example, letting the genetic algorithm evolve the number of rules based on certain constraints. In addition, once the genetic algorithm identifies a near best fit solution, hill climbing search routines could also be used to perform fine tuning of the fuzzy rule set and membership functions.

6. The implication operator used in this research was the Zadeh AND operator. This is the fundamental way in which fuzzy sets are combined; however, several recent research investigated alternative implication operators. Other types of implication operators were developed to compensate for the strict minimum and maximum of the Zedah operators. The compensatory operators generally provide a weaker or less sensitive relationship among attributes when their truth values are separated. Other implication operators include mean, mean^2 , $\text{mean}^{1/2}$, product, bounded sum, Yager, Zimmerman and Zysno, Dubois and Prade. Further investigation could determine the best type of implication operator for a specific application.
7. This research used the centroid method to determine the combined fuzzy output crisp number or to defuzzify the output membership function. Several other types of defuzzification methods are also available and should be investigated. These

include composite maximum, center of evidence maximum, average of maximum plateau, far and near edge of the support set, and center of maximum.

8. Research should be conducted to better understand the types of functions that fuzzy rules and membership functions can adequately model compared to other alternative modeling approaches.
9. Any training data set will include both signal and noise. The genetic algorithm searches for the “best fit” fuzzy rules and membership functions give the training data, regardless of the type and amount of noise present. New techniques should be investigated to pre-filter the noise and then evolve the fuzzy rules and membership functions on the filtered data.
10. This research uses a genetic algorithm search technique to determine the “best fit” fuzzy rules and membership functions to predict a firm’s next-period EVA. The results of this analysis should be compared with a neural network and a fuzzy neural network, as well as the differences evaluated and discussed.
11. A genetic algorithm search technique requires extensive computational resources. New algorithms, computing techniques, and hardware implementation should be investigated to speed the search process.

References

References

- [1] U. S. Department of Commerce, *Critical Technology Assessment of the U. S. Artificial Intelligence Sector*, NTIS Publication No. PB 93-192409, Washington, DC: Government Printing Office, 1994.
- [2] Daniel E. O'Leary, Ed., *Expert Systems in Finance*. New York: North-Holland, 1992, p. 15.
- [3] O'Leary, p. 8.
- [4] Dimitris N. Chorafas, *New Information Technologies: A Practitioner's Guide*. New York: Van Nostrand Reinhold, 1992, pp. 52-53.
- [5] Joseph Giarratano and Gary Riley, *Expert Systems Principles and Programming*, 2nd ed. Boston: PWS Publishing Company, 1994, p. 1.
- [6] Arie Ben-David and Juice Mandel, "Classification Accuracy: Machine Learning vs. Explicit Knowledge Acquisition," *Machine Learning*, vol. 18, p.109, 1995.
- [7] Joseph Giarratano and Gary Riley, pp. 7-8.
- [8] Arie Ben-David and Juice Mandel, p. 109.
- [9] R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, Eds., *Machine Learning, An Artificial Intelligence Approach*. Los Altos, CA: Morgan Kaufmann, 1983, p. 14.
- [10] R. M. Feigenbaum, "A Learning Machine Part 1," *IBM Journal of Research and Development*, vol. 2, pp. 2-13, 1958.
- [11] R. M. Feigenbaum, B. Dunha, and T. North, "A Learning Machine: Part 2," *IBM Journal of Research and Development*, vol. 3, no. 3, pp. 282-287, 1959.
- [12] J. H. Holland, "Adaptive Algorithms for Discovering and Using General Patterns in Growing Knowledge Bases," *Policy Analysis and Information Systems*, vol. 4, no. 2, September 1980.
- [13] R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, p. 14.

- [14] W. W. Bledsoe, "The use of biological concepts in the analytical study of systems," presented at the ORSA-TIMS National Meeting, San Francisco, CA, November 1961.
- [15] H. J. Bremermann, "Optimization through evolution and recombination," In M. C. Yovits, G. T. Jacobe, and G. D. Goldstein, Eds., *Self-organizing systems*, Washington, DC: Spartan Books, pp. 93-106, 1962.
- [16] J. H. Holland, "Outline for a logical theory of adaptive systems," *Journal of the Association for Computing Machine*, vol. 3, pp. 297-314, 1962.
- [17] J. H. Holland, "Some Practical Aspects of a Adaptive Systems Theory," In A. Kent & O. E. Taulbee, Eds., *Electronic Information Handling*, Washington, DC: Spartan Books, pp. 209-217, 1965.
- [18] Charles T. Walbridge, "Genetic Algorithms: What Computers Can Learn from Darwin," *Technology Review*, vol. 92, pp. 46-53, January 1989.
- [19] Peter J. Denning, "Genetic Algorithms," *American Scientist*, vol. 80, pp. 12-14, January - February 1992.
- [20] David E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. New York: Addison-Wesley Publishing Company, pp. 126-129, 1989.
- [21] Lawrence Davis, Ed., *Handbook of Genetic Algorithms*. New York: Van Nostrand Reinhold, 1991.
- [22] Goldberg, pp. 126-129.
- [23] J. H. Holland and J. S. Reitman, "Cognitive Systems Based on Adaptive Algorithms," In D. A. Waterman & F. Hayes-Roth, Eds., *Pattern directed inference systems*, New York: Academic Press, pp. 313-329, 1978.
- [24] Goldberg, pp. 268-270.
- [25] S. F. Smith, *A Learning System Based on Genetic Adaptive Algorithms*. Doctoral dissertation, University of Pittsburgh, 1980.
- [26] Goldberg, pp. 218-221.

- [27] Goldberg, pp. 126-129.
- [28] John J. Grefenstette, Ed., *Genetic Algorithms for Machine Learning*. Boston: Kluwer Academic Publishers, p. 1, 1994.
- [29] Hojjat Adeli and Shih-Lin Hung, *Machine Learning Neural Networks, Genetic Algorithms, and Fuzzy Systems*. New York: John Wiley & Sons, p. 6, 1995.
- [30] Lotfi A Zadeh, "Fuzzy Sets," *Information and Control*, vol. 8, pp. 338-353, 1965.
- [31] E. H. Mamdani, "Applications of Fuzzy Algorithms for a Simple Dynamic Plant," *Proc. IEEE*, vol. 121, no. 12, pp. 1585-1588, 1975.
- [32] E. H. Mamdani and S. Assilian, "An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller," *Int. Journal Man-Machine Studies*, vol. 7, pp.1-13, 1975.
- [33] E. H. Mamdani, "Advances in the Linguistic Synthesis of Fuzzy Controllers," *Int. Journal of Man-Machine Studies*, vol. 8, pp. 669-672, 1976.
- [34] Toshiro Terano, Kiyoji Asai, and Michio Sugeno, *Fuzzy Systems Theory and Its Applications*. New York: Academic Press, pp. 9-16, 1992.
- [35] Earl Cox, *Fuzzy Logic for Business and Industry*. Rockland, Massachusetts: Charles River Media, Inc., pp.38-40, 1995.
- [36] Y. Y. Chen, K. Z. Lin, and S. T. Hsu, "A Self-Learning Fuzzy Controller," *Proc. of IEEE International Conference on Fuzzy Systems*, pp. 189-196, 1991.
- [37] H. R. Berenji, "Learning and Tuning Fuzzy Logic Controllers Through Reinforcements," *IEEE Trans. On Neural Networks*, pp. 724-739, September 1992.
- [38] W. C. Dougherty, R. Rathakrishnan, and J. Yen, "Performance Evaluation of a Self-Tuning Fuzzy Controller," *Proc. of IEEE International Conference on Fuzzy Systems*, pp. 389-397, 1992.
- [39] J. Fei, and C. Isik, "Adaptive Fuzzy Control via Modification of Linguistic Variables," *Proc. of IEEE International Conference on Fuzzy Systems*, pp. 399-406, 1992.

- [40] D. G. Burkhardt and P. P. Bonissone, "Automated Fuzzy Knowledge Base Generation and Tuning," *Proc. of IEEE International Conference on Fuzzy Systems*, pp. 179-188, 1991.
- [41] S. M. Smith and D. J. Comer, "An Algorithm for Automated Fuzzy Logic Controller Tuning," *Proc. of IEEE International Conference on Fuzzy Systems*, pp. 615-621, 1992.
- [42] C. L. Karr, *Analysis and Optimization of an Air-Injected Hydrocyclone*, Doctoral dissertation, University of Alabama, 1989.
- [43] C. L. Karr, "Applying Genetics to Fuzzy Logic," *AI Expert*, vol. 6, no. 3, pp. 38-43, March 1991.
- [44] C. L. Karr, "Genetic Algorithms for Fuzzy Controllers," *AI Expert*, vol. 6, no. 2, pp. 26-33, February 1991.
- [45] C. L. Karr, D. L. Meredith, and D. A. Stanley, "Fuzzy Process Control with a Genetic Algorithm," *Control '90 - Mineral and Metallurgical Processing Conference*, pp. 53-60, February 1990.
- [46] C. L. Karr, "Design of an Adaptive Fuzzy Logic Controller Using a Genetic Algorithm," *Proceedings of the Fourth International Conference on Genetic Algorithms*, pp. 450-457, July 1991.
- [47] C. L. Karr and R. J. Gentry, "Fuzzy Control of pH Using Genetic Algorithms," *Proc. of IEEE International Conference on Fuzzy Systems*, pp. 46-53, February 1993.
- [48] Philips Thrift, "Fuzzy Logic Synthesis with Genetic Algorithms," *Proceedings of the Fourth International Conference on Genetic Algorithms*, pp. 509-513, July 1991.
- [49] Micahel A. Lee and Hideyuki Takagi, "Integrated Design Stages of Fuzzy Systems Using Genetic Algorithms," *IEEE 2nd Int. Conf. On Fuzzy Systems*, pp. 612-617, 1993.
- [50] Micahel A. Lee and Hideyuki Takagi, "Dynamic Control of Genetic Algorithms using Fuzzy Logic Techniques," *Proceedings of the Fifth International Conference on Genetic Algorithms*, pp. 76-83, July 1993.

- [51] Micahel A. Lee, *Automotive Design and Adaptation of Fuzzy Systems and Genetic Algorithms Using Soft Computing Techniques*. Doctoral dissertation, University of California at Davis, 1994.
- [52] Wen-Ruey Hwang, *Intelligent Control Based on Fuzzy Algorithms and Genetic Algorithms*. Doctoral dissertation, New Mexico State University, 1993.
- [53] Chir-Ho Chang, *Designing and Tuning of a Fuzzy Logic Controller via Genetic Input Output Mapping Factors*. Doctoral dissertation, The University of Oklahoma, 1993.
- [54] Daihee Park, *Intelligent Fuzzy Reasoning Models with Applications to Fuzzy Control*. Doctoral dissertation, The Florida State University, 1992.
- [55] Mark G. Cooper and Jacques J. Vidal, "Genetic Design of Fuzzy Controller," *Presented at the Second Int. Conference on Fuzzy Theory and Technology*, pp. 1 - 16, October 1993.
- [56] Mark G. Cooper and Jacques J. Vidal, "Genetic Design of Fuzzy Controller: The Cart and Jointed-Pole Problem," *3rd IEEE Int. Conf. On Fuzzy Systems*, pp. 1332 - 1337. 1994.
- [57] Mark G. Cooper, "Evolving A Rule-Based Fuzzy Controller," *Simulation*, vol. 65, no. 1, 67-73, July 1995.
- [58] Jindrich Liska and Stephen S. Melsheimer, "Complete Design of Fuzzy Logic Systems Using Genetic Algorithms," *3rd IEEE Int. Conf. On Fuzzy Systems*, pp. 1377-1382, 1994.
- [59] Engineering Information Inc., Hoboken, NJ, November 1995.
- [60] G. Bennett Stewart III, *The Quest for Value*. New York: HarperCollins, 1991.
- [61] Bill Richard, "Mastering the New Metrics," *CFO*, vol. 10, no. 10, p. 32, October 1994.
- [62] Bill Richard, p. 37.
- [63] S. Tully, "The Real Key to Creating Wealth," *Fortune*, vol. 128, no. 6, p. 38, September 20, 1993.

- [64] S. Tully, "America's Best Wealth Creators," *Fortune*, vol. 130, no. 11, p. 162, November 28, 1994.
- [65] Tully, p.143, November 28, 1994.
- [66] Tully, p.143, November 28, 1994.
- [67] Raymond H. Myers, *Classical and Modern Regression With Applications*, (2nd ed.). Boston: PWS-Kent Publishing Company, pp. 1-7, 1990.
- [68] Bart Kosko, *Neural Networks and Fuzzy Systems*. Englewood Cliffs, NJ: Prentice-Hall, pp. 327-335, 1992.
- [69] John A. Hartigan, *Clustering Algorithms*. New York: John Wiley & Sons, pp. 84-107, 1975.
- [70] Stewart, p. 2.
- [71] Stewart, pp. 136 - 138.
- [72] Stewart, pp. 137 - 138.
- [73] Stewart, p. 138.
- [74] _____, *Corporate Finance*, New York: Financial World, pp. 22 - 24, October/November 1992.
- [75] Stewart, pp. 153, 215.
- [76] _____, *Industry Norms and Key Business Ratios*, Dun and Bradstreet, Inc. New York, pp. 94, 107, and 118, 1994.
- [77] _____, *Industry Norms and Key Business Ratios*, Dun and Bradstreet, Inc. New York, pp. 94, 107, and 118, 1994.
- [78] _____, *Industry Norms and Key Business Ratios*, Dun and Bradstreet, Inc. New York, pp. 94, 107, and 118, 1994.
- [79] Alexander Wall and Raymond Duning, *Ratio Analysis of Financial Statements*. New York: Harper, New York, 1928.

- [80] Sheldon Gates, *101 Business Ratios*. Scottsdale, Arizona: McLane Publications, 1993.
- [81] John Neter, William Wasserman, Michael H. Kutner, *Applied Linear Statistical Models*, 3rd Ed. Boston: Irwin, Inc. p. 90, 1990.
- [82] Myers, p. 89.
- [83] Myers, p. 182.
- [84] Donald F. Morrison, *Multivariate Statistical Methods*, 3rd Ed. New York: McGraw-Hill Publishing Company, p. 385, 1990.
- [85] J. MacQueen, "Some Methods for Classification and Analysis of Multivariate Observations," *Proc. Fifth Berkeley Symp. Math. Statist. Prob.*, vol. 1, pp. 281-297, 1967.
- [86] Leonard Kaufman and Peter J. Rousseeuw, *Finding Groups in Data, An Introduction to Cluster Analysis*. New York: John Wiley & Sons, p. 113, 1990.
- [87] Goldberg, p. 7.
- [88] Rick L. Riolo, "Survival of the Fittest Bits," *Scientific America*, vol. 267, pp. 114-116, July 1992.
- [89] Davis, pp. 61 - 64.
- [90] Goldberg, p. 10.
- [91] S. Arunkumar and T. Chockalingam, "Genetic Search Algorithms and Their Randomized Operations," *Computers Math. Applic.*, vol. 25, no. 2, pp. 91 - 100, 1993.
- [92] Davis, p. 5.
- [93] Zbigniew Michael, *Genetic Algorithms + Data Structures = Evolution Programs*. New York: Springer-Verlag, pp. 32-42, 1992.
- [94] John H. Holland, "Genetic Algorithms," *Scientific American*, vol. 266, no. 7, p. 66, July 1992.

- [95] Zadeh, pp. 338 - 353, 1965.
- [96] George J. Klir and Tina A. Folger, *Fuzzy Sets, Uncertainty, and Information*. Englewood Cliffs, New Jersey: Prentice Hall, p. 10, 1988.
- [97] Toshiro Terano, Kiyoji Asai and Michio Sugeno, *Fuzzy Systems Theory and Its Applications*. New York: Academic Press, pp. 30-32, 1992.
- [98] Arnold Kaufmann and Madam M. Gupta, *Introduction to Fuzzy Arithmetic, Theory and Applications*. New York: Van Nostrand Reinhold, pp. 9-14, 1991.
- [99] Kaufmann, pp. 14-17.
- [100] Madan M. Gupta, Rammohan K. Ragade, and Ronald R. Yager, Eds., *Advances in Fuzzy Set Theory and Applications*. New York: North-Holland Publishing Company, pp. 153-164, 1979.
- [101] Kaufmann, pp. 14-35.
- [102] Earl Cox, *The Fuzzy Systems Handbook*. New York: AP Professional, p. 143, 1994.
- [103] George J. Klir and Bo Yuan, *Fuzzy Sets and Fuzzy Logic, Theory and Applications*. Upper Saddle River, NJ: Prentice Hall PTR, pp. 336-338, 1995.
- [104] A. Geist, A. Beguelin, J. Dongarra, W. Jiang, R. Manchek, V. Sunderam, *PVM 3 User's Guide and Reference Manual*. Oak Ridge National Laboratory, Oak Ridge, TN, Manual ORNL/TM-12187, p. 1, September 1994.

Vita

Michael William Howard is currently the President of Beta Development Corporation located in Knoxville, Tennessee. As president, he is responsible for managing a venture fund for early-stage, technology-based companies. He is also Director of Technology Transfer and Commercialization for the Tennessee Center for Research and Development (TCRD) in Knoxville. TCRD is a non-profit research and development organization that works with the regions technology research centers in commercializing technology developments and assisting in the transferring of these technologies to industry.

Mr. Howard is presently involved with several small, technology-based companies in providing business and engineering assistance in helping these companies mature. He is also a member and/or officer of several professional societies and serves on the board of several for-profit and non-profit organizations. Mr. Howard is a registered professional engineer in the State of Tennessee.

Mr. Howard was born in Knoxville, Tennessee on December 31, 1957. He attended primary and high school in the public school system in Knoxville where he graduated from Carter High School in June 1975. He graduated in June 1980 from The University of Tennessee at Knoxville with a Bachelor of Science in Electrical Engineering.

After graduating, he began working for Westinghouse Electric Corporation as an Applications Engineer at their Industry Electronics Division in Pittsburgh, Pennsylvania. Concurrently, Mr. Howard attended the three year Westinghouse Sales and Marketing Executive Trainee Program. In May 1981 he entered the Master's program in Industrial Engineering at the University of Pittsburgh and graduated with a Master's degree in May 1983.

In June 1985, following an investment by Westinghouse Electric Corporation in Perceptics Corporation in Knoxville, he began working with Perceptics as Product Manager responsible for there new computer-based image processing products. Mr. Howard successfully managed the market introduction of a complete line of image processing computer systems that resulted in \$1 million in sales in the first year. In June 1987 he became the Director for the Perceptics Biotechnology Division, where he managed the technical development, systems integration and selling of an imaging microscope workstation product line specifically designed for the life and material science research market. The biotechnology division resulted in more than \$3 million

in sales in two years and was recognized as the market leader in the high-end imaging workstation market.

In March 1990, Mr. Howard formed Scientific Imaging Solutions in Knoxville to design and manufacture a low-end imaging microscope workstations for the life and material sciences industry. There he successfully managed the start-up company leading to more than \$1 million in sales with over 15 percent profit in its first year. After successfully managing the new company during its first year, he sold the company during its second year to a larger organization and joined his current employer.