INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®

FORECASTING MACROECONOMIC MODELS WITH ARTIFICIAL NEURAL NETWORKS: AN EMPIRICAL INVESTIGATION INTO THE FOUNDATION FOR AN INTELLIGENT FORECASTING SYSTEM

Dat-Dao Nguyen

A Thesis

in

The Faculty

of

Commerce and Administration

Presented in Partial Fulfilment of the Requirements

for the Degree of Doctor Philosophy at

Concordia University

Montreal, Quebec, Canada

October 1999

© Dat-Dao Nguyen, 1999

National Library of Canada

Bibliothèque nationale du Canada

Acquisitions and Bibliographic Services

Acquisitions et services bibliographiques

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

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

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

0-612-47705-3

Canada

ABSTRACT

FORECASTING MACROECONOMIC MODELS

WITH ARTIFICIAL NEURAL NETWORKS:

AN EMPIRICAL INVESTIGATION INTO THE FOUNDATION FOR

AN INTELLIGENT FORECASTING SYSTEM

Dat-Dao Nguyen, Ph.D.

Concordia University, 1999

This study investigates the foundation of an intelligent system using Artificial Intelligent (AI) technologies to assist decision makers in a specific business problem, namely business forecasting. In time series and macroeconomic modelling, there are many assumptions being imposed on the behavior and functional relationship of the underlying variables. In addition, one may face the complexity in the estimation of these models. This study uses *Artificial Neural Network* (ANN) and other AI technologies in an effective forecasting system in order to overcome the restrictions of traditional modelling and estimation methods.

An ANN has been shown to be a universal function approximator (Cybenko, 1989; Hornik et al., 1989). It requires no prior assumptions on the behavior and functional form of the related variables but it is still able to capture the underlying dynamic and nonlinear relationships among variables in the problem space, ie. a macroeconomic model in this

context. This study integrates the powerful ability of an ANN into an efficient framework incorporating *Recurrent Algorithms* (Jordan, 1986), *Genetic Algorithms* (Holland,1975) in a *Mixture-of-experts Architecture* (Jacobs et al., 1991) to obtain accurate estimation and forecasts. As such, this study addresses the ability of a versatile intelligent technology to solve a general economic forecasting problem involving temporal and non-temporal variables.

Using the contexts provided in the Klein Model I of the US interwar economy in 1921-1941 and the Klein-Goldberger Model of the US economy in 1929-1952, this study investigates the relative performance of the proposed system and traditional methods in modelling and forecasting a mix of economic variables. It extends these frameworks into the future to forecast with more recent data. The study specifies the conditions that will make the implementation of ANN more successful in estimation and forecasting.

This study provides evidence on the effectiveness and efficiency of the proposed system. It asserts empirically the ability of the integrated ANN and GA in estimation and forecasting. The findings should contribute positively to the development of theory, methodology, and practice of using AI tools, particularly ANN and GA, to build intelligent forecasting systems.

ACKNOWLEDGEMENTS

As I look back at the accomplishment of my doctoral study, I realize how much I am indebted with encouragements and supports from mentors, colleagues and family.

I am grateful to my professors in the Department of Economics at Law School. University of Saigon (Vietnam) where I received my first lessons in Economics in early 1970's. In particular, Professors Chau Tieng Khuong, Nguyen Cao Hach, Pham Van Thuyet, and Mai Van Le had definite impact on my intellectual development with their immense knowledge. From the restrictive environment of a developing country, they had provided their students with broader avenues to the industrialized world. It was over there that I determined on my involvement in research endeavor and my contribution to the noble academia. However, at that time I did not realize that it would take me many years later to overcome uncontrollable mundane matters before my dream comes true.

When I restarted my second degree in Decision Sciences and Management Information Systems at Concordia University, I were fortunate to benefit from many mentors. Professor Jamshid Etezadi was the first one who spotted me working late in a computer lab. Later, while involving in some of his projects, I had developed the necessary critic and self-critic mind of a researcher. In my graduate study, I had opportunity to enrich myself in Econometrics with Professor Lorne Switzer, Financial Engineering with Professor Lawrence Kryzanowski, Bayesian Statistics with Professor Fassil Nebebe, Game Theory with Professor James MacIntosh, and Forecasting with Professor Michael Samson. These

distinguished mentors, as well as others, opened many fruitful avenues for my academic development. Later, these fields of study would serve as useful contexts for my research in Information Systems.

Above all is the rich intellectual legacy that I received from Professor Dennis Kira. His greatness is in his tolerance and support to new ideas of the younger generation. With his encouragement, I explored and undertook many researches in Management Information Systems. I learned from him the ability of standing from a general perspective to appreciate the beauty of a landscape instead of getting lost in the jungle of details. Many times, I were very comfortable to consult with him on my half finished papers as I knew that he would look beyond superficial errors to appreciate the main idea of the work. With his company, I have started a successful publication record.

I am grateful for many invaluable comments and helpful suggestions from my dissertation committee. Professor Denis Kira, my dissertation supervisor, as usual gave me his constant encouragement and support, this time on the first major work of my academic career. Professor Prosper Bernard (Université du Québec à Montréal) reminded me to make my work a practical contribution. Professor Lawrence Kryzanowski (Finance, Concordia) offered many editorial comments, besides his invaluable suggestion on the possible inheritance from the best of many seemingly conflicting schools of thought. Professor Fassil Nebebe (Decision Sciences and MIS, Concordia) kept me to be a severe self-critic in order to enhance my work. Professor Paul Mangiameli (University of Rhode Island), my external examiner, who appreciated my early work with a nomination for best paper award at NEDSI 1997 Conference, now shared with me many of his invaluable insights on Machine Learning research. Professor Yogendra Chaubey (Department of Mathematics and Statistics, Concordia), my internal-external examiner, helped me to be consistent in my mathematical expressions. To these mentors, I wish to express my most sincere gratitude. My work has been enhanced significantly with their advices. However, any remaining omissions and/or errors will be of my sole responsibility.

I also received many encouragements from my colleagues in the Ph.D. Program, especially Henry Adobor (now with Quinnipiac University), Alejandro Ramirez (now with Carleton University), and Anamitra Shome (now with Brock University). We motivated each other to excel in our study and research. I would never forget the Winter 1996, when I worked on my first publication. At that time, in the former tiny lab of four PCs in our Ph.D. Student Offices, each of us had been working on a paper for ASAC 1996 Conference in Montreal. Together, we developed a true spirit of friendship. We experienced the same joy and anxiety in our study. This comradeship was indeed making my stay at Concordia the most pleasant souvenir.

My progress in the graduate study has been speeded up with the helpful assistance from Heather Thomson and Theresa Sarazin-Wadey of the Ph.D. Program Office at Concordia University. They truly helped me, as well as my colleagues, in many detailed and timely paperwork. They indeed minimized our mundane distractions so that we could maximize our study effort.

Last but not least, thanks go to my family for their love and support. At his golden age, my father serves as my role model with his constant quest for the Wisdom. My mother always gives me her silent approval on whatever I am doing. Although my folks may not understand fully what I am doing, but they are very understanding. I also tend to agree with the suggestion that one should award an "associate degree" or something similar to recognize the support that a new doctor received from his/her spouse. But I believe that, with or without that "associate degree", my wife Ngoc-Thuyet knows that I always appreciate her loving support. This dissertation is dedicated to my sons, Dat-Tuyen and Dat-Vinh. I wish that, as they grow up in a peaceful and favorable environment, it may take them much less time to accomplish what their father has done.

TABLE OF CONTENTS

Chapter		Page
	LIST OF FIGURES	xiii
	LIST OF GRAPHS	xiv
	LIST OF TABLES	xv
1	**INTRODUCTION**	1
2	**ARTIFICIAL NEURAL NETWORK TECHNIQUE**	8
	2.1. ARTIFICIAL NEURAL NETWORK TOPOLOGY	9
	2.2. ANN AS A UNIVERSAL APPROXIMATOR	12
	2.3. BACKPROPAGATION ALGORITHM	13
	2.4. THEORETICAL PERSPECTIVE OF ANN	16
	2.5. TEMPORAL PATTERN RECOGNITION WITH ANN	20
	2.5.1. Representation of Temporal Information	21
	2.5.2. Storing Temporal Patterns in Recurrent Networks	22
	2.6. MIXTURE-OF-EXPERTS NETWORK TOPOLOGY	23
	2.7. ISSUES IN NEURAL NETWORK TRAINING	26
	2.7.1. Performance Issues	26
	2.7.2. Network Design Issues	28
3	**GENETIC ALGORITHMS IN THE SEARCH FOR THE OPTIMAL NETWORK DESIGN**	38
	3.1. OPERATIONS OF GENETIC ALGORITHMS	39

3.2. REPRESENTATION OF SEARCH SPACE 40

3.3. A SIMPLE GENETIC ALGORITHM 41

3.4. MERITS OF GENETIC ALGORITHMS 43

3.5. GENETIC ALGORITHMS IN ANN OPTIMIZATION 46

3.5.1. Search for Optimal ANN Weights 46

3.5.2. Search for Optimal ANN Architecture 47

4 **FORECASTING WITH ARTIFICIAL NEURAL NETWORKS** 49

4.1. FORECASTING UNIVARIATE TIME SERIES 49

4.1.1. ARIMA Models 51

4.1.2. Forecasting Univariate Time Series with ANN 52

4.2. FORECASTING MULTIVARIATE TIME SERIES 56

4.2.1. VARIMA Models 56

4.2.2. Forecasting VAR Models with ANN 58

4.3 FORECASTING AN ECONOMIC SYSTEM 61

4.3.1. Structural Equations of an Economic System 61

4.3.2. Estimation of Structural Equations 64

4.3.3. Nonlinearity and Dynamics of Economic Variables 66

4.3.4. Major Macroeconomic Models 67

4.3.5. Forecasting Macroeconomic Models with ANN 71

5 **RESEARCH QUESTIONS AND METHODOLOGY** 73

5.1. RESEARCH QUESTIONS 73

5.2. METHODOLOGY 76

6 **ESTIMATION AND FORECASTING OF KLEIN MODEL I** 82

6.1. KLEIN MODEL I REVISITED 82

6.2. PREVIOUS ESTIMATIONS OF KLEIN MODEL I 83

6.2.1. Klein Model I in Reduced Form 83

6.2.2. Single Equation Estimations 84

6.2.3. System Estimations 86

6.3. ANN ESTIMATION OF KLEIN MODEL I 87

6.4. FINDINGS AND DISCUSSION 89

6.4.1. Effect of Selected Network Parameters on Learning 89

6.4.2. Effect of Training Time on Learning 91

6.4.3. Two-stage ANN Estimation of Klein Model I 92

6.4.4. Modular ANN Estimation of Klein Model I 94

6.4.5. Modular ANN Forecasting of Klein Model I 96

6.4.6. Concluding Remarks 102

7 **ESTIMATION AND FORECASTING**

OF KLEIN-GOLDBERGER MODEL 124

7.1. KLEIN-GOLDBERGER MODEL REVISITED 124

7.2. PREVIOUS ESTIMATION OF K-G MODEL 130

7.3. ANN ESTIMATION OF K-G MODEL 131

7.4. FINDINGS AND DISCUSSION 134

7.4.1. Estimation and Forecasting with Modular ANN 134

7.4.2. Estimation and Forecasting with Sets of Modular Variables Selected by GA 135

7.4.3. Estimation and Forecasting with Union Set of Modular Variables Selected by GA 137

7.4.4. Estimation and Forecasting with Union Set of Modular Variables Controlled for ANN Architecture 138

7.4.5. Concluding Remarks 139

8 **CONCLUSION** 151

8.1. SUMMARY OF FINDINGS 151

8.2. IMPLICATIONS OF THIS STUDY 153

8.3. LIMITATIONS OF THIS STUDY 155

8.4. EXTENSIONS OF THIS STUDY IN FUTURE RESEARCH 157

REFERENCES 159

APPENDICES 170

LIST OF FIGURES

Figure		Page
2.1	Artificial Neural Network (5-3-2)	34
2.2	Recurrent Network: Jordan Architecture	35
2.3	Recurrent Network: Elman Architecture	36
2.4	Mixture-of-Experts Network Architecture	37
6.1	Two-stage ANN Estimation of Klein Model I	104
6.2	Modular ANN Estimation of Klein Model I	105
6.3	Moving Window Scheme for Estimation and Forecasting	106

LIST OF GRAPHS

Graph		Page
6.1	Estimation and Forecasting DLC (1950-79)	112
6.2	Estimation and Forecasting DLWp (1950-79)	113
6.3	Estimation and Forecasting DLI (1950-79)	114
6.4	Estimation and Forecasting DLC (1955-84)	115
6.5	Estimation and Forecasting DLWp (1955-84)	116
6.6	Estimation and Forecasting DLI (1955-84)	117
6.7	Estimation and Forecasting DLC (1960-89)	118
6.8	Estimation and Forecasting DLWp (1960-89)	119
6.9	Estimation and Forecasting DLI (1960-89)	120
6.10	Estimation and Forecasting DLC (1965-94)	121
6.11	Estimation and Forecasting DLWp (1965-94)	122
6.12	Estimation and Forecasting DLI (1965-94)	123

LIST OF TABLES

Table		Page
6.1	MSE of In-Sample Learning Across Learning Parameters	107
6.2	Sensitivity Analysis on GA Selection: MSE of Standard Module in Instrumental-Stage of Two-Stage ANN Estimation	108
6.3	Results of Two-Stage ANN Estimation of the Klein Model I	109
6.4	Results of Modular ANN Estimation of the Klein Model I	110
6.5	Modular ANN Estimation and Forecasting 1950-94 Using Klein's Framework	111
7.1	Estimation and Forecasting of KG Model with Full Set of Variables: SSE of In-Sample Estimation	141
7.2	Estimation and Forecasting of KG Model with Full Set of Variables: Residuals in Forecasting 1951-52	142
7.3	Input Variables and Network Architecture Selected by GA: Recurrent Module of Instrumental Stage	143
7.4	Input Variables and Network Architecture Selected by GA: Standard Module of Instrumental Stage	144
7.5	Estimation and Forecasting of KG Model with ANN Topology Selected by GA: SSE of In-Sample Estimation	145

7.6 Estimation and Forecasting of KG Model with ANN Topology Selected by GA: Residuals in Forecasting 1951-52 146

7.7 Estimation and Forecasting of KG Model with Union Set of Modular Variables Selected by GA: SSE of In-Sample Estimation 147

7.8 Estimation and Forecasting of KG Model with Union Set of Modular Variables Selected by GA: Residuals in Forecasting 1951-52 148

7.9 Estimation and Forecasting of KG Model with Union Set of Modular Variables Selected by GA: Controlled for ANN Architecture SSE of In-Sample Estimation 149

7.10 Estimation and Forecasting of KG Model with Union Set of Modular Variables Selected by GA: Controlled for ANN Architecture Residuals in Forecasting 1951-52 150

FORECASTING MACROECONOMIC MODELS WITH ARTIFICIAL NEURAL NETWORKS: AN EMPIRICAL INVESTIGATION INTO THE FOUNDATION FOR AN INTELLIGENT FORECASTING SYSTEM

CHAPTER 1

Introduction

One of the goals of Artificial Intelligence (AI) study is to build intelligent computer systems to assist people in making decisions. The intelligent behavior of a computer program is defined as the ability to learn and understand from experience, make sense out of ambiguous or contradictory messages, recognize the relative importance of different elements in a situation, and to respond quickly and successfully to a new situation (Turban and Aronson, 1998). Besides assisting a user with the powerful computing capacity, an intelligent system, particularly a Decision Support System (DSS), can alert the user to a decision-making opportunity or challenge, recognize what needs to be solved and then solves them with or without a user's interference (Holsapple and Whinston, 1996). A DSS provides the user with advice, expectations, evaluations, facts, and analyses in a decision context. Such a system can facilitate and extend the user's ability in the decision-making process, particularly in acquiring, transforming, and exploring knowledge. Many successful intelligent systems have been reported in various contexts in scientific study as well as in business practice (Medsker, 1994, 1995). The present study investigates the foundations of an intelligent system to assist decision makers in a specific business problem, namely, business forecasting.

A general business forecasting problem, particularly the one dealing with socio-economic variables, usually involves many temporal and non-temporal interactions. It is very often that the value of an economic variable is not only related to its predecessors in time but also to the current and past values of other variables. One cannot accurately forecast an economic variable by focussing solely on its behaviour over time. Therefore, the most important application of econometrics is to build macroeconomic models to incorporate various interrelated variables in the economy.

A time series consists of a set of observations on a variable taken at equally spaced intervals over time. Time series can be studied from two perspectives: analysis and modelling. An *analysis* summarizes the properties of a series and characterizes its salient features in a time domain and/or frequency domain, whereas *modelling* enables the forecast of future values. The distinguishing feature of a time series model, as opposed to an econometric model, is that no attempt is made to formulate the behavioral relationship between a temporal variable and other variables (Harvey, 1993). The movement of a particular temporal variable is explained solely in terms of its own past or its relative position in time. Then forecasts are made by extrapolation.

In contrast, an econometric model takes into consideration the behavior of an economic variable in relation with others. In an economy, it is very often the case that the predicted value of an economic variable is related not only to its predecessors in time but also to the past values of other variables. Most of these models use a Keynesian framework for the determination of national income and its components such as consumption, investment, and government expenditure. They take into account the relationship with other macroeconomic variables such as inflation and unemployment. The purpose of these models is to assist decision makers in structural analysis, forecasting and policy evaluation (Intriligator, 1978). In *structural analysis*, one investigates the underlying interrelationship of economic variables in the system under consideration in order to understand and explain relevant phenomena in the economy. In *forecasting*, one predicts values of related variables beyond the current sample data. In *policy evaluation*, one chooses an alternative on the basis of an economic forecast taking into account the possible effects on and of other economic variables.

Although rigorous econometric methods have been used in model building, there still exists a debate on the appropriateness of an economic model. First, policy issues in economics are related to politics such that the theoretical structure of a model tends to reflect a political standpoint rather than the real world. Second, very often the data can fit into many empirical models but statistical analysis cannot discriminate between rival hypotheses. Finally, economists tend to dismiss or ignore the evidence from empirical models unless it can be supported with traditional economic theory (Karakitsos, 1992). Despite an abundance of econometric models, most of them do not provide accurate estimates and forecast due to the complexity of the economic system, the impossibility of validation with controlled experiments on the economy, and the existence of non-quantifiable factors in economic activities. Econometric models, in which the structures are chosen by hand and parameters are estimated from the data, may help to understand the functioning of the economy qualitatively, but they are not accurate in making quantitative predictions (Moody, 1995).

To identify an appropriate model for multivariate time series in an economic system, one must consider various tests of individual stationarity, joint stationarity and cointegration. Following this, one has to test for normality and independence. Then, depending on whether the series have linear dependence or nonlinear dependence, an appropriate lag order of the model is specified (Cromwell et al., 1994b). Apparently from this process, there are many assumptions being imposed on time series modelling. In addition, when dealing with nonlinear models, one may encounter the complexity of estimation (Mills, 1990).

However, regardless of theoretical controversies on what the economic models should be, in many cases, the practical issue pertaining to any model building is how to obtain accurate forecasts to serve for policy making (Intriligator, 1978). This study focuses on this issue with the use of Artificial Neural Network (ANN) and Genetic Algorithm (GA). These emerging technologies, integrated into an intelligent system, intend to overcome the constraints of traditional modeling and estimation methods in order to obtain better forecasts.

In theory, an ANN has been proved to be a universal function approximator. It can approximate any underlying functional relationship between input (independent, exogenous) variables and output (dependent, endogenous) variables without the need of specifying *a priori* the functional form and imposing prescribed assumptions about the behavior of data. The merit of ANN is in its effective and efficient learning of nonlinear relationships inherent in the data. Similarly, a GA has been shown to be a powerful search tool with its ability to explore a large number of alternatives in the problem space in order to avoid sub-optimality.

In practice, ANN has been implemented in many successful applications in sciences as well as in business (Kryzanowski et al., 1993; Trippi and Turban, 1993; Sharda, 1994; Refenes, 1995; and others). In time series study, ANN has been used to forecast univariate variables (Sharda and Patil, 1990; Foster et al. 1991, Hill et al., 1994; Lachtermacher and Fuller, 1995; and others) as well as multiple variables (Nguyen and Kira, 1997). Literature also reports an early attempt to estimate an economic system with ANN (Caporaletti et al., 1994). However, there are many shortcomings in the mentioned work in its inability to capture the dynamics of simultaneous, contemporaneous variables in the system and its focus on in-sample, ex-post forecasts only.

This study extends previous work with an investigation into the implementation of an integrated system of ANN and GA in order to handle effectively a general family of business forecasting problems, i.e., forecasting with a mix of temporal and non-temporal variables. An econometric model should be a useful context for this investigation. Not only does the study investigate network performance in terms of effectiveness and accuracy for in-sample estimation, but it also projects network forecasts into the future to highlight the behavior of ANN in generalizing on unseen data patterns.

This study advocates the use of Artificial Intelligence (AI) and Machine Learning technology in economic forecasting. The ANN technique, which requires no prior assumptions on behavior and functional forms of the variables, would be a viable alternative in economic forecasting. In particular, the current study proposes an integration of ANN and GA to obtain better approximations and forecasts of a macroeconomic model including temporal as well as non-temporal variables. In this integration, the mixture-of-experts network architecture is implemented to learn the relationship of the mix of economic variables, and GA to build optimal network topology. This integrated framework should serve as an efficient foundation to build intelligent forecasting systems. These systems would be useful not only in Decision Support Systems but also in the emerging field of Data Mining in which AI tools are used to acquire unknown knowledge from a mass of information.

This study provides evidence on the effectiveness and efficiency of the integration of ANN and GA in a mixture-of-experts architecture. The GA helps overcome the sub-optimality of the tedious trial-and-error process in network building. The flexible network architecture offers many alternative network configurations to capture the peculiarities of variables in a business context before aggregating intermediate estimations into final results. The integrated system processes effectively the mix of variables, and produces efficient estimations and forecasts. This study proposes a versatile intelligent tool to solve a general economic forecasting problem. Although there still exist limitations and possible improvements, the findings of this study should contribute positively to the development of theory, methodology, and practice of using AI and Machine Learning tools, particularly ANN and GA, to build intelligent forecasting systems.

The thesis is organized as follows. In the next two chapters, the foundations of Artificial Neural Networks and Genetic Algorithms are reviewed. Chapter 4 reviews forecasting with the ANN technique. Chapter 5 presents research questions and methodology of the undertaken empirical study. Chapter 6 reports and discusses the experiments and findings in estimation and forecasting the Klein Model I. Chapter 7 reports and discusses the experiments and findings in estimation and forecasting with the Klein-Goldberger Model. Chapter 8 concludes with a summary of findings, contributions and limitations of the study, and possible extensions in future research.

CHAPTER 2

Artificial Neural Network Technique

A Decision Support System (DSS) is defined as a computer-based information system that affects or is intended to affect how people make decisions (Silver, 1991). The goal of a DSS is to assist decision makers in identifying and applying some models in order to arrive at a solution to the situation at hand. In a DSS, the main thrust is to allow decision makers to analyze and detect patterns of data, internal as well as external, and to aid them in making strategic decisions. However, the current quantitative approach in DSS research is not sufficient to handle complex decisions. Many aspects of business and economic systems do not lend themselves to quantified measurements. Also, many business and economic structures cannot be represented with mathematical functions supported by a traditional DSS. One may even argue on the appropriateness of some models and then *a priori* assumptions of traditional methods implemented in a traditional DSS.

Artificial Intelligence technology, particularly the Machine Learning approach, has the ability of learning data patterns and improving predicting performance with input data themselves. This technology presents alternative methods of processing information and at the same time avoiding the restrictions and insufficiency of traditional quantitative methods (Hinton, 1992). One of the emerging technologies in AI and Machine Learning is the Artificial Neural Network (ANN) technique. Besides the benefits of a powerful parallel distributed computer system, the main reason for a growing interest in using ANN is in its ability to approximate a nonlinear relationship without imposing *a priori* assumptions on the behavior of the variables under consideration. This chapter reviews the basics of the ANN technique and relevant issues in the implementation of ANN in an intelligent system for forecasting.

2.1. ARTIFICIAL NEURAL NETWORK TOPOLOGY

The Artificial Neural Network (ANN) technique mimics the functioning of human neural system in parallel distributed information processing (Rosenblatt, 1959, 1962; Rumelhart et al., 1986). A neural network contains computing units called neurons (or nodes) arranged into layers in which each node in one layer has weighted connection to nodes in the next layer in a particular configuration (Figure 2.1).

The manner in which the connections among nodes and layers are made defines the flow of information in the network. A network topology consists of nodes as autonomous processing units that are joined by directed arcs. Each arc (or connection) has a numerical weight, w_{ij}, that specifies the influence of node u_j on node u_i. A positive weight indicates a reinforcement whereas a negative weight represents an inhibition to the flow of information.

A subset of nodes are considered as network *input* nodes if they provide external signals to the network and do not recompute their outputs. These nodes have no entering arcs. The *output* nodes produce outputs of the network as a whole. Nodes which are neither input nor output are *intermediate nodes*. These intermediate nodes are necessary to compute non linearly separable functions.

Each node u_i computes a single node output or activation. The output of a node can be an output of the network as a whole and/or the input for other nodes. Node inputs and activations may be discrete, taking on values {0, 1} or {-1, 0, 1}, or they may be continuous, taking on values in the interval [0, 1] or [-1, 1]. In computing bounded continuous output, the output values can be scaled into [0,1]. The computational process of each node is as follows. A node, as a processing unit, receives inputs from a number of other nodes or from an external stimulus. A weighted sum of these inputs constitutes the argument to an activation function or transfer function. The activation of a node is computed from the activation of nodes directly connected to it and the corresponding weights for those connections. The resulting value of the activation function is the output of the node. This output gets distributed (or propagated) along the weighted connections to other nodes. A neural network learns by means of weight adaptation in the training phase with a particular learning rule. The *propagation rule* could be weighted sum, cumulative weighted sum, maximum, minimum, majority, or product. The *activation rule* could be an identity or a threshold function. The *transfer function* usually is a nonlinear, bounded and piecewise differentiable function. However, it could be in one of the following forms: identity, linear, sigmoid, sine, or hyperbolic tangent.

Nodes in a neural network are arranged into layers. Most applications use a 3-layer network consisting of one input, one hidden and one output layer. Hidden nodes with their activation function are needed to introduce nonlinearity into the network. In some cases, more than one hidden layer is necessary to approximate a higher order function. However, a complex network may not only fit the signal but also noise. This leads to overfitting the training data and predicting poorly on the test data.

A network is classified either as *feedforward* if it does not contain directed cycles or *recurrent* if it does contain such cycles. In a recurrent network, node outputs are fed back as inputs for nodes of the previous layer in the subsequent iteration. The recurrent network is particularly useful in dynamic learning and prediction environments.

All knowledge in an ANN is encoded in its interconnection weights among the neurons determined by the learning process. A weight represents the strength of association among connected features, concepts, propositions, or events during a training period (Refenes, 1995). A neural network learns by adaptation in the training phase in which the interconnection strengths are changed appropriately.

In the training phase, a neural network is presented with a training set composed of external input patterns and their associated output patterns (or targets). The learning process can be either unsupervised or supervised. In *unsupervised learning*, there is no performance evaluation available and the network constructs clusters of similar input patterns. In *supervised learning*, the network receives learning feedback on the difference between the target and the network output in order to determine subsequent changes in weights adaptation. The *learning rule* is an algorithm to reduce the differences between network output and actual output by minimizing a cost function or maximizing an objective function. There are many available algorithms and configurations for neural network training. The choice of learning/training algorithm and network configuration depends on data and is problem specific. However, Cybenko (1989) proves that, by using a backpropagation algorithm and sigmoid transfer function, a neural network with one hidden layer can approximate any continuous function.

2.2. ANN AS A UNIVERSAL APPROXIMATOR

The ability of ANN in function approximation is due to its capability of learning the underlying functional relationship from the data itself, therefore, minimizing the necessary *a priori* non-sample information. A multi-layer network can produce a mapping between inputs and outputs consistent with any underlying functional relationship regardless of its true functional form. It eliminates the need for unjustified *a priori* restrictions, such as the Gauss Markov assumption, frequently used to facilitate estimation in regression analysis. In traditional statistics, the appropriateness of the Ordinary Least Squares method is an empirical question, therefore the test of assumptions is a routine part of any application. In contrast, whether these assumptions hold or not, the ANN still yields a similar solution since the image of any underlying mapping can always be projected into a perfectly flexible mapping.

The Kolmogorov's theorem (1957) establishes a perfectly general mapping from R^n - R^m as long as an appropriate transfer function $g(.)$ is chosen.

Kolmogorov Theorem: Any continuous real-valued functions $f(x_1, x_2, ..., x_n)$ defined on $[0,1]^n$, $n \geq 2$, can be presented in the form

$$f(x_1, x_2, \ldots, x_n) = \sum^{2n+1}_{j=1} g_j \left[\sum^{n}_{j=1} \varphi_j(x_i) \right] \qquad (2.1)$$

where g_j are properly chosen continuous functions of one variable, and φ_j are continuous monotonically increasing functions independent of f.

Funahashi (1989), Cybenko (1989), Hornik et al. (1989) extend this theorem to show that $g_j(.)$ can be specified *a priori* as the sigmoid function without sacrificing the flexibility of the mapping.

Cybenko Theorem: Let g be any continuous sigmoidal function [e.g., $g(\xi) = 1/(1 + e^{-\xi})$]. Then, given any continuous real-valued function f on $[0,1]^n$ (or any other compact subset of R^n) and $\varepsilon > 0$, there exist vectors $\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_N, \boldsymbol{\alpha}$, and $\boldsymbol{\theta}$ and a parameterized function $G(\cdot,\mathbf{w},\boldsymbol{\alpha},\boldsymbol{\theta}) : [0, 1]^n \rightarrow R$ such that

$$|G(\mathbf{x},\mathbf{w},\boldsymbol{\alpha},\boldsymbol{\theta}) - f(\mathbf{x})| < \varepsilon \qquad \text{for all } \mathbf{x} \in [0, 1]^n \tag{2.2}$$

where

$$G(\mathbf{x},\mathbf{w},\boldsymbol{\alpha},\boldsymbol{\theta}) = \sum^{N}_{i=1} \alpha_j\, g(\mathbf{w}_j^T\mathbf{x} + \theta_j) \tag{2.3}$$

and $\mathbf{w}_j \in R^n$; $\alpha_j, \theta_j \in R$; $\mathbf{w} = (\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_N)$; $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_N)$, and $\boldsymbol{\theta} = (\theta_1, \theta_2, ...\theta_N)$.

Furthermore, Hornik et al. (1989) show that the underlying true functional form does not need to be continuous. They also show that standard multi-layer feedforward networks using arbitrary transfer functions can approximate any Borel measurable function to any desired degree of accuracy (Hassoun, 1995).

For instance, using log-sigmoidal transfer functions with a backpropagation (BP) learning algorithm, which is discussed in the next section, a 1-3-1 network with 3 hidden nodes produces a response that is the sum of 3 log-sigmoid functions. The response is a superposition of 3 sigmoid functions. Thus, a network with more hidden nodes can approximate better a complex function since the estimated function has a larger number of inflection points (Hagan et al., 1996).

2.3. BACKPROPAGATION ALGORITHM

Backpropagation (Werbos, 1974) is a commonly used algorithm for learning in feedforward networks. It has been estimated that 70% of ANN applications report the implementation of this learning algorithm. The method applies a mean squared error cost function and gradient descent approach for the convergence between the network output and the desired target. Let W be the set of all network weights, the backpropagation finds the network's gradient $\nabla F(.)$, then updates the current W with a small step to form a new weight W^* using

$$W^* = W - \rho \nabla F(.) \tag{2.4}$$

The positive parameter ρ is used to control the step size. To perform gradient descent, one needs a differentiable error function. The backpropagation accomplishes this by using mean squared error and by using nodes with continuous valued activations.

Let a backpropagation network start with a set of training samples with inputs and corresponding desired outputs. The node activations u_i having values in the interval [0,1] are computed by

$$S_i = \Sigma_j \, w_{i,j} \, u_j \tag{2.5}$$

$$u_i = f(S_i) \tag{2.6}$$

where the activation function $f(x)$ is

$$f(x) = 1/(1 + e^{-x}) \tag{2.7}$$

The derivative of this sigmoid function evaluated at x can be expressed as

$$f'(x) = d/dx \, (1 + e^{-x})^{-1} \tag{2.8}$$

$$= [1/(1 + e^{-x})][1 - (1/(1 + e^{-x}))]$$

From this expression, one has the following useful identity for the estimation of derivative,

$$f'(S_i) = u_i \, (1 - u_i) \tag{2.9}$$

The task is to determine the weights $\{w_{i,j}\}$ for the network that minimize the mean squared error between network output and desired target.

The algorithm begins by randomly assigning initial weights to break the symmetry so that many intermediate nodes can take part in computing. Then one chooses a training example and computes the gradient with respect to it. The computation involves a forward pass over the network to compute node activations, followed by a backward pass to compute gradients. Once the gradients are determined, the weights are updated. This process continues until the algorithm converges.

The following representation of a BP algorithm is according to the representation in Lippmann (1987) and Gallant (1993). Specifically:

1. Start with a small positive step size ρ and assign initial random weights $\{w_{i,j}\}$ to all nodes.

2. Repeat until the algorithm converges

   2a. Take the next training example E_i with its correct output C_i

   2b. Forward propagation step: Starting from the input, make a bottom-up pass through the network to compute weighted sum S and activation $u_i = f(S)$ for every node.

   2c. Backward propagation step: Starting with the output, make a top-down pass through the output and intermediate nodes to compute

$$f'(S) = u_i(1 - u_i) \tag{2.10}$$

$$\delta_i = \begin{cases} (C_i - u_i)f'(S) & \text{if } u_i \text{ is an output node} \\ \left(\sum_{m:m>1} w_{m,i}\delta_m\right)f'(S) & \text{for other nodes} \end{cases} \tag{2.11}$$

2d. Update weights

$$w^{*}_{i,j} = w_{i,j} + \rho\delta_i u_j \qquad (2.12)$$

If $\Delta w^{*}_{i,j} = w^{*}_{i,j} - w_{i,j}$ and a momentum α is added, the last steps of the algorithm are modified as

$$2d^{*}. \quad w^{*}_{i,j} = w_{i,j} + \alpha\Delta w^{*}_{i,j} + \rho\delta_i u_j \qquad (2.13)$$

$$2e^{*}. \quad \Delta w^{*}_{i,j} = w^{*}_{i,j} - w_{i,j} \qquad (2.14)$$

It has been proved that

$$-\partial\mathscr{E}(W)/\partial S_i = \delta_i \qquad (2.15)$$

and

$$-\partial\mathscr{E}(W)/\partial w_{i,j} = \delta_i u_j \qquad (2.16)$$

What the backpropagation algorithm does is to perform a gradient descent, where the gradient is estimated from a single training example (Gallant, 1993). Also, this algorithm performs an on-line updating in which the gradient is estimated and weights are updated after every training example.

2.4. THEORETICAL PERSPECTIVE OF ANN

The similarity between ANN techniques and traditional methods in statistics and econometrics has been investigated in the literature (Cheng and Titterington, 1994; Ripley, 1994; Hwang et al., 1994). In this section, the ANN is described from an econometric perspective following the work of Kuan and White (1994).

Consider a simple neural network with one input layer of r nodes and one output layer of v nodes. In this network, an input node i sends a real-valued signal x_i in parallel over connections to an output node j. This signal is amplified by a weight w_{ji} before reaching the output node j. Then, the value of output node j is

$$y_j = \Sigma^r_{i=1} x_i w_{ji} , \qquad i = 1, ..., r ; \ j = 1, ..., v. \tag{2.17}$$

If one adds a bias node with value $x_0 = 1$, then the output value could be represented as

$$f_j(x,w) \equiv \tilde{x}' w_j , \qquad j = 1, ..., v \tag{2.18}$$

or

$$f_j(x,w) \equiv (I \otimes \tilde{x}')w \tag{2.19}$$

where $f = (f_1, ..., f_v)'$, $\tilde{x} = (1, x_1, ..., x_r)'$, $w = (w_1', ..., w_v')$ and $w_j = (w_{j0}', w_{j1}', ..., w_{jr}')'$. The output function f is implemented with MADALINE (Multiple Adaptive Linear) network by Windrow and Hoff (1960) and could be recognized as the systematic part of a standard system of *seemingly unrelated linear equations* (Kuan and White, 1994). When $v = 1$, it could be recognized as the *simple linear model.*

If the neurons become active only after their inputs pass a certain threshold, one has the nonlinear response in ANN. The neurons can switch on or off, or they can smoothly vary from fully off to fully on.

In the first possibility proposed by McCulloch and Pitts (1943), the output of a neuron is given by

$$f_j(x,w) = G(\tilde{x}' w_j) , \quad \begin{cases} G(a) = 1 \text{ if } a > 0 \\ G(a) = 0 \text{ if } a \le 0. \end{cases} \qquad j = 1, ..., v \tag{2.20}$$

G is an activation function of the output, which implements a threshold logic unit. The output node j is activated when $\tilde{x}' w_j > 0$ or, alternatively, $\Sigma^r_{i=1} x_i w_{ji}$ exceeds the threshold $-w_{j0}$.

The other possibility is that neurons can turn on gradually as input activity increases. Then, the activation function could be a smooth sigmoid function, in particular, the logistic function $G(a) = 1/[1 + \exp(-a)]$. This is the *binary logit probability model* in statistics. If G is a normal cumulative distribution function, one has the *binary probit model* (Amemiya, 1981).

For multi-layer configuration, consider a network with one hidden layer of q nodes. The output function at the node h is represented as

$$f_h(x,\theta) = F\,(\beta_{h0} + \Sigma^q_{j=1} G(x'w_j)\,\beta_{hj}\,), \qquad j = 1,\dots, q\,; \qquad h = 1,\dots, v. \quad (2.21)$$

where β_{hj} is the connection weight from hidden node j to output node h. The vector $\theta = (\beta_1',\dots,\beta_v',w_1',\dots,w_q')$, where $\beta_h' = (\beta_{h0},\dots,\beta_{hq})$, collects together all network weights.

If one chooses the activation functions $F(a) = a$ and $G(a) = 1/[1 + \exp(-a)]$, and consider only a network with one output node (i.e., $v = 1$), then the output function can be represented as

$$f(x,\theta) = \beta_0 + \Sigma^q_{j=1} G(x'w_j)\beta_j \qquad (2.22)$$

This equation resembles the *projection pursuit models* in statistics (Friedman and Stuetzle, 1981; Huber, 1985),

$$f(x,\theta) = \beta_0 + \Sigma^q_{j=1} G_j(x'w_j)\beta_j \qquad (2.23)$$

However, in neural networks, the function G is given, whereas, in the projection pursuit models, the functions G_j are unknown and must be estimated from the data (Kuan and White, 1994).

If the input nodes are connected directly to output nodes as well as to hidden nodes, the output function can be expressed as

$$f(x,\theta) = F(x'\alpha + \beta_0 + \Sigma^q_{j=1} G(x'w_j)\beta_j) \qquad (2.24)$$

where α is an $r\times 1$ vector of input-output weights, $\theta = (\alpha',\beta_0, ..., \beta_q, w_1', ..., w_q')'$.

With $F(a) = a$, one has the *standard linear model augmented by nonlinear terms.* The activations of hidden nodes can be viewed as *latent variables* (Kuan and White, 1994).

The functions in the form of

$$f(x,\theta) = \beta_0 + \Sigma^q_{j=1} G(x'w_j)\beta_j \qquad (2.25)$$

can be viewed as universal approximators. These functions can approximate any function $g: R^r - R$ arbitrarily well provided that the neural network has sufficient hidden nodes and properly adjusted parameters (Cybenko, 1989; Hecht-Nielsen, 1989; Hornik et al., 1989).

Similar results hold for network models with general (not necessarily sigmoid) activation functions in L_p spaces with compactly supported measures and in general Sobolev spaces (Hornik et al., 1990, Hornik, 1991)

For networks with many hidden layers, the output of an l-layer network can be represented as

$$a_{hi} = G_h(A_{hi}(a_{h-1})), \qquad i = 1, ...q_h\, ; \quad h = 1, ...l. \qquad (2.26)$$

where a_h is a $q_h\times 1$ vector with element a_{hi} ; $A_{hi}(.)$ is an affine function of its argument, that means $A_{hi}(a) = \tilde{a}'w_{hi}$ for some $(q_h + 1)\times 1$ vector w_{hi} ; G_h is the activation function for nodes of layer h ; $a_0 = x$; $q_0 = r$; and $q_l = v$. In this presentation, $l = 2$ denotes a single hidden layer network.

As opposed to a feedforward network described above, a recurrent neural network can have internal feedback in the architectures proposed by Jordan (1986) and Elman (1988). In Jordan architecture, network output feeds back into the hidden layer, whereas in Elman architecture, hidden layer output feeds back into the hidden layer (Figures 2.2 and 2.3). The Elman network can be represented by the following function

$$f_t(x^t, \theta) = \beta_0 + \Sigma^q_{j=1} a_{tj} \beta_j \tag{2.27}$$

$$a_{tj} = G(x'w_j + a'_{t-1}\delta_j), \qquad j = 1, \ldots, q \ ; \ \ t = 0, 1, 2, \ldots \tag{2.28}$$

where $a_t = (a_{t1}, \ldots, a_{tp})$. In this network, the output depends on the initial value a_0 and the entire history of system inputs $x^t \equiv (x_1, \ldots, x_t)$. In econometrics, this network can be viewed as a *nonlinear dynamic latent variables model* (Kuan and White, 1994).

Other neural networks also have been shown to be similar to statistical methods. For example, probabilistic neural networks are similar to kernel discriminant analysis, Kohonen networks for adaptive vector quantization are similar to k-means cluster analysis, and the Hebbian learning algorithm is closely related to principal component analysis (Kuan and White, 1994). These networks are beyond the scope of this study.

2.5. TEMPORAL PATTERN RECOGNITION WITH ANN

An ANN, if it is configured appropriately, does have the ability of recognizing and storing the temporal nature of patterns. Maren et al. (1990) reviewed many neural network configurations for spatio-temporal pattern recognition as follows.

- Create a spatial representation of temporal data. The sequence of data is presented simultaneously in the input layer of the network.

- Put time-delays into neurons or their connections to handle explicitly the temporal aspect of incoming data. Information taken at one moment in time is shifted to the right down a series of nodes and new information is inserted in the leftmost node. The number of nodes determine the number of time intervals over which information is kept. Each network operation is still done in a single network cycle. There is no explicit integration of temporal information, nor does the information stored in the rightmost nodes degrade with time. When the information in the rightmost node is bumped out of the network, its effect simply disappears.

- Use recurrent connection to create a temporal signal sequence. The connections are set up to feed back node activation to themselves or to nodes in a previous layer.

- Use nodes with time varying activations; nodes are created to keep some residue of the previous signals and allow a decay of historical information.

Within the scope of this study, a combination of the static representation of temporal information and storing temporal patterns in a recurrent network is discussed.

2.5.1. Representation of Temporal Information

In the static representation of temporal information, a sequence of incoming temporal data is represented simultaneously in the input layer of the network. In time series analysis, the value of an economic variable corresponding to each time lag is represented by an input node. For instance, if the variable X has 3 lags X_{t-1}, X_{t-2}, X_{t-3}, then one needs 3 input nodes to capture the values of these lagged values.

In dynamic forecasting, the predicted values of economic variables of concern are used in next period forecasting. Applying to ANN, one can store and generate temporal patterns via recurrent connection. In this configuration, the output just produced by the network is fed back to the input level to represent the state of the network at the preceding moment in time. Also nodes can be created to keep some residue of the previous signals and allow slow decay of historical information.

2.5.2. Storing Temporal Patterns in Recurrent Networks

Jordan (1986) proposes an architecture in which the value of output layer is fed back to a context unit to create the memory traces. Both input units and context units activate the hidden units to produce the next network output. A context unit retains the past value of its input with an exponential decay. It can be considered as a lowpass filter which creates an output that is a weighted average of some of its recent past inputs. Following Gershenfeld and Weigend (1993), the output of a context unit can be represented as

$$y(n) = \Sigma^{n}_{i=0}\, x(n)\tau^{n-1} \tag{2.29}$$

where $0 \leq \tau \leq 1$ is a time constant to control the degree to which past values are factored in. The time constant τ could be set to $1 - 1/D$ where $D>0$ represents the memory depth, i.e., how long a given value fed to the context unit is remembered.

Elman (1990) proposes a different architecture for the internal representation of time. In a network, hidden units develop internal representations for input patterns. These representations re-code patterns in a way that enables the network to produce the correct output for a given input. In Elman architecture, context units remember the previous internal state. As such, hidden nodes have the task of mapping both external inputs and also the previous internal state of some desired output.

The difference between these two recurrent architectures is in the feedback of previous network output. In a Jordan network, context units retain the approximation of previous state. The network output is fed back to the context unit to develop a new internal representation and then the approximation of the next state in the sequence of data (Figure 2.2). In an Elman network, output of hidden nodes is fed back to the context units. The context unit therefore retains an internal representation of the previous state (Figure 2.3). In addition, the Jordan network requires a specification of memory depth whereas in the Elman network the hierarchy is implicit in the structure of hidden unit activations.

2.6. MIXTURE-OF-EXPERTS NETWORK TOPOLOGY

A single network is useful for a simple task but it becomes insufficient to carry out complex ones. In complex situations. one needs a system of networks in which many networks are integrated or interacted which each others in logical or real parallelism. The mixture approach is to build complex models out of simple parts. In this setting, systems of networks can be tightly or loosely coupled (Maren et al., 1990).

A *tightly coupled system* is often the integration of networks of different types in a single network. In this architecture one may not separate out individual networks. As such, one is still facing the limitation of a single network no matter how complex it is. Often, complex problems require multiple stages of processing and therefore multiple networks.

A *loosely coupled system* arranges similar networks in parallel or in hierarchies for fast evaluation of different types of information from the same data. It also has the ability to yield increasingly higher levels of data abstraction. In such systems, functions can be decomposed and assigned to specific networks. Consequently, one network may serve as a preprocessor or filter for another network. Also, an individual network can influence or set the weights of other networks. Each individual network performs some unique task in solving a complex problem. As such one can refine an individual network to achieve a superior performance on a specific task. A system of networks offers improvement in learning or performance in a complex task.

In a loosely coupled system, individual networks may be arranged in one or a combination of the following configurations (Maren et al., 1990):

- *Hierarchy of Networks*: Multiple networks of similar types are used to partition a multi-scale pattern recognition problem into separate problems.

- *Parallel Networks*: A problem is partitioned so that different aspects are processed in parallel by different networks. It is useful when several different analyses must be performed on the same incoming data, or when data can be matched against several different models. Each network is trained to extract different features or to make different distinctions. Results from individual networks can be fused or correlated to obtain desired results.

- *System of Heterogenous Networks*: One cannot expect that one type of network, regardless of how versatile it may seem, to be sufficient to solve all tasks associated with a complex problem. With a system of heterogenous networks, the task is to select the right types of networks and put them together in a useful architecture to solve the problem effectively. Ritter (1989) suggested that one may decompose the function that a mapping network would learn into smaller functions, each of which is recognized or mapped by an individual network. The functions do not need to be known or specified in advance. In addition to the advantage of neural networks in learning the unspecified relationship from training samples, this approach also helps to break down a very complex function into manageable units.

- *Control System*: One network is used to control or assign weights to another. This configuration is useful in adaptive control applications in engineering.

Function approximation with ANN is traditionally based on a superpostion of simple basis functions such as logistic functions. Instead of using solely superposition, one can also use the principle of divide-and-conquer to split an input space into smaller regions which can be fitted with simpler functions by a set of function approximators called *expert networks* (Jacobs et al. 1991; Jordan and Jacobs, 1995). The assumption of the approach is that data can be well described by a collection of functions, each of which is defined over a relatively local region of the input space (Jordan and Jacobs, 1995). The expert networks could be arranged in modular and/or hierarchical systems (Figure 2.4). These systems offer the ability of solving a complex problem since the problem is divided into a set of sub-problems, each of which may be simpler to solve than the original one. *Modular* architecture allocates different expert networks to different regions of the problem space, whereas *hierarchical* architecture divides regions of the problem into sub-regions. With the assumption that a particular type of network (an expert) is appropriate in a region of the input space, the architecture requires a mechanism that identifies the experts or a mixture of experts that most likely produce the correct output from a given associated input. This is accomplished with an auxiliary network, called a *gating network*, to provide the weight of contribution to various experts.

Various algorithms have been proposed to take advantage of the modularity of mixture systems. Simulation experiments indicate that although backpropagation does not converge faster than other algorithms, it provides lower relative error (Jordan and Jacobs, 1995).

2.7. ISSUES IN NEURAL NETWORK TRAINING

Using ANN in a business context, one should be aware of the network performance and how to configure a network appropriately in order to achieve the desired result. This section examines the possibility of sub-optimality in network output and the measurement of network performance. Then it discusses the designing of an effective network architecture in terms of necessary network size, sufficient training set, appropriate training time and other network parameters to assure a satisfactory performance.

2.7.1. Performance Issues

- Local Minima

Like all gradient descent methods, the backpropagation algorithm may not find the global minimum even if it converges. In some problems, one does not even know whether local minima exist or whether a minimum found by gradient descent is a local or a global minimum. However, it has been argued that for multidimensional real world problems, a local minimum must be a local minimum in every dimension. Therefore, the increase of dimensionality seems to help (Gallant, 1993). Poston et al. (1991) show that for a feedforward network using sigmoid activation function with as many hidden units as the number of patterns to learn, it is almost certain that the error function has a unique minimum. In practice, one may avoid local minima by starting the training with different initial random weights. If all of them reach the same minimum for the error function, this value is assumed to be a global minimum of the function (Lachtermacher, 1993).

- Measurement of Network Performance

To measure the performance of a network, the focus is on whether the network has learned input-output patterns in a training set and whether it can generalize its knowledge to predict the input-output relationship in an out-of-sample set. In particular, it concerns the convergence, the generalization and the stability of a network.

Convergence is concerned with the capability of a neural network to learn the relationship of input-output patterns underlying in a data set. With a fixed topology network, as the training time tends to infinity, the error minimized by the gradient descent method tends to zero. With other methods, as the training time tends to infinity, the network can classify the maximum number of possible mappings with an arbitrarily large probability (Refenes, 1995). Therefore, with a given error tolerance margin, an ANN always converges with sufficient training time.

Generalization is concerned with the ability of a network to recognize out-of-sample patterns which it has not learned in the training set. Choosing a network with a structure more complicated than necessary is similar to fitting a high degree polynomial to the lower order data (Refenes, 1995). Also, a long training time may overfit the training data. A complicated and/or overtrained network overfits signals as well as noise of the training set and consequently it performs poorly with out-of-sample patterns.

Stability is concerned with the consistency of results obtained by a neural network when varying the values of the parameters. It is known that a small change in network parameters such as network design, training times, and initial condition may produce large changes in network behaviour. Theoretically, the only criterion for deciding whether neural networks perform better than parametric regressions is if they converge to smaller squared errors of the sample data (Refenes, 1995).

2.7.2. Network Design Issues

- Network size

It has been noted that a 2-layer network cannot accurately represent a function exactly representable by a 3-layer network (Blum and Li, 1991). Certain mappings can be uniformly approximated only with 3- instead of 2-layer network (Sontag, 1990). It has been proposed that the use of a network with an additional number of layers can not hurt because approximation properties of single hidden layer networks carry over to multi-hidden layer networks (Hornik et al., 1989). However, Cybenko (1989) proves that, by using a backpropagation algorithm and sigmoid transfer function, a neural network with one hidden layer is sufficient to approximate any continuous function.

It is known that the learning capacity of the neural network depends on the number of hidden nodes. Lippmann (1987) conjectures that the number of hidden nodes depends on the number of input vectors (i.e., sample size). It has been shown that input dimensionality (i.e., input patterns), together with the number of hidden nodes, also defines the maximum number of separable regions obtainable in the input space (Refenes, 1995). The analytic estimation of the number of hidden nodes necessarily requires an analysis on the dimensionality of input vector space. However, Refenes (1995) points out that feature analysis is not a trivial task since many inputs cannot be regarded as independent and it is impossible to determine the shape of the feature space. Furthermore, the underlying assumptions of feature analysis, such as linearly separable regions of the input space, are not necessary in applying intrinsically nonlinear multi-layer networks.

Nevertheless, many rules of thumb are proposed to select an appropriate number of hidden nodes. For example, Salchenberger et al. (1992) suggest that the number of hidden nodes should be 75 percent of the number of input nodes. Usually more hidden nodes than the minimum number are needed so that the neural network can have a larger learning capacity. If there are too few hidden nodes, the neural network may not be able to generate a function that reflects the underlying problem. Having more hidden nodes than necessary overfits the training set and decreases the ability to generalize to out-of-sample data. A predetermined network with a fixed number of layers and nodes rarely moves to a useful solution (Refenes, 1995). Given a cost function, the only way to determine the number of layers and hidden nodes of an optimal network topology is by trial and error, either by a hand picked process or by an automatic search procedure such as the Genetic Algorithm (Goldberg, 1989).

- Sample Size

Several rules of thumb, based on statistical classification theory, conjecture that the number of connections should be less than one-tenth of the sample size. In fact, as a universal approximator in function mapping, the ANN can always produce a perfect mapping from one universe to another universe, i.e., input set to output set. Also, in a classification problem, one can always merge a large number of regions into a smaller number of classes.

In many AI machine learning problems, the input space represents the whole problem domain. As such, the size of the input set is dictated by the problem. Ideally one should work with the whole population or consider the training set representing the complete population of input space to be mapped into an output domain. From the Kolmogorov Theorem, the mapping with ANN can guarantee any arbitrary degree of accuracy. However, when generalizing on a pattern outside the problem domain, the prediction may not be accurate. Since the set of training patterns is taken from a problem space, the more training patterns that are available, the more information about the universe one has for more accurate generalization.

In a classification problem, consider a set of input patterns represented by d-dimensional Euclidean space R^d. In a single hidden layer network, a hidden node acts as $(d-1)$ dimensional hyperplane that forms 2 decision regions. Mirchandini and Cao (1989) show that the d-dimensional input space is linearly separable into M regions if there exist M disjoint regions whose boundaries are composed of portions of hyperplanes. These regions can be associated with classes, i.e., M regions can be merged into C classes where $C \leq M$. The number of separable regions identifies the minimum number of training patterns T required for training single hidden layer networks. Ideally, the availability of a training vector in each of the M separable regions, i.e., $T = M$, should ensure the separation of the input space into M regions. However, Refenes (1995) notes that, for all but trivial problems, it is impossible to determine what shapes the classes take in the feature space and whether they are disjoint, concave or simple. In addition, networks with more than one hidden layer can find solutions that do not satisfy the assumptions of this analysis.

The effect of sample size on the convergence of an ANN training has not been asserted. However, an abundance of input-output patterns should provide more information on the problem space for generalization purposes. In any case, from the analysis of

Mirchandini and Cao (1989), one can always collapse a space of higher dimensionality into a smaller one. Ideally, the training set should contain the possible minimum and maximum of each dimension of the input vector to assure an accurate generalization (Refenes, 1995).

- *Training time*

Training time is the number of presentations of the data set to the network so that it can adjust the interconnection weights in order to achieve a convergence. The term *epoch* is often used to refer to the number of training cycles after which an update of the connection weights is performed. As the training progresses, the network has a sufficiently large number of free parameters and starts to overfit the training data. However, one can control the training time with cross-validation and premature termination of training. The cross-validation set is used to test the performance of the network on out-of-sample patterns. Network training is terminated when the cross-validation error begins to rise.

- *Transfer Functions*

The transfer function captures the relationship between nodes of different layers in an ANN. In theory, any differentiable function can be used as a transfer function. In practice, the choice of the transfer function is made based on a small number of bounded, monotonically increasing, and differentiable functions such as the sigmoid logistic, hyperbolic tangent, and linear functions, given below

Sigmoid function:

$$f(x) = 1 / (1 + e^{-x}) \tag{2.30}$$

Hyperbolic tangent function:

$$f(x) = (e^{x} - e^{-x}) / (e^{x} + e^{-x}) \tag{2.31}$$

Linear function:

$$f(x) = x \tag{2.32}$$

There are some heuristics on the selection of transfer functions. For instance, the logistic function is used to learn the average behavior in classification and the hyperbolic tangent function is used to learn about the deviation from the average in forecasting (Klimasaukas, 1991). Symmetric functions such as the hyperbolic tangent can yield faster convergence although the learning can become extremely slow if the weights are too small (Refenes, 1995).

In theory, each network node may have a different transfer function. However, most studies report the use of the same transfer function for all hidden nodes in the same layer. The most used transfer function for hidden nodes is the sigmoid function.

For output nodes, some studies report the use of the sigmoid transfer function while others report using a linear function. The sigmoid output is well suited in classification problems having binary target values. Rumenhart et al. (1995) demonstrate the appropriateness of using linear output nodes in forecasting. However, Cottrel et al. (1995) remark on the limitation of linear output nodes in modelling time series with a trend.

- Learning Rate and Momentum

The learning process of ANN is governed by a learning rate and a momentum. A *learning rate* determines the magnitude of a correction term applied to adjust the weight of each node. A large value of the learning rate causes the network to learn more quickly and helps the network escape from a local minimum. However, it may cause the training to be unstable or cause no learning to occur. A *momentum* is the percentage of previous errors applied to weight adjustment in each training case. A large value of momentum causes the network to retain more impact of previous corrections to the current corrections, and helps to prevent the impact of unusual case while continuously correcting for consistent errors.

Figure 2.1

Artificial Neural Network
(5-3-2)

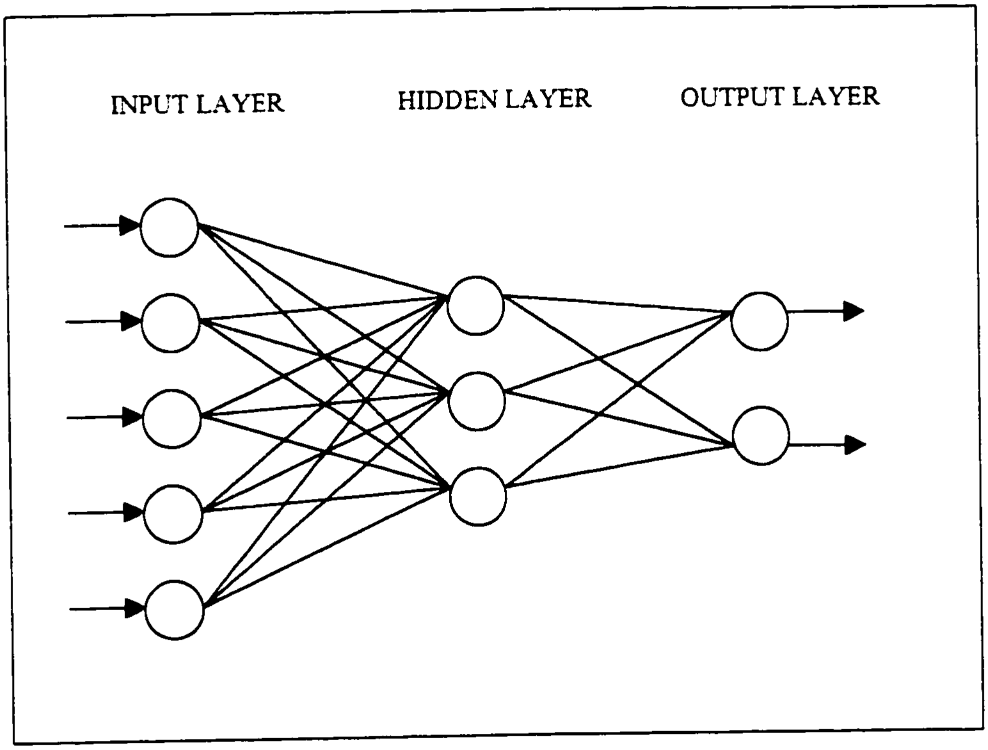


Figure 2.2

Recurrent Network: Jordan Architecture

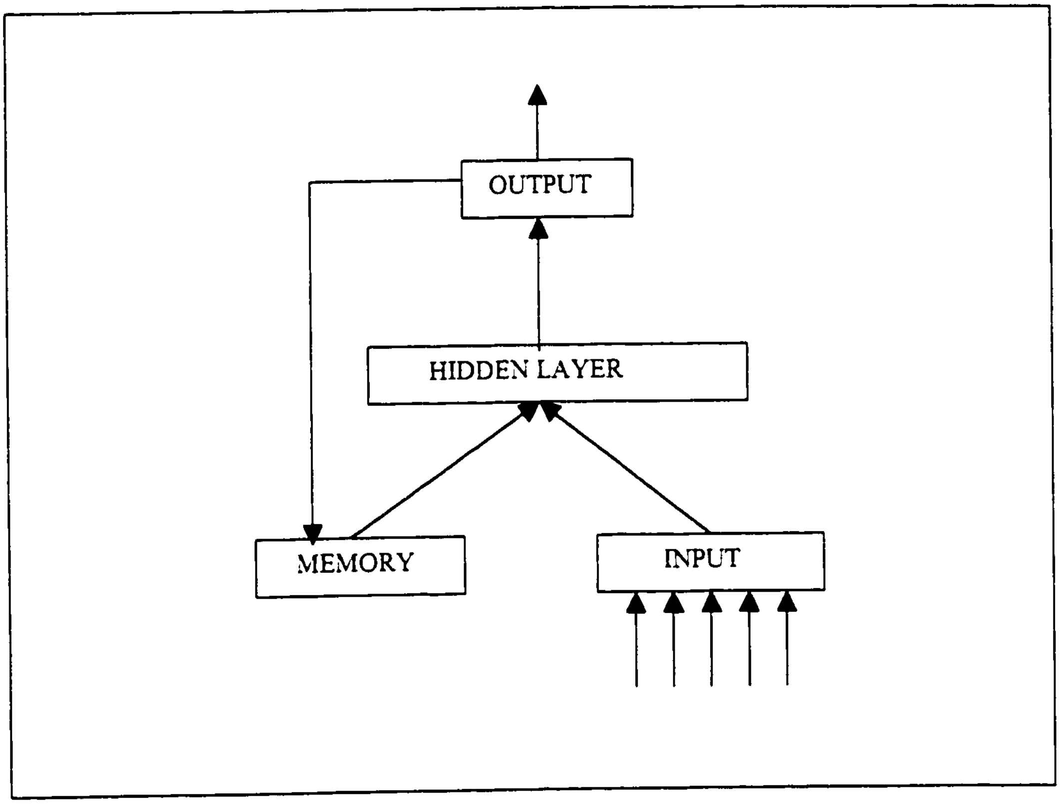


Figure 2.3

Recurrent Network: Elman Architecture

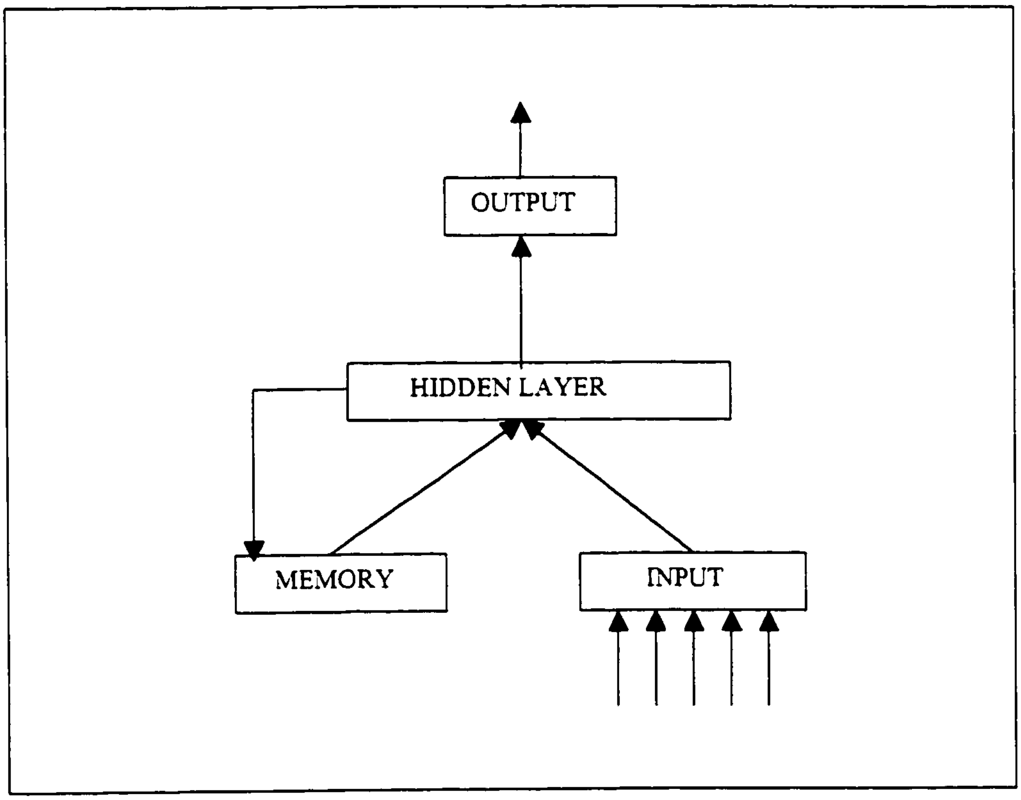
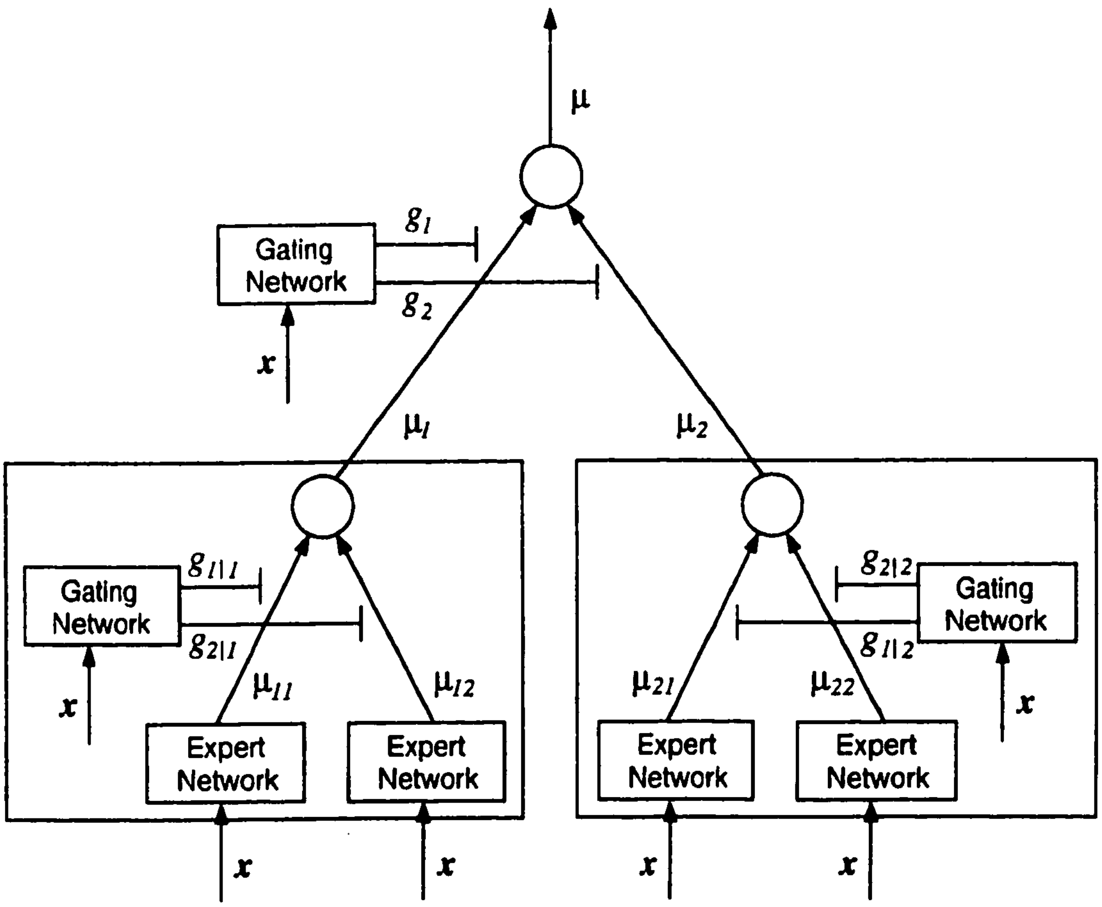


Figure 2.4

Mixture-of-experts Network Architecture



Source: Jordan and Jacobs, 1995

CHAPTER 3

Genetic Algorithms

in the Search for the Optimal Network Design

The choice of network parameters and architectures to achieve a desired network performance is a tedious task involving trial and error. Many heuristics have been proposed to facilitate the task. Nevertheless, the chosen architecture is not necessarily an optimal one that guarantees a satisfactory performance. A Genetic Algorithm (GA) can be implemented to efficiently search for an optimal network design which may produce a better performance for learning and prediction.

Genetic Algorithms are search procedures based on natural selection and genetics. These algorithms mimic the natural process in developing superior entities from a population of entities. The approach was initiated by Holland in the 1960's at the University of Michigan, and was followed by his colleagues and students (Holland, 1975; Goldberg, 1989). The central theme of research on GA has been the robustness or the balance between efficiency and the efficacy necessary for survival in many different environments. The implication is that, if an artificial system can be made more robust, costly redesigns can be reduced or eliminated. GA intends to design artificial systems having important mechanisms of biological systems such as self-repair, self-guidance, and reproduction. GA are theoretically and empirically proven to provide a robust search in complex spaces (Goldberg, 1989).

This chapter begins with a review of the functioning of a standard GA in a search space and its merits over conventional search methods. It then discusses the implementation of GA to ANN optimization.

3.1. OPERATIONS OF GENETIC ALGORITHMS

Using biological terminology, one deals with *chromosomes* (or strings) composed of *genes* (or features) which take values called *alleles* (or values) in a GA. The entire genetic package of chromosomes is called a *genotype* (or structure). A candidate genotype is called a *phenotype* (or candidate solution). A simple GA consists of the following three operators.

Selection (reproduction) is the survival of a fitness test in meeting an objective function within the GA. This operation gives preference to better entities to be copied into a mating pool to breed the next generation.

The selection could be based on a *roulette wheel* where each current string in the population has a lot size in proportion to its fitness. As such, a string with a higher fit value has a higher probability of contributing one or more offsprings in the next generation. This method gives a slow convergence to the population while emphasizing good genetic mixing. The process does not work well when the fitness values are all very similar in a large population. As an alternative to avoid this difficulty, one can select a certain *top percentage* performers (e.g., top 50%) in the population for the mating pool.

Crossover (cross-breeding, mating) is the mating of two entities chosen randomly from the mating pool. In a *simple crossover* (swap tail), a cross site along the string length is chosen at random with a uniform probability, then position values are exchanged between the two strings following the cross site. For instance, the cross-over of two strings 1111111 and 0000000 at the third locus produces the two offsprings 1110000 and 0001111. These offsprings are placed in a new population. The process continues pair by pair until it completes a new population with offsprings of superior parents. Besides the simple crossover method of one cross site, one can have *multi-point crossover* to assure the feature exchange of longer strings.

Mutation is the occasional, but low probability (e.g., .001), alteration in the bit value of randomly selected features of successful structures. The purpose of this operation is to introduce a simple random walk through the string space. This operator is needed to avoid the overzealous search of reproduction and crossover which may lose potential useful features encoded in a bit value. However, a frequent mutation rate will make the GA no better than a random search.

3.2. REPRESENTATION OF SEARCH SPACE

In a GA, all values of a structure represent characteristics that uniquely define a candidate solution in the space of possible solutions. The GA evolves successor populations, or *generations*, from a limited population of initial candidate solutions. The process is accomplished by defining features having a binary value of 0 or 1 to represent the inclusion or exclusion of a particular value of a feature. These features are varied in each new generation with the resulting structure evaluated in terms of its *fitness*. If a structure meets the selected fitness criteria, then the values of its features are retained and bred with other structures. Some of the weakest structures may be discarded, and replaced with new structures.

Applying a GA search to ANN optimization, a neural network architecture can be defined by a structure. The values of this structure represent network characteristics that uniquely define a candidate in the space of all possible architectures. The encoding of features of an ANN may include the number of inputs, the number of layers, the number of nodes in each layer, the number of feedback connections allowed, the degree of connectivity from one layer to another, the learning rate, and the transfer function utilized by the learning rule. All network parameters are encoded in one long stream of bits. This bit string composes of many segments, each of which represents an area in the search space such as input area, hidden node area, hidden layer area, etc.

A *fitness function* in GA search is defined such that it can account for different performance and cost factors. The fitness function can be a weighted sum of performance metrics (Lin and Lee, 1996). In searching for optimal ANN design, the general metrics may include performance factors such as observed learning speed and accuracy of the network, and cost factors such as the size of network in terms of nodes and weights. Since the relative weights on each metric can be modified, the network structure can be tuned for different optimization criteria. Once the optimal network architecture is discovered, this configuration is implemented in a standard ANN for extensive training.

3.3. A SIMPLE GENETIC ALGORITHM

Hassoun (1995) indicates that in order to apply a GA to an arbitrary optimization problem of the form

$$\text{Minimize } y(\mathbf{x}) \qquad (3.1)$$

$$\text{subject to } \mathbf{x} \in \Sigma \subset R^n$$

it is necessary to establish the following:

- A correspondence between the search space Σ and some space of binary strings Ω, i.e., an invertible mapping of the form D: $\Sigma \to \Omega$;

- An appropriate fitness function $f(.)$ such that the maximizers of f correspond to the minimizers of y.

Given a defined problem. an appropriate symbol string representation (in l-bit long) for candidate solutions, and a fitness function $f(x)$, a simple GA can be represented as follows (Davis, 1991; Mitchell, 1996):

1- Generate randomly n strings of l-bit to form an initial population of candidate solutions.

2- Calculate the fitness $f(x)$ of each string x in the population.

3- Repeat until n offsprings have been created:

- Select with replacement a pair of parent strings from the current population using roulette wheel or top percentage method. The probability of selection is an increasing function of fitness.

- With probability p_c (cross-over rate), cross over the pair at a randomly selected point, chosen with uniform probability, to form two offsprings. If no crossover takes places, form two offsprings as exact duplicates of their respective parents.

- Mutate the two offsprings at each locus with probability p_m (mutation rate).

- Place the resulting strings in a new population.

4 - Replace the current population with the new population. If *n* is odd, one new population member is discarded at random.

5- Go to 2.

Each iteration of this process is called a *generation*. Although a GA is typically iterated for a number of generations, there is no exact specification on how many generations a GA should be iterated. The entire set of generations is called a *run*. Since the process is randomized, each random-number seed will generally produce different results from run to run. As such, GA research usually reports best fitness found in a run and the generation at which the best fit entity was discovered. The results are averaged over many different runs of GA on the same problem.

3.4. MERITS OF GENETIC ALGORITHMS

The merits of GA in comparison with conventional search methods are in its ability to meet robustness requirements. Current literature identifies three main types of conventional search methods, namely calculus-based, enumerative, and random search (Goldberg, 1989). Although these methods are useful, they are not robust and efficient in more complex problems.

- *Calculus-based methods*

Calculus-based optimization techniques can be classified into two categories: indirect and direct methods. Indirect methods seek local extrema by solving the set of nonlinear equations resulting from setting the gradient of the objective function equal to zero. This is the multidimensional generalization of the notion of extrema in calculus. Given a smooth, unconstrained function, finding possible peak starts by restricting search to those points with slopes of zero in all directions.

Direct search methods seek local optima by hopping on the function and moving in a direction related to the local gradient. To find the local best, the function is directed in the steepest permissible direction.

Both methods are local in scope since the optimum they seek are the best in the neighborhood of the current point. Once the lower peak is reached, and in order not to miss the global optima, further improvement is necessary such as using a random restart.

Then, calculus-based methods depend upon the existence of derivatives, i.e., well-defined slope values. Many practical parameter spaces have little respect for the notion of a derivative and the smoothness it implies. The real world search is dealing with discontinuities in multimodal and noisy search spaces. As these methods depend on the restrictive requirements of continuity and derivative existence, they are suitable for a very limited problem domain.

- *Enumerative methods*

Within a finite search space, or a discretized infinite search space, the search algorithm starts looking at objective function values at every point in the space, one at a time. Apparently, this method lacks efficiency as many practical spaces are too large for point-to-point search.

- *Random search algorithms*

One may search and save the best solution with random walks and random searches. In the long run, these methods can be expected to do no better than the enumerative scheme.

Apparently, many traditional search techniques require much auxiliary information in order to work properly. For example, gradient techniques need derivatives, calculated analytically or numerically, in order to be able to climb the current peak. Combinatorial optimization requires access to most if not all tabular parameters. In contrast, a GA has no need for this information. To perform an effective search for better and better structures from a population of candidates, they only require payoff values (objective function values) associated with individual strings representing individual choices in a problem space.

In addition, Goldberg (1989) notes that GAs are different from traditional optimization and search procedures in the following aspects:

- GAs work with a coding of the parameter set, and not with the parameters themselves. Since GAs exploit the coding similarities in a very general way, they are unconstrained by the limitations of other methods such as continuity, derivative existence, unimodality, and so forth.

- GAs search from a population of points, not from a single point. Since GAs work simultaneously with a population of strings representing a rich database of points, they investigate many peaks at the same time. As such, GAs can avoid locating false peaks in multimodal search spaces that usually happen in point-to-point searches.

- GAs use probabilistic transition rules, not deterministic rules, to guide their search. While randomized in its search in order to explore the coding of a parameter space, a GA process is not a simple random walk. In fact, GAs use random choice as a tool to guide a search toward regions of search space with likely improvement.

3.5. GENETIC ALGORITHMS IN ANN OPTIMIZATION

GAs have been applied to the optimization of ANN. They are implemented to search for either a set of optimal network weights or an optimal network architecture.

3.5.1. Search for Optimal ANN Weights

GAs have been used to search for the optimal interconnection weights of a multilayer feedforeward network from its weight space without using any gradient information (Montana and Davis, 1989). Unlike the backpropagation rule, a GA can avoid local minimum traps while performing a global search. In this search, a complete set of weights is coded in a string. Starting with a random initial population of such strings, the GA evolves to arrive at a best fit string.

The literature reports on the superiority of a set of network weights selected by GA (Whitley et al., 1990; Sexton et al., 1998). Apparently, the limitation of this application is in its operation only with a fixed network architecture. Since a set of weights is associated with a particular network architecture, one cannot search for both in the same GA. As the information on network architecture and its associated interconnection weights is represented in the same string, a GA operator may arrive at a solution in which the number of weights exceeds the number of connections.

In addition, for an ANN with continuous activation functions, one cannot replace totally the gradient methods. An alternative is to build a hybrid system in which the genetic weight search is followed by a gradient method, or a gradient-descent step can be included in one of the genetic operators (Lin and Lee, 1996).

3.5.2. Search for Optimal Network Architecture

GAs have also been used to conduct the search in the space of all possible ANN architectures. Parameters of a candidate network are encoded into one long binary value stream of many segments. Each of these segments represents an area in the search space such as the number of inputs, the number of hidden layers, the number of nodes in each layer, the number of feedback connections allowed, the degree of connectivity between layers, the learning rate, and the transfer function. Each candidate network is trained and then evaluated with an appropriate fitness or cost function. This cost function incorporates both performance and the simplicity of the network. The search is carried out with GA such that a good building block in one trial architecture is likely to survive and be combined with good building blocks from the others (Lin and Lee, 1996).

Schaffer et al. (1990) propose the use of GA to evolve ANN architecture. Their method of representing a network architecture in a string allows for the possibility of including or eliminating a hidden node/layer and changing network learning parameters during the evolutionary process. The method of optimizing network architecture with ANN has been investigated by many other researchers (Davis, 1991).

Recently, Kira et al. (1997) investigate the relative performance of using GA in a feedforward ANN versus traditional statistical methods in discriminant analysis (classification) and logistic regression (function approximation). Using GA to select an optimal network architecture and the contributive input variables, they conclude that the hybrid of neural genetic network behaves similar to nonlinear, nonparametric stepwise regression without any *a priori* assumption on the functional form of the relationship among data.

Hansen (1998) compares the performance of backpropagation networks designed by GA and heuristics. The test domains are sets of two-dimensional problems having compensatory, conjunctive, and mix decision structures. With a caution on the results being problem-specific, he finds that heuristics produce a simpler architecture and yet perform comparatively well.

CHAPTER 4

Forecasting with Artificial Neural Networks

In time series studies, traditional forecasting models are linear (Granger and Newbold, 1986; Hamilton, 1994). The nonlinearity in time series has only been investigated recently (Granger and Terasvirta, 1993). With a focus on capturing nonlinear relationships, ANN has emerged as a useful and powerful technique with its capability of function approximation without imposing prior assumptions on the behavior of the variables under consideration. This ability should be efficient in approximating the underlying relationship of related factors in economic events and providing accurate forecasts.

This chapter reviews traditional modelling, and forecasting time series and economic models. It discusses the performance and findings of previous works on using ANN as an alternative method in forecasting. It highlights the shortcomings of and possible improvements on previous works which are being addressed in this thesis.

4.1. FORECASTING UNIVARIATE TIME SERIES

In time series studies, persistent forecasting is based on an assumption that the system has a certain momentum such that the future will replicate the past (Intriligator, 1978). The simplest model of this type is the *status quo* forecast, or the *Naive* forecast. This method predicts that the present value of the interested variable will continue through time in the future, i.e.

$$y_{t-1} = y_t \tag{4.1}$$

Another method called *Naive* II predicts the same change from one period to the next

$$y_{t+1} - y_t = y_t - y_{t-1} \tag{4.2}$$

or predicts the same proportionate change

$$(y_{t+1} - y_t) / y_t = (y_t - y_{t-1}) / y_{t-1} \tag{4.3}$$

A related approach is trend extrapolation, based on simple functions of time, or

$$y_t = a + bt \tag{4.4}$$

Similarly, the exponential trend

$$y_t = \mathrm{A}e^{\alpha t} \tag{4.5}$$

is a special case where the prediction is based on a constant relative change

$$(y_{t+1} - y_t) / y_t = (y_t - y_{t-1}) / y_{t-1} = e^{\alpha} - 1 \tag{4.6}$$

and the forecast at time $t + h$ can be written as

$$\ln y_{t+h} = \ln \mathrm{A} + \alpha\,(t+h) \tag{4.7}$$

A general form for most persistent forecasts is the autoregressive model, in which the forecast value is a weighted linear combination of all past values of the variable

$$y_{t+1} = \Sigma^{n}{}_{j=0}\, a_j\, y_{t-j} \tag{4.8}$$

However, most time series do not fluctuate around a fixed level but have some upward or downward trend. As such, a time series could be modelled as

$$y_t = \alpha + \beta t + \varepsilon_t\,, \qquad t = 1, \ldots, T; \quad \varepsilon_t \text{ i.i.d. } (0, \sigma^2) \tag{4.9}$$

The disadvantage of the above model is that it captures only a global trend and assumes that its parameters are constant over time. It is known that the underlying dynamics of an economy as well as the noise distributions for economic series change with time. Therefore, the useable length of a time series is shortened, and using older data induces biases in predictions (Moody, 1995). A more satisfactory modelling method allows parameters to change over time to capture any change in the data. Most forecasting would be more useful if it could forecast the local trend accurately with more recent observations.

One may emphasize the impact of recent observations in forecasting with various smoothing methods. These are recursive procedures in which current estimates of the level and slope of the model are revised as new observations become available. Although these methods are simple and low cost, better accuracy is obtained with more sophisticated techniques (Makridakis et al., 1983).

4.1.1. ARIMA Models

A time series can also be considered as a stochastic process in which each observation is a random variable ordered in time. Since observations in time series are serially correlated, a simple process can be modelled as a first-order autoregressive AR(1) model

$$y_t - \mu = \varphi(y_{t-1} - \mu) + \varepsilon_t \tag{4.10}$$

Since μ is a fixed level, the model can be represented as

$$y_t = \varphi y_{t-1} + \varepsilon_t \tag{4.11}$$

where the random disturbance ε_t is i.i.d. $(0, \sigma^2)$.

If φ is between [-1, 1], the process is called *stationary*, i.e., the observations are fluctuating around a constant level and there is no tendency for their spread to increase or decrease over time. The properties of almost any stationary time series can be reproduced by introducing a sufficiently high number of lags (Harvey, 1993). However, a more parsimonious model can be constructed by taking into account the lagged values of both observed variables and disturbance terms in an autoregressive-moving average (ARMA) process. An ARMA(1,1) is represented as

$$y_t = \varphi y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1} \tag{4.12}$$

If the process is not stationary, such as the random walk

$$y_t = y_{t-1} + \varepsilon_t \tag{4.13}$$

then its first differences, $y_t - y_{t-1}$, are stationary and one can apply ARMA modelling on the transformed series. This general class of models is called autoregressive-integrated-moving average (ARIMA). Box and Jenkins (1976) develop an extensive procedure to fit ARIMA models. The parameters of ARMA models could be estimated by Ordinary Least Squares methods (OLS).

4.1.2. Forecasting Univariate Time Series with ANN

Traditional time series studies impose strong assumptions on the behaviour of the underlying variables and the functional form of their relationship. The assumptions of normality and linearity on the distribution and behavior of variables are necessary to implement traditional estimation methods, such as OLS. In modelling a time series, one should conduct routine tests on stationarity, normality, independence, linearity, model specification, and model order of related variables (Cromwell et al., 1994a). The belief is that once these restrictions are relaxed and an appropriate estimation method is implemented to reflect the true behaviour and the relationship of the series, forecasting may become more accurate. To recognize patterns of the underlying economic variables and to make accurate prediction, the ANN as a nonparametric, nonlinear function approximator, can be a useful and powerful alternative.

The literature reports many studies on the performance of ANN in forecasting univariate time series in comparison with other traditional forecasting methods. The comparisons are based either on benchmark data (such as the ones used in the M-Competition and the Santa Fe Competition) or on results of previous work derived by traditional methods.

Sharda and Patil (1990) use 75 series from a systematic sample of 111 series from 1001 real time series in the M-Competition data set (Makridakis et al., 1982). They find that the ANN performed equally well as the automatic Box-Jenkins (Autobox) procedure. The other 36 series did not provide both methods with enough data for estimation purposes.

Foster et al. (1991) compare the performance of ANN with those of Holt's, Brown's, and the least square methods. Using the M-Competition data set, they find that ANN is inferior for yearly data and comparable for quarterly data.

Hill et al. (1994, 1996) attempt to arrive at a more definitive comparison of ANN and statistical models based on a systematic sample of 111 series taken from the M-Competition data set. The statistical models are those used by Sharda and Patil (1990, 1992), and Foster et al. (1991) plus deseasonalized single exponential smoothing, deseasonalized Holt exponential smoothing, Box-Jenkins and a combination of results derived from traditional forecasting models. The authors find that the performance of ANN is significantly better than those of other methods. In previous work, ANN is used to forecast all periods in the forecast horizon simultaneously, whereas in Hill et al. the first period of the forecast horizon is generated and fedback into the network to forecast the second period. The performance of ANN is especially superior in later periods of the forecast horizon.

On a different benchmark data set, Gershenfeld and Weigend (1993) report the results of the Santa Fe Competition on the performance of ANN in forecasting time series. In particular, the competition provides the Data Set C concerning tick-by-tick currency exchange rates for Swiss Francs and US Dollars, and asks for six forecasts concerning 1 minute, 15 minutes, 60 minutes, the closing value of the day of the last tick, the opening value of the next trading day, and the closing value of the fifth day after the day of the last tick. In this competition, the recurrent ANN technique had a slight improvement over forecasting with the random walk, whereas the feedforward ANN provides predictions worse than chance. However, the authors suggest that univariate time series are not enough to provide information on the series behaviour.

Other studies report the comparison of ANN performance against results derived by traditional methods reported in previous work. Tang et al. (1991) use ANN to forecast three univariate time series in the time horizon of 1, 6, 12 and 24 months. In comparison with the performance of ANN, they find that Box-Jenkins model performs less well as the forecast horizon extends. With small series, the ANN performs reasonably well whereas the Box-Jenkins method does not work well or does not work at all. However, if the series has a long memory (i.e., a deterministic pattern), the Box-Jenkins method can describe it very accurately. However, the Box-Jenkins model is sensitive to noise and, since forecasts are built on previous observations, the method is only good for short-term forecasting.

Kang (1991) compares the performance of ANN and Autobox on 50 series of the M-Competition data set. These series have been designated as most appropriate for the Box-Jenkins technique. Kang finds that the forecast error of ANN is lower when there are trend and seasonal patterns in the series. Kang also notes that ANN often performs better when predicting beyond the first few periods ahead.

Lee and Jhee (1994) use ANN to identify the order of ARMA model. The identification of ARMA order is based on the Extended Sample Autocorrelation Function (ESACF) discussed in Tsay and Tiao (1984). The authors use ANN to determine the degree of matching between noisy ESCAF pattern with the prototype pattern of ARMA model.

Chu and Widjaja (1994) use ANN to identify six demand patterns in time series: stationary, stationary plus seasonal, linear trend, linear trend plus seasonal, quadratic trend, and quadratic trend plus seasonal. Once the demand patterns are identified, their network is trained to recommend one among six exponential smoothing methods: single exponential smoothing, Brown's linear exponential smoothing, Brown's quadratic exponential smoothing, Holt's two parameter linear exponential smoothing, adaptive-response-rate single exponential smoothing and Winters' seasonal methods.

Maasoumi et al. (1994) use ANN to forecast fourteen US macroeconomic series. However, the series are treated as univariate with time lags varying from one to five.

Lachtermacher (1993) and Lachtermacher and Fuller (1995) report the performance of ANN in comparison with the forecasts of previous studies on four stationary non-seasonal, non cyclic, and on four non-stationary non-seasonal, non cyclic times series. The authors use the Box-Jenkins modelling method to identify the order of the ARIMA model. The number of necessary input nodes for network training is initially determined by the number of autoregressive terms and the number of differencing operations of the calibrated ARIMA model. Then these network architectures are adjusted by trial and error to obtain better MSE in forecasting. They find that ANN performed at least as good as ARIMA models.

4.2. FORECASTING MULTIVARIATE TIME SERIES

It has been noted that univariate time series do not provide enough information to explain the behaviour of the series itself. In addition, economic variables are interrelated so that the multivariate time series approach should provide more accurate forecasts. The purpose of multivariate time series is to determine and discover interactions between a given time series with other time series. From this perspective, a given time series may be influenced not only by certain exogenous events occurring at a particular point in time but also by contemporaneous, lagged, and leading values of many other variables (Cromwell et al., 1994b). The following review of multivariate time series uses the representation in Mills (1990), Lütkepohl (1991), Harvey (1993), and Hamilton (1995).

4.2.1. VARIMA Models

Let y_t denote the value of the economic variable of interest in period t. Then the forecast for period $T + h$, made at the end of period T may have the form

$$\hat{y}_{T-h} = f(y_T\,,\; y_{T-1}\,, \dots) \tag{4.14}$$

where $f(.)$ denotes some suitable function of the past observations $y_T\,,\; y_{T-1}\,, \dots$. In many applications, linear functions are used so that

$$\hat{y}_{T-h} = v + \alpha_1 y_T + \alpha_2 y_{T-1} + \dots + \alpha_p y_{T-p} \tag{4.15}$$

However, the value of an economic variable often is not only related to its predecessors in time but also to the past values of other variables. Examples are the relationships among macroeconomic variables investigated in Sims (1980).

Denoting the related variables by y_{1t}, y_{2t}, ..., y_{Kt}, the forecast of $y_{1,T+h}$ at the end of period T may be of the form

$$\hat{y}_{1,T+h} = f_1(y_{1,T}, y_{2,T}, \ldots, y_{K,T}, y_{1,T-1}, y_{2,T-1}, \ldots, y_{K,T-1}, \ldots) \tag{4.16}$$

In general, the forecast of the k-th variable may be expressed as

$$\hat{y}_{k,T+h} = f_k(y_{1,T}, y_{2,T}, \ldots, y_{k,T}, y_{1,T-1}, y_{2,T-1}, \ldots, y_{k,T-1}, \ldots) \tag{4.17}$$

A multivariate model seeks to capture various interrelationships between different series under consideration, and to provide more accurate forecasts of these series.

Consider the first-order vector autoregressive model VAR(1)

$$\mathbf{Y}_t = \mathbf{\Phi} \mathbf{Y}_{t-1} + \boldsymbol{\varepsilon}_t , \qquad t = 1 \ldots T \tag{4.18}$$

In this model, the disturbances are serially uncorrelated but may be contemporaneously correlated

$$\mathrm{E}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') = \mathbf{\Omega} \tag{4.19}$$

Since each variable depends not only on its own past values but also on past values of others, if $y_{k,t}$ denote the k-th element of Y_t in VAR(1), then

$$y_{k,t} = \varphi_{k1} y_{1,t-1} + \varphi_{k2} y_{2,t-1} + \ldots + \varphi_{kN} y_{N,t-1} + \varepsilon_{kt} , \qquad k = 1 \ldots N \tag{4.20}$$

This representation can be extended by bringing in more lags on y_t and introducing lags of ε_t in a vector autoregressive-moving average (VARMA) model. For example, VARMA(1,1) is expressed as

$$\mathbf{Y}_t = \mathbf{\Phi} Y_{t-1} + \boldsymbol{\varepsilon}_t + \mathbf{\Theta} \boldsymbol{\varepsilon}_{t-1} , \qquad t = 1 \ldots T \tag{4.21}$$

or, in general, VARMA(p,q) is expressed as

$$\mathbf{Y}_t = \mathbf{\Phi}_1 \mathbf{Y}_{t-1} + \ldots + \mathbf{\Phi}_p \mathbf{Y}_{t-p} + \boldsymbol{\varepsilon}_t + \mathbf{\Theta}_1 \boldsymbol{\varepsilon}_{t-1} + \ldots + \mathbf{\Theta}_q \boldsymbol{\varepsilon}_{t-q} , \qquad t = 1 \ldots T \tag{4.22}$$

where $\mathbf{\Phi}_i$s are $N \times N$ matrices of AR parameters and $\mathbf{\Theta}_i$s are $N \times N$ matrices of MA parameters.

When applied to differenced observations, the model becomes a vector autoregressive-integrated-moving average (VARIMA) model.

Define a non-singular $N \times N$ matrix $\mathbf{Q}$, such that

$$\mathbf{\Phi} = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^{-1} \tag{4.23}$$

where $\mathbf{\Lambda} = \text{diag}\{ \lambda_1, ..., \lambda_N \}$. The elements $\lambda_1, ..., \lambda_N$ are obtained by solving the equation

$$| \mathbf{\Phi} - \lambda \mathbf{I} | = 0 \tag{4.24}$$

A VARMA(p,q) process can be shown to be *stationary* if the roots of

$$| \lambda^p \mathbf{I} - \lambda^{p-1} \mathbf{\Phi}_1 - ... - \mathbf{\Phi}_p | = 0 \tag{4.25}$$

are less than unity in absolute value. Stationarity implies that a process has a mean that is the same for each time period. It follows from Wold's Decomposition Theorem that, under quite general conditions, every stationary, purely non deterministic process can be well approximated by a finite-order VAR process (Lütkepohl, 1991).

The process is *invertible* if the roots of

$$| \lambda^q \mathbf{I} - \lambda^{q-1} \mathbf{\Theta}_1 - ... - \mathbf{\Theta}_q | = 0 \tag{4.26}$$

are less than unity in absolute value. With the invertibility, weights placed on past observations decline as one moves further to the past, i.e., larger weights are put on more recent observations. If the invertible condition is satisfied, a VARMA process can have a pure VAR representation (Lütkepohl, 1991).

4.2.2. Forecasting VAR Models with ANN

The parameters of a VAR model can be estimated by using either the Ordinary Least Squares (OLS) or Maximum Likelihood (ML) method. In estimation of VARMA process, ML procedures are recommended as they result in consistent, asymptotically efficient, and normally distributed estimators under very general conditions. However, except for some special cases allowing a fairly straightforward estimation, the optimization of the likelihood function may result in a complicated nonlinear optimization problem. Furthermore, the small sample properties of these ML procedures are in general unknown (Mills, 1990).

If the process follow a VAR model and there is no parameter restrictions, then the model has the form of a *seemingly unrelated regression model* (Zellner, 1962) with an equal number of regressors in each equation. In this case, the parameters can be estimated by regressing each variable on the lags of itself and the other variables. OLS estimation can provide consistent and asymptotically efficient estimates. If there are zero constraints on the parameters, then there may be different regressors in different equations. In this case, the estimation technique proposed by Zellner (1962) can still be applied to derive a structural equation system so that OLS estimation of each separate equation remains efficient.

To identify an appropriate model for multivariate time series, one has to conduct various tests, such as testing for individual stationarity, joint stationarity, and cointegration. Then one has to test for normality and independence. Depending on whether the series have linear dependence or non linear dependence, an appropriate lag order of the model is specified (Cromwell et al., 1994b).

If the series are non-stationary, VAR models can be extended to incorporate constraints such as steady-state relationships and cointegration which link the various series together. However, one can avoid non-stationarity by first differencing all variables.

To measure the performance of VAR models, predictors that minimize the forecast mean squared errors (MSE) are the most widely chosen ones. Arguments in favour of using the MSE as a loss function are advocated by Granger (1969) and Granger and Newbold (1986). They show that minimum MSE forecasts also minimize a range of loss functions other than the MSE. For many loss functions, the optimal predictors are simple functions of minimum MSE predictors. Furthermore, applying to an unbiased predictor, the MSE is the forecast error variance which is useful in setting up interval forecasts (Lütkepohl, 1991).

Since many assumptions are being imposed on time series modelling in addition to the complexity of estimating with the ML method, the use of ANN may provide accurate forecasts without prior assumptions. To date, no report exists on using ANN to forecast multivariate time series and its relative performance against the one of VARMA modelling. Refenes et al. (1995) investigate the forecasting of stock prices based on six independent variables. These time series are investigated in the framework of multiple linear regression.

The application of ANN technique and Vector Autoregression (VAR) in forecasting multivariate times series is investigated by Nguyen and Kira (1997). Benchmark data used in the study are taken from Lütkepohl (1991). The data are West German quarterly, seasonally adjusted fixed investment, disposable income and consumption expenditures in billion of Deutchmarks from 1960 to 1982. To evaluate the improvement of using ANN to forecast multivariate time series, results obtained from the ANN are compared with those obtained from a caliberated VAR model (Lütkepohl, 1991). Using a feedforward network to make static forecasts and a recurrent network for dynamic forecasts, the results indicate that the performance of ANNs are equally accurate as those of the traditional VAR method.

This thesis extends the previous study of forecasting multivariate time series with ANNs to a general business forecasting problem. A general business setting usually involves temporal as well as non-temporal economic variables in which economic events have simultaneous, lagged, and/or contemporanous effects on each others. An economic system is a useful context of investigation into the versatile ability of ANNs in dealing with a general forecasting problem. The next section reviews the structure and forecasting of an economic model.

4.3. FORECASTING AN ECONOMIC SYSTEM

An economic system is a set of simultaneous equations which is meant to describe the working of an economy (Bodkin et al., 1991). One can have a whole system as well as a partial system. For many years, macroeconometric models have been constructed as essentially empirical counterparts to the Keynesian system. Only recently have alternative paradigms appeared in econometric models such as the monetarist, radical or Post-Keynesian, rational expectation, time series (Bodkin et al., 1991).

4.3.1. Structural Equations of an Economic System

An economic model organizes the information available about the system under study and postulates the interrelationships among observables. The formulation of an economic model requires the following specifications: (1) the classification of economic variables, (2) the variables that enter a specific equation, (3) any possible lags involved, (4) nonsample information about a single parameter or combinations of parameters, and (5) how many equations there should be and how the system should be closed or made complete (Judge et al., 1985). The equations of such system are called structural equations, and their corresponding parameters are called structural parameters. The system of equations is complete if there are as many equations as the number of endogenous variables.

A system of simultaneous equations may include: (1) *behavioral equations* describing the responses of economic agents in the form of economic relations, (2) *technical equations* involving relations in the system, (3) *institutional equations* and *accounting identities* or *definitional equations*, and (4) *equilibrium conditions*. The institutional equations, accounting identities or definitional equations, and the equilibrium conditions are deterministic and contain neither stochastic terms nor unknown parameters to be estimated. However, they provide important feedback relations for jointly determined variables. The behavioral equations and technical equations specify possible relationships among the endogenous and predetermined variables. They contain stochastic disturbance terms as well as unknown parameters to be estimated.

Simultaneous and structural equations of an econometric model usually contain information on the following variables (Judge et al., 1985):

- *Endogenous*, or jointly determined variables, have outcome values determined through the joint interaction with other variables within the system.

- *Exogenous variables* affect the outcome of the endogenous variables, but whose values are determined outside the system. Exogenous variables are assumed to condition the outcome values of the endogenous variables but are not reciprocally affected because no feedback relation is assumed.

- *Lagged endogenous variables* can be placed in the same category as the exogenous variables since the observed values are predetermined for the current period. The exogenous variables and lagged endogenous variables that may involve any length of lag are called predetermined variables.

- *Non-observable random errors,* also called random shocks or disturbances.

Following Intriligator's representation (1978) for any econometric model summarized by the estimated reduced-form equations, a short-term forecast of values taken by all endogenous variables in the next period is given as

$$\hat{\mathbf{y}}_{t-1} = \mathbf{y}_t \hat{\mathbf{\Pi}}_1 + \hat{\mathbf{z}}_{t+1} \hat{\mathbf{\Pi}}_2 + \hat{\mathbf{u}}_{t+1} \tag{4.27}$$

As such, the prediction of next period values of the endogenous variables consists of two systematic components and one judgmental component.

The first systematic component $\mathbf{y}_t \hat{\mathbf{\Pi}}_1$ indicates the dependence on current values of the endogenous variables which are weighted by the estimated coefficients in $\hat{\mathbf{\Pi}}_1$. This term summarizes the systematic dependence of each endogenous variable on previous values of all endogenous variables due to factors such as serial correlation, constant growth processes, or distributed lag phenomena.

The second systematic component $\hat{\mathbf{z}}_{t-1} \hat{\mathbf{\Pi}}_2$ is based on a prediction of the future values of the exogenous variables $\hat{\mathbf{z}}_{t-1}$ and the estimated coefficients $\hat{\mathbf{\Pi}}_2$. This term reflects the dependence of the endogenous variables on exogenous variables of the model. Since the $\mathbf{z}_{t-1}$ are exogenous variables, they are themselves determined on the basis of factors not explicitly treated in the econometric model. As such, it is reasonable to assume that these variables must be forecasted on the basis of factors other than those of the model itself.

The third component is the judgmental component $\hat{\mathbf{u}}_{t+1}$ called "added factors" which can be interpreted as estimates of future values of the disturbance term. This component summarizes the effect of all other factors including variables omitted from the model. The added factors are based on judgments of factors not explicitly included in the model. The exclusion of these factors may be due to their rare occurrence or difficulty in obtaining data. However, this does not mean that they must be overlooked in formulating a forecast. Indeed, it would be inappropriate to ignore relevant considerations simply because they were omitted from the model.

4.3.2. Estimation of Structural Equations

For a set of equations to be estimated, the disturbances in the system at a given time are likely to reflect some common unmeasurable or omitted factors and therefore could be correlated. The correlation between disturbances from different equations at a given time is known as contemporaneous correlation. It is distinct from the autocorrelation which refers to the correlation over time for the disturbances in the single equation. When contemporaneous correlation exists, it may be more efficient to estimate all equations jointly, rather than to estimate each one separately using least squares. The joint estimation technique is known as *seemingly unrelated regressions estimation* (Zellner, 1962)

When one deals with a statistical model consistent with a system of simultaneous equations in which many variables are interdependent, the classical least squares rule is biased and does not converge to the true parameters even in large size samples. The indirect least squares procedure is viable in the case of just-identified equations. In the case of over-identified structural equations, one can use a generalized least squares procedures such as 2-stage, or 3-stage least square estimators. Literature on the asymptotically equivalent estimators shows that in an econometric model with a high degree of over-identification and low degree of freedom, the ordinary and two-stage least squares estimators will show similar patterns in their statistical behavior (Judge et al., 1985)

Various errors must be taken into account in studying the accuracy of econometric forecasts. First, there is the inaccuracy in the model, which is a simplification of reality such that it omits certain influences and simplifies others. Second, there is the inaccuracy of the data used in the estimation of the model. Third, there is the inaccuracy or bias present in the method of estimation. Fourth, there are errors in the forecast of the exogenous variables and in the added factors. Finally, there are possible inaccuracies in the actual data in which the forecast is compared (Intriligator, 1978).

Following the representation of Intriligator (1978), the absolute error of a short-term forecast can be expressed as

$$\hat{\mathbf{e}}_{t+1} = \mathbf{y}_t\,(\mathbf{\Pi}_1 - \hat{\mathbf{\Pi}}_1) + \mathbf{z}_{t+1}(\mathbf{\Pi}_2 - \hat{\mathbf{\Pi}}_2) + (\mathbf{z}_{t+1} - \hat{\mathbf{z}}_{t+1})\hat{\mathbf{\Pi}}_2 + (\mathbf{u}_{t+1} - \hat{\mathbf{u}}_{t+1}) \qquad (4.28)$$

The first term consists of errors due to incorrect estimation of the coefficient matrix $\mathbf{\Pi}_1$ being weighted by $\mathbf{y}_t$. The second term consists of errors in estimating the coefficient matrix $\mathbf{\Pi}_2$ being weighted by the true value of future exogenous variables $\mathbf{z}_{t+1}$. The third term consists of errors in the forecasting of these future exogenous variables, weighted by the estimated coefficient matrix $\hat{\mathbf{\Pi}}_2$. The fourth term consists of errors in the disturbance terms, where $\mathbf{u}_{t+1}$ are the added factors reflecting other effects not being included in the model.

More recent experience with forecasting using an econometric model has indicated the importance and value of added factors $\mathbf{u}_t$. These added factors, which reflect expert judgment on factors not included in the model, in general significantly improve model performance. Forecasts with such subjective adjustments generally are more accurate than those obtained from the purely mechanical application of econometric models. Combining an econometric model with expert opinion in this way utilizes the best features of each. It combines the explicit objective discipline of the formal econometric model and regression estimators with the implicit subjective expertise of individual experts intimately aware of the real world system. In any cases, an econometric model should be considered as a useful starting point for formulating the forecast. It identifies those factors for which judgmental decision must be made, and it provides a framework to ensure an internally consistent forecast.

4.3.3. Nonlinearity and Dynamics of Economic Variables

Economic variables change over time and the linearity of an economic model is a strong assumption. Current concerns in forecasting are in how to capture the nonlinearity and the dynamics of economic events in economic modelling.

Nonlinearity in economic models can be in various forms. The nonlinearity can be in the variables or in the parameters, or in both. In such cases, a traditional method is to find a transformation, such as a Box-Cox transformation, to convert the model into a linear specification. Judge et al. (1985) note that it is not unusual that parameters entering in a regression model simply reflect one's uncertainty on which model would adequately represent the relationship among the variables. But there are models, called *intrinsically nonlinear models*, which can not be linearly transformed. The estimation of these models is based on minimizing or maximizing an objective function such as the sum of squared errors or the likelihood function (Judge et al., 1985). However, with the current optimization methods, one may encounter estimation complexity when dealing with nonlinear optimization problems (Mills, 1990).

In the literature, the dynamics in forecasting is discussed within a family of statistical models called dynamic regression models (Pankratz, 1991). In this family, an output is linearly related to current and past values of one or more inputs. The crucial assumption of these models is that inputs are not affected by outputs. However, an alternative approach is simultaneous equation modelling which was introduced into econometrics with an investigation into the relationship among a set of macroeconomic time series (Sims, 1980). From this perspective, a given time series may be influenced not only by certain exogenous events occurring at a particular point in time but also by contemporaneous, lagged, and leading values of a second variable or many other variables (Cromwell et al., 1994b). Therefore, the dynamics of an economic model should be studied from a multivariate perspective rather than a univariate one.

4.3.4. Major Macroeconomic Models

Econometric models reported in literature range from small to large scale models. Small models have less than ten endogenous variables such as the Klein interwar and St-Louis models. Medium-size models have between ten to one hundred variables such as Klein-Goldberger and Wharton Models. Large models have more than one hundred variables such as Wharton Mark II, Chase Econometrics, and DRI models (Intriligator, 1978). In a comprehensive work, Uebe (1995) compiles a catalogue of economic models of various countries in the world.

The antecedent of econometric models dates back to the models of general equilibrium which were first developed by Leon Walras as an abstract system and later extended by Vilfredo Pareto for empirical estimation. The last major antecedent is the empirical literature on Keynesian macroeconomic concepts spanning between the publication of Keynes' *General Theory* in 1936 to the publication of Tinbergen's *Business Cycles in the United States of America, 1919-1932* in 1939 (Bodkin et al., 1991).

The early contributions to macroeconometric modelling initiated by Tinbergen (1939) are in his seminal work on an annual model of the US economy. The model is fitted by Ordinary Least Squares to data for the period 1919-1932. There are 31 behavioral equations and 17 identities for a total of 48 equations. All the relationships of the model are linear including the linearizations of non-linear relationships. The model is used to draw conclusions about the US economy based on an examination of individual behavioral equations during the period of study. Sensitivity analysis on the estimated regression coefficients was conducted. The model is also used to evaluate the theories of business cycles. However, there are no partial or full model simulation to test the hypothesized structure of the model. Also no forecasting uses of the model are reported (Bodkin et al., 1991).

In 1950, Klein published a monograph on *Economic Fluctuations in the United States 1921-1941*, which presents three models of the US economy. These models are usually indicated as Klein's interwar models or Model I, Model II and Model III. The principal attempt of these early econometric models is to test hypotheses and to describe the US economy (Klein, 1950). The parameters of these models are estimated with Maximum Likelihood and Ordinary Least Squares methods.

Model I contains three behavioral equations and three identities. The endogenous variables are levels of consumption, net investment, private wage and salary bill, real national income, total property (non wage) income and the stock of capital. All endogenous variables are in constant dollars, and all relationships are strictly linear. The exogenous variables are primarily fiscal with a time trend, such as real government expenditures including net exports, the public wage and salary bill in deflated terms, and real net indirect tax collection.

Model II contains a consumption function and two identities which relate current dollar gross national product to the sum of expenditure components and to disposable income. In the behavioral equation of consumption, per capita consumption is related to current and lagged real, per capita disposable income and to the real per capita level of the money stock.

Model III contains 12 behavioral equations and four identities. There are three behavioral equations to describe the money market and four behavioral equations to describe the housing market. There are also a consumption function, a labor demand function, an investment demand equation for plant and equipment, and a demand for inventories equation. Later, this model was developed further into the Klein-Goldberger Model.

Klein-Goldberger Model was initiated as a project of the Research Seminar in Quantitative Economics at the University of Michigan. It is a medium size model which consists of 15 structural equations, five identities and five tax-transfer auxiliary relationships. The model is based on annual observations from the split sample period 1929-41, 1946-1952, and is estimated by the limited information maximum likelihood technique. This model served as the paradigm for many subsequent model-builders (Bodkin et al., 1991).

The structure of the Klein-Goldberger Model may be viewed as the first empirical representation of the broad basic Keynesian system (Bodkin et al., 1991). A blending of real and money values is achieved as both the constant-dollar magnitudes and their associated price deflators are estimated as part of the model. The dynamic components are added in terms of cumulated investment, time trends, and distributed lags. The model also contains several non-linearities in terms of the variables, which are linearized in an approximate manner to obtain the solution for the entire system (Bodkin et al., 1991).

The framework of the model consists of 20 equations which explain 20 endogenous variables of the model. The list of related variables presented in Klein and Goldberger (1955) contains 63 variables in which 20 are endogenous and 43 are predetermined. Of the 43 predetermined variables, 19 are current exogenous and 24 are lagged. In the model, there exists some degree of disaggregation. For example, among the endogenous activity variables, there are five categories of income, two liquid assets, two interest rates, and three types of prices. The exogenous set contains five demographic and social environmental variables, and nine policy instruments including two types of government expenditures and five types of taxes.

The Klein-Goldberger model was used to provide *ex ante* forecasting for the period 1953 (prepared in February-March 1953), 1954 (prepared in November-December 1953) and 1955 (prepared in December 1954). Forecasts were based on the new estimation with updated sample period and refinements of the model. Goldberger (1959) systematically compares the ex-ante forecasts of the Klein-Goldberger model for 1953, 1954, and 1955 with alternative predictions generated by naive methods and concludes that the forecasting performance of the Klein-Goldberger model is superior. The model almost always and for every variable predicted rightly the direction of change.

4.3.5. Forecasting Macroeconomic Models with ANN

The first attempt to use ANN in modeling nonlinear relationships among economic variables of a structural system was undertaken by Caporaletti et al. (1994) with an in-sample estimation of Klein's Model I. Their ANN contains thirteen input nodes corresponding to seven predetermined variables plus six exogenous variables of the model. The output layer contains a single node corresponding to the remaining endogenous variables. The hidden layer of this configuration contains eight nodes. Three ANNs are constructed and trained, each of which is used to forecast one of three endogenous variables of the model which are consumption, investment and private wage bills. The authors conduct ex-post forecasts and find that results are significantly better than those from traditional estimation methods.

This attempt has the following shortcomings. First of all, with a single output node the network does not account for the contemporaneous and simultaneous effects of endogenous variables. As such, it has a similar drawback of traditional single equation estimation method. In a simultaneous equation system, the appropriate estimation should be based on a multivariate approach.

Then, current values of endogenous variables in this setting are considered as inputs of the network. In addition, there is no feedback to account for the dynamics of the system. As such, this network cannot estimate and forecast a particular endogenous variable without the need of predetermined, current as well as lagged, values of all other endogenous variables.

Lastly, this network architecture does not handle a mix of non-temporal and temporal variables. As such, one cannot effectively account for the contemporaneous and lagged effects of related variables of an economic system.

This thesis addresses these shortcomings with a network architecture which has the ability to account for the simultaneous and contemporaneous effects of the variables in an economic model. Using recurrent network design, the proposed network also accounts for the dynamics of the system. As such, ANNs can provide effectively not only *ex-post* estimations but also *ex-ante* forecasts of an economic system.

CHAPTER 5

Research Questions and Methodology

The aim of this study is to investigate the integration of Artificial Neural Networks and Genetic Algorithms in an effective architecture for an efficient intelligent forecasting system. The proposed system is expected to capture effectively the underlying dynamics and relationships among variables in an economic setting, and then to provide accurate forecasts. This chapter presents the research questions studied herein and discusses methodology used to address these questions.

5. 1. RESEARCH QUESTIONS

A general business forecasting problem has to deal with a multiple of interacting temporal and non-temporal variables. One notes that a macroeconomic model usually contains time series variables and other current variables. As such, it serves as a useful context for developing an intelligent forecasting system using ANN and GA.

A major drawback of traditional econometric forecasting is the requirement of an exactly specified simultaneous equation model. Actually, the functional specification of a model is influenced more by the needs of the econometrician and the proposed estimation procedure than by a knowledge of the true underlying process. The relationship of variables in a model always reflects adherence to an economic school (Karakitsos, 1992).

The choice of a traditional estimation technique is more often made for computational ease rather than for capturing the true functional relationship. Also, traditional estimations have not addressed sufficiently the nonlinear relationships among economic variables (Mills, 1990).

The Artificial Neural Network technology has emerged as a prominent alternative to traditional modelling and forecasting with its nonparametric and nonlinear approach. An ANN is a universal approximator of any functional relationship. Given the existence of interactions among variables in a macroeconomic model, an ANN should be able to approximate well the underlying relationship. The main advantage of ANN is in its ability to learn patterns from data without imposing any strong assumptions on the behavior of related variables as well as on the functional form of the relationships.

This study investigates the implementation of an integrated framework of ANN and other related decision technology in order to capture the underlying dynamics and nonlinear relationships among variables in an economic setting. The proposed framework is designed to learn well patterns of relationships among related variables, and then to provide forecasts which closely reflect reality. In the development of economic policies, macroeconomic variables could be specified, but their functional forms may not necessarily be prescribed *a priori*. Unlike traditional modelling and estimation methods, the ANN technology does not impose strong assumptions on the behaviour and relationship of related variables to fit data in an *a priori* model. In addition, the proposed framework has its own mechanism based on Genetic Algorithms to select the most appropriate ANN topology and related information on system inputs in order to arrive at the most accurate learning and forecasting. This adaptive mechanism of the Machine Learning approach requires a minimum of interference from analysts. As such, this integrated framework should serve well as a foundation for an intelligent forecasting system.

The study first proposes an effective ANN architecture from a mixture perspective to capture temporal and non-temporal patterns of economic variables. Then, it addresses the issues on the effectiveness of GA in the selection of optimal network topology. After that. it discusses the efficiency of the integrated ANN learning and forecasting.

- *Mixture of Networks in Learning Multiple Economic Patterns*

Forecasting in a general business context represented by an economic model of many interacting variables requires a large and complex network. However, in a large single network architecture, the behavior of individual economic variables may not be well examined. In addition, it can be very difficult to recognize the patterns of a mix of temporal and non-temporal variables. As an alternative, *Mixture-of-experts Network* (Jacobs et al., 1991) provides a useful architecture to learn the patterns of multiple economic variables. From this perpective, the study proposes various ANN architectures to estimate and forecast a mix of temporal and non-temporal economic variables effectively.

- *The Effectiveness of Genetic Algorithms in Network Topology Selection*

Since a GA is used intensively in this study to search for the optimal network topology, the first issue that needs to be addressed is the effectiveness of GA's selection. The network parameters to be selected by GA are learning rate, momentum, transfer functions, number of hidden layers and their nodes. In addition, GA is used to designate appropriate contributing inputs for forecasting purposes. With its powerful search mechanism in investigating a large population of alternatives, GA should be useful in network design for a superior performance. GA helps to overcome the suboptimality of a tedious trial-and -error process in network building and variable selection.

- *The Efficiency of ANN Learning and Forecasting*

This study investigates the efficiency of proposed framework over traditional methods. A comparative study is conducted on in-sample learning and out-of-sample forecasting performance across methods. The particular behavior of ANN in estimation and forecasting is analyzed for future effective implementation.

5.2. METHODOLOGY

This section presents the related AI technology used in the study, the research context, and the framework for the evaluation of the experimental results.

- *Related AI Technology*

To implement a mixture of networks in learning multiple economic patterns, an ANN can have a recurrent module to deal with temporal data and a standard module for other non-temporal data. These modules can be refined to learn individual patterns better. Then a gating network captures the relationship between actual output of the economic model and the estimations of modular networks.

To learn temporal patterns of a time series, there are many available networks using various topologies such as time delays, time-varying activations, short-term and long-term memories among others (Maren et al., 1990). Within the scope of this study, the static representation of temporal information and the use of a recurrent network to capture temporal patterns are implemented. In *static representations* of temporal information, a sequence of incoming temporal data is represented simultaneously in the input layer of the network. In *dynamic forecasting*, the predicted values of the economic variables concerned are used in next period forecasting.

This study implements the recurrent learning algorithm proposed by Jordan (1986). In the *Jordan algorithm*, the network output is fed back to the input layer. There is a similar recurrent learning algorithm in the literature, namely the *Elman algorithm*, in which the output of hidden layer is fed back as the next input. The difference between the two algorithms is that the actual network output is fed back to the input layer in the Jordan algorithm whereas the previous internal representation of the output is fed back in the Elman algorithm. The choice of Jordan algorithm over Elman and other time-delayed recurrent algorithms is based on its comprehensiveness in representing lagged information. With this topology, one can have dynamic forecasting, since the previous network output is used in the next forecast period. Furthermore, the algorithm can retain a decayed effect of the previous state to an economic system as the older information has less influence on the current economic state.

Previously, one needed to refine a network topology with a tedious trial and error process. The implementation of a *Genetic Algorithm* is intended to provide an efficient search for optimal network design. In addition, a GA is used to identify the useful inputs (exogenous variables and lagged variables) to predict outputs (endogenous variables) of the economic model. As a GA has been proved to be a powerful search tool with its ability to explore a large population of alternatives, it should help to overcome the suboptimality of the trial-and-error process in network building. Consequently, a GA is used to select optimal network design at each stage of modular estimations in this study.

- *Research Context*

Using macroeconomic models as a general business forecasting context involving a mix of temporal and non-temporal data, this study investigates Klein's Model I of US economy in 1921-1941 (Klein, 1950) and the Klein-Goldberger Model of US economy in 1929-1952 (Klein and Goldberger, 1955; Goldberger, 1959). These two models, representing small and medium-size econometric models, have served as the paradigm for many model-builders for a long period of time (Bodkin et al., 1991). In addition, the availability of data for modelling and estimation in the original work makes these models the subject of many competitive estimation methods. The results from these methods serve well as a benchmark for comparison with those obtained from the ANN technology.

In a comparative study, the relative performance of Klein's Model I and the proposed system are evaluated on *ex-post* forecast for the period 1921-1941. Data for estimation are taken from Klein (1950). Then the framework of Klein's Model I is used to train and validate the *ex-ante* forecast ability of ANN on a moving window scheme from 1950 to 1994. Related data are taken from *National Income and Products Accounts of the United States 1929-1994* (U.S. Department of Commerce, 1998). Within this time horizon, a moving window frame is implemented. In each window, 20 periods are used for estimation, 5 subsequent periods for validation and next 5 periods for testing.

In a similar manner, the relative performance of the Klein-Goldberger Model is compared for in-sample estimations for 1929-1950 . Then *ex-ante* forecast are evaluated for the period 1951-1952. Data for estimations and forecasting are taken from Klein-

Goldberger (1955). The original model involves many data transformation and aggregation that one does not have equivalence in more recent data. As such, this study does not extend the comparative study further ahead. However, the framework of original model and its available data is used to demonstrate the functioning of an intelligent system in selecting input variables and optimal network design.

In any case, Klein's Model I and Klein-Goldberger Model and their available data serve as a useful context to investigate the foundation of an intelligent forecasting system. This system should perform well in any other similar forecasting contexts with more current information on economic variables.

- *Performance Evaluation*

The evaluation of empirical results in this study proceeds as follows. After proposing an appropriate modular ANN framework for each economic model, the study conducts a GA search for optimal network design and a sensitivity analysis to assert the effectiveness of this search. Once the network is refined according to the GA specification, the comparative performance of network estimation and forecasting will be evaluated.

Sum of Squared Errors (SSE) is chosen as the benchmark for comparison because ANN is a nonparametric estimation method. As such, the error measurements do not account for the number of parameters to be estimated as in Mean of Squared Errors (MSE) of traditional parametric regressions. Since this study does not replicate other traditional estimation methods, there is not enough available information on residuals of estimations and forecasts to conduct comparisons across methods based on other error measurements. Also, this study does not impose the assumption of normality on the estimated residuals.

Consequently, significance tests on the results are not conducted herein. However, this study documents extensively information on its estimations and forecasting for future reference.

Sensitivity Analysis of GA Selection

GA is used to search for the optimal number of inputs, the number of hidden layers, the number of hidden nodes, the transfer functions, and learning parameters. Controlling for other factors, the sensitivity analysis on each selection is conducted to evaluate the effectiveness of the GA selection. In particular, experiments are undertaken to address the following issues on ANN performance:

- The effect of learning parameters selected by GA;
- The effect of transfer function selected by GA from a pool of sigmoid logistic, hyperbolic tangent, and linear functions;
- The effect of network architecture selected by GA in terms of number of hidden layers, hidden nodes, and input nodes.

Comparative Analysis of ANN Efficiency

This study evaluates the relative as well as absolute performance of the proposed intelligent forecasting system. *Relative performance* of the framework is evaluated against traditional estimations. The comparison across methods is based on the discrepancy between estimation/forecasting and actual data. The performance is evaluated by Sum of Squared Errors (SSE) between the forecast and actual values. For Klein's Model I, error measurements for the period of 1921-1941 are available in Klein (1950), SAS/ETS (1984), and Greene (1990). For Klein-Goldberger Model, error measurements on estimations and forecasts for the period of 1929-1952 are available in Klein and Goldberger (1955).

Absolute performance of the proposed framework is assessed in the case where comparable traditional estimations are unavailable. The absolute assessment would serve as benchmark for future comparative work. In this case, each module as well as mixture network is trained in 30 runs of 1000 to 10000 epochs, and the minimum and maximum errors are recorded. The varying training times from 1000 to 10000 epochs is to investigate the effects of under-training and over-training on network performance. The choice of 30 runs applied for each network is to provide information for future evaluations such as significance tests and confidence intervals from a traditional statistical perspective. The maximum and minimum error measurements serve as bound of errors on estimation resulting from the proposed framework. This also provides useful information on the performance of ANN in common practice where networks are run just a few times.

CHAPTER 6

Estimation and Forecasting of Klein Model I

6.1. KLEIN MODEL I REVISITED

Klein (1950) modelled the US. interwar economy from 1921 to 1941 in six structural equations. The model has three behavioral and three definitional equations. All terms are measured in 1934 dollars. The model specification and variable descriptions are as follows. For simplicity, time subscripts are omitted unless they indicate the lagged effects.

- Consumption Equation:

$$C_t = \alpha_0 + \alpha_1 P_t + \alpha_2 P_{t-1} + \alpha_3 (Wp_t + Wg_t) + u_{1t} \tag{6.1}$$

where C is consumption, Wp is private wage bill, Wg is government wage bill, and P is non-wage income (profits).

- Investment Equation:

$$I_t = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + \beta_3 K_{t-1} + u_{2t} \tag{6.2}$$

where I is net investment, P is profits, and K_{t-1} is stock of capital at the beginning of the year.

- Private Wages: (Demand of Labor)

$$Wp_t = \gamma_0 + \gamma_1 (Y_t + T_t - Wg_t) + \gamma_2 (Y_{t-1} + T_{t-1} - Wg_{t-1}) + \gamma_3 t + u_{3t} \tag{6.3}$$

where Y is output, T is taxes, and t is time trend (year minus 1931).

- Equilibrium Demand:

$$Y_t + T_t = C_t + I_t + G_t \tag{6.4}$$

This equation defines total output as the sum of goods demanded by consumers C plus goods demanded by business firms I plus goods demanded by the government and foreigners G. The net change in inventories is included in I hence the demand and output cannot differ.

- Income:

$$Y_t = Wp_t + Wg_t + P_t \tag{6.5}$$

This equation defined total output (income) as the sum of profits and wages.

- Capital Stock:

$$K_t = K_{t-1} + I_t \tag{6.6}$$

This equation defined investment as the rate of change of the capital stock.

The system has six endogenous variables C, I, Wp, P, K, Y, and four exogenous variables T, Wg, G and t.

6.2. PREVIOUS ESTIMATIONS OF KLEIN MODEL I

The following is a review of traditional estimation methods. These methods estimate either single equations of the system or the system as a whole. The details of estimation are described in Klein (1950), Theil (1971), and Greene (1990). The results obtained from these traditional estimation methods serve as the benchmark for comparison with those from ANN.

6.2.1. Klein's Model I in Reduced Form

The purpose of the six equation system is to describe six endogenous variables (C, P, Wp, I, Y, K) in terms of four exogenous variables (Wg, T, G, t). In the structural form, each of its equations describes part of the structure of the economy. This implies that some of the equations describe an endogenous variable in terms of other endogenous variables.

An alternative is to write the system in "reduced form", i.e., to solve the system with respect to the endogenous variables (Theil and Boot, 1962). Given the assumption on linearity of the system, the reduced form can be specified as follows

$$\mathbf{y}_t = \mathrm{A}\mathbf{y}_{t-1} + \mathrm{B}\mathbf{x}_t + \mathrm{C}\mathbf{x}_{t-1} + \mathbf{u}^*_t \qquad (6.7)$$

where $\mathbf{y}$ is the column vector of endogenous variables, $\mathbf{x}$ is that of the exogenous variables, and $\mathbf{u}^*$ is that of the reduced-form disturbance. These variables are given by:

$$\mathbf{y} = [\, C \;\; P \;\; Wp \;\; I \;\; Y \;\; K \,]^T \qquad (6.8)$$

$$\mathbf{x} = [\, Wg \;\; T \;\; G \;\; t \,]^T \qquad (6.9)$$

$$\mathbf{u}^* = [\, u^*_C \;\; u^*_P \;\; u^*_{Wp} \;\; u^*_I \;\; u^*_Y \;\; u^*_K \,]^T \qquad (6.10)$$

In the system of reduced form, each of the endogenous variables in year t is described linearly in terms of the same variables lagged one year ($\mathrm{A}y_{t-1}$), the exogenous variables in the same year ($\mathrm{B}x_t$), the exogenous variables lagged one year ($\mathrm{C}x_{t-1}$), and the reduced-form disturbances. It is noted that C, Wp and I do not occur in lagged form, so that the corresponding columns in the coefficient matrix consist of zeros.

6.2.2. Single Equation Estimations

Single-equation Method of Least Squares

This method treats each equation independently of all others in the system. Klein noted that one had to make arbitrary choice of dependent variables for each of the three equations (Klein, 1950). The problem with this method is that it does not account for simultaneous and contemporaneous effects as one takes the values of other endogenous variables as predetermined in the calculation of the equation of interest. At best, this method may serve as a sensitivity analysis given the predetermined values of other endogenous and exogenous variables of the system.

Two-Stage Least Squares (2SLS):

Estimation of the system one equation at a time has the benefit of computational simplicity. However, the method neglects information contained in other equations. Therefore, it is called limited-information method. It has been noted that the Ordinary Least Squares method cannot be applied with overidentified equations. In this case, one has to use 2SLS as an alternative.

Consider a system of simultaneous equations, the nonzero terms in the *j*th equation are

$$\mathbf{y}_j = \mathbf{Y}_j\boldsymbol{\gamma}_j + \mathbf{X}_j\boldsymbol{\beta}_j + \boldsymbol{\varepsilon}_j \qquad (6.11)$$

The procedure of 2SLS consists of the following 2 steps (Greene, 1990):

- In the first stage, ordinary least squares prediction $\mathbf{Y}_j^*$ is obtained from a regression of $\mathbf{Y}_j$ on $\mathbf{X}$;

- In the second stage, the 2SLS estimator is obtained by ordinary least squares regression of $\mathbf{y}_j$ on $\mathbf{Y}_j^*$ and $\mathbf{X}_j$.

Limited Information Maximum Likelihood

In limited information maximum likelihood estimation, one takes into account the absence of certain variables from a particular equation but not the absence of any variables from any other equations (Theil, 1971). Using the reduced form of the system, the joint density of endogenous variables is formulated and maximized subject to the constraints that relate the structure to the reduced form (Klein, 1950).

6.2.3. System Estimations

Three-Stage Least Squares (3SLS):

3SLS involves the application of generalized least squares estimation to the system estimation, each of which has first been estimated using 2SLS. The process is as follows (Greene, 1990).

- In the first stage, the reduced form of the system is estimated. Using ordinary least squares, this results in $\mathbf{Y}_j^*$ for each equation.
- Then the fitted values of the endogenous variables are used to get 2SLS estimations of all the equations in the system. Also residuals of each equation are used to estimate the cross-equation variances and covariances $\mathbf{\Sigma}^*$.
- In the last stage, generalized least squares parameters are obtained for the system.

Full Information Maximum Likelihood

This estimation method assumes that (i) each of the three variables C, Wp, and I is non-autocorrelated, i.e., no correlation between their successive values, and (ii) there is no correlation between the disturbances in any of the structural equations. The estimators treat all equations and all parameters jointly in formulating the likelihood function to be maximized subject to all of the restrictions imposed by the structure. Estimation with full information maximum likelihood was reported in Klein's monograph (Klein, 1950).

Estimated parameters for the three equations for C, Wp, and I obtained from different methods of limited- and full-information estimations are reported in Greene (1990). In this study, the comparison across methods is based on residuals of related estimations reported in Klein (1950) and SAS/ETS (SAS, 1984).

6.3. ANN ESTIMATION OF KLEIN MODEL I

This study uses the Mixture-of-experts Network architecture to design an effective ANN for estimation and forecasting a mix of temporal and non-temporal variables. The network accounts for not only the nonlinearity and dynamics but also the simultaneous and contemporaneous effects of the variables in an economic system. Applied to the Klein Model I, the ANN learning and forecasting proceeds as follows.

If one relaxes the linear restriction on the relationships among variables of the Klein Model I, the system can be specified as:

$$C_t = f(P_t, P_{t-1}, Wp_t, Wg_t) \tag{6.12}$$

$$I_t = f(P_t, P_{t-1}, K_{t-1}) \tag{6.13}$$

$$Wp_t = f(Y_t, T_t, Wg_t, Y_{t-1}, T_{t-1}, Wg_{t-1}, t) \tag{6.14}$$

$$P_t = f(Y_t, Wp_t, Wg_t) \tag{6.15}$$

$$Y_t = f(C_t, T_t, G_t, I_t) \tag{6.16}$$

$$K_t = f(K_{t-1}, I_t) \tag{6.17}$$

Using the formulation of Theil and Boot (1962), Intriligator (1978), the reduced form of the Klein Model I can be specified as :

$$\mathbf{y}_t = f(\mathbf{y}_{t-1}, \mathbf{z}_t) \tag{6.18}$$

where

$$\mathbf{y}_t = [\,\mathrm{P\ Y\ K\ C\ Wp\ I}\,]^T \tag{6.19}$$

$$\mathbf{y}_{t-1} = [\,\mathrm{P}_{t-1}\ \mathrm{Y}_{t-1}\ \mathrm{K}_{t-1}\ \mathrm{C}_{t-1}\ \mathrm{Wp}_{t-1}\ \mathrm{I}_{t-1}\,]^T \tag{6.20}$$

$$\mathbf{z}_t = [\,\mathrm{Wg\ T\ G\ Wg}_{t-1}\ \mathrm{T}_{t-1}\,]^T \tag{6.21}$$

The task of the ANN is to learn the underlying relationship in order to map accurately the input-output patterns of the economic model. One notes that the system has two components: a group of temporal variables and another group of non-temporal variables. Thus, the ANN needs two modules to learn the specific patterns of each type of variable. Endogenous variables of the system have contemporaneous and simultaneous effects on each other. Consequently, the ANN needs a mechanism to aggregate modular estimations into final results and to account for the simultaneous and contemporaneous effect of variables in the system. This task can be performed by a gating network. Using Mixture-of-experts Network architecture in this study, the ANN estimation of Klein Model I is conducted with two-stage and modular architectures.

Two-stage Estimations

Since the endogenous variables are contemporaneously related, it is not accurate to estimate them with a single equation approach. In this study, the relationship of endogenous variables and other variables of the system are estimated in the *instrumental stage*. Although these variables are estimated simultaneously, their contemporaneous effect has not been taken into account. Consequently, these instrumental estimations will be mapped to their actual values to account for this contemporaneous effect in the *final stage* (Figure 6.1).

Modular Estimation

In an economic system, some endogenous variables are affected by their lagged values. In addition, the depth of lagged effects may vary across endogenous variables. Also some variables of the model may be affected by a certain exogenous variable *a priori*. Consequently, the ANN should have different modules at the *instrumental stage* to capture these lagged effects or specified effects separately. Without modular estimation for each effect, it could be very difficult to approximate accurately the mix of temporal and non-temporal variables. Instrumental output results from modular estimations are aggregated at the *final stage* to account for the contemporaneous effect of all endogenous variables in the network outcome (Figure 6.2).

Specifically, in the modular estimation of Klein Model I, the instrumental stage has two modules: a *recurrent module* to estimate P_t, Y_t, K_t taken into account their lagged effect, i.e., P_{t-1}, Y_{t-1}, K_{t-1}, and a *standard module* to estimate C_t, Wp_t, I_t. Then these instrumental estimations P^*_t, Y^*_t, K^*_t, C^*_t, Wp^*_t, and I^*_t are mapped to their actual values P_t, Y_t, K_t, C_t, Wp_t, and I_t to account for their contemporaneous effect.

6.4. FINDINGS AND DISCUSSION

6.4.1. Effect of Selected Network Parameters on Learning

From the initial structure of mixture-of-networks, at each stage and for each module in the ANN estimation, GAs are used to select the optimal network topology. The fitness criterion is a function of the simplicity of the network and its relative performance. The simplicity is measured in terms of number of hidden layers and hidden nodes. The relative performance of a candidate is measured in terms of its discrepancy between network outputs and desired targets.

In the following, the network configuration is represented as I-H1F-H2F-OF where I is the number of input nodes, H1 is the number of nodes in the first hidden layer, H2 is the number of hidden nodes in the second layer, O is the number of output nodes, and F is the transfer function choosing from a pool of logistic sigmoid functions (L), hyperbolic tangent functions (T), and linear functions (Lin). For instance, the notation of 9-7L-6T denotes a network configuration of 9 input nodes, one hidden layer with 7 nodes using sigmoid logistic transfer functions, and 6 output nodes using hyperbolic tangent transfer functions.

Experimenting intensively with the standard network of the instrumental stage in two-stage ANN estimation, one seeks to confirm the effectiveness of GA in the selection of network topology. The best configuration for this module is 9-7L-6T. The algorithm uses a learning rate of .1 and momentum of .5. The GA evolves in 30 generations, each of which has a population size of 30 strings. As such, 900 candidates of ANN architectures are examined for each GA run.

- Learning Rate and Momentum

Using the configuration selected by the GA for the standard module, the effect of various learning rates and momentum on learning error is reported in Table 6.1. The errors are averaged over 30 runs, each run lasted 5000 epochs. This sensitivity analysis asserts the optimal selection of learning rate and momentum by GA. It also systematically confirms the effect of learning rate in interaction with momentum on network error. If the learning rate is controlled, the larger momentum produces the larger error. Similarly, if the momentum is controlled, the larger learning rate produces the larger error. To assure a better generalization, this study uses an early stopping rule in which the training is halted when the cross-validation error begins to rise or the training time reaches a preset number of epochs. Consequently, the training time in this experiment is controlled, and the speed of convergence across learning rate and momentum has not been investigated.

- Transfer function

This study lets the GA select an appropriate transfer function from a pool consisting of sigmoid logistic, hyperbolic tangent, and linear functions. In contrast to heuristics reported in the literature, the linear transfer function for output nodes never emerges as the best choice in this study. GA selects most transfer functions of the form of sigmoid logistic or hyperbolic tangent function. Controlling the learning rate at .1 and momentum at .5 in the standard module, sensitivity analysis on number of hidden nodes and their transfer functions are reported in Table 6.2. By eliminating or adding one hidden node and varying the transfer functions of hidden and output layers, the effectiveness of the GA selected configuration is evaluated in terms of network complexity. The performance chracteristics are averaged over 30 runs of 5000 epochs each. This analysis asserts the optimality of GA selection, as the selected network configuration produces the lowest error in learning.

6.4.2. Effect of Training Time on Learning

Consistent with the documentation in the literature, this study confirms that the longer one trains the network, the smaller the error is. In the experiment, in-sample learning error has been decreased as the training time is increased from 1000 epochs to 10000 epochs. Also this study finds that the network does not learn well if it has not been trained sufficiently (e.g., less than 1000 epochs). This fact is particularly important in recurrent network training. One observes from the experiments that if the recurrent network is not trained sufficiently, it produces a constant estimated output both in-sample and out-of-sample. In the following comparison, results from different training epochs are reported for each network configuration. However, learning in-sample patterns well does not necessarily guarantee a good performance on out-of-sample generalization. As such, in further investigation herein, an early stopping rule is applied. The network training is stopped when the error of prediction on the test set starts increasing or the training time reaches a preset number of epochs. In the following sections, the performance of network is reported along with the training times of 1000, 5000, and 10000 epochs to illustrate the effects of under- and over-training on the network performance.

6.4.3. Two-stage ANN Estimation of the Klein Model I

At each stage of estimation (namely instrumental stage and final stage), the selected network is trained in 30 runs, each lasting from 1000 to 5000 epochs. Each run is started with a different initial set of network weights generated by a different random number seed. At each stage of estimation, minimum and maximum errors are recorded. This results in two streams of data representing instrumental estimations, one with maximum error and the other with minimum error. These streams of instrumental estimations are used in final estimation of system equations. The GA selects a network configuration of 9-7L-6T for the instrumental stage and 6-6L-3T for the final stage. Table 6.3 reports the performance of network with minimum/maximum error at the instrumental stage and the minimum/maximum error at the final stage. The rationale of this recording is to evaluate the case where the network is trained just one time at each stage, what would be the boundary of errors taking into account the best and the worst estimations from 30 runs.

Results in Table 6.3 show that the performance of the ANN is superior to those of traditional methods (Klein, 1950; SAS, 1984) and Caporaletti et al. (1994). The comparison is based on SSE in the estimation of each exogenous variable as well as total SSE in the estimation of the whole system at different training times.

At 1000 training epochs, the Total SSE of the two-stage ANN estimation with maximum error at instrumental stage and maximum error at final stage is 33.87866 or about 50% of the estimation errors of traditional methods. This error measurement is about the same as the one reported in Caporaletti et al.. If the final estimation is obtained with the minimum error at the instrumental stage and minimum error at the final stage, the Total SSE is 18.72112 or about 25% of the estimation errors reported in traditional methods. This error is about 50% of the estimation error reported in Caporaletti et al..

At 5000 training epochs, the Total SSE of the two-stage ANN estimation with maximum error at instrumental stage and maximum error at final stage is 18.046797 or about 25% of the estimation errors of traditional methods, and 50% of the one reported in Caporaletti et al.. If the final estimation is obtained with the minimum error at the instrumental stage and minimum error at the final stage, the Total SSE is 10.774157 or about 20 % of the estimation errors reported in traditional methods, or about 30% of the one reported in Caporaletti et al..

At 10000 training epochs, the Total SSE of of the two-stage ANN estimation with maximum error at instrumental stage and maximum error at final stage is 14.725791 or about 20% of the estimation errors of traditional methods, and 40% of the one reported in Caporaletti et al.. If the final estimation is obtained with the minimum error at the instrumental stage and minimum error at the final stage, the Total SSE is 7.450847 or about 10 % of the estimation errors reported in traditional methods, or 20% of the one reported in Caporaletti et al..

Caporaletti et al. (1994) use an ANN to estimate each endogenous variable of the system, one by one. This single equation estimation approach may not capture well the simultaneous and contemporaneous effects of other endogenous variables in the system as the system approach used in this study.

6.4.4. Modular ANN Estimation of the Klein Model I

Although the two-stage ANN configuration reported in the previous section performs better than traditional estimations and single equation estimations with ANN, it cannot handle a mix of temporal and non-temporal variables existing in many business contexts represented by a simultaneous equation system. Forecasting a mix of variables with a system of standard ANNs is not effective and efficient as one has to rely on predetermined lagged values and/or the values of other variables estimated outside the system to feed in as input. The appropriate network configuration for a mix of variables should have separate modules in the instrumental stage to deal with different aspects of the data.

Using Mixture-of-experts Network architecture, the modular network configuration in this study has two modules: a recurrent module and a standard one. The recurrent module is refined to learn the lagged effect on related temporal variables. The standard module learns the inter-relationship of other variables in the system. Then instrumental estimations from these two modules are processed in the final stage with the mapping of instrumental estimations to desired targets of the system. This mapping accounts for the contemporaneous and simultaneous effects on the final estimation of the endogenous variables.

Specifically, in the instrumental estimation stage, GA selects the configuration of 9-6L-3T for the network of recurrent module and 9-7L-3T for the network of standard module. Each network is trained in 30 runs, each run lasts from 1000 to 10000 epochs. Each run is initiated with a different set of network weights generated randomly. In each module, minimum and maximum errors of estimation are recorded. It results in two streams of data representing instrumental estimations with maximum/minimum errors. These streams of instrumental estimation are used in the final estimation of system equations. GA selects the configuration of 6-6T-3T for the network in the final stage. The final network is trained in 30 runs, each with different initial random weights. Each run lasts from 1000 to 10000 epochs. Results from modular ANN estimation are reported in Table 6.4.

At 1000 training periods, the Total SSE of modular ANN estimation with maximum error at instrumental stage and maximum error at final stage is 17.79388. If the final estimation is obtained with the minimum error at the instrumental stage and minimum error at the final stage, the Total SSE is 6.49236 . These error measurements are about 50 % of errors from the two-stage ANN estimations reported in the previous section.

At 5000 training periods, the Total SSE of modular ANN estimation with maximum error at instrumental stage and maximum error at final stage is 11.79388 versus 18.046797 resulted from the two-stage ANN estimations. If the final estimation is obtained with the minimum error at the instrumental stage and minimum error at the final stage, the Total SSE is 2.728608 versus 10.774157 resulted from the two-stage ANN estimations.

At 10000 training periods, the Total SSE of modular ANN estimation with maximum error at instrumental stage and maximum error at final stage is 10.178641 versus 14.725791 resulted from the two-stage ANN estimations. If the final estimation is obtained with the minimum error at the instrumental stage and minimum error at the final stage, the Total SSE is 1.414566 versus 7.450847 resulted from the two-stage ANN estimations.

In all cases, the results obtained from modular ANN estimations are superior than those of two-stage ANN and traditional methods reported in the previous section. The reason for this improvement is that the temporal effect of lagged endogenous variables on the system is taken into account explicitly in modular estimation.

6.4.5. Modular ANN Forecasting of the Klein Model I

This study uses the variables defined in the Klein Model I to forecast the related endogenous variables for the period from 1950 to 1994. As the US economy grows dramatically, the level of macroeconomic variables in this period increases accordingly. For instance, taking all measurements in 1992 constant dollars, the national income of the US grows from 41.98 billions in 1950 to 6,086.60 billions in 1994. Similarly, the consumption grows from 35.22 billions in 1950 to 4,957.80 billions in 1994 (U.S. Department of Commerce, 1998). It would be difficult for a network to deal with variables whose values increase to an unbound limit and spacing with big gaps. As an alternative, this study considers a more compact space and focuses on the growth rate of related endogenous variables. Consequently, related data in the period are transformed into first differences of their natural logarithmic values to capture their growth rates.

The dynamics of an economic system can be studied either with a rolling window scheme or with a moving window scheme. In a rolling window scheme, the time frame extends (walks forward) from the origin further to the future to capture in formation of the new data. In a moving window scheme, a time window slides forward and the older data are discarded. In time series forecasting, the information of older data may not be useful, particularly in the context of growing US economy. Consequently, the ANN estimation and forecasting studied herein is conducted on a moving window scheme for the yearly growth of related US economic variables from 1950 to 1994.

In the experiments, for each time window , 20 yearly periods are used for training, the next 5 for testing, and the subsequent 5 for validation or forecasting (Figure 6.3). Consequently, a modular network is trained with in-sample information for 20 periods and tested on 5 out-of-sample periods. GA is used to select the appropriate configuration for each module. Each modular network is trained in 30 runs. Each run starts with a different initial set of random weights. The training stops when errors of prediction on the test set starts increasing or the training time reached 5000 epochs. This early stopping rule is applied to ensure an accurate generalization of ANN on future cases. The best network from 30 runs is used to make forecasts for the next 5 out-of-sample periods of the scheme.

SSE of ANN estimation are reported in Table 6.5. The estimated and actual values of related endogenous variables in each time window are illustrated in Graphs 6.1 to 6.12. In the following analysis, the growth rates of consumption, private wages, and net investment are indicated as *DLC*, *DLW*p, and *DLI*, respectively. Their final estimates result from modular network estimations are indicated as *DLC***, *DLW*p**, and *DLI***, respectively. Following Klein (1950) and Klein and Goldberger (1955) who concentrate their analyses on the sign of the forecast residual, the current analysis focuses on the ability of the ANN to pick up the future direction of related variables. The following provides descriptions on the behavior of ANN in learning and forecasting. The experiments intend to observe how the network behaves when it forecasts unseen patterns in the fluctuation of economic variables. There is no benchmarking and significance tests on the accuracy of ANN forecasting in these experiments.

- *Period from 1950 to 1979*

In the period from 1950 to 1979, data from 1950-1969 are used for training, 1970-1974 for testing, and 1975-1978 for forecasting. GA selects a network configuration of 9-4L-3T for the recurrent module, 9-5L-3T for the standard module, and 6-4L-3T for the final stage.

For the growth rate of consumption *DLC*, the network learns well the data patterns in the training period as it captures correctly the changes in direction of the variable. The SSE for this training period is .000331 on the estimation of the growth of this variable. However, in the test/forecasting period, the network projects a slight fluctuation at a lower level when the related variable started fluctuating in an upward trend (Graph 6.1). The SSEs for the testing and forecasting periods are .001794 and .008602, respectively. The network does not experience these high growth levels in an upward trend. Consequently, it produces moderate forecasts at an average level.

For the growth rate of private wages *DLWp*, during the training period, the network learns well the data patterns and follows closely the changes in direction of the variable with a SSE of .000832. In the test and forecast periods, the network picks up the changes in direction with SSEs of .001821 and .001209, respectively. One notes that as the network learns the large fluctuation patterns in the training set and it is able to forecast a moderate level while following the future directions of the data (Graph 6.2).

For the growth rate of investment *DLI*, network forecasting picks up the changes in direction of the variable as it already learned the fluctuated patterns of the data. The SSE for the training and test periods are .017397 and .09676, respectively. However, when the variable fluctuates at a wide levels (in 1974-75), the network has not experienced this new pattern to make a close prediction. Therefore. it produces forecasts at moderate levels. The SSE for this forecast period is .267968 (Graph 6.3).

- *Period from 1955 to 1984*

In the period from 1955 to 1984, data from 1955-1974 are used for training, 1975-1979 for testing, and 1980-1984 for forecasting. GA selects a network configuration of 9-6T-3T for the recurrent module, 9-6T-3T for the standard module, and 6-7T-3Tfor the final stage.

For *DLC*, the network learns well the upward trend in the training set by following correctly the changes in direction of the variable with a SSE of .000682. It is able to pick up the patterns in the test period with a SSE of .000751. When the future data (1980-84) fluctuated in a new, downward pattern, the network produced a dampening forecast at a moderate level (Graph 6.4). As the network has not experienced this pattern, it produces a SSE of .004883 for the forecast period.

For *DLW*p, after learning well the upward trend in the training set with a SSE of .001172, the network forecasts a slight fluctuation at moderate level when the future data (1980-84) fluctuate in a new pattern (Graph 6.5). SSE for the test and forecasting periods are .002552 and .010367, respectively.

For *DLI*, after learning well the fluctuation in the training set with a SSE of .008684, ANN forecasts follow the future data pattern. However the network had not experienced the large changes in the levels of the extreme variation in the forecast period (e.g., the changes in 1982 to 1984). Consequently, when future data start fluctuating widely (1980-84), the network produced forecasts at moderate levels (Graph 6.6). SSE for the test and forecasting periods are .074401 and .758536, respectively. The large error in the forecast period are due to the large changes in levels of data that the network is unable to capture.

- *Period from 1960 to 1989*

In the period from 1960 to 1989, data from 1960-1979 are used for training, 1980-1984 for testing, and 1985-1989 for forecasting. GA selects a network configuration of 9-5T-3L for the recurrent module, 9-6T-3T for the standard module, and 9-3L-3T for the final stage.

For *DLC*, the network learns well the upward pattern of the training set with a SSE of .000752. When future data starts a downward trend (1985-1989), the network had not experienced the large changes in levels to produce closer forecasts. As a result, network forecasts follows the future directions at higher levels (Graph 6.7). SSE for the test and forecast periods are .006123 and .012539, respectively.

For *DLWp*, the network learns well the upward pattern of the training set with a SSE of .001325. When future data starts a downward trend, ANN forecasts follow the trend but at higher levels (Graph 6.8). Similar to learning and forecasting DLC in this period, the network has not learned the large change in levels of future fluctuation in order to provide closer forecasts. SSE for the test and forecast periods are .00773 and .024928, respectively.

For *DLI*, the network learns well the fluctuation in the training set with a SSE of .019973. When future data start dropping (1985) and then fluctuating at a lower level (1985-89), the network forecasts follow the trend but at a higher level (Graph 6.9). SSE for the test and forecast periods are .082321 and .727144. The large error for the forecast period is due to the change in data patterns as they fluctuate at a moderate levels that the network is unable to follow closely.

- *Period from 1965 to 1994*

In the period from 1965 to 1994, data from 1965-1984 are used for training, 1985-1989 for testing, and 1990-1994 for forecasting. GA selects a network configuration of 9-5L-3T for the recurrent module, 9-3L-3Tfor the standard module, and 9-5L-3T for the final stage.

For *DLC*, the network learns well the upward trend in the training set with a SSE of .001085. When future data start a long downward trend, the network has not learned these patterns in order to predict accurately the future level and in some occasions changes of directions, e.g., in 1990 (Graph 6.10). SSE for the test and forecast periods are .002213 and .014209, respectively.

For *DLW*p, the network learns well the upward trend in the training set with a SSE of .001471. When future data has a downward trend, the network does not predict accurately the level and change of directions from the pattern that it has learned (Graph 6.11). SSE for the test and forecast periods are .002426 and .012759, respectively.

For *DLI*, the network learns well the fluctuation in the training set. Since the training set contains patterns of large fluctuations, the network forecast is able to follow the trend of future data, however in a large fluctuating pattern that it has learned (Graph 6.12). SSE for the test and forecast periods are .136035 and .164281, respectively. The large error from these periods is due to the sudden change in levels of fluctuation that the network is unable to capture.

6.4.6. Concluding Remarks

From extensive experiments with GA and ANN on the Klein Model I, and with the caution on problem specific, the following remarks on the efficiency and effectiveness of the proposed system are in order.

- GAs have selected the learning parameters and network configuration efficiently from a pool of candidates. The selected network designs always produce smallest errors between network output and target values.
- ANNs constructed in mixture-of experts network architecture provide effective alternatives to handle efficiently a mix of temporal and non-temporal variables. This architecture offers flexible alternatives to study patterns of the problem space. One can use hierarchical networks to conduct instrumental estimations. One can also partition the problem space into domains and assign them to modular ANN to learn the related patterns.

From the observations of ANN behavior in learning patterns of the training set and producing forecasts on unseen data, the following remarks are related to the performance of the proposed ANN system in generalization, and to the improvement of its forecast ability.

- In contrast to what is stated in literature, the network does not learn and project the extreme and recent trend. It tends to provide a moderate forecast in terms of fluctuation and level.

- If the network has been trained with data having an upward trend and related variable to be predicted fluctuates in a downward trend, network forecasts will be dampened at a middle level.
- If the network has not experienced drastic level changes in the training set, it produces a forecast following the trend but at a higher level for future downward change and lower level for future upward change.
- If the network is trained with the fluctuated pattern, its forecasts follow the future trend but at a moderate change in level. The larger the variation in the training set, the closer the ANN will follow the patterns in the forecast period in terms of directions and levels.
- The network cannot predict accurately a level outside the range of pattern it has learned from the training set. When it encounters such a case, it produces a forecast at an average level of the training set.

Consequently, in order to improve its forecasting ability, a network should experience with the variation in trend (upward/downward, long/short fluctuation) and the possible highest and lowest levels of data patterns. The experiments conducted herein illustrate that the more variations exist in the training set, the closer ANN follows with future fluctuations in terms of directions and levels.

Figure 6.1

Two-stage ANN Estimation of Klein Model I

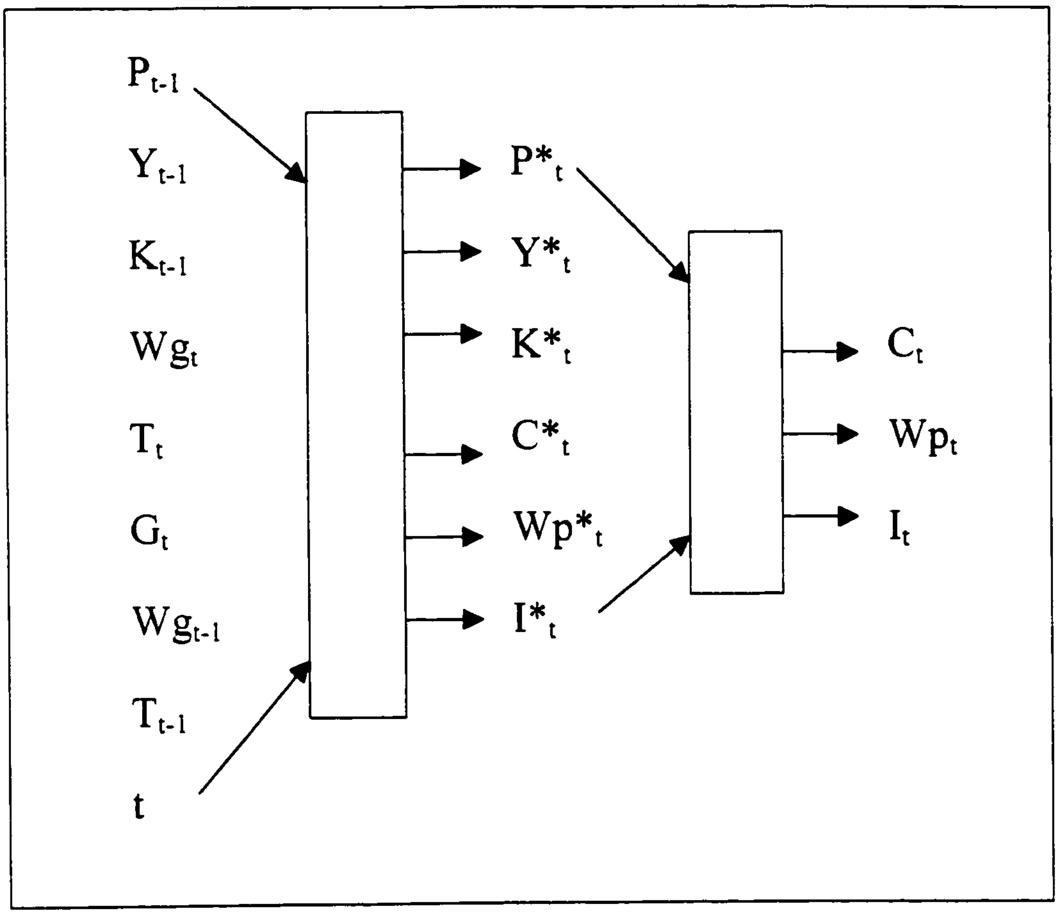


Figure 6.2

Modular ANN Estimation of Klein Model I

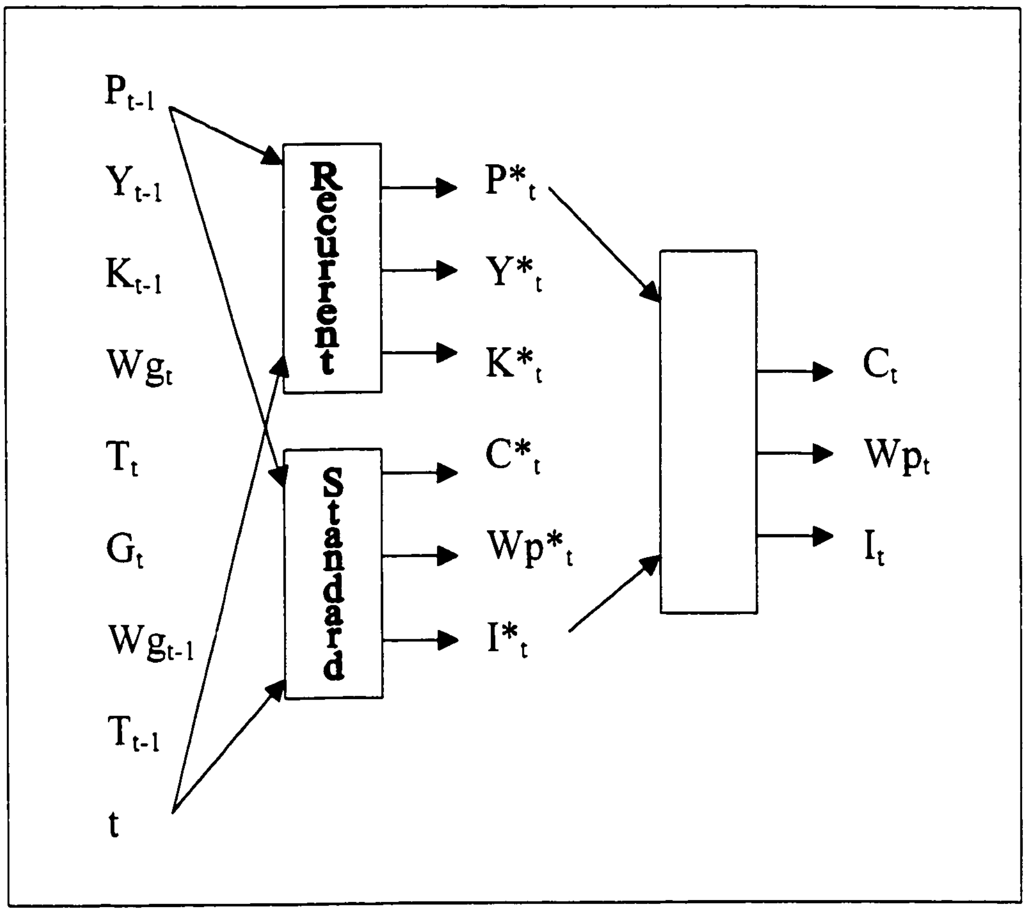


Figure 6.3

Moving Window Scheme
for Estimation and Forecasting

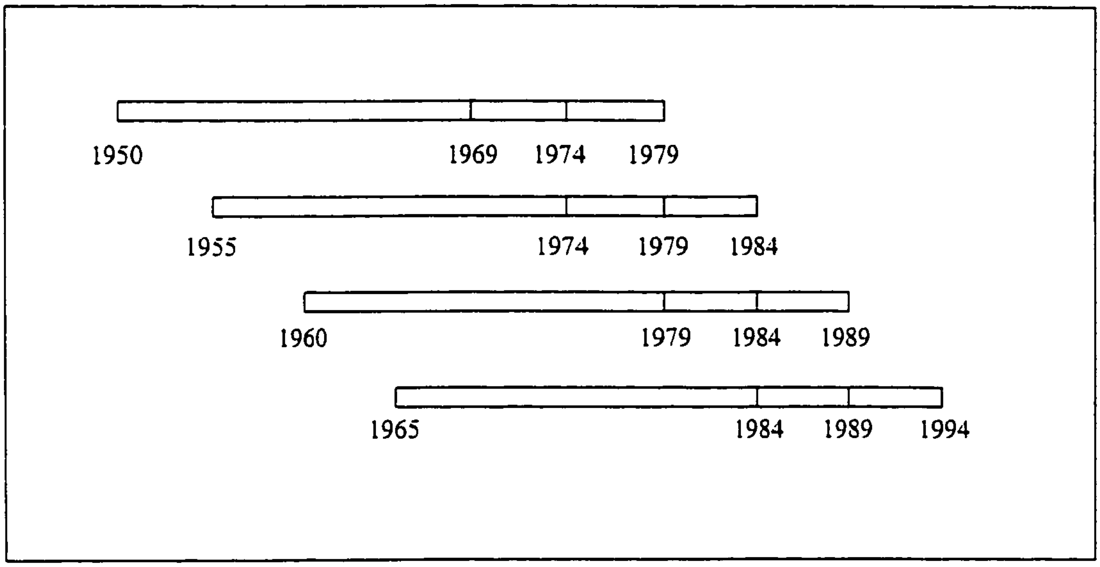


Table 6.1

MSE of in-sample Learning Across Learning Parameters

Momentum	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Learning Rate											
0.1	5.4643	3.393	4.2047	5.0834	3.0245	**2.1962**	4.6336	4.8184	3.9205	8.142	32.4135
0.2	3.014	4.3893	3.2444	3.2789	3.6814	3.9368	5.0605	3.4445	7.7877	12.6784	36.5673
0.3	2.7792	4.913	5.2576	6.8616	4.5582	5.5848	4.5373	7.4952	13.2769	25.309	58.3624
0.4	5.071	6.5417	3.8282	8.0942	6.3748	5.7017	6.8542	9.4965	15.4669	28.0985	49.2605
0.5	5.1242	6.5626	7.5671	5.7506	7.0973	7.1633	9.4145	13.2206	18.6392	35.8238	48.7229
0.6	5.0847	4.0318	6.7895	7.6182	9.1734	6.9509	11.6573	13.677	20.9549	38.2919	83.2423
0.7	4.9103	8.6242	6.7211	8.8542	7.3537	8.8466	12.6237	16.88	21.1843	67.58	54.5003
0.8	8.1558	7.1454	8.6966	10.5346	9.4676	11.0363	13.0733	23.2007	29.7818	55.0628	61.1305
0.9	6.1239	9.4459	10.8051	8.8779	10.4027	13.2878	15.8474	21.5042	30.2498	51.4673	90.2918
1	6.8712	8.9805	10.7231	11.9032	14.9586	15.7748	21.033	23.8109	32.8098	74.0701	133.3779

Table 6.2

Sensitivity Analysis on GA Selection:
MSE of Standard Module
in Instrumental Stage of Two-stage ANN Estimation

ANN Architecture	MSE
9-8L-6T	2.6744
9-8L-6L	6.8746
9-8T-6T	2.8999
9-8T-6L	5.3162
9-7L-6T	**2.1962**
9-7L-6L	6.7239
9-7T-6T	2.5495
9-7T-6L	4.8737
9-6L-6T	3.8810
9-6L-6L	8.0856
9-6T-6T	5.2359
9-6T-6L	6.0632

Legend:

L : Sigmoid logistic transfer function
T : hyperbolic tangent transfer function
MSE : Average MSE of 30 runs, 5000 epochs each

Table 6.3

Results of Two-Stage ANN Estimation
of the Klein Model I

	Total SSE	SSE C	SSE W_p	SSE I
2SLS*	60.97260	21.92525	10.000496	29.04686
LIML*	85.90255	40.88414	10.021920	34.99649
3SLS*	73.60150	18.72696	10.920560	43.95398
ML**	56.26009	22.08910	10.218495	23.95249
CapoANN***	32.4987	9.5356	9.8813	13.0813
1000 training epochs				
Min**** - Min*****	18.72112	6.805794	12.80344	2.431763
Min**** - Max*****	28.53662	9.140769	16.64345	2.752003
Max**** - Min*****	23.33051	4.796138	13.92080	4.613574
Max**** - Max*****	33.87866	8.404020	20.55768	4.916960
5000 training epochs				
Min**** - Min*****	10.774157	3.511433	5.884045	1.378679
Min**** - Max*****	15.164985	4.690529	8.620608	1.853848
Max**** - Min*****	15.770872	2.684769	9.96278	3.123323
Max**** - Max*****	18.046797	4.353278	10.19043	3.503089
10000 training epochs				
Min**** - Min*****	7.450847	2.212344	3.946774	1.291729
Min**** - Max*****	10.129417	3.218656	5.394284	1.516477
Max**** - Min*****	12.985553	2.209983	7.958855	2.816715
Max**** - Max*****	14.725791	3.582101	7.809497	3.334193

Legend:

* SAS (1984) SAS/ETS User's Guide, Version 5
** Klein (1950)
*** Caporaletti et al. (1995)
**** Max/Min error on Instrumental ANN Estimation in 30 runs
***** Max/Min error on Final Stage ANN Estimation in 30 runs

Table 6.4

Results of Modular ANN Estimation
of the Klein Model I

	Total SSE	SSE C	SSE W_p	SSE I
2SLS*	60.97260	21.92525	10.000496	29.04686
LIML*	85.90255	40.88414	10.021920	34.99649
3SLS*	73.60150	18.72696	10.920560	43.95398
ML**	56.26009	22.08910	10.218495	23.95249
CapoANN***	32.4987	9.5356	9.8813	13.0813
1000 training epochs				
Min**** - Min*****	6.49236	2.548467	3.376657	.567239
Min**** - Max*****	10.71094	3.150886	6.57862	.963706
Max**** - Min*****	13.81735	5.598224	6.402702	1.816427
Max**** - Max*****	17.79338	4.621934	10.73451	2.436934
5000 training epochs				
Min**** - Min*****	2.728608	1.714980	.710870	.302756
Min**** - Max*****	4.074969	1.781884	1.539434	.753651
Max**** - Min*****	3.735219	1.952168	1.048673	.734379
Max**** - Max*****	11.100670	2.792316	6.484070	1.824284
10000 training epochs				
Min**** - Min*****	1.414566	.653434	.459314	.301818
Min**** - Max*****	3.357525	1.613617	1.091524	.652384
Max**** - Min*****	2.686912	1.136092	.840595	.710231
Max**** - Max*****	10.178641	1.784696	7.326907	1.067041

Legend:
- * SAS (1984) SAS/ETS User's Guide, Version 5
- ** Klein (1950)
- *** Caporaletti et al. (1995)
- **** Max/Min error on Instrumental ANN Estimation in 30 runs
- ***** Max/Min error on Final Stage ANN Estimation in 30 runs

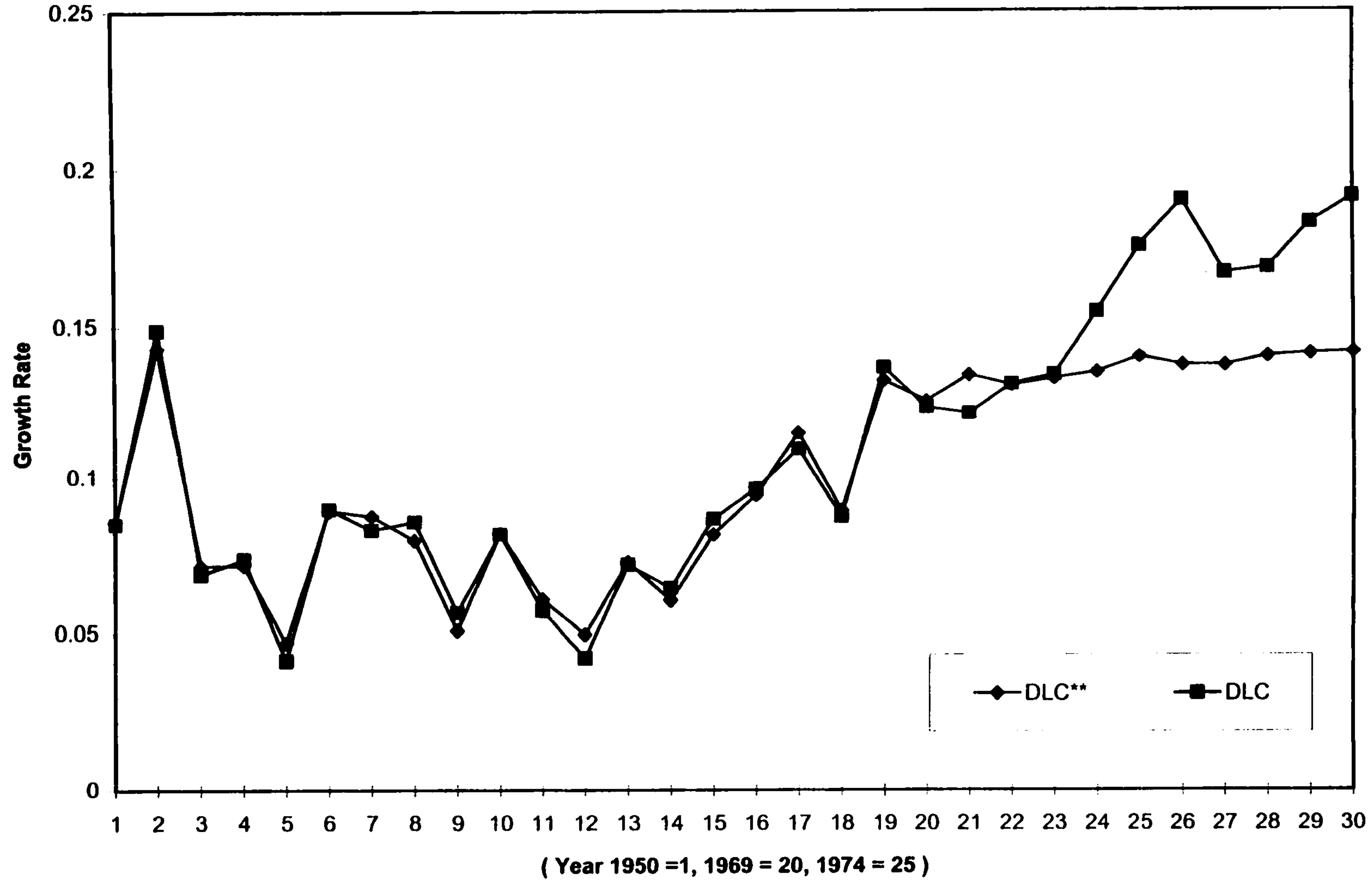
Table 6.5

Modular ANN Estimation and Forecasting (1950-94)
Using Klein's Framework

	DLC	SSE DLWp	DLI
Period of 1950-79:			
Training (1950-69)	.000331	.000832	.017397
Testing (1970-74)	.001794	.001821	.096760
Forecasting (1975-79)	.008602	.001209	.267968
Period of 1955-84:			
Training (1955-74)	.000682	.001172	.008684
Testing (1975-79)	.000751	.002552	.074401
Forecasting (1980-84)	.004883	.010367	.758536
Period of 1960-89:			
Training (1960-79)	.000752	.001325	.019973
Testing (1980-84)	.006123	.007730	.082321
Forecasting (1985-89)	.012539	.024928	.727144
Period of 1965-94:			
Training (1965-84)	.001085	.001471	.028825
Testing (1985-89)	.002213	.002426	.136035
Forecasting (1990-94)	.014209	.012759	.164281

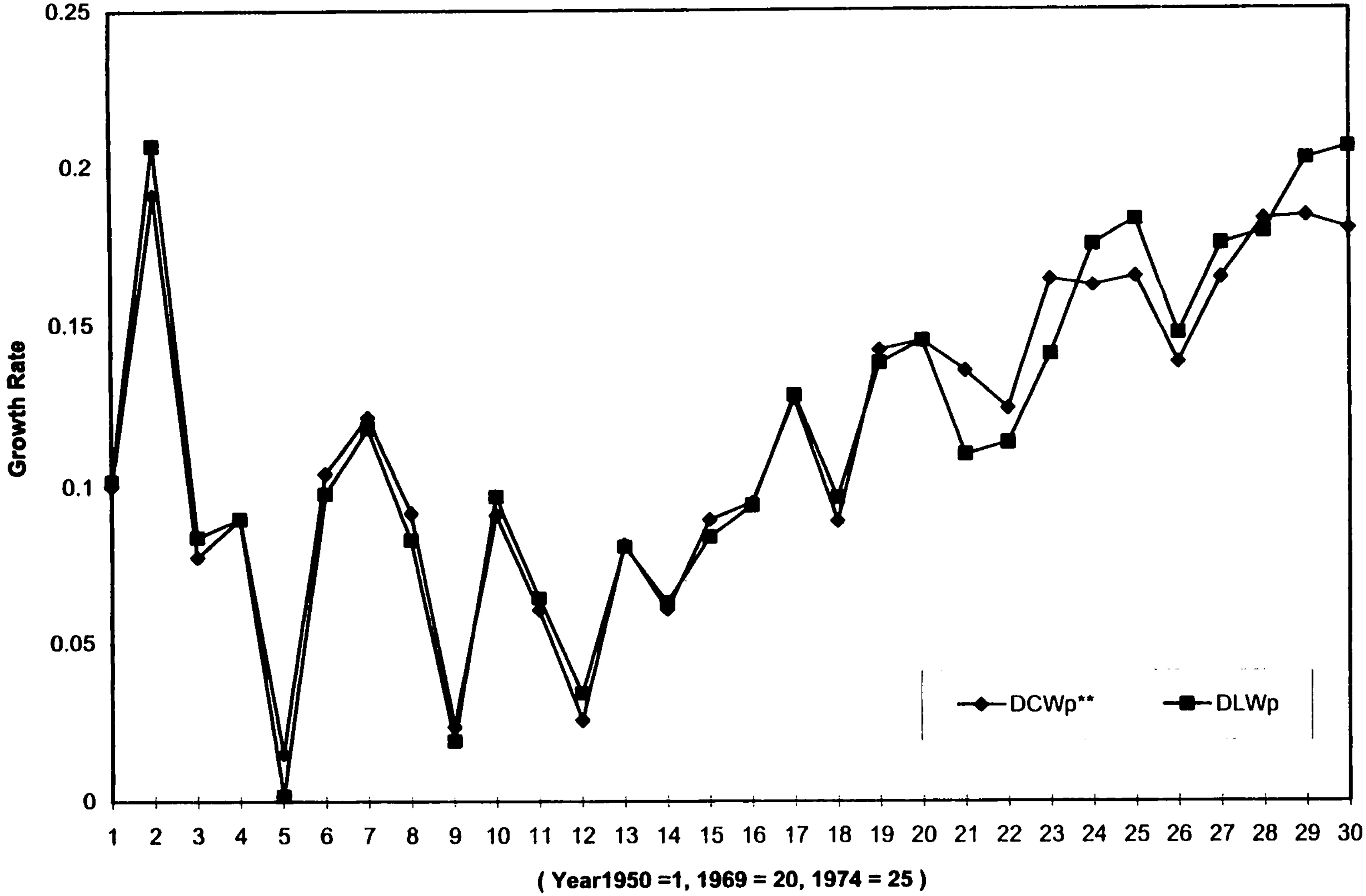
Graph 6.1

Estimation and Forecasting of DLC (1950-79)



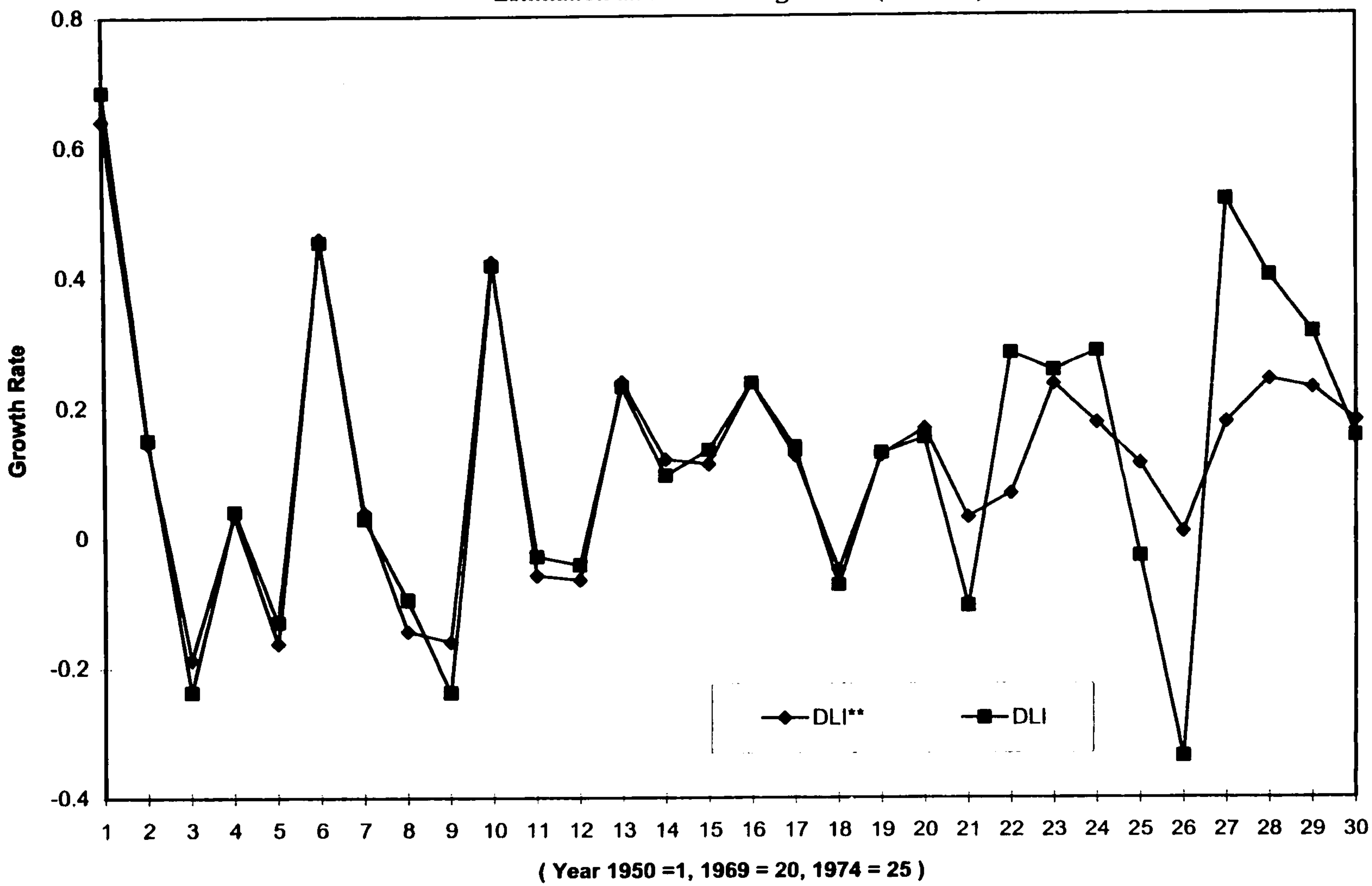
Graph 6.2

Estimation and Forecasting of DLWp (1950-79)



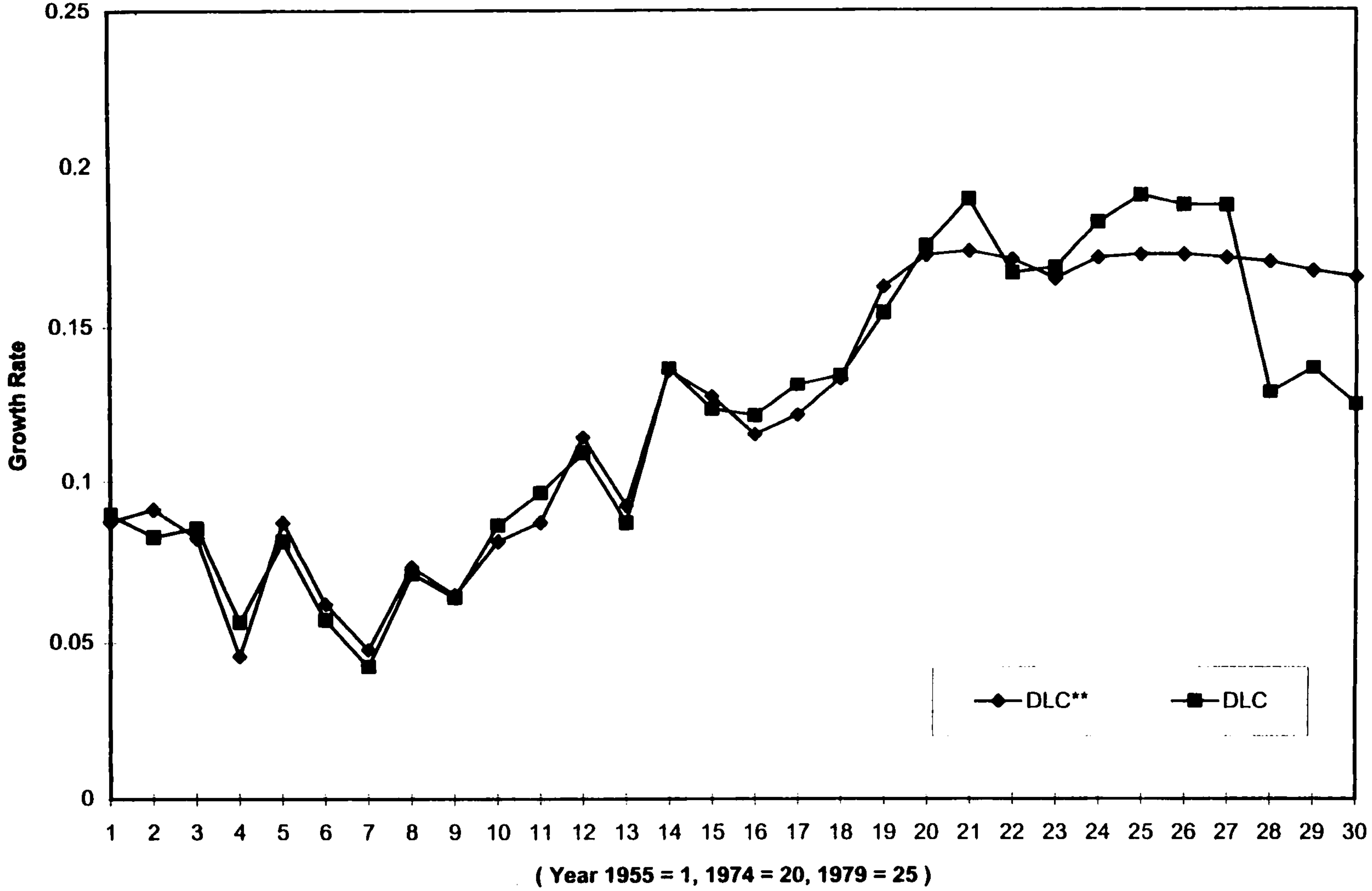
Graph 6.3

Estimation and Forecasting of DLI (1950-79)



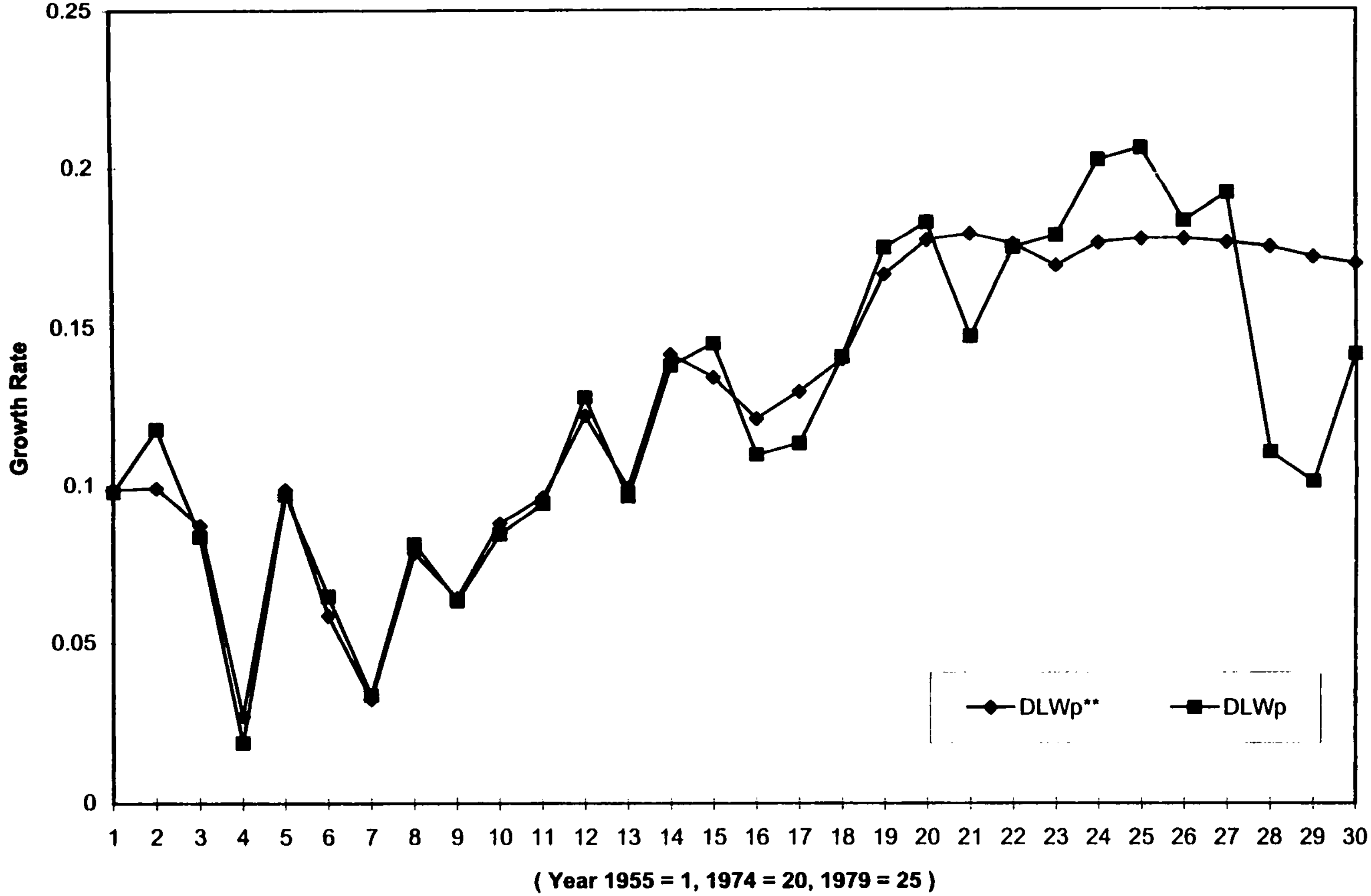
Graph 6.4

Estimation and Forecasting of DLC (1955-84)



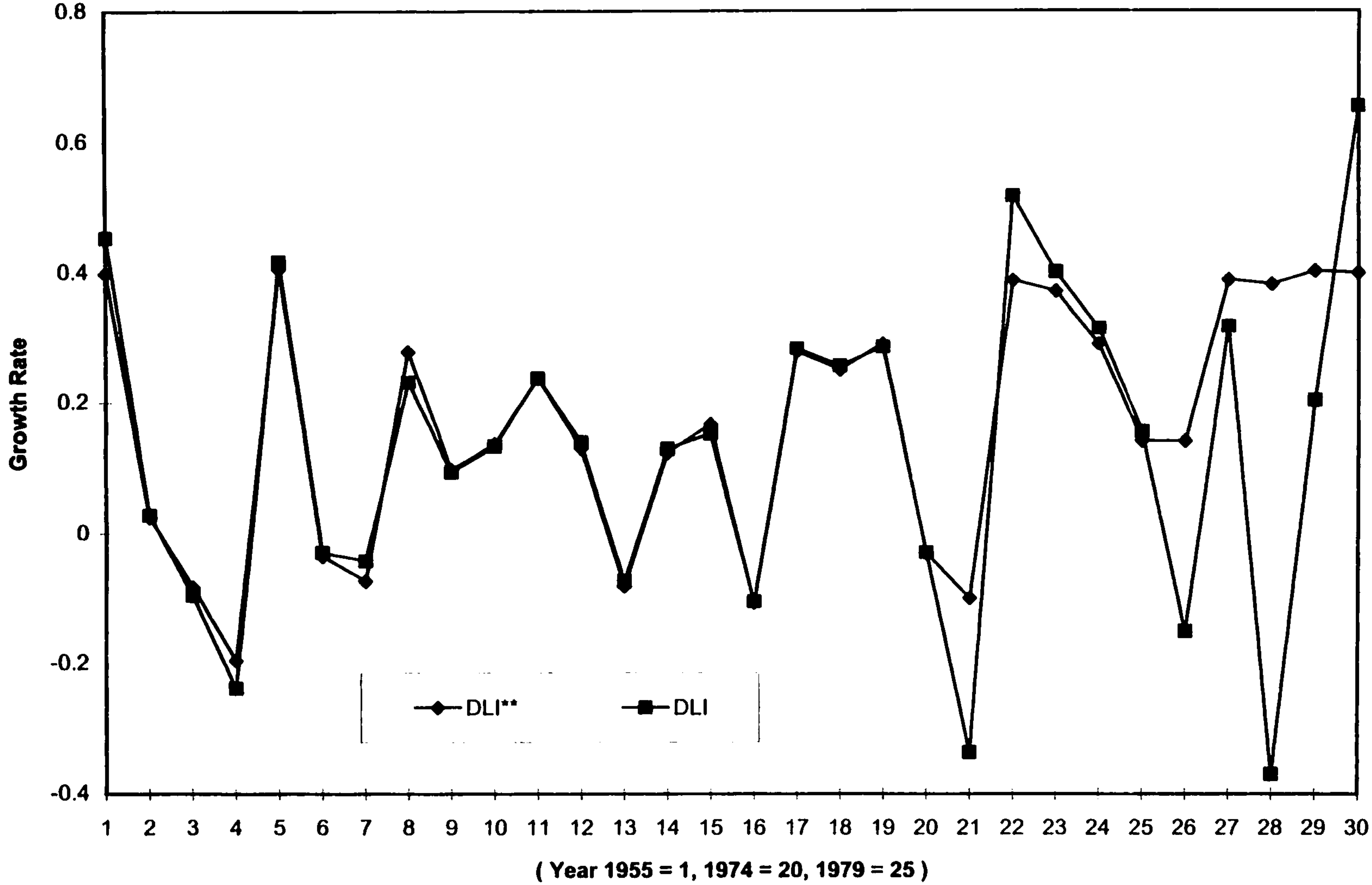
Graph 6.5

Estimation and Forecasting of DLWp (1955-84)



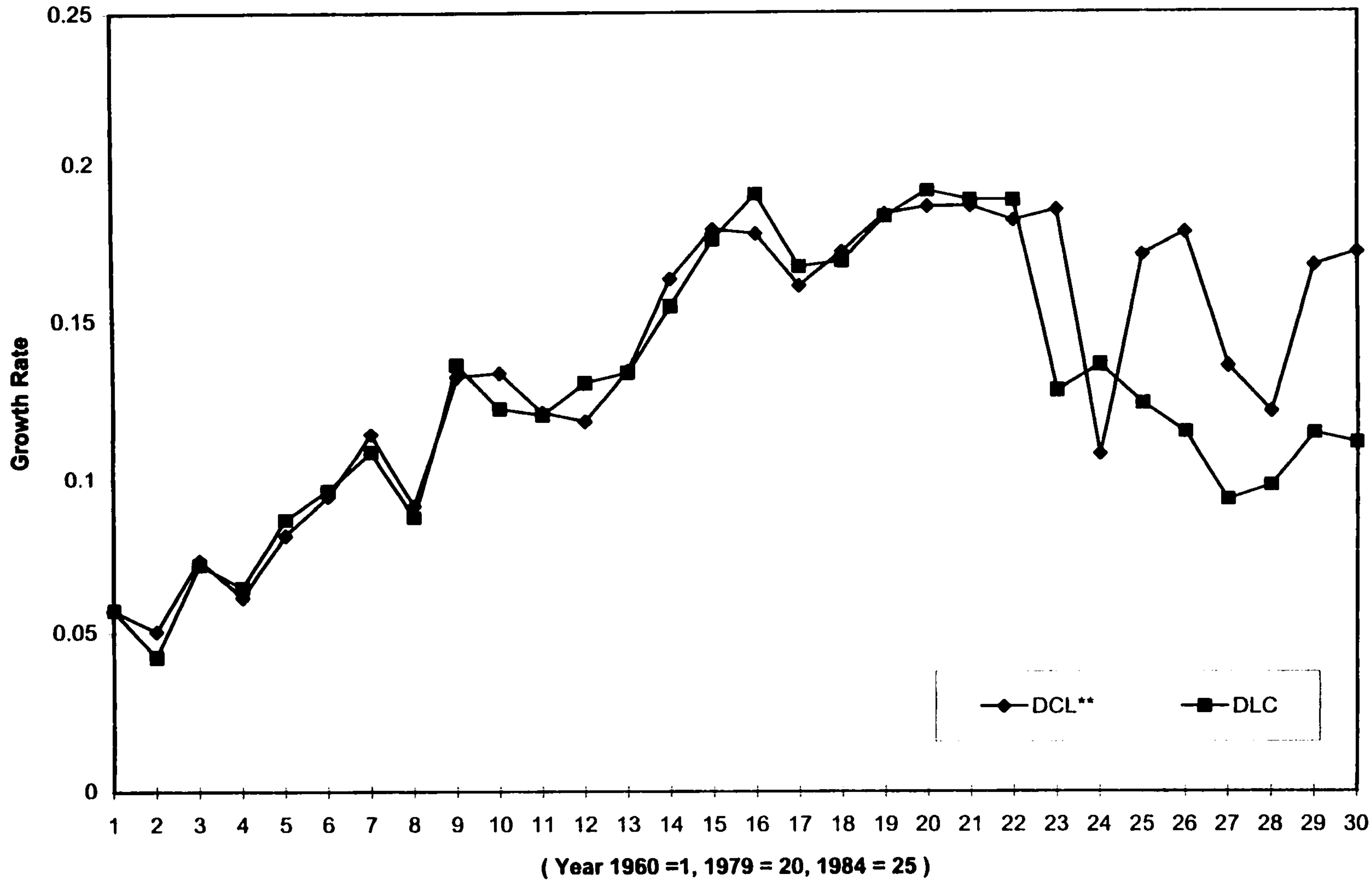
Graph 6.6

Estimation and Forecasting of DLI (1955-84)



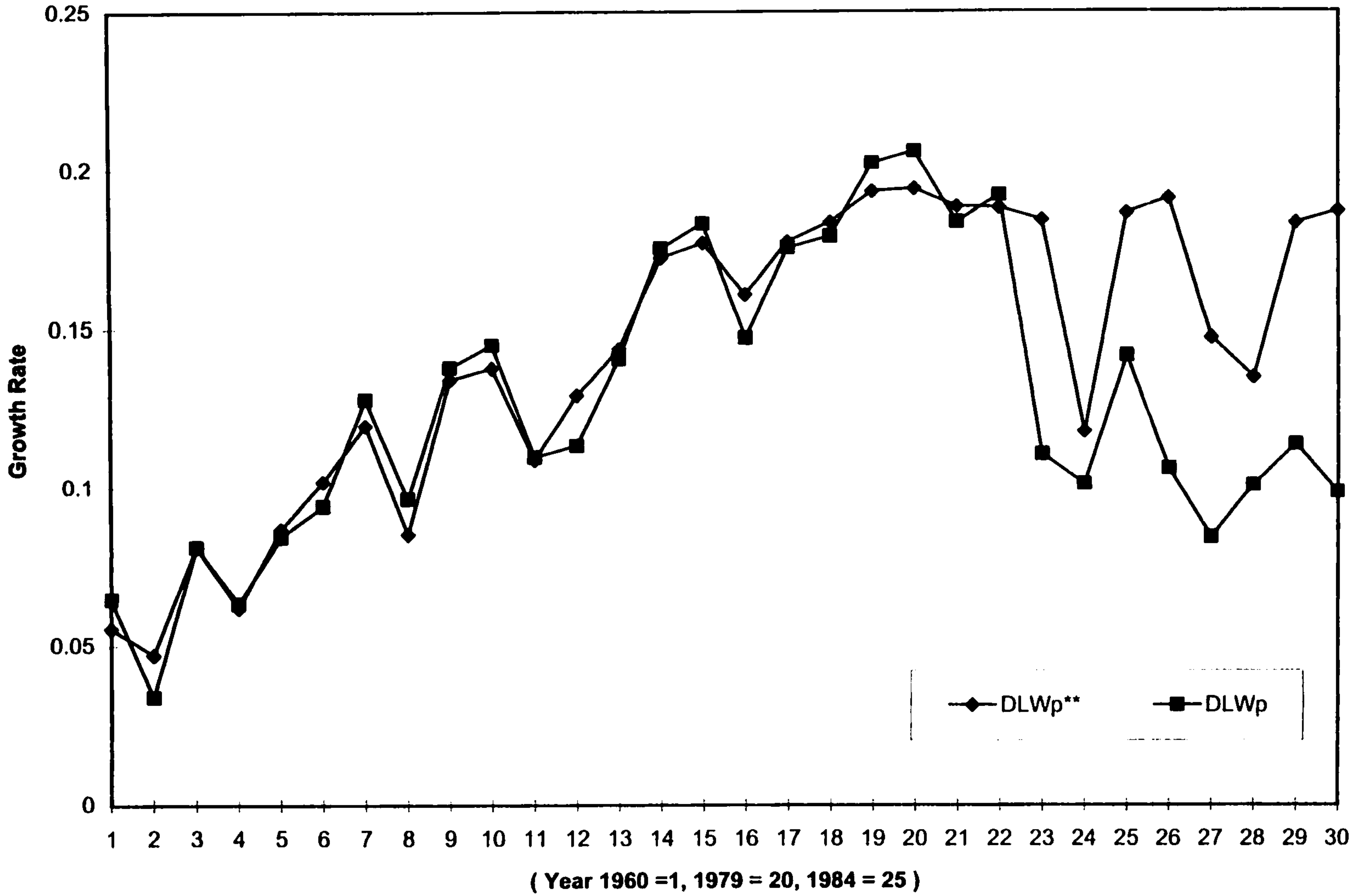
Graph 6.7

Estimation and Forecasting of DLC (1960-89)



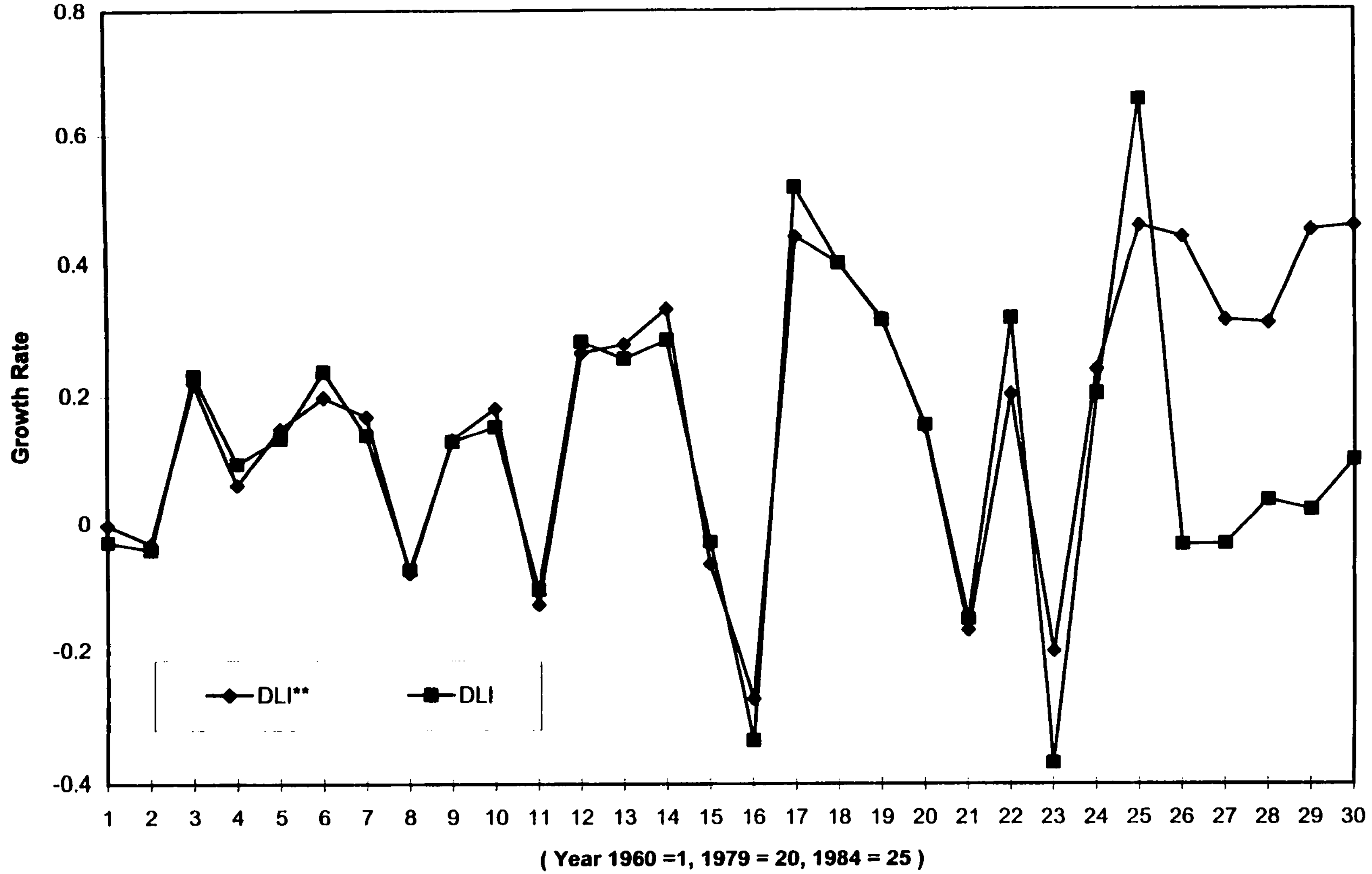
Graph 6.8

Estimation and Forecasting of DLWp (1960-89)



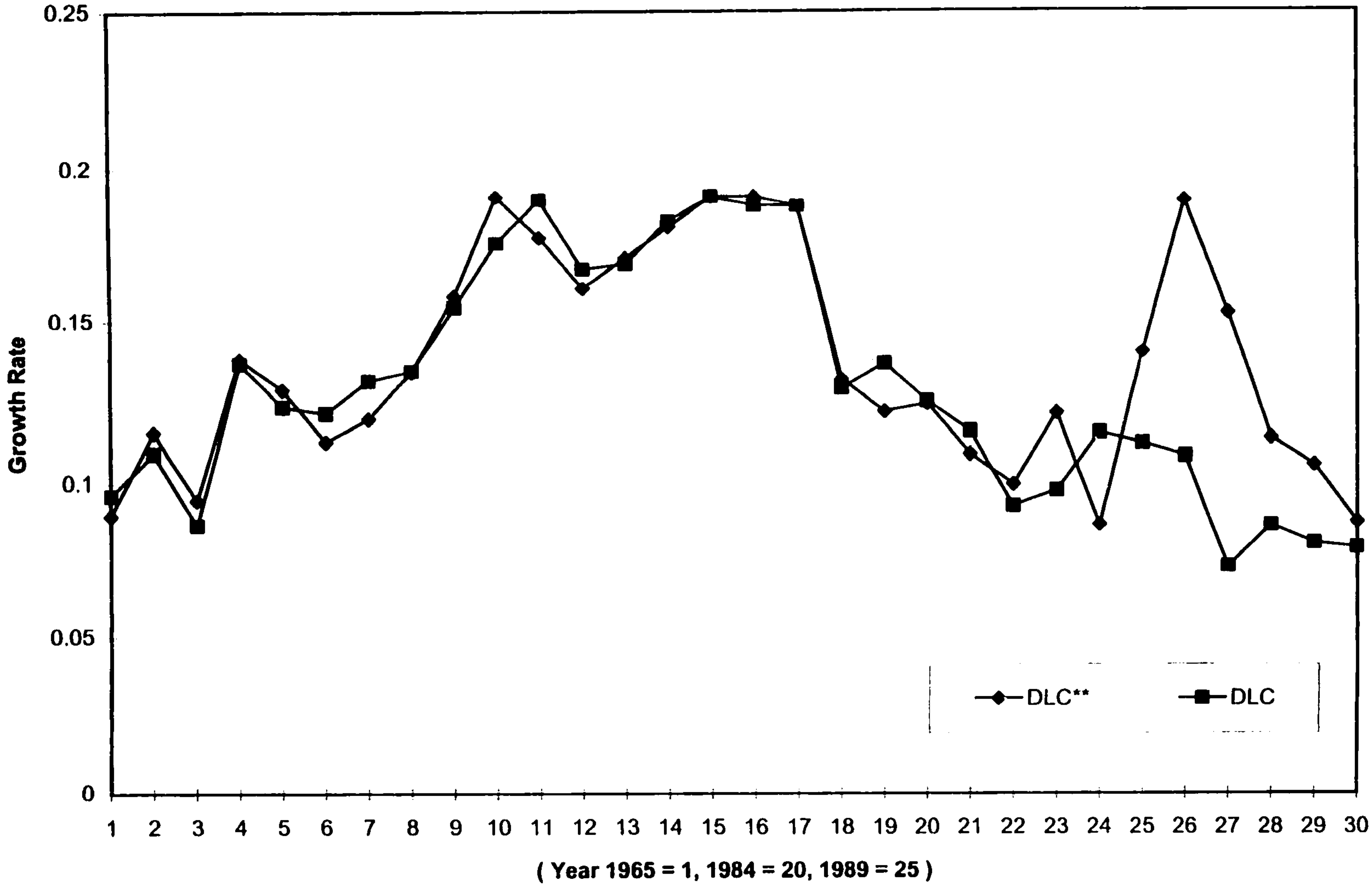
Graph 6.9

Estimation and Forecasting of DLI (1960-89)



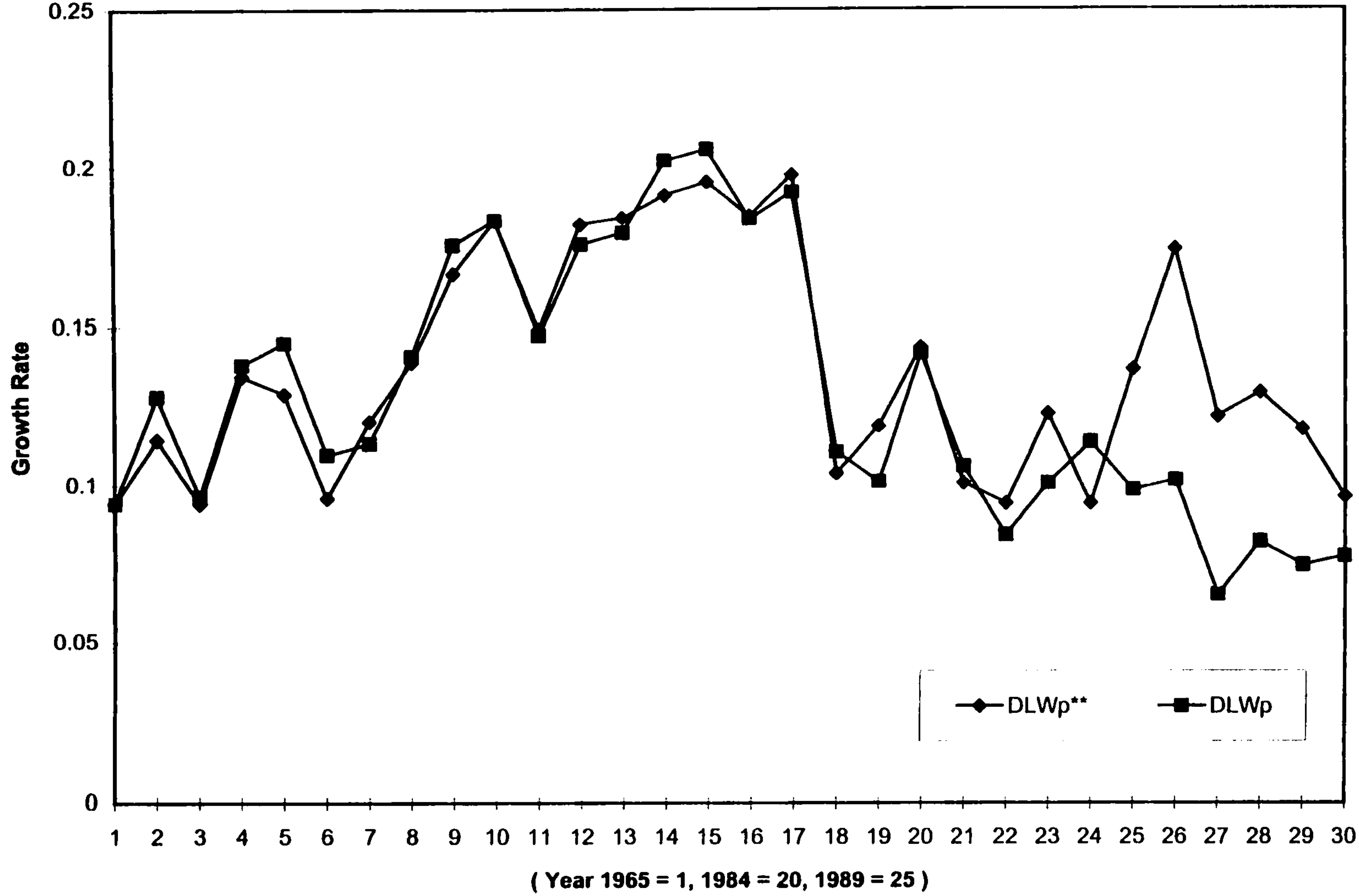
Graph 6.10

Estimation and Forecasting of DLC (1965-94)



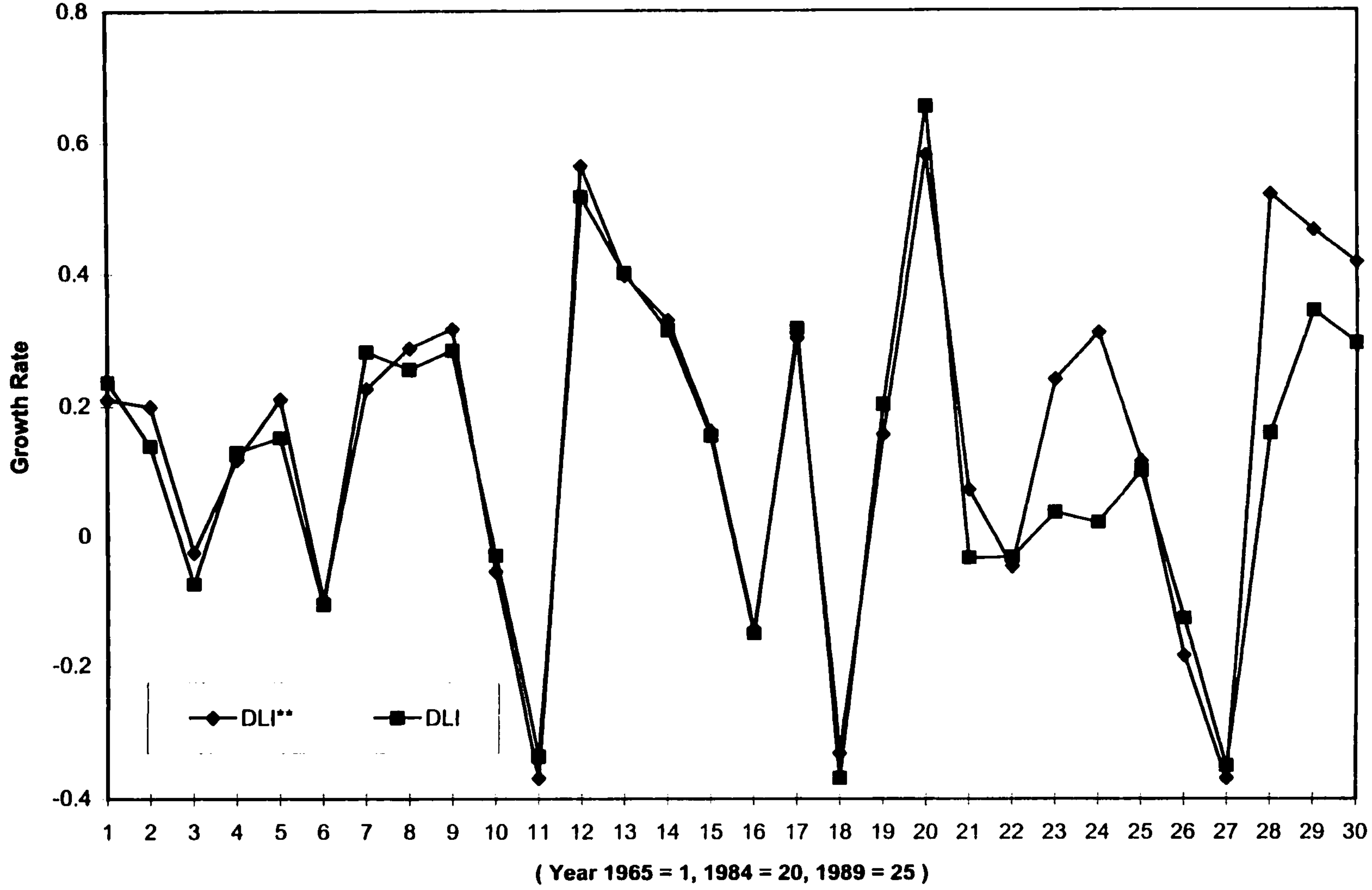
Graph 6.11

Estimation and Forecasting of DLWp (1965-94)



Graph 6.12

Estimation and Forecasting of DLI (1965-94)



CHAPTER 7

Estimation and Forecasting of Klein-Goldberger Model

7.1. KLEIN-GOLDBERGER MODEL REVISITED

Klein and Goldberger (1955) use their model to describe the US economy for the period 1929 to 1950 with the exception of the war years from 1942 to 1945. The model (denoted as the KG Model hereafter) is reformulated and re-estimated many times later to incorporate more recent data until 1954. In the original form, the model has 20 structural equations consisting of 15 behavioral and 5 definitional equations as follows:

(1) Consumption Equation:

$$C_t = \alpha_0 + \alpha_1 (Wp + Wg - Tw)_t + \alpha_2 (P - Sp - Tp)_t + \alpha_3 (A - Ta)_t + \alpha_4 C_{t-1} + \alpha_5 (L_1)_{t-1} + \alpha_6 (N_p)_t + u_{1,t} \tag{7.1}$$

where C = consumer expenditure in 1939 dollars

Wp = deflated private employee compensation

Wg = deflated government employee compensation

$(Wp + Wg - Tw)$ = deflated disposable employee compensation

P = deflated nonwage, nonfarm income

Sp = deflated corporate saving

$(P - Sp - Tp)$ = deflated disposable nonwage, nonfarm income

A = deflated farm income

$(A - Ta)$ = deflated disposable farm income

L_1 = deflated end-of-year liquid assets held by persons

N_p = number of persons in US

(2) Investment Equation

$$I_t = \beta_0 + \beta_1 (P + A + D - Tp - Ta)_t + \beta_2 (P + A + D - Tp - Ta)_{t-1} + \beta_3 (i_L)_{t-1}$$
$$+ \beta_4 K_{t-1} + \beta_5 (L_2)_{t-1} + u_{2,t} \tag{7.2}$$

where I = gross private domestic capital formation in 1939 dollars

D = capital consumption charges in 1939 dollars

i_L = average yield on corporate bonds

K = end-of-year stock of private capital in 1939 dollars

L_2 = deflated end-of-year liquid assets held by enterprises

(3) Corporate Saving Equation

$$(Sp)_t = \gamma_0 + \gamma_1 (Pc - Tc)_t + \gamma_2 (Pc - Tc - Sp)_{t-1} + \gamma_3 B_{t-1} + u_{3,t} \tag{7.3}$$

where Pc = deflated corporate profits

Tc = Deflated corporate income taxes

B = deflated end-of-year corporate surplus

(4) Relation between Corporate Profit and Nonwage Nonfarm Income

$$(Pc)_t = \delta_0 + \delta_1 P_t + \delta_2 P_{t-1} + u_{4,t} \tag{7.4}$$

This empirical, nonstructural relation is intended to maintain the completeness of the model.

(5) Depreciation Equation

$$D_t = \varepsilon_0 + \varepsilon_1 [(K_t + K_{t-1})/2] + \varepsilon_2 (Y + T + D - Wg)_t + u_{5,t} \tag{7.5}$$

where $Y + T + D - Wg$ = private Gross National Product in 1939 dollars

(6) Demand for Labor (Private Wages)

$$(Wp)_t = \zeta_0 + \zeta_1 (Y + T + D - Wg)_t + \zeta_2 (Y + T + D - Wg)_{t-1} + \zeta_3 t + u_{6,t} \qquad (7.6)$$

where t = time trend in years

(7) Production Function

$$(Y + T + D - Wg)_t = \eta_0 + \eta_1[\, h(Nw - Ng) + Ne + Nf\,]_t + \eta_2\,[(K_t + K_{t-1})/2]$$

$$+ \eta_3 t + u_{7,t} \qquad (7.7)$$

where h = index of hours worked per person per year

Nw = number of wage and salary earners

Ng = number of government employees

Ne = number of nonfarm entrepreneurs

Nf = number of farm operators

(8) Labor Market Adjustment Equation

$$w_t - w_{t-1} = \theta_0 + \theta_1(N - Nw - Ne - Nf)_t + \theta_2(p_{t-1} - p_{t-2}) + \theta_3 t + u_{8,t} \qquad (7.8)$$

where w = index of hourly wages

N = number of persons in the labor force

$N - Nw - Ne - Nf$ = unemployment in number of persons

p = general price index

(9) Import Demand Equation

$$(Fi)_t = \iota_0 + \iota_1[(Wp + Wg + P + A - Tw - Tp - Ta)\, p / pi] + \iota_2(Fi)_{t-1} + u_{9,t} \qquad (7.9)$$

where Fi = import of goods and services in 1939 dollars

pi = index of prices of imports

$(Wp + Wg + P + A - Tw - Tp - Ta)$ = deflated disposable income plus corporate saving (by an index of import prices)

(10) Agricultural Income Determination Equation

$$[A\ (p\ /\ pa)]_t = \kappa_0 + \kappa_1[(Wp + Wg + P - Sp - Tw - Tp)\ p\ /\ pa]_t$$
$$+ \kappa_2[(Wp + Wg + P - Sp - Tw - Tp)\ p\ /\ pa]_{t-1}$$
$$+ \kappa_3(p\ /\ pa)_t + \kappa_4(Fa)_t + u_{10.,t} \qquad (7.10)$$

where pa = index of agriculture prices

Fa = index of agricultural exports

$Wp + Wg + P - Sp - Tw - Tp$ = deflated disposable nonfarm income

(11) Relation between Agricultural and Nonagricultural Prices

$$(pa)_t = \lambda_0 + \lambda_1 p_t + \lambda_2(pa)_{t-1} + u_{11,t} \qquad (7.11)$$

(12) Household Liquidity Preference Equation

$$(L_1)_t - \mu_1(Wp + Wg + P + A - Tw - Tp - Sp - Ta)_t = \mu_0[(i_L)_t - i_L{}^0]^{\mu_2} u_{12,t} \qquad (7.12)$$

where i_L = average yield on corporate bonds in percent

$i_L{}^0$ = minimum possible interest rate

(13) Business Liquidity Preference Equation

$$(L_2)_t - \nu_1(Wp)_t = \nu_0 + \nu_2(p_t - p_{t-1}) + \nu_3(i_S)_t + \nu_4(L_2)_{t-1} + u_{13,t} \qquad (7.13)$$

where i_S = average yield on short term commercial paper

(14) Relationship Between Short Term and Long Term Interest Rates

$$(i_L)_t = \xi_0 + \xi_1(i_S)_{t-3} + \xi_2(i_S)_{t-5} + u_{14,t} \qquad (7.14)$$

(15) Money Market Adjustment Equation

$$[(i_S)_t - (i_S)_{t-1}]\ /\ (i_S)_{t-1} = o_0 + o_1 R_t + u_{15,t} \qquad (7.15)$$

where R = excess reserves of banks plus bank holdings of government bonds as a percentage of total reserves

The following Equations 16 to 20 are five definitions and accounting identities of the model.

(16) Definition of GNP

$$C_t + I_t + G_t + Fe_t - Fi_t = Y_t + T_t + D_t \tag{7.16}$$

where G = government expenditures for goods and services in 1939 dollars

Fe = exports of good and services in 1939 dollars

T = deflated indirect taxes less subsidies

(17) GNP = GNI Identity

$$(Wp)_t + (Wg)_t + P_t + A_t = Y_t \tag{7.17}$$

(18) Relation Between Wage Rate, Hours of Work, Employment, and Wage Bill

$$h_t (w_t / p_t)(Nw)_t = (Wp)_t + (Wg)_t \tag{7.18}$$

(19) Definition of Investment

$$K_t - K_{t-1} = I_t - D_t \tag{7.19}$$

(20) Corporate Surplus

$$B_t - B_{t-1} = (Sp)_t \tag{7.20}$$

This model uses 20 *endogenous variables*; namely:

p = price index of gross national product (1939 base = 100)

C = consumer expenditures in 1939 dollars

Wp = deflated private employee compensation

P = deflated nonwage income

Sp = deflated corporate saving

A = deflated farm income

L_1 = deflated end-of-year liquid assets held by persons

I = gross private domestic capital formation in 1939 dollars

D = capital consumption charges in 1939 dollars

i_L = average yield on corporate bonds

K = end-of-year stock of capital in 1939 dollars

L_2 = deflated end-of-year liquid assets held by enterprises

Pc = deflated corporate profits

B = deflated end-of-year corporate surplus

Y = deflated national income

Nw = number of wage- and salary-earners

w = index of hourly wages (1939 base : 122.1)

Fi = imports of goods and services in 1939 dollars

pa = index of agricultural prices (1939 base : 100)

i_S = average yield on short term commercial paper

The model has the following 19 exogenous variables:

Wg = deflated government employee compensation

Tw = deflated personal and payroll taxes less transfers associated with wage and salary income

Tp = deflated personal and corporate taxes less transfers associated with nonwage nonfarm income

Ta = deflated taxes less transfers associated with farm income

Np = number of persons in the US

Tc = deflated corporate income taxes

t = time trend in years

h = index of hours worked per person per year (1939 base : 1.00)

Ng = number of government employees

Ne = number of nonfarm entrepreneurs

Nf = number of farm operators

N = number of persons in the labor force

pi = index of prices of imports (1939 base : 100)

Fa = index of agricultural exports (1939 base : 100)

R = excess reserves of banks as a percentage of total reserves

G = government expenditures for good and services in 1939 dollars

Fe = exports of goods and services in 1939 dollars

T = deflated indirect taxes less subsidies

7.2. PREVIOUS ESTIMATION OF KLEIN-GOLDBERGER MODEL

In the original estimation, Eq. 14 for the empirical relation between short and long term interest rate and Eq.15 for the market adjustment both use single-equation least squares estimation with one lagged endogenous variable. In addition, Eq. 2 for the investment function is treated as a function of predetermined variables alone.

Other equations of the model are estimated by the method of limited information using the following set of *predetermined variables*:

C_{t-1} = lagged consumer expenditures in 1939 dollars

$(L_1)_{t-1}$ = lagged deflated, year-end liquid assets held by households

$(Np)_t$ = number of persons in the US

$(P + A + D - Tp - Ta)_{t-1}$ = lagged deflated, disposable nonwage income plus depreciation and corporate savings

K_{t-1} = lagged year-end stock of fixed capital in 1939 dollars

$(L_2)_{t-1}$ = lagged deflated, year-end liquid assets held by enterprises

$(Pc - Tc - Sp)_{t-1}$ = lagged deflated corporate dividend payments

B_{t-1} = lagged deflated, year-end accumulated corporate savings

$(Y + T + D - Wg)_{t-1}$ = lagged gross national product in 1939 dollars

t = time trend in years

$p_{t-1} - p_{t-2}$ = lagged first differences in the general price index

$(Fi)_{t-1}$ = lagged imports of goods and services in 1939 dollars

$(Wp + Wg + P - Sp - Tw - Tp)_{t-1}$ = lagged deflated nonfarm disposable income

$(Fa)_t$ = index of agricultural exports

$(G + Fe)_t$ = government expenditures plus exports of goods and services

The estimations are conducted over the period 1929-41 and 1946-50 in billions of dollars, millions of persons, and indices based on 1939.

7.3. ANN ESTIMATION OF KG MODEL

This section reports on the use of ANN and GA in an intelligent system to estimate and forecast a mix of temporal and non-temporal economic variables. The Klein-Goldberger Model with its available data serve as a useful context for the experiment. In addition, results from the original work (Klein and Goldberger, 1995) provide useful benchmark for comparative purposes.

Adelman and Adelman (1959) note that as the excess bank reserves R_t are taken as exogenous, the short term interest rate $(i_S)_t$ and long term interest rate $(i_L)_t$ can both be computed without reference to the rest of the system. Therefore, one treats them as predetermined and eliminates the estimation of these variables (Eqs. 7.14 and 7.15) in the system. Similarly, index of agricultural prices *pa* in Eq. 7.11 can be calculated independently by using its lagged values and price index of GNP which is considered as predetermined. As a result, *pa* is treated as predetermined value. Consequently, there are 12 endogenous variables to be estimated in this study.

The ANN used in this study has 12 outputs, namely, C, Y, Sp, Pc, D, Wp, w, Fi, A, L_1, L_2, and I. They represent the estimates of 12 simultaneous structural equations of the system. All of these outputs, except I, have lagged effects. Using a mixture-of-experts network architecture, the modular ANN has a recurrent module to account for these lagged effects and a standard module for I.

The ANN uses the following 45 inputs:

- 11 actual lagged values of 11 endogenous variables having lagged effects;
- Government expenditures for goods and services G, government employee compensation $(Wg)_t$ and its lag $(Wg)_{t-1}$
- Export of goods and services Fe
- Nonwage income (Profit) P and its lag P_{t-1}
- Indirect taxes T, personal and payroll taxes $(Tw)_t$ and its lag $(Tw)_{t-1}$, personal and corporate taxes $(Tp)_t$ and its lag $(Tp)_{t-1}$, corporate income tax $(Tc)_t$ and its lag $(Tc)_{t-1}$, income taxes of agricultural sector $(Ta)_t$ and its lag $(Ta)_{t-1}$

- Lagged year-end accumulated corporate savings B_{t-1}

- Year-end stock of fixed capital K_t and its lag K_{t-1}

- Agricultural export Fa

- Price index of GNP p and its lags p_{t-1}, p_{t-2}, price index of agricultural products pa, price index of import pi

- Population of the US Np, number of people in labor force N, number of wage- and salary-earners Nw, number of government employees Ng, number of farm operators Nf, and number of nonfarm entrepreneurs Ne

- Index of hours work per person per year h

- Short term interest i_L and long term interest i_S

- Time trend t (1929 =0)

In contrast to model specification of traditional estimation methods, as one does not know exacly the interaction of variables in an economic system, the ANN estimation conducted herein does not specify any particular effect of an input (exogenous or lagged endogenous) on a particular equation of the system. However, temporal variables with their lagged effects are treated explicitly.

The experiments are conducted to evaluate the relative performance of ANN in comparison with previous estimations of the KG Model in term of error committed for in-sample estimation and forecasting. The data are taken from Klein and Goldberger (1955). The data from 1929 to 1950 are used to train ANN. Then the network is used to generalize on out-of-sample cases in 1951 and 1952. Results are compared with those of the same period reported in the original work of Klein and Goldberger (1955). Although the network estimated 12 variables, only 10 variables have similar error information reported in the original work. One notes that, in Klein and Goldbeger (1955), error on index of hourly wages w is reported in the difference with its previous value (w_t - w_{t-1}), and the error of private liquid assets L_t is reported in a logarithmic format. To avoid introducing further error in re-estimation and re-transformation of final results, this study compares only the variables having compatible format.

The experiments conducted herein also seek to evaluate the performance of GA in selecting a parsimonious set of input variables. Besides selecting the appropriate ANN configuration for network training, GA is used to select the most relevant variables that contribute to a more accurate approximation of output in the related module.

7.4. FINDINGS AND DISCUSSION

7.4.1. Estimation and Forecasting with Modular ANN

Using the full set of input variables described in the KG Model, GA is implemeted to select only the optimal network configuration for the instrumental stage and the final stage. GA selects a configuration of 45-15T-11T for the recurrent module, and 45-14L-1T for the standard module. For each module, the selected network configuration is implemented and trained in 30 runs, each run lasting for 5000 epochs. Minimum and maximum ANN error from the 30 runs are recorded. This results in two streams of instrumental estimations representing the bound of errors on estimations from 30 runs. The ANN performance is evaluated for learning and forecasting from each stream of instrumental estimations. GA selects a configuration of 12-15L-12T for the gating network at the final stage. Maximum and minimum errors from estimation with the two streams of instrumental estimations are recorded. This study does not impose the assumption of normality on the estimated residuals. Consequently, the significance tests have not been conducted on the final estimations. However, the residuals from these estimations are recorded for future analyses with conventional statistical methods.

Results in Table 7.1 indicate the range of errors for in-sample estimation from a network with minimum/maximum error at the instrumental stage and/or minimum/maximum error at the final stage. Exception in learning C and using instrumental estimation with maximum error to learn Y and Wp, the ANN performance in terms of SSE is superior than those reported in Klein-Goldberger (1955). Total SSE for the variables estimated in Klein-Goldberger (1955) is 344.0039. Total SSE for the ANN estimations with minimum error at the intrumental stage and minimum error at the final stage is 41.4183. Total SSE for the ANN estimations with minimum error at the intrumental stage and maximum error at the final stage is 45.0717. Total SSE for the ANN estimations with maximum error at the instrumental stage and minimum error at the final stage is 184.9078. Total SSE for the ANN estimations with maximum error at the instrumental stage and maximum error at the final stage is 241.8141. In forecasting for 1951 and 1952, with the exception of C, Y and Wp, the residuals resulted from ANN forecasting are lower than those reported in Klein-Goldberger.

7.4.2. Estimation and Forecasting with Sets of Modular Variables Selected by GA

In this experiment, at each stage of instrumental estimation, besides selecting an appropriate configuration for each module, GA also is used to select the most relevant variables that contribute to a more accurate approximation of output. The top ten network topologies for each module are presented in Tables 7.3 and 7.4 . GAs are run in 30 trials, each trial evolves in 30 generations of 30 strings. One notes that hyperbolic tangent transfer function is selected for output nodes of both modules. But for each set of selected input variables, GA also finds an appropriate network configuration that provides the minimum estimation error.

For the recurrent module, the top GA selection contains 33 variables from the input set, i.e., all except D_{t-1},$(L_1)_{t-1}$, Fe, $(Wg)_{t-1}$, P_{t-1}, Tp, Ta, p, p_{t-1}, p_{t-2}, pa, and pi. The corresponding ANN configuration for this module is 33-15T-11T.

For the standard module, the top GA selection contains 38 variables from the input set, i.e., all except Y_{t-1}, w_{t-1}, T, D_{t-1}, Fe, and Tp. The corresponding ANN configuration for this module is 38-9L-1T.

One notes that D_{t-1}, Fe, and Tp are not selected for both modules. The ANN configuration selected for the final stage is 12-15L-12T.

Results reported in Tables 7.5 and 7.6 show that the performance of ANN in learning and forecasting is improved in comparison with those reported in the previous section. Unless instrumental estimation containing maximum error being used in the estimation of C and Wp, the SSE of all estimations are lower than those estimated by Klein-Goldberger and ANN with full input set. Total SSE for the ANN estimations with minimum error at the intrumental stage and minimum error at the final stage is 37.9853. Total SSE for the ANN estimations with minimum error at the intrumental stage and maximum error at the final stage is 55.1164. Total SSE for the ANN estimations with maximum error at the instrumental stage and minimum error at the final stage is 38.5591. Total SSE for the ANN estimations with minimum error at the instrumental stage and minimum error at the final stage is 60.2807. In forecasting for the period 1951 and 1952, residuals of all except those of *C*, *Y* and *Wp* are lower than the forecast from Klein-Goldberger and ANN with full input set.

7.4.3. Estimation and Forecasting with Union Set of Modular Variables Selected by GA

Although each module could function well with a particular set of related input variables, but in a system as a whole, the more informative variables the network has, the more it can learn effectively about the patterns in the problem space. To assess this possibility, the union set of selected variables in two instrumental modules are used for network training. The GA is used to select the appropriate configuration for each module at each stage of estimation.

The union set of variables contains 42 variables, all but D_{t-1}, *Fe*, and *Tp*. The corresponding network configuration is 42-13T-11T for recurrent module, 42-12L-1I for standard module, and 12-12T-12T for the final stage.

Except for using instrumental estimation with maximum error of *C*, *Y* and *W*p, results reported in Table 7.7 show that this union of variables and network selection provide the most accurate ANN performance for in-sample learning experiments. Total SSE for the ANN estimations with minimum error at the intrumental stage and minimum error at the final stage is 12.9667. Total SSE for the ANN estimations with minimum error at the intrumental stage and maximum error at the final stage is 50.3069. Total SSE for the ANN estimations with maximum error at the intrumental stage and minimum error at the final stage is 26.74. Total SSE for the ANN estimations with maximum error at the instrumental stage and maximum error at the final stage is 186.4155. In forecasting for the period 1951 and 1952, results in Table 7.8 show an unsatisfactory ANN performance on *C*, *Y* and *Wp*. However, forecasts for other variables have smaller residuals than those used full set and modular set of variables.

7.4.4. Estimation and Forecasting with Union Set of Modular Variables Controlled for ANN Architecture

To validate the effectiveness of the union set of selected variables, the configuration of gating network for ANN learning with full input set is used as controlled architecture. This final stage has a network configuration of 12-15L-12T. The network is used to map the stream of instrumental estimations (resulted from using the union set of input variables) to their actual output values. Then the results are compared with those using an appropriate gating network configuration (12-12T-12T) reported in the previous section.

Results reported in Tables 7.9 and 7.10 show that the performance of ANN decreases. It is lower then the one using full set with optimal configuration, as well as the one using union set with optimal configuration. Total SSE for the ANN estimations with minimum error at the instrumental stage and minimum error at the final stage is 49.1934. Total SSE for the ANN estimations with minimum error at the instrumental stage and maximum error at the final stage is 53.3079. Total SSE for the ANN estimations with maximum error at the instrumental stage and minimum error at the final stage is 212.4023. Total SSE for the ANN estimations with maximum error at the instrumental stage and maximum error at the final stage is 246.3837. Forecasts for 1951 and 1952 also have larger residuals than those using the union set of variables and corresponding network configuration.

The reason is the streams of data resulted from instrumental estimations with the full set of inputs versus the union set of selected inputs contain different pattern information. These patterns can only be learned well with an appropriate ANN configuration. The results show that the union set contains effective information on patterns of the problem space for a more accurate network estimation.

7.4.5. Concluding Remarks

From experiments with the Klein-Goldberger Model, this study offers the following remarks on the implementation of ANN and GA in an intelligent system for estimation/forecasting and its performance.

- The performance of ANN in approximating C, Y, and Wp is not remarkable. The estimation and particularly forecasting of these variables with ANN may need other alternative architectures. This study is conducted from a multivariate perspective. An alternative could be a *three-stage ANN estimation* in which the streams of instrumental estimation, resulting from recurrent and standard modules, are used to estimate each endogenous variable in a separate instrumental module. Then, these single equation estimations are aggregated into a gating network to account for their contemporaneous and simultaneous effect. With the mixture-of-experts architecture, one can have many alternative ANN designs by dividing a problem space into refined domains to learn modular patterns better before aggregating them into a final result.
- In contrast to the belief that the more information the network has the better its performance becomes, this study shows that only relevant information helps. The

presence of some "irrelevant" information may in fact decrease the performance of the ANN. The effectiveness of information can be assessed by GA in relation to the performance of ANN. The integrated ANN and GA helps in implementing a step-wise selection similar to the traditional method.

- Each data stream of instrumental modular estimations carries different information on patterns of the problem sub-spaces. As one changes the boundary of these sub-spaces, i.e., different partitions of input space into instrumental modules, the content of these data streams will change. Consequently, in the final stage of estimation, one needs an appropriate gating network configuration to effectively aggregate information on patterns of the sub-spaces. This could be overlooked in the case where one has the same number of modules for different partitions of the input space.
- In ANN learning, one may not have any problem with the mis-specification of input variables for a system of equations, and consequently for its endogenous variables. This study does not prescribe the influence of specific inputs on outputs. Apparently, a specific set of input variables will have more effect on a particular output. However, either the GA selects the relevant information and the appropriate configuration in network training, or the network learning algorithm tries its best to produce a mapping with least errors between network outputs and target values. This finding is particularly useful in data mining in which one deals effectively and efficiently with a mass of information without the need of subjective data selection and model specification.

Table 7.1

Estimation and Forecasting of KG Model with Full Set of Variables
SSE of In-Sample Estimation

	Total SSE	C	Y	Sp	Pc	D	Wp	Fi	A	L_2	I
Klein-Goldberger*	344.0390	6.7682	90.1531	3.5079	9.9619	30.8695	9.8488	2.9	26.9645	27.38	135.65
ANN											
Min**-Min***	41.4183	8.2112	12.7286	1.4250	1.9135	.2834	7.0387	.0290	.3252	6.5192	2.9445
Min**-Max***	45.0717	8.7272	9.7940	1.7353	2.5799	.3013	7.5564	.0319	.4248	10.9321	2.9888
Max**-Min***	184.9078	40.9348	76.3586	6.6697	5.6236	2.5672	22.5362	.5102	4.6368	19.5933	5.4774
Max**-Max***	241.8141	59.6348	93.5641	9.3571	7.4562	3.6171	31.5383	.5422	5.5007	22.2050	8.3986

Legend:

- * Klein and Goldberger (1955)
- ** Max/Min Errors on Instrumental ANN Estimation in 30 runs
- *** Max/Min Errors on Final ANN Estimation in 30 runs

Table 7.2

Estimation and Forecasting of KG Model with Full Set of Variables
Residuals in Forecasting 1951-52

	C	Y	Sp	Pc	D	Wp	Fi	A	L_2	I
Klein-Goldberger*										
1951	.3	-7.4	-2.47	-2.6	-2.4	-2.9	-.2	1.31	_	-7.8
1952	2.8	-.9	-1.38	- .7	-4	.1	-.6	2.6	_	-4.3
ANN										
Min**-Min***1951	-2.3611	-11.4601	1.3570	-2.1270	-1.1829	-5.7742	.0910	1.2646	-1.5386	-1.8923
Min**-Min***1952	-11.6512	-17.0769	1.9272	-.4749	-6.3292	-11.2319	-.4830	2.3407	.6544	.0046
Min**-Max***1951	-2.2526	-11.3377	1.2952	-2.1440	-1.2751	-5.6827	.0882	1.0612	-1.7517	-1.7566
Min**-Max***1952	-9.6699	-15.7354	1.8022	-.8215	-5.8353	-11.7808	-.5386	2.3115	1.2078	.6989
Max**-Min***1951	-2.9614	-13.2471	1.0430	-2.4288	-1.6902	-5.9897	-.0596	.4095	-.3035	-2.0369
Max**-Min***1952	-6.3439	-15.6397	1.3405	-.3363	-4.1232	-9.5039	-.4014	1.0060	-.6636	.9742
Max**-Max***1951	-2.9824	-12.8559	1.0910	-2.2609	-1.7229	-6.6215	-.0806	.5765	.0425	-2.2269
Max**-Max***1952	-6.1321	-15.2123	1.4127	-.1422	-4.0794	-10.0603	-.4158	1.0846	-.3649	.8054

<u>Legend</u>:

- * Klein and Goldberger (1955)
- ** Max/Min Errors on Instrumental ANN Estimation in 30 runs
- *** Max/Min Errors on Final ANN Estimation in 30 runs

Table 7.3

Input Variables and Network Architecture Selected by GA: Recurrent Module of Instrumental Stage

ANN Topology	MSE
33-15T-11T	.9107
28-11T-11T	1.0130
29-15T-11T	1.0394
22-15T-11T	1.0512
30-10T-11T	1.0933
24-14T-11T	1.1292
25-13T-11T	1.1529
27-9T-11T	1.2012
21-15T-11T	1.2718
20-15T-11T	1.3470

Table 7.4

Input Variables and Network Architecture Selected by GA
Standard Module of Instrumental Stage

ANN Topology	MSE
38-9L-1T	.1646
20-3L-1T	.1889
37-10T-1T	.1908
23-12L-1T	.2015
19-12L-1T	.2016
27-12L-1T	.2069
15-13L-1T	.2078
22-14L-1T	.2085
21-8L-1T	.2242
43-15L-1T	.2315

Table 7.5

Estimation and Forecasting of KG Model with ANN Topology Selected by GA:
SSE of In-Sample Estimation

	Total SSE	C	Y	Sp	Pc	D	Wp	Fi	A	L_2	I
Klein-Goldberger*	344.0390	6.7682	90.1531	3.5079	9.9619	30.8695	9.8488	2.9	26.9645	27.38	135.65
ANN											
Min**-Min***	37.9853	7.4702	15.1217	.5944	1.6071	.4117	6.0455	.0381	.3822	3.9883	2.3261
Min**-Max***	55.1164	6.4883	23.2577	1.7824	1.4860	.2894	8.8445	.0350	.5132	8.6521	3.7678
Max**-Min***	38.5591	7.3777	12.2153	.8904	2.9993	.5028	6.8019	.0341	4379	5.1236	2.1761
Max**-Max***	60.2807	11.4127	23.9983	1.4292	2.7897	.7101	10.7280	.0410	4458	5.8077	2.9182

Legend:

* Klein and Goldberger (1955)

** Max/Min Errors on Instrumental ANN Estimation in 30 runs

*** Max/Min Errors on Final ANN Estimation in 30 runs

Table 7.6

Estimation and Forecasting of KG Model with ANN Topology Selected by GA:
Residuals in Forecasting 1951-52

	C	Y	Sp	Pc	D	Wp	Fi	A	L_2	I
Klein-Goldberger*										
1951	.3	-7.4	-2.47	-2.6	-2.4	-2.9	-.2	1.31	_	-7.8
1952	2.8	-.9	-1.38	- .7	-4	.1	-.6	2.6	_	-4.3
ANN										
Min**-Min***1951	-1.4300	-11.0110	.6817	-2.1029	-.9600	-5.4738	.1347	.4957	.4972	-1.6092
Min**-Min***1952	-4.9228	-13.6967	1.3729	-.1307	-3.3633	-9.4031	-.2021	-.1190	.4622	1.3002
Min**-Max***1951	.9330	-11.8811	.4373	-2.0414	-.9901	-5.6142	.1511	.2762	.5155	-1.6399
Min**-Max***1952	-4.2809	-14.7904	1.3895	.0024	-3.3295	-9.2398	-.2040	-.1905	-.3647	.7342
Max**-Min***1951	-1.1209	-12.0360	.6959	-2.4679	-.9691	-5.6828	.1647	.6088	1.0821	-1.7739
Max**-Min***1952	-5.8685	-15.9945	1.6596	-.9683	-3.9185	-10.2559	-.4390	1.6359	1.5978	-.0423
Max**-Max***1951	-1.2404	-12.1634	.2963	-2.4318	-1.2033	-5.6329	.1283	.4475	.8572	-1.7123
Max**-Max***1952	-5.9632	-16.3162	.9436	-1.2577	-4.1990	-10.1090	-.4521	1.2070	1.3506	.0128

Legend:
- * Klein and Golberger (1955)
- ** Max/Min Errors on Instrumental ANN Estimation in 30 runs
- *** Max/Min Errors on Final ANN Estimation in 30 runs

Table 7.7

Estimation and Forecasting of KG Model with Union Set of Modular Variables Selected by GA:
SSE of In-Sample Estimation

	Total SSE	C	Y	Sp	Pc	D	Wp	Fi	A	L_2	I
Klein-Goldberger*	344.0390	6.7682	90.1531	3.5079	9.9619	30.8695	9.8488	2.9	26.9645	27.38	135.65
ANN											
Min**-Min***	12.9667	2.7278	4.5386	.0930	.5695	.0743	2.0799	.0210	.3755	1.1639	1.3232
Min**-Max***	50.3069	5.6683	13.2919	.6247	1.3708	.6772	9.4112	.0784	.4536	11.4240	7.3068
Max**-Min***	26.7400	8.7458	4.4653	.3077	1.4007	.1907	3.3700	.0337	.5239	5.4047	2.2975
Max**-Max***	186.4155	51.1057	31.7325	13.4323	10.0909	6.8519	24.7148	.2603	6.2291	36.3675	5.6305

Legend:

* Klein and Goldberger (1955)

** Max/Min Errors on Instrumental ANN Estimation in 30 runs

*** Max/Min Errors on Final ANN Estimation in 30 runs

Table 7.8

Estimation and Forecasting of KG Model with Union Set of Modular Variables Selected by GA:
Residuals in Forecasting 1951-52

	C	Y	Sp	Pc	D	Wp	Fi	A	L_2	I
Klein-Goldberger*										
1951	.3	-7.4	-2.47	-2.6	-2.4	-2.9	-.2	1.31	_	-7.8
1952	2.8	-.9	-1.38	- .7	-4	.1	-.6	2.6	_	-4.3
ANN										
Min**-Min***1951	.0910	-10.8239	.0853	-1.9198	-.8876	-4.7209	.1774	-.5209	.6111	-1.1292
Min**-Min***1952	-2.8713	-13.0432	.6845	.2302	-3.2094	-8.0484	-.1295	-.1329	.4127	2.0349
Min**-Max***1951	-.1585	-10.8305	.0523	-2.0859	-1.0089	-4.8621	.1961	-.5155	.4958	-1.5850
Min**-Max***1952	-3.0532	-12.8284	.3078	.1183	-3.2942	-8.1305	-.1039	-.0385	.3528	1.7080
Max**-Min***1951	-.0993	-10.3580	.1493	-2.0183	-.8448	-4.6587	.1342	-.3508	.3882	-1.0102
Max**-Min***1952	-3.6210	-12.9596	1.2803	-.0201	-3.2298	-8.1562	-.2410	1.1291	-1.2039	1.9728
Max**-Max***1951	-.7779	-10.3347	1.9456	-1.6787	-1.1510	-5.0244	.1336	-.3331	-.0115	-.9428
Max**-Max***1952	-3.7816	-12.3912	2.2744	.5215	-3.4751	-8.3466	-.1752	.3017	-.1750	2.3326

Legend:

* Klein and Goldberger (1955)

** Max/Min Errors on Instrumental ANN Estimation in 30 runs

*** Max/Min Errors on Final ANN Estimation in 30 runs

Table 7.9

Estimation and Forecasting of KG Model with Union Set of Modular Variables Selected by GA, Controlled for ANN Architecture:
SSE of In-Sample Estimation

	Total SSE	C	Y	Sp	Pc	D	Wp	Fi	A	L_2	I
Klein-Goldberger*	344.0390	6.7682	90.1531	3.5079	9.9619	30.8695	9.8488	2.9	26.9645	27.38	135.65
ANN											
Min**-Min***	49.1934	11.5397	16.2121	.9741	1.8868	.3610	11.3719	.0359	.4625	3.4181	2.9313
Min**-Max***	53.3079	6.6571	21.0021	2.7883	1.1443	.2045	9.4921	.0323	.5731	7.9929	3.4212
Max**-Min***	212.4023	56.2737	46.5118	11.6787	15.1477	5.7292	33.0505	.2709	4.6864	36.5566	2.4968
Max**-Max***	246.3837	63.4066	58.9659	12.9470	16.2110	9.3780	35.6931	.4329	6.7543	38.4353	4.1596

<u>Legend</u>:

- * Klein and Goldberger (1955)
- ** Max/Min Errors on Instrumental ANN Estimation in 30 runs
- *** Max/Min Errors on Final ANN Estimation in 30 runs

Table 7.10

Estimation and Forecasting of KG Model with Union Set of Modular Variables Selected by GA, Controlled for ANN Architecture: Residuals in Forecasting 1951-52

	C	Y	Sp	Pc	D	Wp	Fi	A	L_2	I
Klein-Goldberger*										
1951	.3	-7.4	-2.47	-2.6	-2.4	-2.9	-.2	1.31	–	-7.8
1952	2.8	-.9	-1.38	- .7	-4	.1	-.6	2.6	–	-4.3
ANN										
Min**-Min***1951	-1.5442	-12.4929	.1497	-2.2004	-.9688	-5.6022	.1546	-.4916	.5370	-1.6726
Min**-Min***1952	-4.5060	-14.7398	.6128	-.0586	-3.2855	-8.9460	-.1543	-.1031	.2694	1.4181
Min**-Max***1951	-.8572	-11.9438	.3105	-2.0200	-1.0280	-5.7050	.1478	-.4833	.7924	-1.7446
Min**-Max***1952	-3.8471	-14.1498	.8151	.1312	-3.3560	-9.0553	-.1635	-.0547	.3065	1.3434
Max**-Min***1951	-2.7039	-12.1875	.5880	-2.4692	-1.4687	-6.5048	.0113	-.1884	-.0130	-1.7902
Max**-Min***1952	-6.5829	-15.0689	.8528	-.6367	-4.1668	-10.3155	-.4032	.7780	.1775	.9993
Max**-Max***1951	-3.0435	-13.3865	.6505	-2.4262	-1.7041	-6.7869	-.0661	.1212	.1571	-1.8670
Max**-Max***1952	-6.8550	-16.2664	.8432	-.5760	-4.2941	-10.5894	-.4673	.8883	.0549	.8640

<u>Legend</u>:

* Klein and Goldberger (1955)

** Max/Min Errors on Instrumental ANN Estimation in 30 runs

*** Max/Min Errors on Final ANN Estimation in 30 runs

CHAPTER 8

Conclusion

8.1. SUMMARY OF FINDINGS

- *On the Effectiveness and Efficiency of Mixture of ANN Learning*

This study provides evidence that an ANN in a mixture-of-experts network architecture has a superior performance to traditional estimations. Its performance also is superior to the one of a large single network. The advantage of the framework is that it can account for patterns of different types of variables, in this case, the temporal and non-temporal variables of an economic system. By means of instrumental estimations resulting from either intermediate or modular expert-networks, it facilitates the indepth study of particular input or group of inputs in the problem space. At the same time, the framework considers the simultaneous as well as contemporaneous effects in the economic system from a multivariate perspective.

- *On the Effectiveness of GA Selection*

With controlled experiments, this study provides evidence that GA has the ability to select the optimal parameters for efficient network training. Sensitivity analyses also show that the network configuration selected by the GA in terms of number of hidden nodes/layers and transfer function is the most effective one.

In fact, the integration of GA and ANN has the ability of a stepwise selection of input variables for a nonlinear, nonparametric approximation of a problem space. It selects the most relevant input and then a corresponding function approximator (i.e., an appropriate ANN architecture) to learn accurately patterns of the problem space.

- *On the Efficiency of ANN Estimation/Forecasting*

This study confirms the ability of ANN in function approximation to any degree of accuracy. With ANN, one can overcome the restrictions in model specification and variable selection of traditional estimation methods. Its in-sample learning errors are absolutely lower than any other traditional estimation methods.

However, the merits of a Decision Support System in general, and a forecasting system in particular, is to use its stored knowledge about the past in order to provide useful suggested solution for the future, e.g., an accurate forecast. Using ANN in a forecasting system and providing it with abundant information about the problem space, the network should learn patterns from past data and approximate the future well.

Evidence from this study shows that, whenever the network learns a particular pattern well, its prediction follows future patterns closely, provided that these patterns are not extremely different from those that the network has stored in its knowledge. In the case of predicting new emerging patterns, this study highlights some behavior of ANN when forecasting future patterns that it has not learned. In contrast to the remarks reported in literature that the network learns and projects the recent extreme trend, this study finds that the network provides a moderate projection into the future instead.

8.2. IMPLICATIONS OF THIS STUDY

- *Implications for Theory*

This study empirically shows that ANN, as a computing tool, can overcome restrictions of traditional methods in the estimation and forecasting of simultaneous, structural equation systems. One of the strong assumptions in econometric and traditional statistics is the linear relationship of variables in a model. In addition, there are many assumptions being imposed on the behavior of variables and time series in order to make available estimation methods feasible (Mills, 1990). Using ANN as a universal function approximator, without any constraints and prescriptions on data and functional models, one can approximate the underlying relationship between input-output patterns well to arrive at an accurate estimation.

This study also shows that GA, as a search tool, can overcome the sub-optimality in model identification and variable specification. With the integration of ANN and GA, one can build forecasting systems without the need for a specific theoretical model. Although expert's knowledge on the problem domain is valuable in model identification and variable specification, it is not necessary in order to build an effective and efficient forecasting system. With an abundance of information, GA should be able to search through the mass of data to select the most appropriate variables which positively contribute to the accurateness of forecasting. In addition, it also selects the most appropriate ANN configuration in order to approximate well the patterns of selected data.

- *Implications for Methodology*

This study adheres to the Machine Learning field in Artificial Intelligence study in building Decision Support Systems, particularly intelligent forecasting systems. The study provides an effective integration of ANN and GA to exploit the powerful computing and search ability of these emerging technologies. In a mixture-of-experts network architecture, modular networks are used to address the peculiarity of individual patterns. Then the GA is implemented to select the appropriate input for the task and corresponding network topology to learn these patterns.

Particularly in business forecasting, modular networks are implemented in order to handle a mix of temporal and non-temporal data patterns without the need to prescribe data behavior and *a priori* model. The mis-specification of forecasting model may not be a serious issue as the estimation errors at the early instrumental stages are rectified at the final stage of the integrated network. However, too much error in instrumental estimation definitely affects the performance of the network as it has to try harder to detect signals from a noisy stream of data. A GA integrated with an ANN acts as a nonlinear stepwise selection of relevant variables. GA should be able to select input variables which contribute the most to network performance.

In highlighting the behavior of network errors in estimation and forecasting, the experiments also show that one should train networks with an abundance of information about the problem space. ANN should learn the maximum and minimum levels of related variables in order to generalize well on unseen cases.

- *Implications in Practice*

In proposing a unified framework of GA and ANN for building intelligent forecasting systems, this study applies emerging technologies in Artificial Intelligence, particularly those in Machine Learning area to solve business problems. Previously, ANNs were used mostly to learn physical patterns. Recently, ANN and GA are used as a computing tool to overcome the restrictions of traditional estimation methods, particularly, the assumptions on linearity and the functional relationships among variables. The investigation undertaken in this study supports the ability of ANN and GA as a versatile intelligent technology to solve a general family of business problems, namely forecasting problems involving many temporal and non-temporal variables.

In particular, this study presents an effective implementation of modular networks to solve a mix of economic data with the integration of recurrent and standard networks. It also proposes the use of maximum and minimum errors to represent a bound of error on network learning and to serve as an efficient means to assert the accuracy of network estimation.

The proposed integration not only serves as a foundation to build effective Decision Support Systems, but it is also useful in the emerging field of Data Mining. In this field, Artificial Intelligence tools are used to discover and acquire unknown knowledge from the mass of information. Without *a priori* restrictions on the behavior of data and the prescribed functional relationships, the proposed network is able to discover unknown patterns in socio-economic information.

8.3. LIMITATIONS OF THIS STUDY

This study is conducted from a Machine Learning perspective in building Decision Support Systems. The focus is on the aggregation of AI tools in an effective intelligent forecasting system and to study its behavior. As such, it did not investigate other possible improvements of traditional estimation and forecasting methods besides those reported in the literature.

This study uses the framework and available data of Klein and Klein-Goldberger Models as research context. Even these models represent a general context involving temporal and temporal economic events, the findings of this study may still be problem-specific. In addition, the study does not impose the assumption of normality on the behavior of estimation residuals for significant tests with conventional statistical methods. However, the findings and documentation of this study should serve as useful benchmark for future comparative studies on performance of alternative methods.

This study is conducted from a multivariate perspective to estimate simultaneous, contemporaneous variables of an economic system. All endogenous variables are estimated simultaneously in appropriate modules at the instrumental stage or final stage. As such, some peculiarities of individual variables may not be modeled and studied intensively. There are other possibilities in building modular networks. For instance, in a *three-stage ANN estimation*, one can build modular networks to estimate single equations of the system before aggregating them with a final network to account for simultaneity.

This study uses an early stopping rule in training to assure a better generalization of the network. The selection of network parameters and architecture is based on the simplicity of network configuration and on the accuracy of the network estimation. As such, the speed of convergence was not addressed in GA search. In a large problem setting, where the time to reach convergence is critical, GA should account for this issue explicitly when selecting the appropriate learning parameters.

8.4. EXTENSIONS OF THIS STUDY IN FUTURE RESEARCH

- *Identification of Lag Effect with GA*

In this study, temporal information is represented in a static manner, i.e., the value of an economic variable corresponding to each time lag is represented as an input. If an arbitrary number of l time lags of a variable are represented in l network inputs, one can use GA to search for the most important lagged impact in forecasting. The memory of a time series is the longest lag selected by GA which has positive contribution to network estimation. The integration of ANN and GA will help to overcome the linearity in correlograms of traditional methods of time series model identification.

- *Modular Estimation of Temporal Variables having Different Lags*

According to the specification of the original model, instrumental estimations in this study focus on temporal variables of one lag. In a different context, one can have a modular estimation of a mix of temporal variables having different lags. Consequently, one can have many recurrent modules at the instrumental stage, each of which will handle variables having the same number of memory length.

- *Modular Estimation of Single Equation of the System*

The estimations in this study are from a multivariate perspective in which network output (endogenous variables) are computed simultaneously. An alternative is modular *three-stage estimation* in which modular networks are implemented to learn the peculiarity in patterns of individual variables before aggregating instrumental estimations of endogenous variables at the final stage. This approach will require many modules in the integrated network, and the corresponding network design and computing. In any case, the instrumental estimation of single equations with ANN should provide similar information to traditional single equation estimation methods.

- *Seasonal Patterns of Economic Data*

This study uses yearly data of the US economy. In a situation involving forecasting with seasonal patterns, e.g., the Wharton model (Evans and Klein, 1967), one can use indicator variables for seasonal effects. An indicator also is appropriate for event studies in investigating the reaction of an economic system given an unusual external shock. With the indicator variable as input, the network will be useful in sensitivity (what-if) analyses.

This study has investigated and presented a comprehensive foundation for an effective decision support system to assist users making better decisions in a general business forecasting problem. Although there still exist limitations and other possible improvements, the findings of this study should contribute positively to the development of theory, methodology and practice of using Machine Learning and AI tools to build intelligent forecasting systems. With all of its limitations, this endeavor has contributed to the noble goal of Artificial Intelligence study in building and introducing smart machines to assist human beings in their daily activities, in this case, solving a class of business problems.

References

Adelman I. and Adelman F.L. (1959). The Dynamic Properties of the Klein-Goldberger Model. *Econometrica*, 27(4), 596-625.

Amemiya T. (1981). Qualitative Response Models: A Survey. *Journal of Economic Literature*, 19, 1483-1536.

Bodkin R.G., Klein L.R., and Marwah K. (1991) *A History of Macroeconomic Model-Building*. Edward Elgar Publishing Ltd.

Box G.E.P and Jenkins G.M. (1976). *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.

Blum E.K. and Li L.K. (1991). Approximation Theory and Feedforward Networks. *Neural Networks*, 4, 511-516.

Caporaletti L.E., Dorsey R.E., Johnson J.D., and Powell W.A. (1994). A Decision Support System for In-sample Simultaneous Equation Systems Forecasting Using Artificial Neural Systems. *Decision Support Systems*, 11, 481-495.

Cheng B. and Titterington D.M. (1994). Neural Networks: A Review From A Statistical Perspective. *Statistical Science*, 9(1), 2-54.

Cottrel M., Girard B., Girard Y., Mangeas M., and Muller C. (1995). Neural Modeling of Time Series: A Statistical Stepwise Method for Weight Elimination. *IEEE Transactions of Neural Networks* 6(6), 1355-1364

Chu C.H. and Widjaja D. (1994). Neural Network System for Forecasting Method Selection. *Decision Support Systems*, 12, 13-24.

Cromwell J.B., Labys W.C., and Terraza M. (1994a). *Univariate Tests for Time Series Models* (Sage University Paper Series on Quantitative Applications in the Social Sciences, Series no. 07-099). Thousand Oaks, CA: Sage.

Cromwell J.B., Hannan M.J., Labys W.C., and Terraza M. (1994b). *Multivariate Tests for Time Series Models* (Sage University Paper Series on Quantitative Applications in the Social Sciences, Series no. 07-100). Thousand Oaks, CA: Sage.

Cybenko G. (1989). Approximation by Superpositions of A Sigmoid Function. *Mathematics of Control, Signals, and Systems*, 2, 303-314.

Davis L. (Ed.) (1991). *Handbook of Genetic Algorithms*. New York: Van Nostrand Reinhold.

Elman J.L. (1988). *Finding Structure in Time*. CRL Report 8801, Center for Research in Language, UC San Diego.

Evans M.K. and Klein L.R. (1967). *The Wharton Econometric Forecasting Model*. Economics Research Unit, Wharton School, University of Pennsylvania.

Foster B., Collopy F., and Ungar L. (1991). *Neural Network Forecasting of Short, Noisy Time Series*. ORSA/TIMS National Meeting May 1991.

Friedman J.H. and Stuetzle W. (1981). Projection Pursuit Regression. *Journal of the American Statistical Association*, 76, 817-823.

Funahashi K. (1989). On the Approximate Realization of Continuous Mappings by Neural Networks. *Neural Networks*, 2, 355-363.

Gallant S.I. (1993). *Neural Network Learning and Expert Systems*. Cambridge, MA: The MIT Press.

Gershenfeld N.A. and Weigend A.S. (1993). The Future of Time Series: Learning and Understanding. In A.S. Weigend and N.A. Gershenfeld (Eds.) *Times Series Prediction: Forecasting the Future and Understanding the Past.* Reading, MA: Addison-Wesley.

Goldberg D.E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning.* Reading, MA: Addison-Wesley Publishing

Goldberger A.S. (1959) *Impact Multipliers and Dynamic Properties of the Klein-Goldberger Model.* Amsterdam: North-Holland.

Granger C.W.J. (1969). Prediction with A Generalized Cost of Error Function. *Operations Research Quarterly*, 20, 199-207

Granger C.W.J. and Newbold P. (1986). *Forecasting Economic Time Series*, 2nd ed.. San Diego, CA: Academic Press.

Granger C.W.J. and Terasvirta T. (1993). *Modelling Nonlinear Economic Relationships.* Oxford University Press.

Greene W.H. (1990). *Econometric Analysis.* New York: Macmillan Publishing Co.

Hamilton J.D. (1994). *Time Series Analysis.* Princeton, NJ: Princeton University Press.

Hagan M.T, Demuth H.B., and Beale M. (1995) *Neural Network Design.* Boston, MA: PWS Publishing.

Hansen J.V. (1998). Comparative Performance of Backpropagation Networks Designed by the Genetic Algorithms and Heuristics. *International Journal of Intelligent Systems in Accounting, Finance & Management*, 7, 69-79.

Harvey A.C. (1993). *Time Series Models.* Cambridge, MA: The MIT Press.

Hassoun M.H. (1995) *Fundamentals of Artificial Neural Networks*. Cambridge, MA: The MIT Press.

Hecht-Nielsen R. (1989). The Theory of Back-Propagation Neural Networks. *Proceedings of the International Joint Conference on Neural Networks*, I, 593-606. New York: IEEE Press.

Hill T. , Marquez L. O'Connor M., and Remus W. (1994). Artificial Neural Network Models for Forecasting and Decision Making. *International Journal of Forecasting*, 10, 5-15.

Hill T. , O'Connor M., and Remus W. (1996). Neural Network Models for Time Series Forecasts. *Management Science*, 42(7), 1082-1092.

Hinton, G.E. (1992). How Neural Networks Learn from Experience. *Scientific American*, 267, 144-151.

Holland J. (1975). *Adaptation in Neural and Artificial Systems*. Ann Arbor, MI: The University of Michigan Press.

Holsapple C.W. and Whinston A.B. (1996). *Decision Support Systems*. Minneapolis: West Publishing Co.

Hornik K. (1991). Approximation Capabilities of Multilayer Feedforward Nets. *Neural Networks*, 4, 231-242.

Hornik K., Stinchcombe M., and White H. (1989). Multi-Layer Feedforward Networks Are Universal Approximators. *Neural Networks*, 2, 359-366.

Hornik K., Stinchcombe M., and White H. (1990). Universal Approximation of an Unknown Mapping and Its Derivatives Using Multilayer Feedforward Networks. *Neural Networks*, 3, 551-560.

Huber P.J. (1985). Projection Pursuit. *Annals of Statistics*, 13, 435-475.

Hwang J.N., Lay S.R., Maechler M., and Martin R.D. (1994). Regression Modeling in Back-Propagation and Projection Pursuit Learning. *IEEE Transactions on Neural Networks*, 5(3), 342-353.

Intriligator M.D. (1978) *Econometric Models, Techniques, and Applications*. Englewood Cliffs, NJ: Prentice Hall.

Jacobs R.A., Jordan M.I., Nowlan S.J., and Hinton G. (1991) Adaptive Mixtures of Local Experts. *Neural Computation*, 3, 79-87.

Jordan M.I.(1986). *Serial Order: A Parallel Distributed Processing Approach*. UC San Diego, Institute for Cognitive Science Report 8604.

Jordan M.T. and Jacobs R.A. (1995). Modular and Hierarchical Learning Systems. In M. Arbib (Ed.), *The Handbook of Brain Theory and Neural Networks*, Cambridge, MA: MIT Press.

Judge G.G., Griffiths W.E., Hill R.C. Lütkepohl H., and Lee T.C. (1985) *The Theory and Practice of Econometrics, 2nd Ed.* New York: John Wiley and Sons.

Kang S. (1991). *An Investigation of the Use of Feedforward Neural Networks for Forecasting*, Ph.D. Dissertation, Kent State University.

Karakitsos E. (1992) *Macrosystems, The Dynamics of Economic Policy*. Oxford: Blackwell Publishers.

Kira D.S., Doreen D., and Nguyen D.-D. (1997). An Application of Artificial Neural Networks and Statistical Methods in Qualitative Evaluation of Small Business Loans. *Proceedings of VIII International Symposium on Applied Stochastic Models and Data Analysis*. Naples, June 1997.

Klein L.R. (1950) *Economic Fluctuation in the United States, 1921-1941.* New York: John Wiley.

Klein L.R. and Goldberger A.S. (1955) *An Econometric Model of the United States, 1929-1952.* Amsterdam: North-Holland.

Klimasaukas C.C. (1991). Applying Neural Networks. Part 3: Training A Neural Network. *PC-AI*, May/June, 20-24.

Kolgomorov A.N. (1963). On the Representation of Continuous Functions of Many Variables by Superposition of Continuous Functions of One Variable and Addition. *American Mathematical Society Translation*, 28, 55-59.

Kryzanowski L., Galler M., and Wright D.W. (1993). Using Artificial Neural Network to Pick Stocks. *Financial Analysts Journal*, July/August, 21-27

Kuan C.M and White H. (1994). Artificial Neural Networks: An Econometric Perspective. *Econometric Reviews*, 13(1), 1-91.

Lachtermacher G. (1993). *Backpropagation in Time Series Analysis*, Ph.D. Thesis, Department of Management Sciences, University of Waterloo, Canada.

Lachtermacher G. and Fuller J.D. (1995). Backpropagation in Time-series Forecasting. *Journal of Forecasting*, 14, 381-393.

Lee J.K. and Jhee W.C. (1994). A Two-stage Neural Network Approach for ARMA Model Identification with ESCAF. *Decision Support Systems*, 11, 461-479.

Lin C.T. and Lee C.S.G. (1996). *Neural Fuzzy Systems, A Neuro-Fuzzy Synergism to Intelligent Systems*. Upper Saddle River, NJ: Prentice-Hall.

Lippmann R.P. (1987). An Introduction to Computing with Neural Nets. *IEEE ASSP Magazine*, 4(2), 4-22.

Lütkepohl H. (1991). *Introduction to Multiple Time Series Analysis*. Berlin: Springer-Verlag.

Makridakis S., Andersen A., Carbone R., Fildes R., Hibon M., Lewandowski R., Newton J., Parzen E., and Winkler R. (1982). The Accuracy of Extrapolation (Time Series) Methods: Results of A Forecasting Competition. *Journal of Forecasting*, 1, 111-153.

Makridakis S., Wheelwright S.C., and McGee V.E. (1983). *Forecasting: Methods and Applications*, 2nd Edition. New York: John Wiley & Sons.

Maren A., Harston C., and Pap R. (1990). *Handbook of Neural Computing Applications*. San Diego: Academic Press.

Maasoumi E., Khotanzad A., and Abaye A. (1994). Artificial Neural Networks for Some Macroeconomic Series: A First Report. *Econometric Reviews*, 13(1), 105-122.

McCullock W.S. and Pitts W.H. (1943). A Logical Calculus of the Ideas Immanent in Nervous Activity. *Bulletin of Mathematical Biophysics*, 5, 115-133.

Medsker L.R. (1994). *Hybrid Neural Network and Expert Systems*. Boston: Kluwer Academic Publishers.

Medsker L.R. (1995). *Hybrid Intelligent Systems*. Boston: Kluwer Academic Publishers.

Mirchandini M.L. and Cao W. (1989). On Hidden Nodes in Neural Nets. *IEEE Transaction on Circuits Systems*, 36(5), 661-664.

Mills T.C. (1990). *Time Series Techniques for Economists*. Cambridge: Cambridge University Press.

Mitchell M. (1996). An Introduction to Genetic Algorithms. Cambridge, MA: The MIT Press.

Moody J. (1995). *Economic Forecasting: Challenges and Neural Network Solutions*. Keynote talk presented at the International Symposium on Artificial Neural Networks, Hsinchu, Taiwan, December 1995.

Montana D.J. and Davis L.D. (1989). Training Feedforward Networks Using Genetic Algorithms. *Proceedings of the International Joint Conference on Artificial Intelligence*.

Nguyen D.-D. and Kira D. (1997) On Using Artificial Neural Networks and Vector Autoregressive Method in Multiple Time Series Forecasting. *Proceedings of DSI'97 Annual Conference of Decision Sciences Institute*. San Diego, November 1997.

Poston T., Lee C., Choie Y., and Kwon (1991). Local Minima and Backpropagation. *Proceedings of the IEEE/International Joint Conference of Neural Networks*, 2, 173-176.

Refenes A.P. (Ed.) (1995). *Neural Networks in the Capital Market*. Chichester: John Wiley & Sons.

Ripley B.D. (1994). Neural Networks and Related Methods for Classification. *Journal of Royal Statistical Society*, B, 56(3), 409-456.

Ritter H. (1989). Combining Self-organizing Maps. *Proceedings of the First Internal Joint Conference on Neural Networks*. Washington, D.C, June 1989.

Rosenblatt F. (1959). The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychological Review*, 65, 386-408.

Rosenblatt F. (1962). *Principles of Neurodynamics*. New York: Spartan.

Rumelhart D.E., McClelland J.L., and The PDG Research Group (1986). *Parallel Distributed Processing, Explorations in the Microstructure of Cognition; Vol.1: Foundations*. Cambridge, MA: The MIT Press.

Rumelhart D.E., Durbin R., Golden R., and Chauvin Y. (1995). Backpropagation: the Basic Theory. In Y. Chauvin and D.E. Rumelhat (Eds.), *Backpropagation: Theory, Architectures, and Applications*. New Jersey: Lawrence Earlbaum.

SAS Institute (1984). *SAS/ETS User Guide*, Version 5.

Salchenberger L.M., Cinar E.M., and Lash N.A. (1992). Neural Networks: A New Tool for Predicting Thrift Failures. *Decision Sciences*, 23(4), 899-916.

Sexton R.S, Dorsey R.E., and Johnson J.D. (1998). Toward Global Optimization of Neural Networks: A Comparison of the Genetic Algorithn and Backpropagation. *Decision Support Systems*, 22, 171-185.

Sharda R. and Patil R. (1990). Neural Networks As Forecasting Experts: An Empirical Test. *Proceedings of the 1990 International Joint Conference on Neural Networks Meeting*, 2, 491-494. Alamitos, CA: IEEE Press.

Sharda R. and Patil R. (1992). A Connectionist Approach to Time Series Prediction: An Empirical Test. *Journal of Intelligent Manufacturing*.

Sharda R. (1994). Neural Networks for the MS/OR Analyst: An Application Bibliography. *Interfaces*, 24, 116-130.

Silver M.S. (1991). *Systems That Support Decision Makers: Description and Analysis.* New York: John Wiley & Sons.

Sims C.A. (1980). Macroeconomics and Reality. *Econometrica*, 48(1) 1-48.

Sontag E. (1990). *Feedback Stabilization Using Two-Hidden Layer Nets.* Rutgets Center for Systems and Control Technical Report SYCON-90-11.

Tang Z., De Almeida C., and Fishwick P.A. (1991). Time Series Using Neural Networks Vs. Box-Jenkins Methodology. *Simulation*, 57(5), 303-310.

Theil H. (1971). *Principles of Econometrics.* New York: John Wiley.

Theil H. and Boot J.C.G. (1962). The Final Form of Econometric Equation Systems. *Review of the International Statistical Institute*, 30, 136-152.

Tinbergen J. (1939). *Statistical Testing of Business-Cycle Theories.* New York: Agathon Press, 1968 (Originally published in Geneva by the Economic Intelligence Service of the League of Nations in 1939)

Trippi R.R. and Turban E. (1993). *Neural Networks in Finance and Investing.* Chicago: Probus.

Turban E. and Aronson J.E. (1998). *Decision Support Systems and Intelligent Systems (5th Ed.).* Upper Saddle River, NJ: Prentice Hall.

Tsay R.S. and Tiao G.C. (1984). Consistent Estimates of Autoregressive Parameters and Extended Sample Autocorrelation Function for Stationary and Nonstationary ARMA Models. *Journal of American Statistical Association*, 79, 84-96.

U.S. Department of Commerce, Bureau of Economic Analysis (1998). *National Income and Product Accounts of the United States, 1929-94*. Washington, DC: U.S. Goverment Printing Office.

Uebe G. (1995). *World of Economic Models, A Catalogue of Typical Specifications of Economic Models*. Brookfield: Avebury.

Werbos P. (1974). *Beyond Regression: New Tool for Prediction and Analysis in the Behavioral Sciences*. Ph.D. Thesis. Harvard University.

White H. (1989). Some Asymptotic Results for Learning in Single Hidden-Layer Feedforward Network Models. *Journal of the American Statistical Association*, 84, Theory and Methods, 1003-1013.

Whitley D., Starkweather T., and Bogart C. (1990). Genetic Algorithms and Neural Networks: Optimizing Connections and Connectivity. *Parallel Computing*, 14, 347-361.

Widrow B. and Hoff M.E. (1960). Adaptive Switching Circuits. *Institute of Radio Engineers WESCON Convention Record*, Part 4, 96-104.

Zellner A. (1962). An Efficient Method of Estimating Seemingly Unrelated Regression and Test of Aggregation Bias. *Journal of the American Statistical Association*, 57, 348-367.

APPENDICES

Appendix 1

Data for Klein Model I Estimation

time	C	P	Wp	I	K	Y	Wg	G	T
1920	39.8	12.7	28.8	2.7	182.8	43.7	2.2	2.4	3.4
1921	41.9	12.4	25.5	-0.2	182.6	40.6	2.7	3.9	7.7
1922	45.0	16.9	29.3	1.9	184.5	49.1	2.9	3.2	3.9
1923	49.2	18.4	34.1	5.2	189.7	55.4	2.9	2.8	4.7
1924	50.6	19.4	33.9	3.0	192.7	56.4	3.1	3.5	3.8
1925	52.6	20.1	35.4	5.1	197.8	58.7	3.2	3.3	5.5
1926	55.1	19.6	37.4	5.6	203.4	60.3	3.3	3.3	7.0
1927	56.2	19.8	37.9	4.2	207.6	61.3	3.6	4.0	6.7
1928	57.3	21.1	39.2	3.0	210.6	64.0	3.7	4.2	4.2
1929	57.8	21.7	41.3	5.1	215.7	67.0	4.0	4.1	4.0
1930	55.0	15.6	37.9	1.0	216.7	57.7	4.2	5.2	7.7
1931	50.9	11.4	34.5	-3.4	213.3	50.7	4.8	5.9	7.5
1932	45.6	7.0	29.0	-6.2	207.1	41.3	5.3	4.9	8.3
1933	46.5	11.2	28.5	-5.1	202.0	45.3	5.6	3.7	5.4
1934	48.7	12.3	30.6	-3.0	199.0	48.9	6.0	4.0	6.8
1935	51.3	14.0	33.2	-1.3	197.7	53.3	6.1	4.4	7.2
1936	57.7	17.6	36.8	2.1	199.8	61.8	7.4	2.9	8.3
1937	58.7	17.3	41.0	2.0	201.8	65.0	6.7	4.3	6.7
1938	57.5	15.3	38.2	-1.9	199.9	61.2	7.7	5.3	7.4
1939	61.6	19.0	41.6	1.3	201.2	68.4	7.8	6.6	8.9
1940	65.0	21.1	45.0	3.3	204.5	74.1	8.0	7.4	9.6
1941	69.7	23.5	53.3	4.9	209.4	85.3	8.5	13.8	11.6

<u>Source</u>: Klein (1950)

Appendix 2

Min-Min Residuals of 2-Stage ANN Estimation (1000 epochs)
for Klein Model I

e(C)	e(Wp)	e(I)
0.360311	1.528929	-0.011954
0.736206	0.236169	-0.015743
0.432469	-0.921835	-0.738235
0.471651	-0.154280	0.552365
-0.141462	0.080263	-0.211895
-0.504999	0.044576	-0.242458
-0.656745	0.064043	-0.153537
-0.005357	0.176919	0.818062
0.769580	-0.475278	-0.593064
0.463753	0.061498	-0.022514
0.648824	-0.638130	-0.117130
-0.417905	-0.193125	0.217180
-0.028877	0.782393	0.355985
-0.808826	-0.121927	0.032569
0.340850	0.041501	-0.206643
-1.445934	0.922378	-0.000433
0.739304	-0.421848	-0.036950
-0.246384	-0.362567	-0.033045
-0.254919	0.473433	0.021982
0.001221	1.108990	0.237232
-0.284557	-2.361256	-0.376782

Appendix 3

Min-Max Residuals of 2-Stage ANN Estimation (1000 epochs)
for Klein Model I

e(C)	e(Wp)	e(I)
1.131532	1.326138	-0.002029
0.990185	0.071567	-0.070164
0.518731	-1.048158	-0.753569
0.370768	-0.085101	0.558064
-0.212926	0.019933	-0.186135
-0.666010	-0.073943	-0.206856
-0.697150	0.167238	-0.127334
-0.028325	0.369691	0.862541
0.705283	-0.445370	-0.559408
0.544342	0.144094	-0.001322
0.246229	-0.862785	-0.219805
0.535513	-0.330671	0.406732
0.281460	0.670372	0.384700
-0.862144	-0.223800	0.008756
-0.025460	0.133927	-0.248217
-1.547301	1.163852	0.013228
0.761105	-0.239933	-0.007471
-0.558758	-0.265086	0.007400
0.007346	0.671572	0.088566
0.497673	1.030972	0.176928
-0.000011	-3.002785	-0.504329

Appendix 4

Max-Min Residuals of 2-Stage ANN Estimation (1000 epochs)
for Klein Model I

e(C)	e(Wp)	e(I)
0.603784	1.211908	-0.029487
0.248547	-0.123946	-0.033460
0.426629	-1.239141	-1.022296
0.428698	0.030565	0.979262
0.412455	0.368841	-0.176416
-0.445688	-0.044268	-0.392181
-0.655154	0.284563	0.100200
-0.215588	0.332192	1.091555
0.461467	-0.485215	-0.646675
0.232548	0.046072	-0.046239
0.407209	-0.282631	-0.212958
-0.098970	-0.683748	0.602312
0.146111	0.638638	0.294565
-0.475555	-0.011964	-0.158459
-0.025521	-0.030570	-0.106616
-1.279369	1.070682	0.039741
0.686471	-0.673717	-0.251615
-0.052872	-0.011382	0.167677
-0.334978	0.145594	0.109050
0.489670	1.088654	0.089477
-0.077098	-2.563268	-0.395570

Appendix 5

Max-Max Residuals of 2-Stage ANN Estimation (1000 epochs)
for Klein Model I

e(C)	e(Wp)	e(I)
1.428430	1.422968	-0.089117
0.797047	-0.338658	0.208019
0.116807	-1.560468	-1.027363
0.333525	-0.027247	0.958460
0.134150	0.401960	-0.378302
-0.702364	0.044115	-0.679446
-0.601379	0.410497	-0.024183
0.005393	0.590012	1.072173
0.653697	-0.303624	-0.711460
0.386528	-0.236277	0.099969
0.334398	-0.680943	-0.237040
-0.910963	-0.349424	0.515142
-0.043835	1.017673	0.311334
-0.414088	0.075097	-0.119260
0.122283	-0.116679	-0.022701
-1.225455	1.081592	-0.021539
0.645043	-0.613121	-0.244938
0.129940	-0.053970	-0.024577
-0.316377	0.297125	0.156231
0.266541	1.178806	0.100504
-0.958049	-3.238358	-0.302918

Appendix 6

Min-Min Residuals of 2-Stage ANN Estimation (5000 epochs)
for Klein Model I

e(C)	e(Wp)	e(I)
0.062116	1.129910	0.038300
0.249245	0.053872	-0.054155
0.248013	-0.643339	-0.471605
0.168993	-0.302134	0.402436
0.057581	0.331087	-0.004532
-0.063848	0.347761	-0.000001
-0.558582	0.030347	-0.184448
-0.192689	0.010457	0.574360
0.610671	-0.476262	-0.701729
0.004467	0.235864	0.063952
0.251375	-0.145973	-0.016217
-0.055601	-0.080574	0.091278
-0.019398	0.615303	0.079599
-0.703521	-0.297773	-0.038236
0.513309	-0.055808	-0.179666
-1.007842	0.645541	-0.047772
0.809308	-0.715179	0.010137
0.071861	-0.178115	0.112594
-0.236024	0.063509	-0.093839
-0.118217	0.781506	0.158605
-0.235057	-1.229771	-0.178250

Appendix 7

Min-Max Residuals of 2-Stage ANN Estimation (5000 epochs)
for Klein Model I

e(C)	e(Wp)	e(I)
0.553999	1.380148	0.088436
0.580803	-0.028439	-0.155651
0.455254	-0.959138	-0.494438
0.254088	-0.203211	0.383461
-0.042783	0.167745	0.047748
-0.430265	0.131296	0.000001
-0.619685	0.178640	-0.114881
-0.230431	0.240025	0.577911
0.439304	-0.436284	-0.699063
0.526722	0.278768	-0.010620
0.155992	-0.582642	-0.103285
-0.053644	-0.066353	0.590821
-0.210110	0.740799	0.164535
-0.742751	-0.060602	-0.085934
0.436282	0.110829	-0.280275
-1.092551	0.880878	0.044553
0.813260	-0.623043	0.101022
-0.052223	-0.438239	0.069829
-0.012891	0.172034	-0.054948
-0.089325	0.764118	0.172824
-0.025768	-1.572427	-0.263300

Appendix 8

Max-Min Residuals of 2-Stage ANN Estimation (5000 epochs) for Klein Model I

e(C)	e(Wp)	e(I)
-0.000002	1.218117	0.038090
-0.291698	0.005653	-0.080600
0.089539	-1.038236	-0.745341
0.206534	0.074125	0.808249
0.460356	0.465101	-0.120296
-0.166933	0.003297	-0.312733
-0.444129	0.248891	-0.061857
-0.333092	0.167744	0.658472
0.317573	-0.613694	-0.983926
0.043972	0.283167	0.239033
0.196363	0.038837	-0.171935
-0.181924	-0.782532	0.414693
0.300900	0.465504	0.097843
-0.238326	-0.097847	-0.113709
0.329993	-0.115760	-0.054570
-0.713672	0.900275	0.054066
0.714345	-0.979927	-0.269446
-0.225887	-0.046630	0.030215
-0.451269	-0.057756	0.090918
0.544106	1.016193	0.117468
-0.000011	-1.723325	-0.078288

Appendix 9

Max-Max Residuals of 2-Stage ANN Estimation (5000 epochs)
for Klein Model I

e(C)	e(Wp)	e(I)
0.481748	1.330736	0.058435
0.184628	-0.053716	-0.093990
0.352979	-0.979940	-0.805108
0.336073	0.012849	0.891584
0.344561	0.393198	-0.099703
-0.499090	-0.053316	-0.294166
-0.643161	0.204088	-0.014160
-0.266637	0.150082	0.718235
0.356033	-0.580514	-0.986077
0.238644	0.260301	0.157189
0.435663	-0.031658	-0.114862
-0.561952	-0.169333	0.539826
-0.068497	0.750034	0.070008
-0.468131	-0.134374	-0.134349
0.246940	-0.242572	-0.084843
-0.788795	1.003570	0.032262
0.879453	-0.754307	-0.265257
-0.031475	-0.539565	0.013914
-0.338823	-0.055109	0.091847
0.254524	1.209293	0.151306
-0.668338	-1.68168	-0.021147

Appendix 10

Min-Min Residuals of 2-Stage ANN Estimation (10000 epochs) for Klein Model I

e(C)	e(Wp)	e(I)
-0.000002	0.723181	0.057511
0.208641	-0.070903	-0.092158
0.035634	-0.666288	-0.333336
0.074332	0.113306	0.291929
0.053656	0.407003	0.107788
0.019331	0.238676	0.000001
-0.354343	0.103357	-0.166295
-0.292744	0.193196	0.477417
0.373615	-0.594563	-0.731973
0.045158	0.254415	0.096862
0.166109	-0.081524	-0.104947
0.125414	-0.492245	0.379726
-0.026138	0.618269	0.038363
-0.536548	-0.208162	0.001915
0.522269	0.053395	-0.245009
-0.662982	0.350501	0.041065
0.817292	-0.858021	-0.006487
-0.237526	0.036988	0.130850
-0.166319	0.082083	-0.072048
-0.035957	0.443970	0.124021
-0.091335	-0.729016	-0.122857

Appendix 11

Min-Max Residuals of 2-Stage ANN Estimation (10000 epochs) for Klein Model I

e(C)	e(Wp)	e(I)
-0.000002	0.569923	0.084074
0.314213	-0.384761	-0.142667
0.329156	-0.599912	-0.500758
-0.023656	-0.125243	0.430328
0.005652	0.396299	0.031803
-0.062074	0.367578	0.000001
-0.56565	0.033319	-0.136818
-0.310273	0.182816	0.519713
0.600898	-0.420422	-0.738924
0.092182	0.041193	0.057670
0.149458	-0.264957	-0.096484
-0.084447	-0.167080	0.339554
-0.032063	0.879871	0.037860
-0.544479	-0.248785	-0.019008
0.637965	0.154687	-0.145178
-0.869968	0.517932	-0.027068
0.780602	-0.589157	0.121701
-0.024460	-0.194186	0.084917
-0.340387	0.079039	-0.138804
0.068802	0.659935	0.113322
-0.000011	-1.409879	-0.107594

Appendix 12

Max-Min Residuals of 2-Stage ANN Estimation (10000 epochs)
for Klein Model I

e(C)	e(Wp)	e(I)
-0.000002	1.371143	0.058930
-0.282940	0.211024	-0.076564
-0.019477	-0.788072	-0.703065
0.153822	-0.054797	0.753031
0.398074	0.419546	-0.149399
-0.194975	0.035074	-0.293776
-0.329772	0.153928	-0.104863
-0.232194	0.059811	0.613941
0.333530	-0.384213	-0.875565
0.023945	0.245569	0.173824
0.027505	0.033642	-0.122289
0.119246	-0.444721	0.542046
0.002251	0.450617	-0.010493
-0.303199	-0.317108	-0.182206
0.429328	-0.289939	-0.065969
-0.625014	0.708221	0.068312
0.764199	-1.068333	-0.285365
-0.082397	0.131524	0.067485
-0.333067	-0.034696	0.097396
0.128029	0.896816	0.077941
-0.484706	-1.394814	-0.002635

Appendix 13

Max-Max Residuals of 2-Stage ANN Estimation (10000 epochs)
for Klein Model I

e(C)	e(Wp)	e(I)
0.832742	1.019306	0.052095
-0.041916	-0.019273	-0.158488
0.208897	-0.911737	-0.768882
0.029036	0.179861	0.870799
0.396372	0.474332	-0.027972
-0.219717	-0.027357	-0.181903
-0.570766	0.253683	0.021222
-0.359124	0.248479	0.696631
0.510181	-0.619798	-1.005892
0.230164	0.166341	0.151674
0.069444	0.012150	-0.149997
-0.027719	-0.465693	0.538826
0.004425	0.769556	0.065354
-0.419090	-0.170105	-0.130351
0.153175	-0.241611	-0.058766
-0.672836	0.649131	0.089134
0.828309	-0.868774	-0.219017
0.105473	-0.076938	0.051488
-0.390565	-0.123678	0.047330
0.101387	1.003479	0.057652
-0.592921	-1.428986	-0.027808

Appendix 14

Min-Min Residuals of Modular ANN Estimation (1000 epochs)
for Klein Model I

e(C)	e(Wp)	e(I)
-0.000002	0.603565	-0.014728
0.081356	0.171218	-0.120236
-0.085139	-0.204626	-0.183856
0.012949	-0.187914	0.344976
0.036925	0.372816	-0.031605
-0.102731	0.139978	0.000001
-0.077392	-0.047060	0.100376
-0.159745	-0.098933	-0.047272
0.449340	-0.392098	-0.319323
-0.218460	0.282774	0.084471
0.196760	-0.054062	-0.095893
-0.506384	-0.110207	0.423931
0.176487	0.527143	-0.122048
-0.328716	-0.110132	-0.125192
0.604716	-0.047931	0.177792
-0.353816	0.086204	-0.114408
0.580025	-0.457920	0.008708
0.103474	-0.133144	0.052739
-0.616083	-0.134019	-0.095071
0.023201	0.949738	0.034695
-0.767421	-1.019505	0.007913

Appendix 15

Min-Max Residuals of Modular ANN Estimation (1000 epochs)
for Klein Model I

e(C)	e(Wp)	e(I)
0.129381	1.054781	-0.041023
0.046589	0.381211	-0.045182
-0.203296	-0.450582	-0.304437
0.069464	-0.182887	0.331285
0.361983	0.343881	-0.226285
-0.000719	0.177862	0.000001
-0.233890	0.021741	0.091873
-0.278954	-0.302928	0.020232
0.577529	-0.164185	-0.351823
-0.131111	0.300405	0.038732
-0.038321	-0.163269	-0.101643
0.085318	0.332499	0.651431
-0.108498	0.288279	-0.001917
-0.354087	-0.438406	-0.228015
0.389358	-0.406535	0.121549
-0.673572	0.486015	-0.083337
0.490880	-0.655544	0.216345
0.529198	0.106580	-0.006001
-0.512140	-0.136976	-0.089705
0.000946	1.043610	-0.074546
-0.977122	-1.536866	0.057993

Appendix 16

MaxMin Residuals of Modular ANN Estimation (1000 epochs)
for Klein Model I

e(C)	e(Wp)	e(I)
0.816213	0.587423	0.129159
0.432800	-0.250359	-0.244209
0.187661	-0.883939	-0.531542
0.485800	0.386316	0.896346
-0.387796	0.014261	-0.254696
-0.073820	0.531156	0.000001
-0.630447	0.189935	-0.019065
-0.199857	0.017579	0.049069
0.596656	-0.571633	-0.570529
0.276619	0.159860	0.070354
0.029035	-0.188839	0.034409
0.464625	-0.519922	0.341798
0.105244	0.507750	-0.093369
-0.939109	-0.139846	-0.148583
0.088012	0.482724	-0.010258
-0.899776	0.688415	0.070125
1.082635	-0.944027	0.010750
-0.070385	0.131646	0.070880
-0.185835	-0.460157	-0.189244
0.011162	0.827309	0.210215
-0.555835	-1.139451	-0.130650

Appendix 17

Max-Max Residuals of Modular ANN Estimation (1000 epochs)
for Klein Model I

e(C)	e(Wp)	e(I)
0.259367	1.243156	-0.005069
0.581043	0.052100	-0.105022
0.401665	-0.855047	-0.641505
0.223589	0.233904	0.936935
-0.153837	0.052000	-0.280264
-0.114645	0.209333	0.000001
-0.590793	0.270055	0.018823
-0.293565	0.238084	0.106316
0.499026	-0.770299	-0.730441
0.253338	0.307752	0.067264
0.279901	-0.039009	0.163869
-0.184671	-0.530016	0.371322
0.591663	0.444628	0.002940
-0.642184	-0.419923	-0.275334
-0.029153	-0.055015	-0.023066
-1.000140	0.814471	0.111379
0.683724	-0.775665	0.087309
-0.117313	-0.011455	-0.024367
-0.374998	-0.112226	-0.136135
0.846649	1.132874	0.323284
-0.075938	-2.077424	-0.351396

Appendix 18

Min-Min Residuals of Modular ANN Estimation (5000 epochs)
for Klein Model I

e(C)	e(Wp)	e(I)
0.611974	0.228693	0.018888
0.119835	0.136390	-0.115698
-0.148860	-0.274667	-0.156250
0.031077	0.001558	0.242256
0.074252	0.208894	-0.087944
-0.080747	0.194875	0.000001
0.018185	0.018747	0.165254
-0.123868	0.090100	-0.021626
0.187038	-0.408013	-0.249061
-0.079731	0.184396	-0.003646
0.052084	-0.014351	-0.013839
0.132788	-0.147249	0.245303
-0.007549	0.300745	0.001974
-0.537322	-0.077749	-0.097363
0.258002	0.073445	0.086402
-0.092853	-0.121842	-0.067265
0.196339	-0.125946	0.118385
0.141090	-0.175502	0.024785
-0.294386	0.007132	-0.100170
0.240379	0.281479	0.011908
-0.804195	-0.096951	-0.041381

Appendix 19

Min-Max Residuals of Modular ANN Estimation (5000 epochs)
for Klein Model I

e(C)	e(Wp)	e(I)
0.292741	0.372728	0.009595
0.141388	0.378352	-0.086600
-0.081596	-0.383432	-0.218257
-0.036672	-0.011771	0.302669
0.023756	0.235864	-0.171801
-0.170812	0.237638	- 0.000001
0.012658	0.011919	0.063523
-0.147038	-0.058925	0.007101
0.393916	-0.296292	-0.364731
-0.227669	0.097913	0.054439
0.087820	0.095665	-0.040231
-0.068140	0.001440	0.595848
0.225243	0.169271	-0.023272
-0.449580	-0.297836	-0.251366
0.330646	-0.018092	0.069505
-0.052116	-0.159276	0.009614
0.375748	-0.114597	0.059036
0.090092	-0.038028	-0.011725
-0.500562	-0.049471	-0.054887
0.226891	0.438633	0.045919
-0.762279	-0.731759	-0.056340

Appendix 20

Max-Min Residuals of Modular ANN Estimation (5000 epochs)
for Klein Model I

e(C)	e(Wp)	e(I)
-0.000002	-0.000006	0.078127
-0.296181	0.292579	-0.228636
0.284455	-0.064340	-0.200206
0.142134	0.160608	0.491481
-0.114282	-0.425734	-0.289754
-0.266267	0.356542	0.000001
-0.390414	-0.040499	-0.021877
-0.072770	0.116898	0.053286
0.753326	-0.224042	-0.412562
-0.224491	0.056730	0.074944
-0.088053	0.088840	-0.104568
0.047900	-0.156218	0.287874
-0.048260	0.054667	0.081865
-0.172919	0.104498	-0.021332
0.503204	-0.224910	-0.079248
-0.205054	0.106708	0.027847
0.471974	-0.507004	0.089468
0.108551	-0.022113	0.011401
-0.569322	0.350184	-0.035438
-0.086861	0.102875	0.034682
-0.000010	-0.238988	-0.112488

Appendix 21

Max-Max Residuals of Modular ANN Estimation (5000 epochs)
for Klein Model I

e(C)	e(Wp)	e(I)
-0.000002	1.154083	0.071174
-0.101730	0.106511	-0.159366
-0.227203	-0.895860	-0.571419
0.365652	0.440733	0.751692
0.020930	-0.023352	-0.331870
-0.094663	0.496683	0.000001
-0.344684	0.148836	0.006088
0.083850	-0.015670	0.063514
0.286544	-0.567315	-0.579277
-0.178413	0.458280	0.142740
-0.079813	-0.189888	-0.182311
0.413889	-0.052084	0.490136
-0.176765	0.342758	-0.004748
-0.485908	-0.691253	-0.215469
0.657664	0.156678	0.117645
-0.270145	0.125694	0.072041
0.624181	-0.804008	0.022326
-0.140583	0.311604	0.084184
-0.466634	-0.180036	-0.226288
0.233856	0.761299	0.172241
-0.845660	-1.145669	-0.068599

Appendix 22

Min-Min Residuals of Modular ANN Estimation (10000 epochs)
for Klein Model I

e(C)	e(Wp)	e(I)
0.253885	-0.000006	0.024519
-0.032352	0.175431	-0.133144
-0.019508	-0.110147	-0.188630
0.093245	-0.138804	0.240967
0.003390	0.100450	-0.091276
-0.134748	0.029073	0.000001
0.109284	0.090498	0.205106
-0.010579	0.097165	-0.032179
0.035609	-0.145531	-0.216011
-0.110825	0.029115	-0.000590
0.164793	-0.123943	-0.035892
0.070265	-0.146549	0.220213
-0.061707	0.297220	0.001303
-0.263305	-0.049535	-0.062260
0.093642	0.009064	0.102008
-0.009940	-0.110791	-0.040599
0.210027	0.113194	0.056014
0.004337	0.086457	0.022536
-0.261819	-0.175489	-0.097723
0.130081	0.115566	-0.024987
-0.539798	-0.376984	0.110514

Appendix 23

Min-Max Residuals of Modular ANN Estimation (10000 epochs)
for Klein Model I

e(C)	e(Wp)	e(I)
0.535539	-0.000006	0.055440
0.024570	-0.093844	-0.142596
-0.083507	-0.031522	-0.173226
-0.154462	-0.084132	0.245029
0.105322	0.267164	-0.151604
-0.027128	0.126897	0.000001
-0.011275	-0.106226	0.115537
0.026977	0.030549	0.067492
0.156474	-0.184354	-0.367931
-0.193970	0.090692	0.056200
0.092149	0.045265	-0.106592
-0.191713	-0.083891	0.452128
0.228374	0.137476	-0.059950
-0.347949	-0.102998	-0.148146
0.334773	0.158173	0.118539
-0.162292	-0.180653	-0.034309
0.346146	-0.184315	0.016583
-0.003811	0.009229	0.028692
-0.316827	-0.033861	-0.053760
0.184258	0.366695	-0.068567
-0.780734	-0.816590	0.307033

Appendix 24

Max-Min Residuals of Modular ANN Estimation (10000 epochs)
for Klein Model I

e(C)	e(Wp)	e(I)
-0.000002	0.060207	0.038813
0.183140	0.048585	-0.109421
-0.239547	-0.078733	-0.372685
0.176703	0.135412	0.495812
0.148905	-0.240462	-0.400224
-0.084779	0.268564	0.000001
-0.299155	0.017187	-0.069502
0.197192	-0.076286	0.040118
0.116634	-0.070233	-0.183529
-0.106171	0.095399	0.001786
0.021177	0.017471	-0.007948
0.221537	-0.143225	0.223761
-0.049713	0.319225	-0.131623
-0.528907	-0.111436	0.005153
0.443469	-0.075267	-0.033252
-0.231696	0.077747	0.073250
0.365067	-0.302738	0.112654
-0.057804	0.009572	0.032111
-0.290895	0.116000	-0.116552
0.098381	0.166386	-0.051101
-0.135379	-0.616635	0.088600

Appendix 25

Max-Max Residuals of Modular ANN Estimation (10000 epochs)
for Klein Model I

e(C)	e(Wp)	e(I)
0.299455	1.251431	0.084537
-0.122513	-0.039418	-0.102895
0.064935	-0.887960	-0.564079
0.015231	0.463167	0.489807
0.025778	0.022497	-0.251136
-0.008463	0.621255	0.000001
-0.399165	0.151247	0.111382
0.229209	-0.048301	-0.032062
0.103145	-0.667683	-0.245289
-0.008167	0.462507	0.021618
-0.074313	-0.100334	-0.038139
0.108065	-0.172043	0.318682
0.089115	0.035475	0.004662
-0.461684	-0.457977	-0.226685
0.549617	0.279149	0.071677
-0.151790	0.055453	0.144434
0.248357	-0.999252	-0.076589
-0.018353	0.362862	0.175749
-0.477575	-0.128099	-0.295685
0.398529	0.846004	0.212878
-0.663264	-1.221032	-0.064454

Appendix 26

Data for Estimation and Forecasting 1950-94

Year	DLP	DLY	DLK	DLC	DLWp	DLI
1949	-0.187463	-0.013859	0.015166	0.019758	-0.014742	-0.584414
1950	0.275980	0.110646	0.029429	0.084756	0.101850	0.685246
1951	-0.061500	0.187720	0.033206	0.148975	0.206876	0.152099
1952	-0.047932	0.082325	0.025444	0.068572	0.084448	-0.236952
1953	0.036220	0.071896	0.025847	0.073385	0.090335	0.041349
1954	0.025319	0.020847	0.022161	0.041702	0.001638	-0.129853
1955	0.264778	0.086541	0.033921	0.089702	0.097976	0.453765
1956	0.056022	0.105160	0.033767	0.082813	0.118122	0.029287
1957	0.003859	0.087112	0.029749	0.085536	0.083658	-0.094936
1958	-0.120886	0.052936	0.022857	0.056550	0.018989	-0.237240
1959	0.231974	0.073565	0.033725	0.081506	0.097060	0.417284
1960	-0.030911	0.058718	0.031712	0.057218	0.064821	-0.028826
1961	0.004469	0.053085	0.029502	0.042363	0.033910	-0.041623
1962	0.151500	0.073562	0.035978	0.071604	0.081367	0.231199
1963	0.086005	0.060779	0.038154	0.064207	0.063431	0.095797
1964	0.151208	0.084225	0.041942	0.086346	0.084485	0.134719
1965	0.199654	0.098249	0.050745	0.096352	0.094296	0.236896
1966	0.098571	0.112407	0.055340	0.109285	0.128060	0.139752
1967	-0.012300	0.103108	0.048848	0.087138	0.096480	-0.072712
1968	0.082343	0.136031	0.052884	0.136646	0.138017	0.130260
1969	0.010172	0.132774	0.058292	0.123114	0.145065	0.152973
1970	-0.063915	0.123393	0.049745	0.121159	0.109792	-0.104576
1971	0.225255	0.123212	0.062354	0.131139	0.113404	0.282041
1972	0.219404	0.135421	0.075166	0.134199	0.140700	0.255693
1973	0.319792	0.168097	0.091904	0.154546	0.175187	0.284707
1974	0.195026	0.179375	0.081888	0.175353	0.183028	-0.028571
1975	0.073311	0.171321	0.054667	0.189897	0.147135	-0.335972
1976	0.259401	0.157953	0.085441	0.166835	0.175387	0.516800
1977	0.236412	0.164633	0.115279	0.168517	0.178999	0.400141
1978	0.240646	0.192469	0.138914	0.182655	0.202182	0.313849
1979	0.198904	0.200432	0.141014	0.190938	0.205861	0.154991
1980	-0.015887	0.197519	0.107244	0.187931	0.183468	-0.149976
1981	0.037540	0.203438	0.130649	0.187794	0.191942	0.316590
1982	-0.207460	0.121068	0.081270	0.128585	0.110578	-0.369216
1983	0.218016	0.103439	0.091271	0.136565	0.101205	0.202405
1984	0.139361	0.139965	0.155126	0.124572	0.141566	0.654250
1985	-0.084149	0.104196	0.130248	0.115623	0.105980	-0.032400
1986	-0.113130	0.084479	0.111865	0.092932	0.084348	-0.031277
1987	0.390856	0.091362	0.104202	0.097749	0.100668	0.037000
1988	0.302114	0.109166	0.096369	0.115074	0.113679	0.022074
1989	-0.008226	0.114189	0.096763	0.111918	0.098556	0.100652
1990	0.153868	0.107528	0.078163	0.108138	0.101653	-0.126137
1991	0.079658	0.075235	0.051606	0.073734	0.065562	-0.350419
1992	0.116873	0.084810	0.057282	0.086903	0.082250	0.158818
1993	0.156833	0.070895	0.075556	0.081239	0.074804	0.343403
1994	0.172779	0.071800	0.093174	0.079793	0.077606	0.294085

Appendix 27

Residuals of Estimation and Forecast 1950-79
for Klein Model I

Year	e (DCL)	e (DLWp)	e (DLI)
1950	0.000738	-0.001426	-0.044557
1951	-0.005900	-0.015727	-0.004795
1952	0.002655	-0.006407	0.049802
1953	-0.001855	-0.000026	-0.006110
1954	0.005391	0.013401	-0.032477
1955	-0.000596	0.006221	0.005217
1956	0.004558	0.003442	0.010444
1957	-0.006058	0.008438	-0.048820
1958	-0.005485	0.004558	0.076789
1959	0.000269	-0.005541	0.006309
1960	0.003502	-0.003901	-0.028387
1961	0.007364	-0.008343	-0.022295
1962	0.000828	0.000759	0.008097
1963	-0.003809	-0.002379	0.024439
1964	-0.005001	0.005446	-0.021213
1965	-0.002188	0.000658	-0.000035
1966	0.005340	-0.000847	-0.013322
1967	0.002007	-0.007023	0.023174
1968	-0.004303	0.004135	-0.003249
1969	0.002267	-0.000148	0.014801
1970	0.012940	0.025829	0.135730
1971	-0.000378	0.010513	-0.214043
1972	-0.001172	0.023581	-0.020789
1973	-0.019488	-0.012935	-0.108934
1974	-0.035291	-0.017880	0.142211
1975	-0.052494	-0.008966	0.344684
1976	-0.029520	-0.010677	-0.340602
1977	-0.028325	0.004167	-0.158736
1978	-0.041593	-0.018065	-0.085826
1979	-0.049425	-0.025896	0.024245

Appendix 28

Residuals of Estimation and Forecast 1955-84
for Klein Model I

Year	e (DCL)	e (DLWp)	e (DLI)
1955	-0.002164	0.000850	-0.054964
1956	0.008566	-0.018810	-0.003124
1957	-0.003055	0.003654	0.013206
1958	-0.010855	0.008410	0.042159
1959	0.005842	0.001741	-0.011444
1960	0.004919	-0.005973	-0.005666
1961	0.005497	-0.001275	-0.030789
1962	0.001978	-0.002900	0.046374
1963	0.000778	0.000833	0.003482
1964	-0.004844	0.003507	0.002803
1965	-0.009025	0.001907	-0.000645
1966	0.004895	-0.005761	-0.008252
1967	0.005230	0.002955	-0.007493
1968	-0.000701	0.003696	-0.005368
1969	0.004181	-0.010749	0.013492
1970	-0.006080	0.011504	-0.001560
1971	-0.009788	0.016418	-0.004532
1972	-0.001036	-0.000700	-0.005859
1973	0.008199	-0.008168	0.005007
1974	-0.002859	-0.005288	-0.001485
1975	-0.016205	0.032386	0.236747
1976	0.004225	0.000990	-0.129273
1977	-0.003452	-0.009298	-0.029039
1978	-0.011043	-0.025313	-0.024846
1979	-0.018352	-0.027833	-0.013399
1980	-0.015337	-0.005429	0.291121
1981	-0.016259	-0.014977	0.071936
1982	0.041779	0.064944	0.750751
1983	0.030892	0.071159	0.199207
1984	0.041030	0.028847	-0.255537

Appendix 29

Residuals of Estimation and Forecast 1960-89
for Klein Model I

Year	e (DCL)	e (DLWp)	e (DLI)
1960	-0.000200	-0.009236	0.026840
1961	0.008192	0.013314	0.010655
1962	0.001587	-0.000044	-0.012142
1963	-0.003001	-0.001353	-0.033862
1964	-0.005114	0.002353	0.015326
1965	-0.002264	0.007701	-0.038580
1966	0.005827	-0.008211	0.028585
1967	0.003897	-0.011122	-0.005164
1968	-0.003586	-0.003833	0.002983
1969	0.011099	-0.007275	0.028255
1970	0.000777	-0.000895	-0.022644
1971	-0.011912	0.015829	-0.017203
1972	0.000539	0.003071	0.022240
1973	0.008463	-0.003099	0.046261
1974	0.003306	-0.006116	-0.035267
1975	-0.012639	0.013577	0.062828
1976	-0.006045	0.001905	-0.076288
1977	0.003030	0.004406	0.000249
1978	0.000850	-0.009013	0.002164
1979	-0.005053	-0.011839	-0.003422
1980	-0.001897	0.004902	-0.017246
1981	-0.006369	-0.003933	-0.114719
1982	0.056265	0.073582	0.169569
1983	-0.028199	0.016631	0.036731
1984	0.046016	0.044713	-0.196877
1985	0.062016	0.084903	0.472642
1986	0.043137	0.062843	0.344533
1987	0.024405	0.034136	0.272264
1988	0.052079	0.069402	0.428637
1989	0.059363	0.088249	0.356641

Appendix 30

Residuals of Estimation and Forecast 1965-94
for Klein Model I

Year	e (DCL)	e (DLWp)	e (DLI)
1965	-0.006206	0	-0.026946
1966	0.006265	-0.013497	0.060224
1967	0.007762	-0.002184	0.049135
1968	0.001355	-0.003643	-0.011668
1969	0.005260	-0.016184	0.058061
1970	-0.008497	-0.013782	0.008301
1971	-0.011580	0.006839	-0.055531
1972	-0.000302	-0.001895	0.032081
1973	0.003596	-0.008895	0.031953
1974	0.015585	-0.000077	-0.024380
1975	-0.012736	0.001817	-0.033250
1976	-0.006312	0.006377	0.047490
1977	0.002016	0.004752	-0.003330
1978	-0.002067	-0.011169	0.015631
1979	0	-0.010553	0.007109
1980	0.003007	0.000940	0.007239
1981	0.000153	0.005548	-0.014126
1982	0.002826	-0.006747	0.037783
1983	-0.015079	0.017497	-0.045004
1984	-0.000619	0.001877	-0.073358
1985	-0.006759	-0.005229	0.104969
1986	0.006731	0.009948	-0.014236
1987	0.023367	0.021972	0.203480
1988	-0.027936	-0.019383	0.288400
1989	0.028208	0.037959	0.015322
1990	0.081425	0.072223	-0.057132
1991	0.078814	0.056161	-0.018800
1992	0.026684	0.046921	0.361429
1993	0.024278	0.042868	0.122214
1994	0.008140	0.018685	0.122866

Appendix 31

Data for Estimation and Forecasting Klein-Goldberger Model

time	Y	C	I	G	Fe	Fi	W1	W2	P
1928	64.30	58.4				3.9	35.20	3.89	19.80
1929	69.73	62.2	14.9	7.8	5.0	4.1	37.78	4.23	23.00
1930	62.18	58.6	10.1	8.6	4.3	3.6	35.41	4.59	18.86
1931	54.08	56.6	5.9	9.3	3.6	3.2	32.40	5.19	13.80
1932	43.40	51.8	1.1	8.8	2.9	2.7	27.23	5.49	8.96
1933	42.75	51.1	1.6	8.6	2.9	2.8	26.44	5.90	7.90
1934	48.82	54.0	3.5	10.0	3.1	2.7	29.08	6.59	10.76
1935	57.18	57.2	6.7	10.0	3.3	3.3	31.07	6.91	14.19
1936	64.42	62.8	9.3	11.8	3.4	3.5	35.14	8.27	17.10
1937	70.64	65.0	11.4	11.3	4.1	4.0	38.85	7.59	18.65
1938	65.88	63.9	6.3	12.6	4.2	3.1	35.88	8.47	17.09
1939	71.58	67.5	9.9	13.0	4.2	3.3	39.27	8.55	19.28
1940	79.20	71.3	13.7	13.7	4.9	3.6	42.35	8.67	23.24
1941	93.80	76.6	17.1	20.8	5.0	4.0	49.13	9.58	28.11
1945	129.30	86.3	8.3			4.6	61.01	26.68	30.29
1946	114.48	95.7	20.3	18.9	7.5	4.0	61.88	14.85	28.26
1947	113.84	98.3	19.3	15.6	9.4	4.0	64.91	11.00	27.91
1948	120.32	100.3	22.7	17.9	7.2	4.6	66.75	10.87	32.30
1949	116.65	103.2	18.0	20.0	7.4	4.6	65.84	12.17	31.39
1950	126.35	108.9	26.8	19.0	7.1	5.4	70.85	12.72	35.61
1951	136.15	108.5	27.6	27.5	8.6	5.2	75.38	15.21	37.58
1952	138.14	111.4	24.3	33.4	8.3	5.5	78.65	16.82	35.17

Source: Klein and Goldberger (1955)

Appendix 31

Data for Estimation and Forecasting Klein-Goldberger Model (Contd.)

time	A	D	Pc	Sp	T	Tw	Tp	Tc	Ta
1928	4.47	9.50	6.16	0.84		-0.30	1.23	1.10	0.08
1929	4.69	9.84	6.98	1.02	6.33	-0.30	1.69	1.15	0.09
1930	3.35	9.81	4.30	-1.15	6.11	-0.37	1.18	0.73	0.10
1931	2.70	9.89	0.42	-3.95	8.33	-1.56	0.31	0.48	0.09
1932	1.75	8.96	-2.54	-5.67	9.54	-1.28	-0.13	0.40	0.08
1933	2.49	8.90	-2.69	-5.55	9.85	-1.25	0.06	0.58	0.05
1934	2.41	9.62	-0.01	-3.49	9.46	-1.21	0.32	0.78	0.02
1935	5.00	8.50	2.51	-1.40	8.22	-1.29	0.92	0.98	0
1936	3.95	9.18	4.24	-1.79	10.25	-2.10	1.69	1.42	0.02
1937	5.48	8.87	5.37	-0.64	8.39	0.38	2.14	1.46	0
1938	4.37	8.89	3.73	-0.45	9.23	0.19	1.62	1.03	0
1939	4.51	9.02	5.23	-0.01	10.71	0.16	1.70	1.46	0
1940	4.88	9.30	8.47	1.67	11.48	0.23	3.08	2.83	-0.02
1941	6.37	10.10	12.37	1.17	12.14	0.93	7.76	7.14	0
1945	8.96	12.07	11.50	0.21		8.72	10.96	7.95	0.54
1946	9.76	11.32	10.09	0.06	12.23	3.19	8.44	6.23	0.49
1947	9.32	13.46	12.08	1.19	11.72	3.99	9.17	7.04	0.45
1948	9.85	13.69	15.33	4.18	9.99	3.61	8.95	7.17	0.55
1949	7.31	14.75	14.09	3.92	12.60	2.50	7.27	6.03	0.36
1950	7.52	16.25	17.09	2.27	13.39	2.60	11.22	9.88	0.37
1951	7.98	17.06	18.72	2.23	13.79	7.15	14.64	11.88	0.33
1952	7.42	19.35	16.51	1.90	14.51	8.63	13.72	10.14	0.38

Source: Klein and Goldberger (1955)

Appendix 31

Data for Estimation and Forecasting Klein-Goldberger Model (Contd.)

time	B	K	Fa	pr	pa	pi	Np	N	Nw
1928	0	0		122.6	157				
1929	1.02	5.1	162	120.9	156	143.6	121.8	49.4	37.0
1930	-0.13	5.4	138	116.3	132	123.6	123.1	50.1	35.0
1931	-4.08	1.4	134	105.0	92	101.4	124.0	50.7	32.1
1932	-9.75	-6.5	118	94.2	68	82.9	124.8	51.3	28.8
1933	-15.30	-13.8	131	90.7	74	79.2	125.6	51.8	30.3
1934	-18.79	-19.9	102	95.5	95	90.7	126.4	52.5	33.5
1935	-20.19	-21.7	95	97.7	115	96.2	127.3	53.1	34.9
1936	-21.97	-21.6	88	98.3	120	96.6	128.1	53.7	37.9
1937	-22.61	-19.1	102	102.7	128	104.7	128.8	54.3	39.1
1938	-23.07	-21.7	116	100.9	102	97.1	129.8	55.0	37.8
1939	-23.07	-20.8	100	100.0	100	100.0	130.9	55.6	39.2
1940	-21.40	-16.4	72	101.5	105	99.2	132.0	56.2	40.9
1941	-20.23	-9.4	66	109.5	129	107.4	133.4	57.5	45.4
1945	-15.56	-11.4		140.3	217				
1946	-15.49	-2.4	174	152.6	246	158.8	141.4	61.0	49.2
1947	-14.31	3.4	183	168.3	289	183.0	144.1	61.8	49.3
1948	-10.13	12.4	157	180.6	300	189.8	146.6	62.9	50.2
1949	-6.21	15.6	186	179.4	262	180.2	149.0	63.7	48.9
1950	-3.94	26.1	155	183.6	269	196.5	151.7	64.7	50.7
1951	-1.71	36.6	184	197.5	318	244.5	154.4	66.0	54.7
1952	0.19	41.5	164	202.4	303	232.1	157.0	66.6	56.0

Source: Klein and Goldberger (1955)

Appendix 31

Data for Estimation and Forecasting Klein-Goldberger Model (Contd.)

time	Ng	Nf	Ne	h	w	il	is	L1	L2
1928					117.1		4.85	36.5	15.0
1929	3.6	5.6	4.8	1.158	118.4	5.21	5.85	33.6	16.0
1930	3.7	5.7	4.8	1.109	119.9	5.09	3.59	33.9	16.0
1931	4.1	5.8	4.8	1.079	114.1	5.81	2.64	37.6	14.9
1932	4.3	5.9	4.7	1.010	105.9	6.87	2.73	40.4	16.7
1933	5.8	6.0	4.6	1.032	93.8	5.89	1.73	39.7	16.2
1934	6.8	6.1	4.6	0.990	102.7	4.96	1.02	40.5	16.8
1935	7.1	5.9	4.7	1.030	103.3	4.46	0.76	42.1	17.1
1936	8.0	5.7	4.8	1.067	105.5	3.87	0.75	45.4	18.1
1937	7.1	5.5	4.9	1.040	117.2	3.94	0.94	45.2	17.1
1938	8.1	5.3	4.9	0.941	125.6	4.19	0.81	46.2	17.7
1939	7.9	5.2	5.0	1.000	122.1	3.77	0.59	49.6	19.4
1940	7.7	5.1	5.0	1.015	124.7	3.55	0.56	51.6	22.0
1941	8.5	5.0	4.9	1.054	134.5	3.34	0.54	54.3	23.7
1945					195.1		0.75	110.1	52.0
1946	9.2	4.8	5.5	1.097	217.2	2.74	0.81	108.3	43.4
1947	7.2	5.0	5.9	1.077	241.2	2.86	1.03	102.4	38.3
1948	7.2	4.7	6.1	1.059	263.7	3.08	1.44	96.5	35.7
1949	7.6	4.7	6.1	1.033	276.9	2.96	1.48	98.3	37.2
1950	7.8	4.4	6.2	1.056	286.9	2.86	1.45	97.9	38.4
1951	9.6	4.1	6.3	1.056	309.9	3.08	2.17	93.3	37.9
1952	10.4	4.0	6.3	1.057	326.2	3.19	2.33	95.2	38.1

Source: Klein and Goldberger (1955)

Appendix 32

Min-Min Residuals of Estimation and Forecasting KG Model with Full Set of Variables

Year	e (C)	e (Y)	e (Sp)	e (Pc)	e (D)	e (Wp)	e (Fi)	e (A)	e (L2)	e (I)
1929	-0.138942	0.326519	-0.175257	-0.034862	-0.050945	0.254737	-0.006565	-0.034566	0.009451	-0.124954
1930	0.754900	-0.421936	0.303316	-0.093037	-0.034092	-0.243973	0.018442	0.024768	-0.053043	0.394156
1931	-0.200464	-0.701876	0.019892	0.444676	0.069884	-0.606429	-0.033693	-0.052901	1.252817	-0.146593
1932	0.178580	1.011625	0.334252	0.335768	0.159983	0.100017	0.029716	0.333772	-0.452571	0.645815
1933	0.792105	1.594776	0.229912	0.464811	-0.060474	0.739558	-0.070968	-0.251933	0.077767	0.139570
1934	-0.130722	-0.923626	-0.138829	-0.417300	-0.318601	-0.060743	0.103571	0.130960	-0.191085	-0.558405
1935	1.212144	0.829266	-0.338633	-0.617306	0.266829	1.288261	-0.055498	-0.135636	-0.544410	-0.277868
1936	-0.498544	0.246832	-0.314545	-0.179624	0.038239	0.856902	0.040142	-0.039372	-0.669487	-0.058233
1937	-0.480728	0.058845	0.218143	0.399907	0.099830	-0.979906	-0.002654	0.152063	0.047938	0.039262
1938	-0.390932	-0.863871	0.204743	-0.044785	0.014562	-0.111075	0.044465	-0.019665	0.643819	0.702256
1939	-0.547592	-0.214224	0.384673	0.458333	0.053118	-0.736545	-0.001967	0.051112	0.221065	-0.403444
1940	0.130702	1.005917	-0.516313	0.029434	0.126072	0.112311	-0.031125	0.030949	-0.143652	-0.527216
1941	0.446280	-0.617528	0.202414	-0.381744	-0.104780	0.296079	0.001390	-0.051592	-0.012323	0.556681
1946	-0.350238	-0.676541	-0.047813	0.041543	0.028310	0.200917	-0.001792	-0.051224	-1.517752	0.066610
1947	1.056873	1.282856	0.173014	-0.093923	-0.086315	-0.571453	0.005281	0.025777	0.976466	-0.026654
1948	-0.168317	-0.992073	-0.345274	0.193731	0.020647	-0.159843	0.019658	-0.274022	0.003762	-0.172471
1949	-0.057887	0.416452	-0.451301	0.042549	0.077110	0.888653	-0.018036	0.019606	-0.437808	0.155117
1950	-1.736723	-1.154573	0.065602	-0.470276	-0.129704	-0.891351	-0.049266	0.010704	0.139757	-0.786628
1951	-2.361130	-11.460097	1.356991	-2.128957	-1.182948	-5.774173	0.090981	1.264614	-1.538584	-1.892336
1952	-11.651205	-17.076874	1.927192	-0.474941	-6.329169	-11.231902	-0.483044	2.340769	0.654372	0.004577

Appendix 33

Min-Max Residuals of Estimation and Forecasting KG Model with Full Set of Variables

Year	e (C)	e (Y)	e (Sp)	e (Pc)	e (D)	e (Wp)	e (Fi)	e (A)	e (L2)	e (I)
1929	-0.444099	-0.040783	-0.220848	0.199200	-0.041035	0.356776	-0.009420	-0.002198	0.034269	0.035465
1930	1.022574	-0.095119	0.465235	-0.203944	-0.034169	-0.209732	0.002712	0.040537	0.075884	0.262063
1931	-0.238905	-0.457090	-0.231734	0.334621	0.104857	-0.655684	-0.031899	-0.266649	1.421375	-0.154813
1932	0.300666	0.848432	0.398978	0.522847	0.144741	0.147850	0.043928	0.379866	-0.545526	0.807115
1933	0.996771	1.503178	0.305835	0.657705	-0.029668	0.858954	-0.056026	-0.132148	0.033849	0.281979
1934	-0.167793	-0.478493	-0.241929	-0.659416	-0.246177	0.110210	0.103593	0.001861	-0.406295	-0.695850
1935	1.369252	1.267891	-0.320763	-0.399544	0.268830	1.544681	-0.048273	-0.043265	-0.530710	-0.464506
1936	-0.504777	0.269766	-0.099588	-0.332789	-0.027578	0.411243	0.052534	-0.024443	-0.463140	-0.088090
1937	-0.588730	-0.266564	0.113627	0.275707	0.047133	-1.132612	-0.001435	0.071314	0.027871	0.115551
1938	-0.478323	-0.741069	0.221665	0.078281	0.024322	-0.143134	0.017878	0.003430	0.507048	0.574098
1939	-0.691788	-0.464399	0.335708	0.396510	0.031684	-0.728870	-0.004924	0.046937	0.080803	-0.149644
1940	0.655658	1.052464	-0.608958	-0.058333	0.042216	0.311377	-0.033561	0.000280	-0.312660	-0.309793
1941	0.132498	-0.616994	0.241771	-0.298842	-0.075432	0.239923	0.022584	-0.015447	0.223857	0.170580
1946	0.017262	-0.168087	-0.090846	0.266659	0.113657	0.284127	-0.033534	-0.077196	-1.931769	0.514947
1947	-0.113889	0.268376	0.254825	-0.374521	-0.210865	-0.557835	0.038955	0.165268	1.911464	-0.470668
1948	0.449084	-0.678803	-0.407121	-0.223418	0.050322	-0.205376	0.028775	-0.381814	-0.061210	-0.264835
1949	-0.011684	0.780161	-0.384725	0.480026	0.157236	0.712872	-0.032726	0.003187	-0.375850	0.261862
1950	-1.571318	-1.091669	0.037414	-0.389374	-0.190022	-0.83481	-0.058136	0.035310	-0.047293	-0.583476
1951	-2.252556	-11.337706	1.295190	-2.143975	-1.275095	-5.682734	0.088220	1.061181	-1.751677	-1.756609
1952	-9.669913	-15.735428	1.802189	-0.821517	-5.835299	-11.780775	-0.538647	2.311465	1.207789	0.698898

Appendix 34

Max-Min Residuals of Estimation and Forecasting KG Model with Full Set of Variables

Year	e (C)	e (Y)	e (Sp)	e (Pc)	e (D)	e (Wp)	e (Fi)	e (A)	e (L2)	e (I)
1929	0.038617	0.129711	0.050505	0.035511	0.130049	0.035514	-0.003752	0.031073	1.063482	-0.039824
1930	0.287615	0.750172	-0.072357	0.177016	0.058188	0.307903	0.016173	-0.025651	-0.160261	0.178414
1931	-1.810125	-3.553099	0.290357	-0.226785	-0.338032	-2.016369	-0.126242	0.035082	0.843343	-0.559575
1932	0.763446	2.200758	0.673619	0.828144	0.220756	0.755245	0.101169	0.519368	-1.288758	1.191668
1933	1.547964	3.046185	0.598690	1.052827	0.296524	1.642952	0.010549	-0.194612	-0.772206	0.804935
1934	-1.039673	-2.312653	-1.288811	-1.353837	-0.364991	-0.642864	0.146122	-0.029306	-1.315878	-0.672179
1935	0.748261	1.579602	-0.140998	-0.010378	0.381513	1.734321	-0.012647	-0.352284	-0.335926	0.138790
1936	1.498668	3.226385	1.063132	0.419864	-0.280605	1.686469	-0.089289	0.779520	0.395935	-0.260087
1937	1.373421	1.294639	0.458945	0.610055	0.115067	-0.214109	-0.419399	-0.566041	2.097269	-0.271434
1938	-0.345225	-2.826548	-0.248707	-0.573209	-0.009027	-0.559432	0.132508	-0.228966	0.038501	0.154400
1939	-2.783951	-2.870596	-0.611355	-0.236810	-0.104721	-2.030571	0.152368	0.290581	-0.727766	-0.321067
1940	0.019473	-0.229106	-0.693750	-0.457944	-0.063846	-0.118131	0.267193	0.015198	-1.651123	0.523485
1941	-0.119066	0.035098	0.035003	0.054153	-0.007877	0.259698	-0.020990	-0.028288	0.037072	0.031483
1946	-0.064044	-0.244008	-0.019183	0.112946	0.008035	0.479119	-0.005903	-0.109755	-1.483481	0.096544
1947	0.230685	0.663403	0.016871	-0.150230	0.019101	-0.712399	0.017713	-0.136925	1.022773	-0.027303
1948	2.975917	0.340210	-0.872682	0.260308	0.987374	1.195702	0.294590	-1.450560	1.178353	0.717192
1949	-0.944553	0.681657	-0.228533	0.159290	-0.415252	-0.119716	-0.110765	0.498672	-0.735316	-0.169048
1950	-3.336493	-3.426271	1.005529	-0.792532	-0.873071	-1.447359	-0.257440	0.867065	-0.796111	-1.222911
1951	-2.961403	-13.247069	1.042973	-2.428799	-1.690219	-5.989749	-0.059626	0.409489	-0.303496	-2.036951
1952	-6.343947	-15.639741	1.340537	-0.336345	-4.123182	-9.503928	-0.401455	1.006031	-0.663656	0.974204

Appendix 35

Max-Max Residuals of Estimation and Forecasting KG Model with Full Set of Variables

Year	e (C)	e (Y)	e (Sp)	e (Pc)	e (D)	e (Wp)	e (Fi)	e (A)	e (L2)	e (I)
1929	0.170411	0.066669	0.043619	0.022633	0.061912	0.117125	0.014509	0.071015	0.474667	0.140237
1930	0.023844	-0.070572	0.031802	0.030682	0.024734	-0.020683	0.013757	0.070651	-0.145980	-0.142109
1931	-2.520784	-3.999914	0.122489	-0.563554	-0.508691	-2.648470	-0.189835	0.109135	1.217777	-1.233897
1932	1.020015	3.101179	0.729589	1.070323	0.297774	0.950693	0.120023	0.460718	-0.633168	1.419315
1933	1.782915	3.923023	0.657613	1.282954	0.365244	1.817481	0.028059	-0.252123	-0.130536	1.012071
1934	-0.905231	-1.559742	-1.232692	-1.182026	-0.330929	-0.561403	0.156541	-0.077204	-0.721766	-0.560328
1935	1.377526	1.896984	-0.259729	-0.249185	0.469182	2.191707	-0.035394	-0.522061	-0.216362	0.549689
1936	2.207034	3.576895	1.376702	0.616533	-0.236711	1.789005	-0.070154	0.681238	0.569081	-0.265698
1937	1.944588	1.671302	0.696558	0.896397	0.190535	0.213709	-0.398741	-0.628014	2.040528	-0.105325
1938	-1.462801	-3.503405	-0.642582	-0.842379	-0.067390	-1.579169	0.113489	-0.039422	0.356707	-0.079567
1939	-1.973267	-2.422072	-0.273844	0.013702	-0.047423	-1.775604	0.175572	0.182185	-0.603405	-0.259591
1940	-1.730084	-1.298244	-1.036731	-0.489135	-0.051518	-0.411459	0.226567	0.232565	-2.252552	0.434772
1941	0.319037	0.662334	-0.037452	-0.219768	-0.071277	0.398042	-0.004610	-0.108193	0.175685	-0.008381
1946	-0.203670	-0.284878	-0.240629	0.106978	0.210412	0.310983	-0.050336	-0.086357	-1.717299	0.269582
1947	0.624675	0.808872	0.610852	-0.179526	-0.335514	-0.598103	0.103853	-0.027320	1.854991	-0.431248
1948	3.808948	0.700760	-0.899309	0.519399	1.279708	1.033897	0.302411	-1.582447	1.168690	0.758979
1949	-1.498607	-0.469397	-0.705016	-0.006414	-0.271541	0.321095	-0.153063	0.611597	-1.203811	0.069426
1950	-3.368681	-3.036104	1.051645	-0.626005	-0.909419	-2.082552	-0.278627	1.038771	-0.448141	-1.413291
1951	-2.982384	-12.855856	1.091040	-2.260951	-1.722899	-6.621516	-0.080575	0.576543	0.042532	-2.226953
1952	-6.132132	-15.212334	1.412710	-0.142179	-4.079383	-10.060347	-0.415803	1.084641	-0.364877	0.805356

Appendix 36

Min-Min Residuals of Estimation and Forecasting KG Model with ANN Topology Selected by GA

Year	e (C)	e (Y)	e (Sp)	e (Pc)	e (D)	e (Wp)	e (Fi)	e (A)	e (L2)	e (I)
1929	-0.208881	-0.039341	0.083880	-0.012435	-0.146688	-0.011693	0.013673	0.001167	0.070116	-0.007078
1930	0.856806	-0.156624	0.007805	-0.138277	0.033652	0.191276	0.009085	-0.054939	-0.144755	0.?19928
1931	-0.412267	-0.815809	0.054170	0.392027	0.062098	-0.614468	-0.048243	-0.066007	0.993724	-0.265781
1932	0.209186	1.013425	0.382288	0.371657	0.117161	0.116672	0.034787	0.323290	-0.647550	0.669361
1933	1.003725	1.780010	0.213835	0.493074	0.068451	0.904713	-0.065009	-0.331226	0.233928	0.173229
1934	-0.262573	-1.117203	-0.257823	-0.470071	-0.483143	-0.339315	0.119140	0.185800	-0.188086	-0.469101
1935	1.071969	0.934620	0.030529	-0.369742	0.246637	0.973934	-0.063484	-0.063196	-0.450698	-0.006207
1936	-0.178132	0.592039	-0.125275	-0.153869	0.077332	0.967170	0.028183	-0.039965	-0.345052	0.047560
1937	-0.502022	0.235748	-0.126428	0.284987	0.158417	-0.827070	-0.010138	0.104719	0.032666	-0.034880
1938	-0.431506	-0.984492	0.136753	-0.088955	0.029559	0.010640	0.055693	-0.033132	0.476222	0.516542
1939	-0.489601	-0.325415	0.102200	0.277781	0.071490	-0.678970	0.033436	0.026273	0.210424	-0.532917
1940	0.074931	0.765630	-0.106856	0.260773	0.109905	-0.215761	-0.058630	0.049909	0.009584	-0.269716
1941	0.384001	-0.546437	0.074589	-0.409207	-0.080044	0.388780	0.007331	-0.045210	-0.105607	0.431286
1946	-0.204288	-0.795765	0.007756	-0.023668	0.010791	0.103810	0.006850	-0.045870	-1.196997	0.048631
1947	0.702724	1.586178	0.053021	0.043126	-0.058042	-0.358372	-0.009372	0.042967	0.568507	0.035691
1948	-0.206631	-1.036286	-0.383346	0.176046	0.036140	-0.102036	0.015720	-0.313612	-0.010841	-0.266206
1949	0.230397	0.465341	-0.303315	-0.041206	0.008812	0.699246	-0.008390	0.020820	-0.301757	0.164137
1950	-1.661070	-1.143396	0.073750	-0.484526	-0.110218	-0.904726	-0.052600	0.020731	0.113138	-0.770951
1951	-1.429993	-11.011038	0.681737	-2.102950	-0.960021	-5.473803	0.134746	0.495685	0.497224	-1.609186
1952	-4.922842	-13.696664	1.372895	-0.130693	-3.363281	-9.403136	-0.202060	-0.119003	0.462233	1.300224

Appendix 37

Min-Max Residuals of Estimation and Forecasting KG Model with ANN Topology Selected by GA

Year	e (C)	e (Y)	e (Sp)	e (Pc)	e (D)	e (Wp)	e (Fi)	e (A)	e (L2)	e (I)
1929	-0.169677	0.170032	0.026339	0.036515	-0.148299	0.185023	-0.020267	-0.078829	0.074659	-0.169815
1930	0.669756	0.046048	0.136753	-0.001576	0.029734	-0.236046	0.044182	0.122577	-0.067622	0.637653
1931	-0.641973	-1.380030	-0.156104	0.200471	0.078027	-0.946759	-0.061096	-0.181168	0.997407	-0.575927
1932	0.527755	1.479749	0.531271	0.466177	0.098678	0.466198	0.064997	0.382991	-0.598552	0.717877
1933	1.386107	2.295166	0.395145	0.627680	0.087444	1.343859	-0.031182	-0.287802	0.165870	0.224222
1934	0.094330	-0.546440	-0.627787	-0.662442	-0.332914	-0.167160	0.115292	0.021852	-0.284909	-0.444109
1935	0.726796	0.941456	0.095307	0.013918	0.306331	1.226982	-0.033529	0.063419	-0.565755	-0.005707
1936	-0.370843	0.576613	-0.037038	-0.142718	0.070604	0.743797	0.009867	-0.051234	-0.386095	0.063204
1937	-0.443344	-0.000054	-0.201482	-0.052117	0.063072	-0.437509	-0.001101	0.051082	0.010355	-0.139679
1938	-0.203993	-1.638987	0.180931	-0.198208	-0.040713	-0.523009	0.007361	-0.119976	0.721776	0.575722
1939	-0.090271	0.022470	0.197041	0.315616	-0.004259	-0.561771	0.047397	0.024478	0.322242	-0.290069
1940	-0.411053	0.598546	-0.288536	0.128843	-0.007424	-0.064588	-0.064627	0.113580	-0.469971	-0.686622
1941	0.642409	-0.328664	0.140574	-0.160461	0.006354	0.319860	0.025723	-0.059302	0.128279	0.638808
1946	0.261166	-0.467869	-0.122026	0.034276	0.010649	0.171125	-0.007090	-0.033840	-1.820918	0.061919
1947	-0.531262	1.285992	0.332639	0.001133	-0.070850	-0.448322	0.004081	0.117728	1.520755	0.040952
1948	0.154140	-1.447892	-0.613902	-0.140859	0.058587	-0.432526	0.015505	-0.413951	-0.003722	-0.409522
1949	0.323628	1.046831	-0.409760	0.087690	0.028681	1.201809	-0.014777	0.030998	-0.459403	0.228487
1950	-1.190573	-2.032091	0.101080	-0.401419	-0.135176	-0.999757	-0.039659	0.034212	0.096262	-0.832938
1951	-0.933029	-11.88114	0.437287	-2.041447	-0.990069	-5.614184	0.151143	0.276164	0.515485	-1.639932
1952	-4.280875	-14.79036	1.389514	0.002426	-3.329496	-9.239836	-0.203997	-0.190485	-0.364725	0.734248

Appendix 38

Max-Min Residuals of Estimation and Forecasting KG Model with ANN Topology Selected by GA

Year	e (C)	e (Y)	e (Sp)	e (Pc)	e (D)	e (Wp)	e (Fi)	e (A)	e (L2)	e (I)
1929	-0.427325	0.030193	0.061796	0.046188	-0.125200	-0.038713	-0.013420	0.007367	-0.153662	0.094052
1930	0.941996	-0.004604	-0.014935	-0.100882	0.060462	0.262974	0.021602	-0.062942	-0.266215	0.149832
1931	-0.605367	-0.078589	0.115263	0.561972	0.077251	-0.635716	-0.036197	-0.029928	1.005743	-0.157080
1932	0.131419	0.724619	0.383799	0.383732	0.053895	-0.024388	0.034234	0.252545	-0.413591	0.663721
1933	0.993575	1.713459	0.292657	0.650039	0.154631	0.931937	-0.054642	-0.424134	0.113080	0.316956
1934	-0.429447	-0.791977	-0.388234	-0.855512	-0.524165	-0.405233	0.108788	0.217820	-0.408000	-0.457098
1935	0.956162	0.479451	-0.085525	-0.258664	0.328725	1.246376	-0.013460	-0.007857	-0.635025	-0.017113
1936	-0.146874	0.118696	-0.116504	-0.268894	-0.119198	0.600490	0.002038	0.039053	-0.513069	0.049632
1937	0.019318	0.019317	-0.025346	0.383139	0.151401	-0.751505	-0.030479	0.036600	0.254126	-0.057435
1938	-0.527945	-0.453959	0.235000	-0.315316	0.087606	-0.190940	0.052525	-0.102496	0.522494	0.257591
1939	-0.358627	-0.448362	0.133799	0.156865	0.041070	-0.658084	0.054150	0.023129	0.466442	-0.388976
1940	0.215831	0.391522	-0.139763	0.289083	0.072753	0.146117	-0.074814	-0.005266	-0.153559	-0.145052
1941	0.018416	-0.063687	0.033836	-0.346805	-0.041214	0.236978	0.036595	-0.002153	-0.061717	0.231675
1946	0.433224	-0.216526	-0.052645	0.099428	0.008389	0.173818	-0.006809	-0.077658	-1.302267	0.248721
1947	-0.774358	0.395138	0.167888	-0.082453	-0.038269	-0.476299	-0.002012	0.088865	0.755115	-0.228246
1948	0.139796	-0.863633	-0.351870	-0.133967	0.025528	-0.218933	0.032999	-0 333497	-0.090858	-0.237264
1949	0.968556	1.252496	-0.461700	0.379956	-0.011905	0.949116	-0.026206	0.052340	-0.345725	0.317713
1950	-1.271597	-2.189409	0.078562	-0.759333	-0.109941	-1.030580	-0.024569	0.002215	0.156252	-0.894753
1951	-1.120895	-12.035979	0.695929	-2.467885	-0.969092	-5.682795	0.164757	0.608769	1.082105	-1.773923
1952	-5.868513	-15.994462	1.659629	-0.968328	-3.918516	-10.255873	-0.438963	1.635928	1.597830	-0.042985

Appendix 39

Max-Max Residuals of Estimation and Forecasting KG Model with ANN Topology Selected by GA

Year	e (C)	e (Y)	e (Sp)	e (Pc)	e (D)	e (Wp)	e (Fi)	e (A)	e (L2)	e (I)
1929	-0.746577	-0.480900	-0.116596	0.083533	-0.129589	0.072463	-0.058039	0.031904	-0.138955	-0.174461
1930	0.859030	0.389382	0.292758	0.048649	0.053784	-0.060547	0.077597	-0.076274	-0.095744	0.453741
1931	-0.856195	-1.257082	-0.179330	0.134856	0.082993	-0.906390	-0.035481	-0.115610	0.959663	-0.462798
1932	0.430305	1.246164	0.452532	0.452106	0.055424	0.395547	0.042521	0.325255	-0.485870	0.666499
1933	1.353297	2.321468	0.350522	0.732323	0.172752	1.392417	-0.043247	-0.342400	0.044429	0.312589
1934	-0.071434	-0.728905	-0.575736	-0.987684	-0.653505	-0.167317	0.118645	0.155863	-0.566499	-0.368809
1935	1.647248	1.646859	0.034916	-0.166298	0.341405	1.485496	-0.040245	-0.057304	-0.603122	0.095964
1936	-0.526070	0.195410	-0.125803	-0.227515	-0.065765	0.682842	-0.012676	0.066102	-0.375141	-0.161338
1937	0.052910	0.231254	-0.034379	0.153218	0.174131	-0.481329	-0.001095	0.055171	0.197493	0.035122
1938	-0.142855	-1.235499	0.206378	-0.107119	-0.006166	-0.796424	0.028056	-0.043099	0.616044	0.633775
1939	-0.619972	-0.848532	0.147823	0.216151	-0.031596	-0.964019	0.039472	0.022172	0.499090	-0.403864
1940	-0.015569	0.689847	-0.176890	0.264221	-0.029594	0.639529	-0.051871	0.004171	-0.199961	-0.455577
1941	0.208647	0.100154	0.019317	-0.332745	0.006659	0.009427	0.030575	-0.021404	-0.117028	0.489865
1946	0.349110	-0.518597	-0.069927	-0.037381	0.017766	0.362420	-0.011053	-0.128000	-1.467944	0.095218
1947	-0.968129	0.720074	0.236809	0.266775	-0.041018	-0.712238	-0.001789	0.161915	0.871082	-0.080737
1948	0.645015	-1.292092	-0.552389	-0.446213	0.019834	-0.334167	0.037565	-0.348034	-0.092204	-0.318338
1949	0.804669	1.498727	-0.385021	0.149796	0.061729	1.134068	-0.019909	0.029436	-0.320956	0.235685
1950	-1.341055	-2.079027	0.099487	-0.640455	-0.253391	-0.927637	-0.053929	0.011887	0.219873	-0.754025
1951	-1.240425	-12.163359	0.296282	-2.431761	-1.203291	-5.632937	0.128329	0.447471	0.857221	-1.712347
1952	-5.963164	-16.316170	0.943641	-1.257705	-4.198991	-10.109045	-0.452100	1.207051	1.350581	0.012796

Appendix 40

Min-Min Residuals of Estimation and Forecasting of KG Model with Union Set of Modular Variables Selected by GA

Year	e (C)	e (Y)	e (Sp)	e (Pc)	e (D)	e (Wp)	e (Fi)	e (A)	e (L2)	e (I)
1929	-0.774856	-0.406865	-0.065980	0.030974	-0.037104	0.190783	-0.012362	-0.010417	-0.337061	0.023469
1930	0.669520	-0.541420	-0.001142	-0.243639	0.084369	-0.253822	-0.017993	0.053389	-0.538642	-0.412793
1931	-0.115305	-0.225611	-0.067511	0.181075	-0.072519	-0.430935	-0.011422	-0.117652	0.645398	-0.207864
1932	-0.106011	0.531587	0.116089	0.058583	-0.008244	-0.227515	0.026638	0.335370	-0.118707	0.504852
1933	0.599509	1.265469	0.001856	0.225886	0.008415	0.593695	-0.069500	-0.328040	0.097382	0.045656
1934	-0.144497	-0.232506	-0.080040	-0.417527	-0.038401	-0.404064	0.101897	0.097483	-0.082183	-0.019176
1935	0.193288	-0.398952	-0.061160	0.020077	0.167631	0.083461	-0.015025	-0.058619	-0.036131	-0.257333
1936	-0.787240	0.047911	-0.064702	-0.045977	-0.072991	-0.034489	-0.032492	0.056816	0.030108	0.225584
1937	-0.074928	-0.095299	-0.071535	-0.301982	-0.118444	-0.395219	-0.002301	0.034954	0.099480	-0.050379
1938	0.045442	-0.322024	-0.101340	-0.061926	0.046535	0.664304	-0.003978	-0.130666	-0.130483	0.355685
1939	-0.518768	-0.398428	0.066692	0.110413	-0.043079	-0.572967	-0.004794	0.163672	-0.135281	-0.610119
1940	0.112468	-0.025348	-0.126471	-0.099027	0.017945	-0.228754	-0.011303	-0.088012	0.024109	0.027591
1941	-0.077333	-0.129231	-0.000470	-0.035281	0.006107	-0.054683	-0.009164	0.007151	-0.022802	-0.097904
1946	0.352826	-0.135670	0.038711	0.143184	0.023906	0.334279	0.031037	-0.118223	-0.493491	0.067611
1947	0.124995	0.602986	0.032754	0.120143	0.020430	0.033352	0.026770	0.149480	0.054828	0.272790
1948	0.293834	0.455017	-0.130933	0.075544	-0.000094	-0.175240	0.016149	-0.181122	0.081521	0.098561
1949	-0.013759	0.162462	-0.007888	0.060301	0.016946	0.239689	0.022470	0.012496	0.076878	0.173775
1950	-0.308401	-1.012544	0.026764	-0.287300	-0.076511	-0.188646	-0.022515	-0.016442	0.075494	-0.322598
1951	0.091042	-10.823874	0.085297	-1.919776	-0.887578	-4.720904	0.177353	-0.520912	0.611108	-1.12916
1952	-2.871329	-13.043236	0.684521	0.230158	-3.209438	-8.048361	-0.129490	-0.132894	0.412711	2.034894

Appendix 41

Min-Max Residuals of Estimation and Forecasting of KG Model with Union Set of Modular Variables Selected by GA

Year	e (C)	e (Y)	e (Sp)	e (Pc)	e (D)	e (Wp)	e (Fi)	e (A)	e (L2)	e (I)
1929	-0.739467	-0.431653	0.076464	0.028721	-0.022036	0.084464	-0.018056	-0.034558	0.246416	-0.102531
1930	0.579146	-0.650982	0.135616	-0.144052	-0.048540	-0.610256	0.092681	0.078003	-0.295880	-0.662365
1931	0.042651	-0.888258	-0.041443	0.445834	-0.042084	-0.614596	0.059213	-0.162932	0.769712	-0.274396
1932	0.049472	0.438097	0.190618	0.164045	0.023714	-0.166630	0.007685	0.375562	-0.345287	0.504052
1933	0.848544	1.229042	0.050655	0.342309	0.131141	0.711724	-0.088819	-0.338610	0.167365	0.100145
1934	0.439518	-0.034935	-0.046746	-0.396066	-0.317784	0.329277	0.101627	-0.060084	-0.945159	-0.364673
1935	0.553361	2.042466	-0.081836	-0.038198	0.473273	1.501629	-0.126176	0.048334	-0.759492	0.691469
1936	0.489627	0.958830	0.007432	-0.385302	-0.043467	1.154117	-0.033197	0.185196	-0.477514	-0.573267
1937	-0.422676	0.301162	-0.256023	0.128631	0.319604	-0.535493	-0.110302	-0.062851	0.904345	1.413690
1938	-0.782717	-1.161166	0.056093	0.117372	0.014033	-0.491713	-0.003970	-0.153571	0.554223	0.532790
1939	-0.514236	-0.853598	0.105900	0.203228	-0.047915	-0.935314	0.039138	0.012637	0.184505	-0.647263
1940	0.778056	0.499898	-0.002938	-0.100054	-0.173596	0.463610	0.072414	0.029563	-0.440710	-0.354555
1941	-0.032632	-0.004277	0.012241	0.021137	0.003192	-0.012530	-0.003272	0.003181	0.010730	0.012411
1946	0.003812	-0.488636	-0.081075	0.166571	0.038113	0.356450	0.032954	0.009239	-1.856989	0.166839
1947	-0.226277	-0.342953	0.296156	-0.415380	-0.140779	-1.108036	-0.015558	-0.028654	1.960601	-0.783450
1948	0.932559	0.362861	-0.245257	0.053805	-0.068615	0.064835	0.065117	-0.299058	0.030148	-0.488729
1949	-0.652919	1.070421	-0.559416	0.459160	0.356171	1.156155	-0.078138	0.068286	-0.434966	1.123974
1950	-0.559714	-1.023088	0.058829	-0.451071	-0.200949	-0.331895	-0.003922	-0.014504	-0.045309	-0.779357
1951	-0.158485	-10.830481	0.052295	-2.085953	-1.008959	-4.862132	0.196070	-0.515515	0.495786	-1.585043
1952	-3.053221	-12.828400	0.307804	0.118345	-3.294223	-8.130560	-0.103902	-0.038482	0.352820	1.707980

Appendix 42

Max-Min Residuals of Estimation and Forecasting of KG Model with Union Set of Modular Variables Selected by GA

Year	e (C)	e (Y)	e (Sp)	e (Pc)	e (D)	e (Wp)	e (Fi)	e (A)	e (L2)	e (I)
1929	-0.923911	0.327732	0.068679	0.167410	-0.090371	-0.060628	-0.019212	0.038798	-0.519395	0.058317
1930	1.631339	-0.337318	-0.037142	-0.112657	0.049777	0.139244	0.049705	-0.224197	-0.235078	-0.500778
1931	-0.504144	-0.462908	0.012766	0.387624	0.040359	-0.858937	-0.051010	-0.072624	1.034511	0.506101
1932	-0.042195	0.451616	0.158300	0.212322	-0.062618	-0.139054	0.021378	0.340407	-0.206973	0.435454
1933	0.767867	1.269409	0.019312	0.394336	0.103952	0.789036	-0.075061	-0.438690	0.380723	0.019860
1934	0.316841	-0.007344	-0.034416	-0.442038	-0.091810	0.122192	0.102240	0.079305	-0.725137	-0.483455
1935	0.926743	0.706742	0.054943	-0.247327	0.252098	0.820244	-0.054440	0.059522	0.065781	-0.320293
1936	-0.791592	0.207510	-0.154486	-0.516118	-0.043723	0.851924	0.032872	0.164851	-0.971603	0.064388
1937	-0.078384	-0.319283	0.030323	0.220905	-0.040998	-0.560835	-0.004347	-0.025879	0.418641	0.109446
1938	-0.479319	-0.691073	0.114610	0.228888	0.169910	-0.227408	0.032871	-0.034194	0.361777	0.845549
1939	0.352814	0.442438	-0.067909	0.161913	0.096349	-0.172931	-0.040444	-0.008604	0.164556	-0.663044
1940	-0.124013	-0.079066	-0.000049	-0.103393	-0.187471	-0.050024	0.011463	-0.036879	-0.162336	0.200686
1941	0.048766	0.037578	-0.000393	0.010777	-0.011934	0.102178	0.000159	-0.000196	0.010924	-0.018192
1946	0.116582	-0.041295	0.013448	-0.018460	-0.062653	0.062318	0.012632	-0.149798	-1.387511	-0.082442
1947	-0.693410	0.168995	0.061580	0.176528	0.000766	-0.299076	-0.017592	0.104903	0.131656	0.028281
1948	1.229045	-0.641739	-0.470559	-0.390393	0.093854	0.125633	0.021346	-0.289741	0.204297	0.055234
1949	-0.443500	0.411607	0.055253	0.152295	-0.030685	0.146176	-0.009066	0.018998	0.057965	0.024645
1950	-0.468779	-0.526735	0.029117	-0.375362	-0.031399	-0.118929	-0.061971	0.022999	-0.012812	-0.195586
1951	-0.099342	-10.358038	0.149338	-2.018296	-0.844800	-4.658748	0.134194	-0.350765	0.388193	-1.010179
1952	-3.620993	-12.959611	1.280348	-0.020094	-3.229831	-8.156233	-0.240974	1.129133	-1.203947	1.972850

Appendix 43

Max-Max Residuals of Estimation and Forecasting of KG Model with Union Set of Modular Variables Selected by GA

Year	e (C)	e (Y)	e (Sp)	e (Pc)	e (D)	e (Wp)	e (Fi)	e (A)	e (L2)	e (I)
1929	-0.296729	0.512683	-0.268385	0.208691	0.214188	0.544146	0.044302	-0.451856	-0.023057	0.012026
1930	0.653746	0.202549	0.288375	0.014718	-0.073123	-0.378948	-0.013624	0.440751	0.162739	0.463106
1931	-1.520006	-1.892817	0.358201	0.316857	-0.416079	-1.548821	0.020375	0.337071	0.951168	0.171908
1932	0.376491	1.318601	0.311853	0.499162	0.191222	0.303333	0.051788	0.306362	-0.636750	0.774686
1933	1.109484	2.042824	0.206759	0.675900	0.258149	1.129826	-0.045221	-0.423051	-0.142654	0.310833
1934	0.459751	0.627239	-0.014122	-0.198418	-0.014662	0.486355	0.075842	-0.153288	-0.003142	-0.166226
1935	3.241250	3.893219	-0.479363	0.291250	0.358106	2.952465	0.047805	-0.610386	-0.177257	0.566775
1936	-1.162301	-0.564695	0.533846	-0.540821	-0.275346	0.004268	-0.022239	0.674794	-1.195972	-0.835309
1937	0.061584	-0.451157	-0.059747	0.042526	0.053315	-0.600069	0.004409	0.013034	0.558230	0.260292
1938	-0.259745	-0.526065	-0.061236	0.207544	-0.044257	-0.344584	0.026721	-0.069428	0.571137	0.910877
1939	-0.818336	-0.814512	0.357901	0.347387	-0.076894	-1.169063	-0.006242	0.102391	-0.068816	-0.507517
1940	0.308566	0.442380	-0.096446	-0.095656	0.034496	0.601157	0.004227	-0.052228	-0.118103	0.074094
1941	-0.033090	0.122585	0.005530	-0.025221	-0.008707	0.020166	-0.000344	0.012091	0.010163	0.006455
1946	3.149373	0.757793	1.814513	2.166445	1.805579	2.441724	0.210680	-0.674040	-4.594408	-1.020779
1947	0.309993	1.057575	0.593550	0.026105	-0.361287	-0.760822	0.181907	-0.249786	0.509692	-0.244054
1948	1.654712	-0.504258	-0.814977	-0.519599	0.308491	-0.065895	0.009838	-0.943350	2.944524	0.083985
1949	-4.457904	-1.574934	-2.099615	-1.918901	-1.645323	-1.599842	-0.402175	1.775565	1.609208	1.163511
1950	-1.173132	-0.531709	1.905710	-0.048363	-0.339479	-0.492044	-0.065979	0.123337	-0.513304	-0.141694
1951	-0.777885	-10.334708	1.945659	-1.678782	-1.151031	-5.024417	0.133624	-0.333099	-0.011542	-0.942779
1952	-3.781561	-12.391205	2.274465	0.521549	-3.475106	-8.346632	-0.175204	0.301701	-0.175001	2.332572

Appendix 44

Min-Min Residuals of Estimation and Forecasting of KG Model with Union Set of Variables, Controlled for ANN Architecture

Year	e (C)	e (Y)	e (Sp)	e (Pc)	e (D)	e (Wp)	e (Fi)	e (A)	e (L2)	e (I)
1929	-0.472209	-0.076008	0.023255	0.253708	-0.068813	0.153865	-0.004078	-0.070388	-0.320127	0.003277
1930	0.579718	0.373551	0.051995	-0.207910	0.016117	-0.474552	0.033001	0.132876	-0.364212	0.271492
1931	-0.084646	-0.220034	-0.045061	0.254714	0.065983	-0.478199	-0.031645	-0.137075	1.058292	-0.390099
1932	0.192111	0.781644	0.317208	0.360422	0.056274	0.047990	0.037547	0.354871	-0.362651	0.624575
1933	0.876208	1.476986	0.207784	0.530796	0.102333	0.851697	-0.059759	-0.342661	0.001338	0.166081
1934	0.280384	-0.410076	-0.130609	-0.470748	-0.397125	0.331270	0.121856	0.039994	-0.374917	-0.266320
1935	1.311944	0.681797	-0.189064	-0.314277	0.279254	1.704998	-0.033457	-0.016969	-0.469184	0.031705
1936	-0.449532	-0.188982	-0.101009	-0.162877	0.011074	0.624683	-0.003426	-0.002337	-0.021740	-0.098949
1937	0.049469	-0.056077	0.074258	0.150994	0.185099	-0.552049	-0.031247	0.042076	0.392624	-0.061134
1938	-0.238127	-0.627543	0.083645	-0.172330	0.091376	0.116109	0.064917	-0.022470	0.474587	0.609985
1939	-1.301781	-1.149534	0.314098	0.332390	0.037438	-1.500736	0.013789	0.078230	-0.151015	-0.394874
1940	0.842258	1.478024	-0.270346	0.149079	0.102575	0.237265	-0.041993	-0.083577	-0.052032	-0.534020
1941	-0.089861	-0.384862	0.054470	-0.258191	-0.138096	0.211732	0.018317	0.020781	-0.013311	0.648205
1946	0.175061	-0.521039	-0.057860	-0.027620	-0.016075	0.149709	0.021727	-0.076278	-1.012175	0.118148
1947	-0.735738	0.945866	0.226326	0.315436	-0.025063	-0.859753	-0.036218	0.204147	0.342002	-0.159016
1948	0.908893	-0.015305	-0.642753	-0.519496	0.048228	0.322708	0.036040	-0.334566	0.026604	-0.315694
1949	0.694081	0.450098	-0.309102	0.112552	0.011579	1.440373	-0.014985	0.016164	-0.151611	0.307404
1950	-1.953436	-2.671548	0.107450	-0.566686	-0.158993	-1.071298	-0.045063	0.018441	0.004678	-0.857596
1951	-1.544243	-12.492857	0.149678	-2.200438	-0.968812	-5.602237	0.154634	-0.491593	0.537038	-1.672651
1952	-4.506018	-14.739831	0.612765	-0.058567	-3.285484	-8.945967	-0.154335	-0.103067	0.269404	1.418100

Appendix 45

Min-Max Residuals of Estimation and Forecasting of KG Model with Union Set of Variables, Controlled for ANN Architecture

Year	e (C)	e (Y)	e (Sp)	e (Pc)	e (D)	e (Wp)	e (Fi)	e (A)	e (L2)	e (I)
1929	-0.215709	0.205097	0.028779	0.081208	-0.095598	0.239924	-0.027817	-0.067807	-0.032522	-0.016983
1930	0.767058	-0.168739	0.133055	-0.078540	-0.017219	-0.370251	0.046947	0.147280	-0.122247	0.413989
1931	-0.357114	-1.033281	-0.088117	0.282938	0.133050	-0.794960	-0.067931	-0.238705	1.195346	-0.545243
1932	0.331733	1.162737	0.413896	0.348045	-0.053322	0.322685	0.070145	0.420251	-0.435638	0.620094
1933	1.043032	1.895527	0.308257	0.532073	0.009114	1.149401	-0.023031	-0.285825	-0.026090	0.163582
1934	0.200958	-0.139038	-0.415190	-0.553962	-0.103370	0.105778	0.117508	-0.027832	-0.293437	-0.265712
1935	0.849591	1.047589	-0.076260	-0.006882	0.266275	1.299522	-0.037021	0.052289	-0.523285	0.006231
1936	-0.709153	0.447149	-0.105478	-0.134237	-0.008485	0.579967	0.006069	-0.014794	-0.241888	0.177450
1937	-0.612541	-0.104546	-0.098441	-0.034830	-0.031029	-0.517015	0.010986	0.007856	0.063187	-0.171469
1938	0.296365	-1.101703	0.196381	-0.156177	-0.016461	-0.141208	0.010323	-0.150748	0.734105	0.511513
1939	-0.784248	-0.772657	0.444792	0.290296	-0.040923	-1.197101	0.008199	0.125190	0.152700	-0.309554
1940	0.304408	1.156903	-0.503438	0.050607	0.058082	0.487254	-0.021453	-0.028934	-0.269505	-0.666065
1941	0.404684	-0.429478	0.179021	-0.102286	-0.016059	0.164609	0.016783	0.019865	-0.017749	0.631407
1946	0.237775	-0.657086	-0.214652	-0.023860	0.042999	0.086454	-0.010814	-0.072677	-1.607577	0.096570
1947	-0.522916	1.553898	0.701378	0.114559	-0.147869	-0.398091	0.011795	0.153917	1.489696	-0.011910
1948	0.373401	-1.350739	-0.882625	-0.189850	0.114660	-0.224243	0.005768	-0.395842	-0.100051	-0.487983
1949	0.220074	1.029886	-0.623246	0.041282	0.047185	1.276959	-0.007877	0.047419	-0.632011	0.321499
1950	-1.256682	-2.127466	0.263308	-0.384645	-0.217209	-1.171457	-0.051534	0.015081	0.295023	-0.926637
1951	-0.857201	-11.943846	0.310496	-2.020028	-1.028016	-5.705012	0.147813	-0.483270	0.792375	-1.744564
1952	-3.847075	-14.149811	0.815064	0.131159	-3.356040	-9.055312	-0.163475	-0.054703	0.306467	1.343366

Appendix 46

Max-Min Residuals of Estimation and Forecasting of KG Model with Union Set of Variables, Controlled for ANN Architecture

Year	e (C)	e (Y)	e (Sp)	e (Pc)	e (D)	e (Wp)	e (Fi)	e (A)	e (L2)	e (I)
1929	0.486825	0.172786	-0.359122	0.253221	0.031759	0.901084	-0.039199	-0.227202	0.359743	0.527635
1930	-0.162649	-0.219982	0.410177	-0.245486	-0.076706	-1.127724	0.029055	0.205114	0.014400	-0.445912
1931	-0.457315	-1.606340	-0.092418	0.183598	-0.164179	-0.936026	-0.067338	0.065826	0.992585	-0.052633
1932	0.253875	1.058126	0.371982	0.432823	0.099077	0.177465	0.043585	0.321637	-0.336765	0.634286
1933	0.976359	1.764072	0.252982	0.605621	0.196578	0.997010	-0.054193	-0.423829	0.137826	0.168207
1934	0.391396	0.184143	-0.228048	-0.529073	-0.168520	0.452946	0.132808	-0.072375	-0.453116	-0.399014
1935	3.099496	2.621228	-0.051467	-0.090897	0.335692	2.587120	-0.008485	-0.182098	-0.159029	-0.185509
1936	-0.804086	0.800673	-0.035859	0.051717	0.082990	0.641715	0.101952	0.368367	-1.320192	0.517430
1937	-1.497437	-1.020638	0.024839	0.022806	0.192205	-1.537698	-0.037224	-0.067396	0.230090	-0.320228
1938	-0.803019	-2.708453	-0.000744	-0.531956	-0.036228	-1.087420	-0.012719	-0.138198	0.539330	0.051095
1939	-0.809570	-0.475477	0.307806	0.482089	-0.010026	-0.978656	-0.058951	0.049851	0.239225	-0.194386
1940	0.319682	2.825787	-0.333952	0.370061	-0.006820	0.900610	0.006916	0.123707	-0.326679	0.016841
1941	0.544432	-1.595746	0.062398	-0.614263	-0.032557	0.024724	0.017788	-0.090737	0.155698	0.009299
1946	3.766331	2.177951	2.433491	2.865073	1.715752	3.191602	0.341025	-0.808749	-4.413027	-0.146186
1947	0.327251	0.661137	0.093799	0.124513	-0.012991	-0.742741	-0.019247	-0.225409	0.136382	0.016998
1948	1.053424	-0.912606	-1.600724	-1.102075	-0.118355	-0.358414	-0.001909	-0.985229	3.288350	0.050759
1949	-4.003261	-1.284262	-1.449573	-1.741428	-1.454765	-1.373104	-0.278160	1.535629	1.477155	0.266132
1950	-3.001250	-2.297411	0.556391	-0.802008	-0.620610	-1.917467	-0.177888	0.218670	-0.561411	-0.941533
1951	-2.703941	-12.187460	0.587980	-2.469233	-1.468735	-6.504786	0.011332	-0.188416	-0.013041	-1.790262
1952	-6.582945	-15.068879	0.852824	-0.636755	-4.166802	-10.315535	-0.403184	0.778050	0.177550	0.999280

Appendix 47

Max-Max Residuals of Estimation and Forecasting of KG Model with Union Set of Variables, Controlled for ANN Architecture

Year	e (C)	e (Y)	e (Sp)	e (Pc)	e (D)	e (Wp)	e (Fi)	e (A)	e (L2)	e (I)
1929	0.473332	0.415065	-0.377827	-0.062577	0.143799	0.740977	0.044752	-0.417385	0.434355	0.382630
1930	-0.801782	-0.950355	0.547847	0.042462	-0.476393	-1.244045	-0.044294	0.429279	0.093311	-0.539067
1931	-1.042390	-2.073156	0.033624	-0.034417	-0.538712	-1.910391	-0.118276	0.287854	1.391578	-0.611186
1932	0.432742	1.712038	0.406470	0.596098	0.236514	0.455713	0.068270	0.202772	-0.635228	0.917594
1933	1.152491	2.409851	0.293951	0.766714	0.305659	1.267222	-0.029635	-0.532940	-0.130477	0.448194
1934	0.066784	-0.030823	-0.311739	-0.711301	-0.250582	0.402605	0.132023	-0.100172	-0.075631	-0.129278
1935	2.850724	3.461876	-0.458486	0.349308	0.537648	2.743454	0.033614	-0.644927	-0.468336	0.444497
1936	-1.195664	-0.443861	0.284563	-0.146587	-0.035989	-0.107236	0.037306	0.724371	-1.329881	-0.142804
1937	-0.112122	-0.026086	-0.009990	0.007100	0.300987	-0.750909	-0.012883	-0.211816	-0.166259	0.041232
1938	-0.562437	-2.281836	-0.159770	-0.508828	0.173010	-0.384158	0.019906	-0.104044	0.575602	0.653116
1939	-0.570602	-0.928785	0.434112	0.419186	0.141273	-0.995204	-0.039894	0.066892	0.327118	-0.350546
1940	0.482967	1.796490	-0.299386	0.160235	0.056970	0.508658	-0.013733	0.122763	-0.342474	-0.338728
1941	-0.169870	-0.648869	-0.020421	-0.308715	-0.239764	0.221307	0.006044	-0.017271	0.113519	0.301777
1946	3.994176	2.273387	2.223074	2.879050	2.066166	3.361768	0.376957	-0.908854	-4.745253	-0.186881
1947	1.032489	1.528378	0.617705	0.358652	-0.197777	-0.856167	0.083375	-0.335606	0.721141	-0.145850
1948	1.344066	-1.358094	-1.771045	-0.929137	0.380327	-0.358719	0.070062	-1.130408	2.741784	-0.095685
1949	-4.411746	-1.202582	-1.644886	-2.048881	-1.715185	-1.197880	-0.409943	1.560962	1.615674	0.477890
1950	-3.342276	-3.486856	0.626901	-0.759628	-0.861274	-2.196710	-0.255170	0.555463	-0.354754	-1.010379
1951	-3.043549	-13.386542	0.650488	-2.426175	-1.704136	-6.786944	-0.066131	0.121257	0.157068	-1.8670'8
1952	-6.854956	-16.266389	0.843217	-0.576001	-4.294149	-10.589407	-0.467269	0.888259	0.054926	0.864026