(2)

NAVAL POSTGRADUATE SCHOOL

Monterey, California

AD-A273 001

DTIC ELECTE NOV 22 1993 E

THESIS

AN EVALUATION OF ARTIFICIAL NEURAL NETWORK MODELING FOR MANPOWER ANALYSIS

by

Brian James Byrne

September 1993

Principal Advisor: George W. Thomas

Approved for public release; distribution is unlimited

REPORT DOCUMENTATION PAGE

Form Approved OMB Np. 0704

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE September, 1993

3. REPORT TYPE AND DATES COVERED

4. TITLE AND SUBTITLE An Evaluation Of Artificial Neural Network Modeling For Manpower Analysis

5. FUNDING NUMBERS

6. AUTHOR(S) Brian J. Byrne

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey CA 93943-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey CA 93943-5000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(maximum 200 words)*

This thesis evaluates the capabilities of artificial neural networks in forecasting the "take-rates" of the Voluntary Separations Incentive/Special Separations Benefit (VSI/SSB) programs for male, Marine Corps Enlisted Personnel in the grades of E-5 and E-6. The Artificial Neural Networks models are compared with the forecasting abilities of a classical regression model. The data are taken from the Headquarters Marine Corps Enlisted Master File which contains military and personal background on each enlisted member of the United States Marine Corps. The classical regression model is a casual model constructed based upon the theory of occupational job choice. The neural network models are presented with all available data elements. Empirical results indicate that artificial neural networks provide forecasting results at least as good as, if not better than, those obtained using classical regression techniques. However, artificial neural networks are limited in their usefulness for policy analysis.

14. SUBJECT TERMS Neural Networks, Modeling Techniques, Voluntary Separation Programs, VSI, SSB, Marine Corps Separations Incentives.

15. NUMBER OF PAGES 146

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

Approved for public release; distribution is unlimited.

An Evaluation of Artificial Neural Network Modeling for Manpower Analysis

by

Brian J. Byrne
Captain, United States Marine Corps
B.S., University of Colorado, 1985

Submitted in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE IN MANAGEMENT

from the

NAVAL POSTGRADUATE SCHOOL
September 1993

Author: Brian J. Byrne

Approved by: George W. Thomas
George W. Thomas, Primary Advisor

Timothy R. Hill
Timothy Hill, Associate Advisor

David R. Whipple, Chairman
Department of Administrative Sciences

ABSTRACT

This thesis evaluates the capabilities of artificial neural networks i forecasting the "take-rates" of the Voluntary Separations Incentive/Special Separations Benefit (VSI/SSB) programs for male, Marine Corps enlisted personnel in the grades of E-5 and E-6. The Artificial Neural Networks models are compared with the forecasting abilities of a classical regression model. The data are taken from the Headquarters Marine Corps Enlisted Master File which contains military and personal background on each enlisted member of the United States Marine Corps. The classical regression model is a causal model constructed based upon the theory of occupational job choice. The neural network models are presented with all available data elements. Empirical results indicate that artificial neural networks provide forecasting results at least as good as, if not better than, those obtained using classical regression techniques. However, artificial neural networks are limited in their usefulness for policy analysis.

TABLE OF CONTENTS

I. INTRODUCTION

A. BACKGROUND

Prior to implementing a policy, government decision makers attempt to understand its potential impact. This is especially true in today's environment of tight fiscal constraints. The pressure to understand these implications is even greater in the Department of Defense due to the attention focused on the rising cost of the "peace dividend." DoD policy analysts must be able to forecast accurately the cost of implementing new policy.

When these policy changes are related to manpower and personnel, forecasting the impacts of policies becomes particularly complicated. This is because policy analysts must be able to predict the decisions of individuals. For example, if the United States Navy wishes to increase the re-enlistment rate of boiler technicians, one of the tools available is the selective re-enlistment bonus (SRB). However, policy makers need to know how large a bonus is required to meet required personnel staffing levels. Re-enlistment is an individual choice and the factors affecting this choice vary from one individual to the next.

Manpower and personnel policy analysts have several tools to assist them in forecasting the decisions of individuals.

Currently, econometric models based on classical regression techniques are the primary tool of choice. However, there is another tool emerging that has the potential to replace or augment classical regression techniques in forecasting personnel decisions. This tool, artificial neural networks (ANN), bases its design on the functions of the human brain. Current research in the manpower and personnel fields provides supporting evidence that artificial neural networks can yield better forecasting results than classical regression techniques in some applications.

The potential applications for artificial neural networks are spread across many disciplines. Artificial neural networks are being employed in the medical industry for cancer research and illness diagnosis [Ref. 1]; in the business world for stock market predictions [Ref. 2] and credit risk assessments [Ref. 3]; in criminology for fingerprint classification and handwriting analysis; and, in DoD, for target recognition and aircraft systems control. [Ref. 4]

Artificial neural networks (ANNs) offer two benefits to the user [Ref. 5:pp. 3-5]. First, many of the assumptions associated with classical regression techniques are not needed when employing neural networks. ANN model developers claim that analysts do not have to worry about multicollinearity between independent variables, heteroskedasticity, or serial correlation. Additionally, the relationships between the independent and dependent variables do not have to be linear.

The second benefit of neural networks is in variable selection. There need not be an underlying theory to support the choice of independent variables. A researcher can apply his common sense in selecting independent variables. It is important to caution here that variables should still be chosen carefully as in the applicat n of classical regression techniques. However, supporters of ANN claim that users do not have to be theoretical experts in the field where they are applying ANNs. ANNs are designed to seek out patterns that are complex and not readily noticeable to the analyst. If a pattern does not exist between dependent and independent variables, then the network "learns" to disregard the independent variable's values.

This thesis examines the applicability of ANNs to a current manpower related issue, the downsizing of the active duty force. Manpower and personnel policy makers are implementing separation incentive programs to entice service members to leave the service. The United States Marine Corps currently utilizes two of these programs, the Voluntary Separations Incentive (VSI) and the Special Separations Bonus (SSB), to shape its force structure. Accurately forecasting which individuals will accept these programs is paramount for two reasons. First, in budgeting limited financial resources, it is essential to predetermine, as well as possible, the number of service members who will select the programs once they are offered them. The second reason centers on the eligible population for the program. Since the programs are designed to shape the force structure, the categories of personnel targeted to receive these programs must be carefully selected. Further complications to the implementation process arise from the congressional mandate requiring that all those who apply for the programs are accepted.

This thesis provides information for policy makers to use in reaching more informed decisions about the implementation of the VSI and SSB programs. Additionally, it adds another tool to the arsenal of policy analysts when attempting to determine the impacts of policy changes.

B. THESIS OBJECTIVES

This thesis has two objectives. The first is an evaluation of the effectiveness of ANNs as a tool for manpower analysts. ANNs are currently receiving a great deal of positive publicity [Ref. 6:p. 3]. Developers and researchers are treating neural networks as one of the most important advances in computer technology in recent years. However, before it can be effectively utilized as a forecasting tool for manpower and personnel decisions, it should be examined with a critical eye and evaluated under different conditions.

The other objective of this thesis is to examine several of the possible combinations of choices that can be made when constructing a neural network. Builders of ANNs must possess a knowledge of the different facets of a network and make conscious decisions as to the form of the network. The architecture, topology, learning rule, and transfer functions of a network must be clearly defined by the developer. There are many possible combinations and each produces a slightly different network. How these decisions are reached is currently the topic of much of the research surrounding neural networks.

C. RESEARCH QUESTIONS

This thesis attempts to answer the following questions:

1. Do artificial neural networks provide better forecasting results than classical regression techniques in forecasting the "take-rates" of voluntary incentive programs offered to Marine Corps enlisted personnel?

2. Which artificial neural network "topology" provides the "best" forecasts of these "take-rates"?

D. SCOPE AND LIMITATIONS

This thesis has two limitations. The first concerns the data set. This study examines only Marine Corps enlisted personnel offered the separation incentive programs. Officers are eligible for the programs but the number of applicants provide too small a sample.

The other limitation deals with the algorithms underlying the classification processes that are utilized by an ANN. These algorithms are complex and quite lengthy. An examination of the mathematical formulas that support ANN is beyond the scope of this thesis.

E. RESEARCH METHODOLOGY

This thesis uses a data set developed by members of the Enlisted Career Force Planning Cell at Headquarters, Marine Corps. The data are taken from two places. A list of all enlisted members who voluntarily separated from the Marine Corps under the incentive programs is combined with information on each individual available from the Marine Corps Enlisted Master File. The data are from fiscal year 1993.

These data are used to construct three models to forecast the "take-rates" of the separations programs offered. The first model is a variation of a binomial LOGIT model which was developed by Noblit (1993). It contains independent variables that represent the lifetime earnings potential of an individual along with proxy variables for his ability, preferences, and family background.

The other two models are ANN models constructed using a commercially available software package, NeuralWorks Professional II/Plus developed by NeuralWare Incorporated. This software is chosen because of its flexibility in creating a neural network and its user friendly interface.

One neural network model is constructed and adjusted to reach its peak forecasting potential. The topology of the initial model contains one hidden layer with thirty hidden neurons. The other ANN model is similar in design except that it contains two hidden layers, one with twenty hidden neurons and the other with ten hidden neurons. The two ANN models are compared using criteria such as the correctly predicted takers and correctly predicted non-takers of the programs.

F. ORGANIZATIONAL DESCRIPTION

The thesis is divided into three parts. The first part is comprised of Chapters II and III. These two chapters lay the foundation for the rest of the thesis. Chapter II provides background information on the functions of both types of neural networks, biological and artificial. Chapter II also contains a section which describes the decisions designers face when constructing an ANN.

A review of current literature on ANNs is presented to demonstrate the nature of its components. Individuals inexperienced with neural networks can gather the necessary knowledge from the literature review to comprehend the steps involved in the building a neural network.

The other aspect of the literature review builds the conceptual framework for the decision to take a separations incentive. Literature dealing with the theory of occupational choice is reviewed. Although the separations programs are relatively new, there exist many studies on the impacts of the SRB. The goal of the SRB is to entice military members to choose a military career over a civilian occupation. The separation incentives have the opposite goal. They entice service members to seek civilian rather than military occupations. In essence, the separation incentives are simply the SSB in reverse.

Chapter III describes the VSI and SSB programs along with the implementation plan used by the Marine Corps in employing these two programs. This is followed by a discussion of the theoretical framework that Noblit (1993) uses in developing his model. The economic theory of occupational choice establishes the structure for the model. The chapter concludes with a description and critique of Noblit's model.

The middle section of the thesis, Chapters IV, V, and VI, is analytical in nature. Chapter IV describes the data used in this comparison and the development of the three analytical models. Chapter V presents the results of the classical regression model which is compared to the later neural network results. Chapter VI furnishes the results of the two neural networks employed in this comparison. Additionally, part of the chapter is devoted to discussing the choices made during the development of the ANN and the reasoning behind the selection of a network's characteristics.

The last section of the thesis contains Chapters VII and VIII. Chapter VII compares the forecasting capabilities of all three models and their usefulness to military manpower and personnel policy analysts. Chapter VIII contains concluding comments and recommendations for further research.

II. NEURAL NETWORKS

A. INTRODUCTION

Forecasting the future is a difficult task. However, it is a necessity if manpower and personnel planners are to determine the effects of policy decisions on the shape and size of the military force. Planners must be able to determine the impacts of advertising on recruitment and bonuses on reenlistment. It is not enough to know increases in advertising and bonuses will increase enlistments and reenlistments. To derive the greatest benefit from instituting manpower and personnel policies, the incremental effects of a policy must be determined.

Just as a carpenter relies on his tools to build a house, manpower and personnel planners rely on tools to forecast the future consequences of policy changes. Among these tools are econometric models based upon regression analysis. Recently, a new tool for forecasting the impacts of manpower policy changes has emerged. The design of this new tool, artificial neural networks, is based on the human brain. Recent studies provide supporting evidence that artificial neural networks produce the same or better results than classical regression techniques while allowing the researcher to relax the assumptions required when using regression.

This chapter includes a brief description of the elements of neural networks and a review of the options a designer must choose among when developing an artificial neural network as well as a section concerning special data requirements. The chapter concludes with a discussion of several techniques available to measure the success of forecasting tools.

B. NEURAL NETWORK COMPONENTS

1. Background

A neural network is a collection of neurons working together to produce a result based upon information provided to the network. Neural networks are either biological or artificial. The brain of an animal is an example of a biological neural network. The six senses of an animal allow it to detect important aspects of its surroundings and react to them accordingly. Current theory labels this ability to react to a changing environment as thinking. [Ref. 6:pp. 9-11]

Artificial neural networks are an attempt to simulate biological neural networks. Biological neural networks are efficient at associative reasoning, learning, and thought. Artificial neural network designers hope to capture these qualities in order to enable computers to simulate more closely the processes of the brain. This will allow computers to adjust their functions to match the ever-changing environment in which they operate without having to undergo the complex steps of reprogramming. [Ref. 6:pp. 9-11]

2. Biological Neurons

The basic element of a neural network is the **neuron.** The biological neuron is a complex cell comprised of four basic components. The **soma** is the center of the cell. The **dendrites** are the input channels to the soma. The **axon** is the output channel from the soma. The **synapses** are the connections between different neurons within the network. It is important to note that synapses are not part of the cell but regions between the axon of a sending neuron and the dendrites of a receiving neuron. [Ref. 6:pp. 30-31]

Biological neurons use a complicated and not yet fully understood process in their operations. However, the basics are simple to comprehend. The neuron receives an input from a synapses. This input travels through the dendrites to the soma. Inside the soma, the input accumulates over time. When the soma decides it has received enough inputs, it activates and sends a signal to other neurons. The axons carry this signal to the synapses which connect with the dendrites of other neurons.[Ref. 7:pp. 3-5]

3. Artificial Neurons

a. *Overview Of The Imitation Process*

Artificial neurons attempt to imitate this process. The process in an artificial neuron consists of four phases. During the first phase, the **processing element** (in an artificial neural network, a processing element is analogous to the biological neuron), receives inputs from either the outside world or other processing elements through input paths (dendrites). Prior to entering the processing element, these inputs receive a weight corresponding to the strength of their signal. Once the processing element receives all of the weighted inputs, these values move on to the next phase. [Ref. 8:pp. 4-5]

The second phase of the process is activation of the processing element. During this phase, the processing element decides how to treat the weighted inputs. An **activation function** transforms all the weighted inputs into a single value, the activation value. In simple processing elements, the activation value is the sum of all of the weighted inputs. In more complex processing elements, the activation value can be a function of the weighted inputs and the previous state of the processing element. This permits the processing element to "self-excite", allowing previous inputs to have a continuing effect. Once the processing element determines the activation value this value moves to the next phase.[Ref. 6:pp. 84-85]

During the third phase, the processing element converts the activation value into the output of the neuron. The **transfer function** of the processing element performs the conversion process of the activation value into the output value. Transfer functions can take on one of several forms. In developing an artificial neural network, the form of the transfer function is left up to the developer of the neural network. [Ref. 6:pp. 84-85]

The final phase of the process is the output phase. The processing element sends the value of its output as determined by the transfer function's action on the activation value to one of two places. The output can be sent to other processing elements within the network or to the outside environment as a product of the network. Figure 1 contains a breakdown of the four phases.

b. Types Of Transfer Functions

The **threshold function** and the **saturation function** are two types of transfer functions. The threshold function is an "all or nothing" approach. The activation value must be above some minimum value, the **threshold**, for the processing element to contribute anything to the whole network. In networks that use the threshold function, if the activation value is below the threshold, the processing element's output is zero. Otherwise the processing element's output is one. [Ref. 6:pp. 85-86]

The output values of a saturation function are continuous, unlike those of a threshold function. However, in a saturation function, excitation above some maximum firing level has minimal additional effect. Two commonly used saturation function are the **sigmoid function** and the **hyperbolic tangent function**. Figure 2 contains a graph of

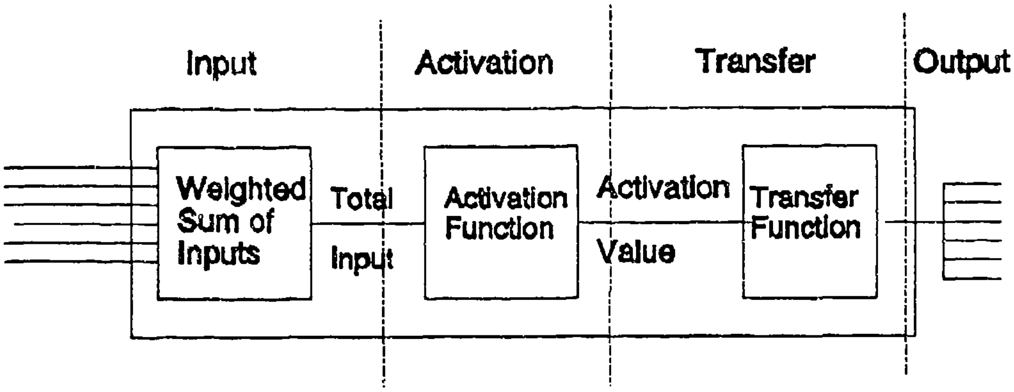

Figure 1. Phases of the Artifical Neuron Process

Source: Russell (1993)

each type of transfer function. These functions are advantageous because they provide a continuous, monotonic value for any input and are bounded by a high and low value. [Ref. 6:pp. 85-85]

4. Layers

A neural network consists of many neurons, biological or artificial, organized into groups defined as **layers.** There are three types of layers. The first, the input layer, contains neurons that receive inputs from the outside world. In biological neural networks, these input neurons are located in the eyes, ears, and other sensory organs of animals. Input neurons in artificial neural networks receive data from the keyboard or other computer paraphernalia. [Ref. 8:pp. 4-5]

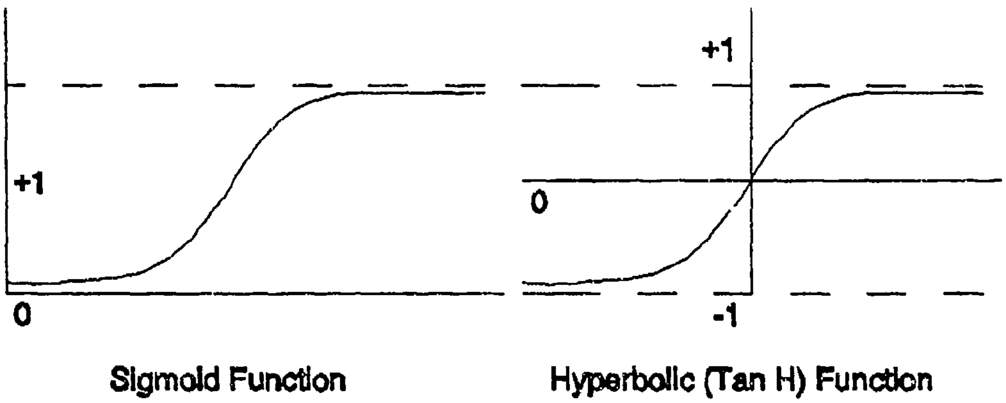

Figure 2. Common Saturation Functions

Source: Russell (1993)

A second type of layer is the output layer. In the output layer, output neurons send information back to the outside world. In a human being, this information is transmitted through muscle activity such as voice activation, emotional responses, or body movement. Output neurons in the artificial world send their signals via computer monitors or other output devices. [Ref. 6:p. 79]

If a neuron is not in the input or output layer, it falls into the hidden layer. Neural networks can have one hidden layer, no hidden layer, or thousands of hidden layers. What happens within the hidden layer is not observable to the outside world. The neurons of the hidden layer are connected only to other neurons of the network. [Ref. 6:p. 79]

5. Networks

A neuron is the basic element of a neural network. Neurons are organized in groups called layers. These layers form networks. Biological neurons make up biological neural networks. The human brain is an example of a biological neural network. It contains approximately one hundred billion neurons each connected to approximately ten thousand other neurons. [Ref. 9:pp. 3-4]

This vast network functions in parallel which means that many connections can occur simultaneously. This allows the brain to make roughly ten thousand trillion connections per second. The Cray Computer, the fastest computer developed to date, is serial which means it can only make one connection at a time. The Cray can make fifty million interconnections per second. The end result is that the human brain is about twenty million times faster than the "fastest" serial computer. [Ref. 9:pp. 3-4]

The speed and parallel nature of the human brain allow it to be excellent at problem solving or pattern recognition but slow at complicated mathematical problems. The serial nature of computers facilitates the solving of mathematical problems but prohibits the computer from excelling at pattern recognition or pattern solving. Artificial neural networks attempt to bridge this gap. [Ref. 6:p. 14]

a. *Network Types*

There are many different kinds of artificial neural networks. However, each type falls into one of two categories, **feed-forward** or **feedback** networks. The distinction between the two centers on whether or not the network pays attention to its results. If a network uses its previous result in determining the succeeding output, then the network is classified as a feedback network. On the other hand, feed-forward networks are so-called because the output of each processing element flows forward in the network and does not interact with neurons in either previous layers or the same layer. [Ref. 6:p. 148]

Feedback networks often are constructed without a hidden layer. Every neuron is allowed to connect to every other neuron and to itself. In order for a feedback network to produce a result, it must continually adjust the processing element weights until the weights reach a steady state. It is impossible to predict how long this will take. However, most systems are able to accomplish the adjustments after a few iterations. Examples of feedback networks include Hopfield Networks, Content-Addressable Memories, and Traveling-Salesman Problem Networks. [Ref. 6:p. 153]

Feed-forward networks produce results in the following manner. Each processing element receives input from processing elements in the previous layer. In the input layer these inputs come from the outside world. In the hidden layer, these inputs come from the input layer or another hidden layer. In the output layer, these inputs come from a hidden layer.

The processing element determines the weight of the input, calculates an activation value, and then decides on an output. This output is passed on to the next layer. The output flows from the input layer to the hidden layer(s) to the output layer to the outside world. In essence, feed-forward networks are hierarchial. Processing elements connect only with processing elements in the next layer. Unlike the processing elements of feedback networks, feed-forward network processing elements do not connect with themselves or processing elements in the same layer. Figure 3 contains a diagram of a feed-forward network with three inputs, two hidden layers each containing two processing element, and one output layer. [Ref. 6:p. 148-150]

b. Network Learning

Determining the weight of the connection between each processing element is the biggest feat that networks must accomplish. Networks accomplish this by a process called **learning**. Learning is the process by which the connection weights are adjusted to minimize the error between the desired output and the actual output. Initially, the weights of each connection are set randomly. The network is then presented with **training pairs**. A training pair is one observation from

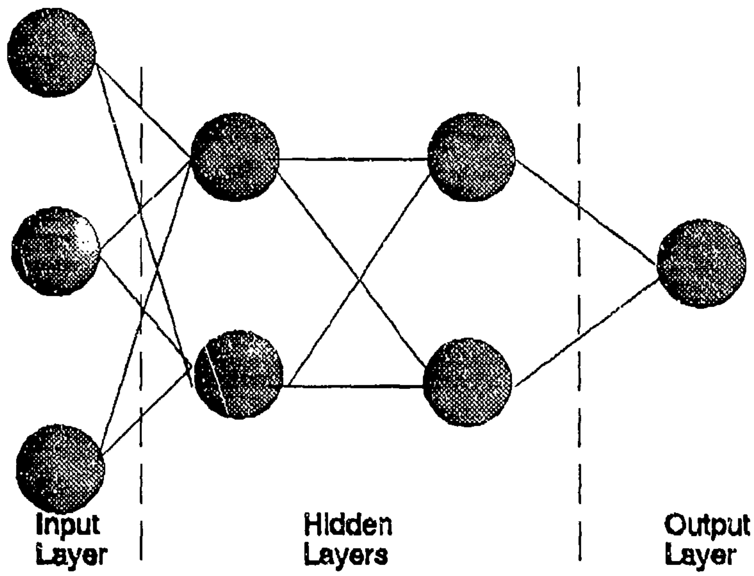

Figure 3. Feed-Forward Neural Network

Source: Russell (1993)

the data set containing all input values and the desired output. Based on the input, the network predicts an output value. This predicted value is then compared to the actual output. [Ref. 6:p. 149]

Differences between the actual and predicted values cause the network to adjust the connection weights. The weights are adjusted by a **learning rule.** The learning rule determines how the connection weights are adjusted throughout the network. Popular learning rule algorithms include Perceptron, Adaline, Madaline, Boltzman, Counter Propagation, and Back Propagation. The mechanics of these algorithms are beyond the scope of this thesis. Detailed descriptions and the underlying mathematical formulas may be found in Wassermann (1989), Beale and Jackson (1991), and Muller and Reinhardt (1990). [Ref. 6:p. 151]

One of the largest differences between feedback and feed-forward networks is that feedback networks do not learn, they are constructed. Feed-forward networks learn based upon the presentation of input and the desired output. Feed-forward networks simulate learned behavior of an animal while feedback networks simulate instinctive behavior.[Ref. 6:p.153]

C. NETWORK DESIGN

Researchers who use artificial neural networks face many decisions in the construction of a artificial neural network. What is the appropriate network architecture? How many hidden layers should the network contain? What transfer function is suitable for the network? What are the best learning rates? What is the appropriate learning rule for the network? These questions only scratch the surface of the decisions faced by researchers when building an artificial neural network. This section examines several of the decisions faced by artificial neural network developers. Although researchers devote much time and energy to determining the appropriate design for an artificial neural network, there are endless possibilities.

1. Network Architecture

Network architecture refers to the type of network to be used. Is a feedback or a feed-forward design the most appropriate? To answer this question, the network designer must understand the applications for which the network will be used. Military manpower and personnel planners attempt to forecast, or predict, the decisions military service members will make. Feed-forward neural networks are the most successful for this application. [Ref. 8:p. 36]

Klimasauskas (1991) suggests that back-propagation is the best architecture for neural networks designed for forecasting or classification. Forecasting stock market moves, predicting future job performance, bankruptcy forecasting, and credit risk determination are several of the uses for back-propagation architecture. [Ref. 3:p. 29]

Yoon and Swales (1991) used neural networks to predict stock price performance. The researchers collected data from two sources used regularly by investors, Business Week and Fortune magazines. The artificial network design used a back-propagation architecture. The results of the neural network model were compared to forecasts made using multiple discriminate analysis. Yoon and Swales concluded that neural networks using a back-propagation architecture significantly improve the predictability of stock price performance. [Ref. 2:pp. 156-162]

Raghupathi, Schkade, and Raju (1991) employed neural networks to forecast the bankruptcy rate of private companies. Data on 102 companies, 51 that filed for bankruptcy and 51 that did not, were used to train a neural network with a back-propagation architecture. The authors concluded that neural networks provide a suitable model to predict bankruptcies. [Ref. 10:pp. 147-155]

Wiggins, Engquist, and Looper (1992) compared neural networks and classical regression techniques in forecasting the re-enlistment rates of Air Force Enlisted personnel. Several neural network models were developed using three different architectures: back-propagation, probabilistic neural network (PNN), and learning vector quantization (LVQ). The results of these three models were compared to the results of a probit model. Wiggins et al. concluded that the back-propagation neural network provided the best results, especially in cases with large sample sizes. [Ref. 11:p. 52]

At this point, it is important to clarify an issue. Earlier it was stated that popular learning rule algorithms include back-propagation. Many times the network architecture and the learning algorithms the network use share the same name. This adds some confusion. However, keep in mind that the name of a network can describe how the network functions.

2. Network Topology

The topology of a neural network refers to the number of layers and quantity of processing elements within each layer. An artificial neural network has one input and one output layer. The input layer contains one processing element for each input variable. The number of input variables in a network depends upon the configuration of the input data. Later in this chapter, further discussion addresses changing sample data into the appropriate form for neural networks.

The output layer is similar to the input layer. One processing element in the output layer is required for every piece of output data. Output data faces the same constraints as the input data. Therefore, a transformation back to original form may be required for proper interpretation.

Developers of artificial neural networks must decide on the number of hidden layers in a network and the quantity of processing elements in each hidden layer. Klimasauskas (1991) states that most research problems employing a neural network technique can be solved using a topology with one hidden layer. If better performance is expected with more than one hidden layer, it is best to connect each hidden layer to all prior layers. This facilitates learning and increases the ability of the network to recognize complex patterns. However, if traditional statistical methods have provided satisfactory results, then connecting the input layer directly to the output layer is often very effective. [Ref. 12:p. 20]

To summarize this concept, start without a hidden layer and examine the results. Then add one hidden layer and determine which network provides the best results. Adding additional layers should not improve network capabilities but may if the data contain complex patterns.

Lapedes (1989) furnishes an argument for utilizing no more than two hidden layers when using a back-propagation network. Using calculus, he supports the hypothesis that increasing the number of hidden layers provides no increase in the learning capabilities of an artificial neural network.[Ref. 13:p. 456]

Once the decision about the number of hidden layers is reached, the designer must determine the appropriate quantity of processing elements in each hidden layer. If the number of processing elements in each layer is too large, the network will not learn the generalities of the input data but, rather, will memorize them and the network will be unable to handle inputs it has not already seen. [Ref. 14:p. 18]

On the other hand, if the number of processing elements is too small, the amount of time required to train the network drastically increases. An excellent starting point is five processing elements in each hidden layer. [Ref. 14:p. 18]

Since the decision about network topology is more an art than a science, the best alternative is to start with a simple design, say one hidden layer with five processing elements. After the network has learned to a sufficient degree, adjust the number of layers and processing elements to determine if any improvement in results is detected.

3. Network Transfer Functions

The transfer function within a processing element transforms the activation value to the output value of the processing element. The activation value is a function of the weighted inputs to the processing element. The two most often employed transfer functions are the hyperbolic tangent function and the sigmoid function. [Ref. 12:p. 23] Other choices include a linear function, a sine function, or a threshold function. The function choice is dependent upon the nature of the input data and what the network is trying to learn.

When employing neural networks in forecasting problems, many of the input variables will be dummy variables equal to one if a condition exists and zero if it does not exist. In these situations, the best choice of a transfer function is the hyperbolic tangent or sigmoid function. The choice of which of these two to use depends on the forecasting goal. [Ref. 12:p. 20]

Forecasting goals fall into one of two categories, average or exceptional behavior. Average behavior is defined such that the probabilities of choosing from a list of alternatives are roughly similar. For example, in forecasting the response rate of surveys, the chance a person will respond or will not respond are fairly similar. In stock prices predictions and bankruptcy forecasting, the goal is to find the exceptions. Researchers attempt to find the stocks that will increase greatly in price or the company that stands the highest probability of failure. In forecasting exceptional behavior, the hyperbolic tangent function works best. In average behavior cases, the sigmoid function works best. [Ref. 12:p. 20]

It is important to note that each processing element, regardless of its layer location, must possess a transfer function. It is not necessary that every processing element in the network have the same transfer function. Several different types of transfer functions may be used in a single network. For example, the neurons of the input layer have only one input each, the value of the input. Transforming these values using a hyperbolic tangent or sigmoid function serves no purpose. Therefore, the use of a linear transfer function is appropriate at this level. [Ref. 8:pp. 106-108]

4. Network Learning Rules

Learning is the process a network undergoes to match the desired output to the actual output. By adjusting the weights of each connection between the processing elements, a network accomplishes learning. **Learning rules** specify how connection weights are adjusted. Widrow and Hoff (1960) developed a learning rule, the **Delta-Rule**, which is the basis for several of the well known learning rules used extensively today. [Ref. 5:pp. 15-24]

a. *The Delta Rule*

The **Delta-Rule** states that the change in connection weight $\mathbf{W_j}$ is equal to the **learning rate** times the **error** times the **output** from the processing element $\mathbf{X_j}$. [Ref. 15:p. 218]

$$\Delta W_j = LearningRate * Error * X_j$$

The **learning rate** is a value between zero and one. The developer of the network determines the learning rate and can adjust it as the network continues learning. An **iteration** is the presentation of one training set, both the independent and dependent variables, to the network. Initial iterations through the data should be done with a relatively high learning rate. As the number of iterations increase, the learning rate should be decreased. This allows the network to reach a global error minimum. If the learning rate does not decrease as the iterations increase, the error will oscillate around the minimum achievable error for the network. [Ref. 5:p. 17]

The **error** is the difference between the actual output of the network based on the inputs and the desired output that is associated with those inputs. Each connection weight is assigned an equal part of the total error. If there are **N** weights, then the error term is calculated as follows:

$$ERROR = \frac{Desired - Actual}{N+1}$$

The **N + 1** term accounts for the weight of the bias element contained in the network. [Ref. 5:p. 16]

Widrow and Huff developed this learning rule for the simple, single-layered neural network called **Perceptrons.** Rumelhart, McCelland, and Williams (1986) developed a more complicated learning rule based upon the Widrow-Huff rule. This new development was designed for complex neural networks that employ multiple layers.

b. Variations Of The Delta Rule

Two variations to the delta rule, the **Cumulative Delta Rule** and the **Normalized Cumulative Delta Rule**, were developed to increase the efficiency of neural networks. Before reviewing these rules, two new terms need introduction, **epoch size** and **momentum.**

Epoch size is the number of training pairs that are presented to the network before adjusting the connection weights. Simply put, the epoch size is the number of iterations the network undergoes before any of the connection weights are adjusted. [Ref. 5:p. 30]

Momentum is a factor added to the delta rule to allow the network to minimize the error across the entire network and not become "stuck" in a local minimum. The factor is based on previous changes in the connection weights and is between zero and one. If the weight changes are large, then the momentum tends to be large but as the changes become smaller, so does the momentum factor. Momentum assists in speeding up the convergence to a minimum error.[Ref. 16:p. 92]

The Cumulative Delta Rule updates the weights after each epoch. Cumulative weight changes are stored and then propagated through the network after the presentation of a builder defined number of iterations or epoch size. These weight changes will contain both increases and decreases in connection weights thus allowing for a net, cumulated effect to occur on the weights. The benefit here is that this decreases processing time because the error is not propagated back through the network after each iteration. [Ref. 5:p. 30]

The Normalized Cumulative Delta Rule is similar to the Cumulative Delta Rule in the determination of the change in the connection weights. The difference is that the changes in the weights are **normalized.** To normalize the weight changes, the learning rule algorithms transforms the connection weight vector into a unity length vector dy dividing it by its magnitude. The advantage to making this modification is that it further reduces the training time of the network. [Ref. 16:p. 115]

The Delta-Rule and its two variations, the Cumulative Delta and Normalized Cumulative Delta Rules, are possible choices for the learning rule used by an artificial neural network. Of these three rules, network designers favor using the Normalized Cumulative Delta Rule when working with back-propagation networks. [Ref. 12:p. 22] Of the three choices, it is the quickest and requires the least amount of CPU processing time.

D. DATA CONSIDERATIONS

The type of data fed to the network is as important as the network development. As in classical regression techniques, all of the data presented to a model must be in numerical form. Neural networks put a further limitation on the data. Networks work best with data in the range of -1 to +1 or 0 to 1. Each input variable for the neural network model must be converted to this form. [Ref. 3:p. 28]

In modeling manpower and personnel decisions, independent variables are usually either continuous numeric variables or "dummy" variables. The "dummy" variable is the easiest for a neural network to handle because it has been previously coded as a zero or a one. The value of the variable is one if a certain characteristic is present and zero if it is not. For example, if an individual is a male, the variable MALE is set equal to one. If the individual is a female, the variable MALE is set to zero.

Continuous numerical variables pose a slightly more complicated problem. These variables must be re-scaled to fit in the range from -1 to 1 or 0 to 1. Although this is not extremely difficult, it can be time consuming unless the neural network software being utilized has the capability of completing this transformation transparent to the operator.

Fortunately, NeuralWorks Professional II/Plus possess this capability.

Output falls under the same restrictions as the input data. Since the decision to stay or leave the military is simply a choice between only two items, then a dummy variable can be used as the output variable. The variable can be assigned a value of one if the observation left the service and a zero otherwise. [Ref. 3:p. 28]

In processing the output data for delivery to the network, NeuralWorks scales the values into a smaller range. The range for a two condition dummy output variable is either -0.8 to 0.8 or 0.2 to 0.8. The purpose behind this is to allow the network to determine results for new input data outside the range of the original data.

E. MEASURES OF SUCCESS

In making a comparison between models to determine which is the "best" at forecasting and provides the "best" fit to the data, researchers must develop some type of measurement criteria. This section reviews current measures.

Hill, O'Conner, and Remus (1991) compared neural networks and classical regression techniques using time series data. A variety of models were created using weekly, monthly, and yearly time series data. In determining the best model for forecasting using time series data, Hill et al. used the absolute percentage forecast error (APE). This measure was chosen based on its applicability to time series data. Additionally, APE is a measure used quite often and appears to be the most robust measure available. [Ref. 17:p. 12]

APE is calculated based upon the predicted value of the forecast, the actual value of the forecast, and the total number of residuals. The formula to determine the APE of a certain model is:

$$APE=\frac{1}{N}*(\sum(\frac{P_i}{A_i}))*100$$

Where **N** is the number of residuals, $\mathbf{P}_i$ is the predicted value, and $\mathbf{A}_i$ is the actual value of the forecast. [Ref. 7:p. 55]

Sharda and Wilson (1993) review several different measures of success when evaluating model performance. The authors do not specify which is the "best" metric to use. They do conclude, however, that more than one measure should be used when comparing forecasting techniques.

Two of the metrics proposed by Sharda and Wilson are easily calculated and effective measures for classification experiments. Since this study focuses on "classifying" a military person as a "taker" or "non-taker" of the separations incentive programs, the classification metrics are the best suited for this research.

Prior to examining the actual measures, it is important to classify the possible results of a forecasting model. Forecasting models have the potential to provide four results:

true positive, true negative, false positive, or false negative. If the model predicts the observation will take the program and he does, this is a true positive. If the model predicts the observation will not take the program and he does not, this is a true negative. A false positive occurs when the model predicts the observation will take the program but in reality he does not. A false negative occurs when the model predicts the observation will not take the program but he does.

These four possibilities are represented by the following symbols:

O_t = True positive
(predicted taker and actual taker)

O_n = True negative
(predicted non-taker and actual non-taker)

F_t = False positive
(predicted taker and actual non-taker)

F_n = False negative
(predicted non-taker and actual taker)

The first metric, M_1, provides a measure of success in predicting the correct category. M_1 is defined in two parts. The first measure is the probability of correctly predicting a "taker" (M_{1t}) and the second is the probability of predicting a "non-taker" (M_{1n}).

The values of these measures are defined by the following equations:

$$M_{1t} = \frac{O_t}{(O_t + F_t)}$$

AND

$$M_{1n} = \frac{O_n}{O_n + F_n}$$

The other measure, M_2 is an inefficiency measure. It symbolizes false classifications. Like M_1, M_2 is composed of two values, one for the model's ability to forecast "takers" and the other to represent the model's ability to forecast "non-takers". The variables are the same as defined for M_1. The two values of M_2 are:

$$M_{2t} = \frac{F_t}{O_t}$$

AND

$$M_{2n} = \frac{F_n}{O_n}$$

The goal of any forecasting model would be to maximize M_1 and minimize M_2. [Ref. 18:pp. 649-656]

Wiggins, Engquist, and Looper (1992) use a simulated R^2 when comparing the results of an artificial neural network model to previously developed models using classical regression techniques. The benefit of this measure is its ability to identify models that do not provide better estimates than provided by examining the sample means. The simulated R^2 will produce a negative value in these instances. In instances where perfect prediction occurs, the simulated R^2 takes on the value of one similar to the coefficient of determination used in measuring the effectiveness of classical regression models. [Ref. 11:pp. 14-15]

The simulated R^2 formula is:

$$R^2 = 1 - \frac{\sum (P_i - A_i)^2}{\sum (AVG - A_i)^2}$$

Where:

P_i is the predicted take-rate for observation i,

A_i is the actual decision made by observation i, and

AVG is the average take-rate over the whole sample.

Prior to developing a neural network, measures of success must be determined. This is critical in research where the goal is model comparison. Any of the previously mentioned methods are effective measures in determining which is the appropriate model to use. However, it should be kept in mind that it is best to use several measures (Sharda and Wilson (1993).

F. CONCLUSION

A review of current literature supports the hypothesis that artificial neural networks provide forecasting results at least as good if not better than results provided by classical regression techniques. The network architecture that provides the best structure is back-propagation. Researchers in the manpower and personnel fields feel that this architecture is the best suited for forecasting individual decisions, especially for large data sets.

Since the application of ANNs is relatively new, no "standard" technique or rule exists for constructing a neural network. Little of the literature addresses the different aspects of network design. Researchers fail to comment on the number of hidden layers, the number of processing elements within each hidden layer, the learning rules, the learning rates, and the transfer functions of their networks.

Based on the literature review, this thesis builds two different models and compares the results of each to the results of a logit model. When developing the neural network models, the research employs a two-step approach.

During the first phase, a network with one hidden layer and thirty processing elements in the hidden layer is constructed. In this phase, the different possibilities for learning rule, learning rate, epoch size, and transfer function are adjusted to obtain peak performance.

Klimasauskas (1991) provides excellent information on the choice of learning rules, learning rates, epoch sizes, and transfer functions. Each of these facets of a neural network are optimized. The optimization process is described in later chapters.

Phase two changes the network topology while holding the other aspects of the network constant. A hidden layer is added and the amount of hidden neurons is adjusted. The first hidden layer contains twenty neurons and the second hidden layer contains ten hidden neurons.

The comparison stage of the research employs several of the success measures previously discussed. Each of the three models are evaluated using three criteria: the number of true positive, the number of false positive, and the simulated R^2.

III. BACKGROUND ON THE VSI AND SSB PROGRAMS

A. INTRODUCTION

When the draft ended in 1973, military manpower policy makers faced the problem of meeting force manpower needs without the assistance of a draft. Policy was developed and implemented to attract and retain high quality personnel. To ensure these policies achieved the required goals, manpower analysts conducted numerous studies analyzing retention and turnover behavior.

Present day requirements necessitate a smaller military force for two reasons. The first reason is the reduced threat brought on by the end of the cold war. Since the collapse of the Soviet Union in 1991, a constant topic of debate centers on the proper size of the United States Military.

Increased public concern over government spending is the other reason. The large federal budget deficit and resulting debt have brought all federal programs under scrutiny for possible reductions in cost.

Currently, two programs, the Voluntary Separation Incentive (VSI) and the Special Separation Bonus (SSB), are being offered to entice career military service members to leave the service. This section of the thesis provides background information on these two programs, the theoretical framework of occupational choice, and an examination and critique of a classical regression forecasting model built to predict the "take-rates" of these two programs among United States Marine Corps Enlisted personnel.

B. BACKGROUND

The 1992 Department of Defense (DoD) Authorization Act provides the military services with the VSI and SSB programs. The Marine Corps views these programs as positive force management tools which assist in shaping the career force structure. Essentially, these programs provide monetary incentives to military service members who chose to leave the service. This section provides a detailed description of each program, program eligibility requirements, and a description of the program implementation process used by the Marine Corps [Ref. 19:p. 1].

1. Voluntary Separation Incentive Program Description

The VSI program is an annuity payment made over several years. The payment continues for a period of time equal to twice the number of years the service member has served on active duty. The amount of the payment is calculated using the following formula:

ANNUAL PAYMENT = (.025) X (Final Monthly Base Pay) X (12) X (Years of Active Service)

Fractions of years of service are calculated as they are for retirement purposes. Table 1 presents four of the possible

VSI payments a Marine Corps Enlisted Member could receive. [Ref. 19:p. 2]

TABLE 1.

VSI PAYMENTS FOR SELECTED MARINE ENLISTED PERSONNEL

RANK	YEARS OF SERVICE	BASE PAY	VSI PAYMENT	NUM OF YRS
E-4	6	$1,082.10	$1,974.78	12
E-5	10	$1,426.50	$4,279.50	20
E-6	12	$1,706.70	$6,144.12	24
E-6	14	$1,760.10	$7,392.42	28
E-7	16	$2,022.90	$9,709.92	32

Source: Author

Footnote: These payments are for FY 93 separations and represent before tax amounts.

Service members who chose VSI incur an obligation to serve in the ready reserve. The reserve commitment lasts as long as the service member continues to receive VSI annuity payments. If at any time the member is separated from the ready reserve, VSI payments stop. However, VSI payments continue if either of the following conditions apply:

1) The member dies (payments will continue to the member's beneficiaries); or

2) The member becomes ineligible for continued service in the ready reserve for any reason that is not under his control (i.e., age, promotion failure, medical, etc.). [Ref. 19:p. 1]

2. Special Separation Benefit Program Description

SSB differs from VSI mainly in that SSB is a lump sum payment given upon transfer from the active force to the ready reserve. The lump sum is computed using the following formula:

PAYMENT = (.15) X (Final Monthly Base Pay) X (12) X (Years of Active Service)

Fractions of years of service are computed in the same way as for a member who receives non-disability separation pay. Table 2 presents five of the possible SSB payments a Marine Corps Enlisted Member could receive. [Ref. 19:p. 3]

TABLE 2.

SSB PAYMENTS FOR SELECTED MARINE ENLISTED PERSONNEL

RANK	YEARS OF SERVICE	BASE PAY	SSB PAYMENT
E-4	6	$1,082.10	$11,848.68
E-5	10	$1,426.50	$25,677.00
E-6	12	$1,706.70	$36,864.72
E-6	14	$1,760.10	$44,354.52
E-7	16	$2,022.90	$58,259.52

Source: Author

Footnote: These payments are for FY 93 separations and represent before tax amounts.

Members who elect separation under SSB incur a three year obligation to serve in the ready reserve. [Ref. 19:p. 3]

3. Separations Benefits

The transition benefits for separating members changed as the programs were implemented. During fiscal year 1992, members separated under VSI were entitled to the same transition benefits as those members who **voluntarily** left the service, but members who left the service under the SSB program received the same benefits as members who were **involuntarily** separated from the service. The important point here is that the non-pecuniary benefits associated with each program were different in fiscal year 1992. For example, a service member choosing SSB is entitled to two years of commissary and exchange privileges after his discharge. An individual selecting VSI does not receive any commissary or exchange privileges. [Ref. 19:pp. 2-3]

In fiscal year 1993 this changed. The 1993 DoD Authorization Act standardized the non-pecuniary benefits members received when voluntarily leaving the service under the VSI and SSB programs. [Ref 20:pp. 2-4]

4. Program Eligibility Requirements

If a member decides to separate from the service, he can chose either the VSI or SSB program. The two programs are designed to be offered simultaneously. In order to qualify for the programs, Congressional Acts require that service members meet the following criteria:

1) Serves on active duty for more than six years as of 5 Dec. 1991;

2) Completes their initial term of enlistment;

3) Serves at least five years of continuous active duty immediately preceding the date of separation;

4) Are not eligible for retirement; and

5) Are a regular officer or enlisted, or a reserve officer on the active duty list. [Ref. 21:p. 3]

In addition to these basic requirements, each service could further delineate eligibility requirements. The Marine Corps chose to further limit participation in the program. Marines in the following categories are ineligible for the bonuses:

1) Those who have received a re-enlistment bonus in fiscal years 1991 or 1992;

2) Those who are pending administrative action that could result in a involuntary discharge (i.e., court-martial);

3) Those who have been previously denied re-enlistment during fiscal year 1992;

4) Those who require a waiver for re-enlistment; and

5) Those who have previously refused to extend or re-enlist to avoid accepting orders to a new duty station. [Ref. 19:p. 3]

5. Marine Corps Program Implementation

Implementation of the separation programs by the Marine Corps is fairly straightforward. The programs are offered in three phases. Both officers and enlisted personnel are eligible for each phase of the program. However, since the focus of this thesis is on enlisted personnel, only the implementation plan for enlisted personnel is reviewed.

Phase I for both fiscal year 1992 and 1993 targets Marine enlisted personnel with military occupational specialties that are no longer needed by the Marine Corps. These specialties are for equipment that the Marine Corps no longer possesses in its inventory. For example, a Marine with a military occupational specialty that deals with operating or maintaining the F-4 aircraft is ineligible for the program, because the Marine Corps has phased out the F-4 aircraft and has replaced it with the newly acquired F-18. [Ref. 22:p. 2]

Additionally, during Phase I of fiscal year 1993, the programs are opened to enlisted members who are not selected for promotion to E-7 or E-8 but are in the promotion zone. Also, E-5s and E-6s who have more than eleven or fifteen years of service, respectively, are eligible for the programs. [Ref. 23:p. 3]

Phase II for both fiscal years is designed to elicit volunteers from miliary occupational specialties that are over staffed. The programs levy specific grade and years of service requirements. Enlisted personnel in the ranks of E-4 through E-7 are eligible. For example, Marines with the MOS 0151 (Administrative Clerk) who are E-5s and have more than eleven years of service are eligible to apply for the separations incentives. The goal is to reduce overages in these skills in order to put promotions back in line with the rest of the enlisted force. [Ref. 22:pp. 2-3]

Phase III for fiscal year 1992 targets E-6s and E-7s in crowded military occupational specialties. For example, E-6s and E-7s who possess the MOS 0369 (Infantryman) can apply for the programs. Again, the goal is to shape the force and relieve the overcrowding pressure that has led to a slow-down in promotions for several different communities. During fiscal year 1993, Phase III is only open to officers. [Ref. 20:p. 2]

C. THEORETICAL FRAMEWORK

The decision process a service member goes through to determine if he/she will stay in or leave the military is unique to the individual and involves many non-observable factors. Individual preferences for military or civilian life are difficult to measure and weigh heavily in the decision to stay or leave the military. The economic theory that best fits this decision is the theory of occupation choice.

This theory states that an individual will select the occupation that provides the largest expected lifetime utility. This utility includes both pecuniary and non-pecuniary aspects. Occupations with higher levels of compensation attract more individuals than those with low compensation (holding all else constant). However, other factors contribute to the decision about which occupation to choose. These factors include individual preferences for occupational tasks and working conditions (i.e., risk, cleanliness, management philosophy, family separation time, etc.). [Ref. 24:p. 258]

If an individual decides the military is the occupation best suited to maximize his utility, he has examined not only the actual monetary compensation he can receive, but the non-monetary aspects of military life as well. Intangible factors such as the individual's preference for military order and discipline, family separations, patriotic feelings, the stigma (positive or negative) associated with military service, and other elements contribute to the occupational choice decision.

The population targets for the VSI and SSB programs are service members who have finished their initial obligations at a minimum and have chosen to remain in the service. They are experienced with the military lifestyle. Since they have chosen to remain in the military, it is evident that this lifestyle is preferred over the life style they perceive they would live if they leave the military for another occupation.

Most military personnel have limited experience with the civilian labor market but have family and friends who are in the civilian labor force. These contacts provide information useful in determining if the service member's preferences are better suited in a civilian occupation. Although these contacts do not provide perfect information, they do provide some information to the individual. The decision to remain in the military comes about because the individual believes his utility is maximized in the military environment.

Previous studies support the theory that the potential lifetime earnings of an individual are based upon three components: observable ability indicators (education, experience, mental aptitude, etc.), socioeconomic factors (race, gender, marital status, etc.), and non-observable components of ability (assertiveness, loyalty, responsibility, etc.). These studies include Tinney (1991), Berndt (1988), and Cymrot (1987).

Tinney (1991) develops a mathematical equation that summarizes the occupational choice theory. He states that the potential lifetime earnings **E** of an individual **i** in occupation **j** is a function of observable ability indicators **OA**, demographic characteristics **D**, and non-observable ability indicators **NA**. This relationship is represented by:

$$E_{ij} = f(OA_i,\ D_i,\ NA_i)\,. \tag{1}$$

Tinney further states that the present value **V** to an individual **i** in occupation **j** is a function of the present value of lifetime earnings E_{ij}, observable individual preferences and family background effects OP_i, and non-observable individual preferences and family background effects NP_i. This may be represented by:

$$V_{ij} = g(E_{ij},\ OP_i,\ NP_i)\,. \tag{2}$$

The theory of occupational choice states that individual **i** chooses occupation **j** when **V** is maximized. [Ref. 25]

Noblit (1993) applies this concept to the stay or leave decision facing members of the Marine Corps and develops the following mathematical equations:

$$V_{im} = g(E_{im}, OP_i, UP_i) \qquad (3)$$

$$V_{ic} = g(E_{ic}, OP_i, UP_i). \qquad (4)$$

where the subscript "m" indicates the occupation of being in the Marine Corps and the subscript "c" represents a civilian occupation. [Ref. 21:pp. 23-45]

Based on these equations, Noblit stated that if $V_{im} > V_{ic}$ then the individual remains in the service. If the opposite is true, and $V_{ic} > V_{im}$, then the individual separates from the service [Ref. 21:pp. 23-45]. Therefore, in order for an individual to take either of the separation bonuses, the present value of the dollar amount of the bonus B_i plus the present value of the non-monetary separations benefits SB_i must be at least equal to the difference between the present values of a job in the Marine Corps and a job in the civilian sector, or:

$$B_i + SB_i > V_{im} - V_{ic} \qquad (5)$$

Equation (5) must be satisfied for an individual to take either of the separation bonuses.

As a member continues his active duty service and is closer to reaching the minimum retirement criteria, the present value of military retirement increases. At the same time, the number of years to accumulate a civilian retirement decreases. Additionally, an individual becomes less likely to leave the service the longer he remains because he grows more accustomed to the military lifestyle and less likely to undertake a career change.

These two factors cause a widening of the gap between the present value of a military occupation and the present value of a civilian occupation as the length of service increases. If the goal of the services is to attract senior members to voluntarily leave the service, the present value of the separation benefits (both pecuniary and non-pecuniary) must be a positive function of years of service. Since the non-pecuniary aspects of each program are constant among service members, the monetary value of each separation incentive program must increase with the individual's years of service.

D. THE NOBLIT MODEL

In building his model to predict the "take-rates" of the VSI and SSB programs, Noblit reviews current literature on the effects of the Selective Reenlistment Bonuses, voluntary termination from the service, reenlistment behavior of careerists, and retention behavior. These studies laid the foundation for his model. [Ref. 21:pp. 11-34] This section provides a description of the model and a brief discussion of the variable choice.

1. Functional Form

The decision to take either of the separation bonuses is, in essence, a stay or leave decision by the individual. Therefore, the dependent variable of Noblit's model is dichotomous. When modeling a dichotomous dependent variable, researchers may choose a model from several options with different functional forms. [Ref. 26:pp. 510-529] Noblit's model is a binomial LOGIT which uses the maximum likelihood estimation technique.

The binomial LOGIT model has the following functional form:

$$P_i = \frac{1}{1+e^{-(\beta_0+\beta_i X_i+\epsilon_i)}}$$

P_i is the probability that individual i has the characteristic defined by $P_i = 1$. The "betas" are the parameter estimates. X_i represents the vector of explanatory variables. The epsilon is the stochastic error term associated with the model.

2. Explanatory Variable Selection

The theoretical model of occupational choice states that an individual chooses the occupation that maximizes his lifetime utility. In choosing between a military occupation and a civilian one, an individual examines the present value of military compensation, the present value of civilian compensation, and the present value of non-pecuniary factors that determine the individual's taste for military service or civilian service. Noblit's model develops proxy variables for each of these elements that contribute to the stay or leave decision.

a. Present Value Of Military Compensation

Developing proxy variables for the present value of military compensation is the easiest element of the stay or leave decision to model. Military compensation is based upon three categories. First, there are the entitlements that all Marines are eligible for: base pay, housing allowances and subsistence allowance. Base pay is determined by the individual's rank and years of service. Housing allowances and subsistence allowances are computed based upon rank and marital status.

Second, total military compensation includes reenlistment bonuses paid to enlisted service members. These bonuses are a function of reenlistment contract, rank, base pay, and the current stock of personnel relative to requirements in similar military occupational specialties. Not all enlisted members are entitled to or receive re-enlistment bonuses.

The last element of military compensation centers around proficiency pay. Marines receive extra pay if they possess certain skills. These skills include airborne or scuba qualified, flight crew member, explosive ordnance disposal, and foreign language capable. Marines serving as embassy guards, recruiters, and drill instructors receive proficiency pay as well.

Noblit chose four explanatory variables to act as proxies for the present value of military compensation. These variables represented the individual's rank, marital status, and whether or not the individual received either a reenlistment bonus or proficiency pay. [Ref. 21:p. 40]

Each of these variables are factors which determine the total monetary compensation an individual receives for active military service. A service member can calculate the present value of his expected military compensation because this compensation is fixed by law and is based upon rank, years of service, marital status, job proficiency, and reenlistment bonuses. If the individual assumes he receives promotions and longevity raises at the same rate as his predecessors, then calculating the present value of his military pecuniary benefits is straightforward.

An interesting point here is that Noblit chose to leave years of service out of his model. This was done mainly because of the high correlation between rank and years of service. The military is a hierarchical organization and members, with few exceptions, reach the upper grades only after serving for a particular amount of years. [Ref. 21:p. 40] *b. Present Value Of Civilian Compensation*

Civilian compensation is a function of observable ability indicators and socioeconomic factors. Noblit develops four proxy variables to represent the present value of civilian compensation. These variables represent the education level, gender, race, and military occupation specialty of the individual. The large number of different occupational specialties within the military necessitated the grouping of individuals into four similar occupational categories. The groupings were: 1) Admin/Supply, 2) Combat, 3) Combat Service Support (non-technical), and 4) Combat Service Support (Technical). [Ref. 21:pp. 41-43]

c. Non-Pecuniary Job Choice Factors

Noblit creates six proxy variables to represent the individual's satisfaction with miliary life. These variables represent promotion rates, type of duty, and job satisfaction.

Two explanatory variables represent promotion rates. The first variable is the length of time the individual has spent in his present grade. The other explanatory variable is a promotion index. This promotion index indicates if the individual is being promoted at a faster or slower rate than his peers. It is calculated by dividing the individual's time-in-service at his last promotion by the average time-in-service at the last promotion for Marines in the same grade and occupational specialty. If the index is greater than one, the individual is being promoted more slowly than his peers. If it is less than one, the opposite is true. [Ref. 21:pp. 43-44]

Two explanatory variables represent type of duty. Marines are classified into three categories of duty: Atlantic Operating Forces, Pacific Operating Forces, and supporting forces. Noblit develops one variable to indicate the assignment of Marines to units in the Atlantic Operating Forces. Another variable represents Marines who are not assigned to the operating forces but to the supporting establishments. Marines in the Pacific Operating Forces are treated as the omitted condition in the model. [Ref. 21:p. 44]

Two explanatory variables represent job satisfaction. One variable indicates whether or not the individual is in his primary military occupational specialty. The other measures how much time remains on the individual's enlistment contract. This is a good proxy variable for job security. Those close to their end of active service may fear they can not re-enlist thus causing job dissatisfaction. [Ref. 21:p. 45]

Table 3 contains a list of the types of variables Noblit uses in his model.

E. CRITIQUE OF THE NOBLIT MODEL

Noblit's model is well specified and built upon a strong theoretical foundation. Each proxy variable represents factors that, in theory, are important in deciding to stay or leave the military. Noblit chose the proper functional form for the model. The only area for improvement is the choice of proxy variables. The present value of military compensation is well represented. However, both the present value of civilian compensation and the non-pecuniary aspects of military life could be represented more thoroughly with the addition of several other proxy variables.

TABLE 3.

EXPLANATORY VARIABLES USED TO BY NOBLIT PREDICT THE "TAKE-RATES" OF THE VSI AND SSB PROGRAMS AMONG MARINE CORPS ENLISTED PERSONNEL

Variable Description	Proxy For:
Rank	Mil. Compensation
Proficiency Pay	Mil. Compensation
Marital Status	Mil. Compensation
Bonus	Mil. Compensation
Educational Level	Civ. Compensation
Gender	Civ. Compensation
Race	Civ. Compensation
Mil. Occ. Specialty	Civ. Compensation
Time In Grade	Non-Pecuniary Factor
Promotion Index	Non-Pecuniary Factor
Work In Primary MOS	Non-Pecuniary Factor
Job Location	Non-Pecuniary Factor
Time Left On Contract	Non-Pecuniary Factor

Source: Author.

In addition to the proxies Noblit specified to predict civilian earnings, there are three other possibilities. First, the mental ability of an individual has an impact on his potential to earn more money in the civilian sector. The GCT score of the individual can be used as a proxy variable for the individual's mental capability.

Age is another factor that impacts civilian earnings. Age tends to act as a proxy for the amount of experience individuals possess and their level of maturity. Civilian earnings is a function of work experience. Work experience is a function of age. It is understood that military and civilian experience differ in some respects. An individual may not be able to apply all of his military skills in a civilian job, but some of the skills developed while in the Armed Forces are transferable.

The last possibilities of proxies for the present value of civilian earnings potential centers around the additional occupational specialties a Marine has been assigned. Assignment to jobs such as recruiting duty, embassy duty, instructor duty, and joint duty are highly competitive. Only the best and the brightest receive these assignments. Proxy variables designed to capture this information may reveal information about the non-observable capabilities of the individual that would have an impact upon civilian earnings.

To assess the present value of non-pecuniary aspects of military life, two additional variables should be added to

Noblit's model. The Marine Corps is highly physical and places great demands on individual to maintain certain standards. These standards include both weight and athletic ability. Individuals who are not athletically inclined or tend to gain weight more easily than others may put forth more effort to maintain the standards than the average Marine. This extra effort may not result in any extra reward for the individual which in turn may lead to higher dissatisfaction with the service. The individual's personal fitness test score should be added to the model as proxy variables to account for this potential source of job dissatisfaction.

The other variable needed in Noblit's model centers around the amount of time an individual spends away from his family in a deployed status. Marines deployed for long periods of time tend to show greater dissatisfaction for military life. The amount of time an individual has spent deployed will act as an excellent proxy for this negative non-pecuniary aspect of life in the Marine Corps.

Noblit's model forms an excellent starting point. However, the addition of five explanatory variables captures more of the aspects that make up the decision to stay or leave the military. The five variables are contained in Table 4. The cost of adding these five variables is relatively low. The variables Noblit used as well as the five additions are all available through the Marine Corps Enlisted Master File.

TABLE 4.

ADDITIONAL EXPLANATORY VARIABLES USED TO PREDICT THE "TAKE-RATES" OF THE VSI AND SSB PROGRAMS AMONG MARINE CORPS ENLISTED PERSONNEL

Variable Description	Proxy For:
GCT Score	Civ. Compensation
Age	Civ. Compensation
Additional MOS Type	Civ. Compensation
Deployed Time	Non-Pecuniary Factor
Physical Fitness Test Score	Non-Pecuniary Factor

Source: Author.

F. CONCLUSION

The theory of occupational choice sets the framework for the model Noblit built to predict the "take-rates" of the VSI and SSB programs for Marine Corps enlisted personnel. The model is appropriate for the data and the explanatory variables are excellent proxies for use in modeling the stay or leave decision. Five additional variables are suggested to capture more of the factors that weigh in the decision. Utilization of the model will allow manpower planners to implement the separations incentive programs with more efficiency.

IV. DATA AND METHODOLOGY

A. INTRODUCTION

One method of determining the predictive power of artificial neural networks is to compare the forecasting capabilities of an artificial neural network with some other established prediction tool. Since military manpower and personnel planners use classical regression techniques in forecasting the effects of policy changes, this research compares the forecasting abilities of neural networks with classical regression techniques.

This comparison involves four steps. First, an appropriate data set on the behavior of interest is obtained. The second step is to develop a classical regression model based upon previous research with the aid of a statistical software package, SAS. The third step is to build two neural network models using a commercially available software product, NeuralWorks Professional II/Plus. The fourth, and last step, is to compare the results of each of the neural network models to the regression model.

This chapter of the thesis outlines the data set and variables used to develop all three models and then provides a description of each model's structure.

B. DATA

The research utilizes a data set obtained from the Master Files of Enlisted Personnel maintained at Headquarters, Marine Corps, Washington, D. C. It contains the population of eligible enlisted personnel who are qualified to leave the Marine Corps under the VSI and SSB programs during fiscal year 1993. The data set was developed by manpower planners within the Enlisted Force Planning Section of the Manpower Management Division.

The original data set contains 19,212 observations. After deleting observations with missing values and removing observations that contain values which make little logical sense, there are 17,535 observations. Further restrictions placed on the data result in a final subset of males who meet the following criteria:

1) are an E-5 or an E-6;
2) have between eight years of service (YOS) and fourteen years of service, inclusive;
3) are at least 26 years old but not older than 34 years old; and
4) have at least a high school education.

The resulting data set contains 8,042 observations. Table 5 provides a summary of the deletions performed on the population of eligible enlisted members who qualify for the VSI and SSB programs during fiscal year 1993.

The primary purpose behind restricting the data is to obtain an appropriately homogeneous data set for analyzing VSI and SSB program subscribers. Restricting the data to the

TABLE 5.

DELETIONS FROM THE POPULATION OF MARINE ENLISTED PERSONNEL ELIGIBLE TO SEPARATE FROM THE MARINE CORPS UNDER THE VSI AND SSB PROGRAM

Reason for Deletion	# of Observations Meeting Criteria	
Initial Population Size:		19,212
Missing/Illogical variable values:	1,677	
Female:	961	
E-4	287	
E-7	3,905	
YOS < 8	727	
YOS > 14	6,692	
AGE < 24	6	
AGE > 36	2,417	
Final Data Set Size :		8,042

Footnote: Numbers do not sum to total due to possibility of an observation meeting more than one criteria for deletion (i.e., female who is an E-4).
Source: Author

above criteria ensures the development of a theoretically sound classical regression model. During the research, separate linear probability models are estimated on appropriate subsets divided by gender, rank, and years of service. A chow test is performed on each model to determine coefficient stability between models. The resulting "F" statistic supports the conclusion that separate models are required to estimate the impacts of the explanatory variables for each of the subsets.

The data are divided into three subsets. The first subset is labeled the "test" data subset and its purpose is twofold. First, it is used to validate each of the **trained** models. Second, it provides data which has not been seen by any of the models and is used to generate test statistics that are used as comparison metrics. The test data subset is created with the assistance of the random number generator available in the SAS software package. The end result is a randomly selected test data subset which contains approximately ten percent of the observations from the original data. This amounts to 819 observations of which 80 (9.8 percent) are takers of the VSI and SSB programs.

The other two subsets are "training" data subsets and contain observations that **train** all of the models. The term **train** means to estimate the explanatory variables' coefficients in the case of the classical regression model and to determine the weights of each processing element in the ANN models. The classical regression model's training data subset amounts to every observation not placed in the test data subset: 7,223 observations, of which 654 (9.1 percent) are takers of the VSI and SSB programs.

When training an ANN, it is important that the training data set possess a relatively equal quantity of all possible outcomes. In the full data set, the ratio of takers to non-takers of the VSI and SSB programs is approximately nine to one. To ensure the network learns relationships and not probability distributions, the training data subset used by the ANN models contains 1,400 observations representing 654 takers and 746 non-takers. Observations are chosen from the ANN training set in a random method similar to the one used for the test data subset.

In his research, Noblit used the entire population of Marine Enlisted personnel who qualified for separation from the Marine Corps under the VSI and SSB programs during fiscal year 1992. The data for this research differs slightly for two reasons. First, the data set is restricted for the structural homogeneity purposes discussed above thus producing a subset and not the population. The second difference concerns the difference in fiscal years. Noblit used fiscal year 1992 data while this research uses fiscal year 1993 data.

C. VARIABLE DESCRIPTION

This section is divided into two parts. The first section lists those variables used in constructing the regression model. The second section lists those variables used in building the neural network models. Every variable used in the classical regression model is used in the ANN models. However, the ANN models contain several additional variables.

When constructing classical regression models, researchers must use great care in selecting explanatory variables. Researchers choose the "best" variables which act as proxies for certain characteristics. Causal modelers must be cautious that little multicollinearity exists between the explanatory variables. This is not the case with ANN models. ANN models have been postulated to possess the capability to handle multiple measures of the same effect and account for the inter-relationships between explanatory variables [Ref. 7]. Therefore, ANN model developers often present all available variables to the ANN even if there exists substantial multicollinearity between them. The usefulness of the resultant ANN models for policy purposes is discussed in Chapter VII.

The explanatory variables are classified into three groups of proxy variables that represent: 1) the potential value of military earnings, 2) the potential value of civilian earnings, and 3) the non-pecuniary aspects of military life versus civilian life.

1. Explanatory Variables Used In The Classical Regression Model

The variables in this model are chosen based upon previous research conducted by Noblit (1993). Appendix A contains a comparison of the variables used by Noblit and the variables used in this research.

a. Military Compensation

This research uses four variables as proxies for military compensation. They represent rank and marital status. Noblit specified his model to include these, plus proxy variables to represent proficiency pay and re-enlistment bonuses. This research did not use variables to represent proficiency pay or re-enlistment bonuses because the data did not contain accurate information for these conditions. Table 6 contains a summary of each proxy variable and its expected sign.

TABLE 6.

PROXY VARIABLES OF MILITARY COMPENSATION FOR CLASSICAL REGRESSION MODEL

Variable Name	Variable description	Expect Sign
SSGT	= 1 if obs is an E-6; 0 otherwise	-
SINGLE	= 1 if obs is not married; 0 otherwise	+
SINWDEP	= 1 if obs is single and has dependents	+
MARRIED	= 1 if obs is married w/o children	+
MARWCHLD	= 1 if obs is married and has children	NA*

* This is treated as the base case
Source: Author

(1) Rank. The variable which represents rank, **SSGT**, is set equal to one if the observation is an E-6 and is set equal to zero otherwise. The hypothesized sign is negative because an individual who is an E-6 earns more money than an E-5 and, therefore, would be less likely to leave the service.

(2) Marital Status. Marital status is represented by four dummy variables, **SINGLE**, **SINWDEP** (single with dependents), **MARRIED** (no children), and **MARWCHLD** (married with children). The variable corresponding to the observation's status is given the value of one while the others are coded as zero. Married military members with children are expected to be less likely to undertake an occupational change because of their family responsibilities. Therefore, it is hypothesized that the variables **SINGLE**, **SINWDEP**, and **MARRIED** will have positive coefficients when compared to the base case of being married and having children (**MARWCHLD** = 1).

b. Civilian Compensation

Thirteen variables act as proxies for the potential earnings of an individual in the civilian job market. They represent race, military job, education level, intelligence level, and experience. Noblit's model used all of these factors except the last three. Table 7 contains a summary of each variable and its expected sign.

(1) Race. Race is represented by three dummy variables, **WHITE**, **BLACK**, and **OTHMIN** (other minorities). The variable corresponding to the observation's racial group is given the value of one and the other variables receive the value of zero. Previous studies indicate caucasians face better civilian employment opportunities. Therefore, the expected sign of the coefficients of the variables **BLACK** and

OTHMIN should be negative when the variable **WHITE** acts as the base case.

(2) Military Occupational Specialty. An observation's military occupational specialty (MOS) is represented by four dummy variables, **FIGHTER**, **ADMIN**, **NONTECH** (non-technical MOS), and **CSSTECH** (technical MOS). Individuals who have technical MOSs receive more formalized training which

TABLE 7.

PROXY VARIABLES OF CIVILIAN COMPENSATION FOR CLASSICAL REGRESSION MODEL

Variable Name	Variable description	Expect Sign
BLACK	= 1 if obs is Afro-American; 0 otherwise	-
OTHMIN	= 1 if obs is hispanic, asian or other	-
WHITE	= 1 if obs is a caucasian; 0 otherwise	NA*
FIGHTER	= 1 if obs has a combat MOS; 0 otherwise	-
ADMIN	= 1 if the obs is an administrator	-
NONTECH	= 1 if the obs has a non-technical MOS	-
CSSTECH	= 1 if the obs has a technical MOS	NA*
YR3COLL	= 1 if obs has 3 or more yrs of college	+
SOMCOLL	= 1 if obs has 2 or less yrs of college	+
HISCHGRD	= 1 if obs is a high school graduate	NA*
INTEL	Continuous value equal to GCT score	+
AAGE	Continuous value equal to obs age	+
BONUSMOS	= 1 if obs is a recruiter/drill instructor	-

* These are treated as the base case.
Source: Author

is easily transferred to the civilian marketplace. These individuals can make a transition to the civilian sector more smoothly and have greater opportunities in a civilian market than those with other MOS types. Therefore, it is hypothesized that the signs of the coefficients for the variables **FIGHTER**, **ADMIN**, and **NONTECH** will be negative when the variable **CSSTECH** acts as the base case. Appendix B contains a list of MOSs and their occupational categories.

(3) Education Level. Education level is represented by three dummy variables, **HISCHGRD** (high school graduate), **SOMCOLL** (two or less years of college), and **YR3COLL** (three or more years of college). Previous studies provide evidence that individuals with higher education levels receive more compensation than those with limited formal schooling. Military members with more formal education have better opportunities in the civilian marketplace than those with less education. Therefore, it is hypothesized that if the variable **HISCHGRD** acts as the base case, the variables **SOMCOLL** and **YR3COLL** will have positive signs.

(4) Intelligence Level. The variable **INTEL** is the value of the observation's GCT score. This score is a composite of several of the separate ASVAB scores. It acts as a proxy for an individual's mental ability. Individuals who have higher mental abilities face better civilian opportunities and, therefore, would tend to make a transition

to civilian life at a higher rate than those with lower mental ability. The hypothesized sign for this variable is positive.

(5) Age. The variable **AAGE** is the observation's age. This acts as a proxy variable for the individual's experience level. Older individuals have a broader experience base and greater opportunities to receive additional training, both general and specific. Previous studies indicate that individuals with more experience receive higher compensation. Thus older individuals within our data (in the range 26 to 34 years old) have better opportunities in the civilian market place than the younger individuals. Therefore, the hypothesized sign is positive.

(6) Bonus Skills. The variable **BONUSMOS** is a dummy variable assigned the value of one if the observation is a recruiter or a drill instructor. The selection process for these two occupations is extremely competitive. If a Marine is chosen to hold either of these jobs, the Marine Corps has provided the individual with a message implying that he fits well in the organization. This increases the individual's confidence that he possesses the skills and abilities needed to complete a military career successfully. Therefore, a negative sign for this variable is hypothesized. A drill instructor or recruiter is anticipated to be less likely to leave the Marine Corps under these programs.

c. Non-Pecuniary Aspects of Military Life

Ten variables are used to capture non-pecuniary aspects of military life. These variables include measures of current job satisfaction and job duty type. Noblit's research used each of the same measures with one exception, physical fitness test class. Table 8 contains a summary of each variable and its expected sign.

TABLE 8.

PROXY VARIABLES OF NON-PECUNIARY ASPECTS OF MILITARY LIFE FOR CLASSICAL REGRESSION MODEL

Variable Name	Variable description	Expect Sign
INMOS	= 1 if obs works in his primary skill	-
SLOWTRK	= 1 if obs is promoted slower than peers	+
FASTTRK	= 1 if obs is promoted faster than peers	-
AVGPROMO	= 1 if obs is promoted with his peers	NA*
JUNIOR	= 1 if the obs has 2 yrs or less TIG	-
MOSEAS	Continuous value equal to time remaining on active duty	-
PFT_1ST	= 1 if the obs scores well on the PFT	-
FMFLANT	= 1 if obs duty station is East Coast	+
FMFPAC	= 1 if obs duty station is West Coast	+
NONFMF	= 1 if obs is not in operating force	NA*

* These are treated as the base case.
Source: Author

(1) Job Satisfaction Variables. Job satisfaction is represented by seven variables which capture information on current job, promotability, job security, physical fitness,

and reward factors. The first variable, **INMOS**, is a dummy variable and is given the value of one if the observation is currently working in his primary MOS. The hypothesis is that an individual who is filling a position other than what he is trained for, has a lower job satisfaction level and, therefore, wants to explore other alternatives including leaving the service. The expected sign is negative. An individual working in his MOS will be less likely to leave the Marine Corps.

The next three variables are dummy variables and capture the rate at which the Marine is being promoted in relationship to his peers. The dummy variables are **SLOWTRK** (representing a slower promotion rate), **AVGPROMO** (representing an average promotion rate), and **FASTTRK** (representing a faster promotion rate). The observation's time-in-service at his last promotion is compared to the average for his occupation. An index is created by dividing the time-in-service at last promotion by the average time-in-service for Marines of the same pay grade and occupational category. Those with a high index (time-in-service at last promotion is greater than the average) are on the slow track and those with a low index (time-in-service at last promotion is less than the average) are on the fast track. The variable **AVGPROMO** acts as the base case. The expected sign is positive for **SLOWTRK** since low promotion opportunities are likely to lead to higher job dissatisfaction which results in voluntary separation from the service. The expected sign for **FASTTRK** is negative as fast promotion rates tend to lead to job satisfaction and retention.

The fifth variable representing job satisfaction, **JUNIOR**, is a dummy variable which is assigned a value of one if the observation has two years time-in-grade (TIG), or less, and zero otherwise. This captures the "motivational" aspects associated with promotions. The hypothesis is that those who have recently been promoted are happy with the extrinsic awards they have received and therefore, are less likely to separate voluntarily from the service. Thus the expected sign is negative. An individual recently promoted is less likely to attempt separation under the VSI and SSB programs.

The sixth variable representing job satisfaction captures information on job security. The variable, **MOSEAS**, is a continuous variable which represents the amount of time, in months, until the observation's end-of-active-service (EAS) date arrives. The hypothesis is that those closer to their EAS are considering an occupational change and are more likely to separate. The sign of this variable should be negative. The larger the variable (further from EAS), the less likely the observation will take the program.

The last variable representing job satisfaction is a dummy variable, **PFT_1ST**, and represents the score on the physical fitness test. Twice a year, all Marines are required to take a physical fitness test (PFT). The score on the test is categorized into three classes based on the individual's score and his age. Those who receive a "first class" are normally in top physical condition and have little difficulty completing any physical activity required in the execution of their duties. Those who fall into the "second" or "third" class are in moderate or low physical condition and may experience difficulty in completing physical activity associated with their duties. The variable **PFT_1ST** receives a value of one if the observation has a first class score and a value of zero otherwise. The hypothesis is that individuals with a first class score have less job dissatisfaction due to their physical abilities and, therefore, are less likely to separate. Thus the expected sign is negative. Marines with first class PFT score are less likely to separate from the service under the VSI and SSB programs.

(2) Job Duty Type. Job duty type is represented by three dummy variables, **NONFMF**, **FMFLANT**, and **FMFPAC**. Operating Forces in the Fleet Marine Force (FMF) experience long deployments and busy training cycles. East Coast units in the FMF fall under the operational command of the Commander, Fleet Marine Forces, Atlantic (FMFLANT). West Coast units in the FMF fall under the operational command of the Commander, Fleet Marine Forces, Pacific (FMFPAC). All other non-operating forces are classified as being non-FMF. The variable corresponding to the observation's status is assigned a value

of one while the other variables are assigned a value of zero. Individuals in the FMF experience frequent deployments and intense training cycles unlike those in non-FMF billets who experience a more stable job environment. These differences can lead to higher job dissatisfaction amongst Marines assigned to FMF units. Therefore, the expected sign of the variables **FMFLANT** and **FMFPAC** are positive. Thus those in the FMF tend to separate under the VSI and SSB programs at a higher rate than those in the non-FMF.

2. Explanatory Variables For The Artificial Neural Networks

The ANN models are presented with all variables contained in the data. Thus, the ANN models use every variable which the classical regression model uses including the omitted variables, plus nine others. Of these nine, one impacts upon military compensation, five are proxy variables for potential civilian compensation, and three measure non-pecuniary aspects of military life.

a. Military Compensation

The continuous variable, **YOS**, measures the number of years of active service an individual has completed. This information directly impacts upon military pay but is not used in the classical regression model due to its high collinearity with the other explanatory variables. The hypothesis is that as years-of-service increase, the tendency to leave the Marine

Corps decreases because both pay and the present value of retirement benefits increase. Thus the expected sign for **YOS** is negative. The probability of separation from the service decreases as years-of-service increase.

b. Civilian Compensation

Five additional variables are measures of the individual's mental capacity. These are represented in Table 9. The first variable, **AFQTPREC**, is a continuous variable that reflects the percentile ranking of the individual's Armed Forces Qualification Test (AFQT) score. Those individuals with a high value for this variable scored better on the AFQT than their counterparts. The hypothesis is that those Marines in a higher percentile have better civilian opportunities than their counterparts and stand a higher probability of voluntarily leaving the service. Thus the expected sign is positive.

The last four variables, **ASVABRAR**, **ASVABRMK**, **ASVABRPC**, and **ASVABRWK**, are continuous variables which represent the raw scores received on the arithmetic reasoning, math knowledge, paragraph comprehension, and word knowledge portions of the Armed Services Vocational Aptitude Battery (ASVAB), respectively. The hypothesis is the same as for the variable **AFQTPREC**. Thus the expected sign for these variable is positive.

c. Non-pecuniary Benefits

Three additional variables are added to the ANN models that represent some aspect of job satisfaction. These are represented in Table 10. The first, **DCTB_YRS**, is a continuous variable which reveals the length of time a service member has been located at his current duty station. The longer he remains at one station, the more rooted in the community he becomes. His spouse and children, (if he has any), become content with their environment and the thought of disrupting this environment through a change in duty station may lead to increased job dissatisfaction. Thus, the expected sign is positive. The longer an individual has been at the same duty station, the higher the probability that he separates under the VSI and SSB programs.

TABLE 9.

PROXY VARIABLES OF CIVILIAN COMPENSATION FOR NEURAL NETWORK MODELS

Variable Name	Variable description	Expect Sign
AFQTPREC	Percentile ranking of AFQT score	+
ASVABRAR	Arithmetic Reasoning score on ASVAB	+
ASVABRPC	Paragraph Comprehension score on ASVAB	+
ASVABRWK	Word Knowledge score on ASVAB	+
ASVABRMK	Math Knowledge score on ASVAB without family	+

Source: Author

The second job satisfaction variable, **ACC_DEP**, represents the amount of time, in months, an individual has been deployed away from his duty station. Longer deployment times lead to higher job dissatisfaction and a desire to leave the service. Thus the expected sign is positive. As **ACC_DEP** gets larger, the probability of voluntarily leaving the service increases.

TABLE 10.

PROXY VARIABLES OF NON-PECUNIARY ASPECTS OF MILITARY LIFE FOR NEURAL NETWORK MODEL

Variable Name	Variable description	Expect Sign
DCTB_YRS	Number of years at same duty station	+
ACC_DEP	Number of months spent deployed	+
DAUS_DR1	Number of years since assigned to duty station without family	+

Source: Author

The last job satisfaction variable, **DAUS_DR1**, is a continuous measure of the length of time since the service member has deployed to a duty station where his dependents are not authorized. The longer this interval, the higher the probability the individual may be assigned to such a duty station. This is a good indicator of upcoming family separation time. Increased chances of family separation generate higher levels of job dissatisfaction. Thus the expected sign for this variable is positive. The larger the value of **DAUS_DR1**, the higher the probability the Marine will separate voluntarily under the VSI and SSB programs.

3. Conclusion

There are twenty-two variables used in the classical regression model. The ANN models use thirty-seven variables. The difference exists for two reasons. First, the classical regression model requires omitted conditions when using more than one dummy variable to represent a certain condition. For example, if race is the condition, an observation can be classified as white, black, and other. Only two variables are required to classify all possibilities properly. The ANN models require dummy variables for every possible category. This accounts for six of the additional variables.

The other nine variables contained in the ANN model but not in the classical regression model are additional measures of the same effects which are measured by other variables contained in both models. Adding them to the classical regression model causes severe multi-collinearity problems. The "kitchen sink" approach is used to present every available variable to the network. This simulates the approach of researchers who are unfamiliar with the theory behind the behavior they are investigating and are unable to determine relationships that exist within the data.

D. MODEL SPECIFICATION

This research builds three models, one classical regression model and two ANN models. The classical regression model is based upon the theory of occupational choice and is an extension of previous research conducted by Noblit (1993). The two ANN models differ mainly in their basic topology. This section provides a description each network and the methods employed to reach solutions.

1. Classical Regression Model

The classical regression model is a binomial logistic regression model. Logistic regression models are best suited to forecasting problems when the dependent variable is dichotomous. The dependent variable in this research (for every model) is **VSISSB** which equals one if the Marine separated from the Marine Corps under the VSI and SSB programs and equals zero otherwise.

The model has the following function form:

$$P_i = \frac{1}{1+e^{-\beta_0+\beta_1 X_{1i}+\beta_2 X_{2i}+\ldots+\beta_n X_{ni}+\epsilon_i}}$$

The value of $\mathbf{P}_i$ equals the probability that individual **i** will separate from the Marine Corps under the VSI and SSB programs, given his individual values for the set of explanatory variables (X_1, X_2, ..., X_n). B_0 is the constant term, e_i is the stochastic error, and B_i, through B_n are the coefficients for each respective explanatory variable.

2. NEURAL NETWORK MODELS

The two artificial neural network models are constructed using a commercially available software product, <u>NeuralWorks Professional II/Plus</u>. This program is chosen for two reasons. First, it provides a large degree of versatility in constructing an artificial neural network. Every aspect of the network may be adjusted to suit the needs of the developer. The other reason stems from the support provided by the software developers. The documentation provided with the software and the technical support available make this product an industry leader.

Each network is similar in design. The main differences between the two networks center on network topology. One model has one hidden layer with thirty hidden neurons while the other has two hidden layers, the first layer having twenty hidden neurons and the second layer having ten hidden neurons.

Each network employs the back-propagation learning algorithm using the normalized cumulative delta learning rule discussed earlier. The network continues training until one of two conditions is met. First, the root mean square error reaches 0.001 or the network has seen 3,500,000 training cases. Continuing on after either of these points is unnecessary because a minimal error has been reached or the network "memorizes" the training data and can not generalize to the test data.

E. COMPARISON TECHNIQUES

In determining which of the three models provides the best results, two methods are used. First, each model is run on a validation set of data which none of the models have seen before. The model which correctly predicts the largest number of actual outcomes receives the label as "best" model. Additionally, the other criterion used is the simulated R^2 discussed in Chapter II which was developed by Wiggins and Engquist.

F. CONCLUSION

In summary, this research attempts to compare the forecasting capabilities of artificial neural networks to classical regression techniques. This is accomplished in four steps. First, a homogeneous data set is obtained and divided into subsets to test and train each of the models. Second, a classical regression model is developed based upon previous research. Third, two artificial neural networks are constructed, each with a different topology. Fourth, the predictive capabilities of each model is compare using previously defined metrics.

V. RESULTS OF THE LOGISTIC REGRESSION MODEL

This chapter presents a descriptive analysis of the data, the results of the classical regression model, an interpretation of the results, and the forecasting abilities of the model when applied to the test data subset.

A. DATA DESCRIPTION

Table 11 contains the descriptive statistics for the entire data set. Included are the means and standard deviations for each continuous variable as well as the percentage distribution of the categorical variables. A little more than nine percent (734 out of 8,042) of the individuals contained in the data separated from the Marine Corps under VSI and SSB. Forty-one percent are minorities and 6.7 percent have post high school education. A little more than 31.0 percent are in a technical MOS and approximately 50.0 percent are assigned to units not in the operating forces. The average Marine in this sample is twenty-nine years old, has a GCT of 105 and is twenty-one months from his End-of-Active-Service date.

Appendix C contains tables representing additional characterist. 's of the data set. These tables include cross tabulations of race by promotion status and occupational specialty. Information in the tables indicates the possible

TABLE 11.

SIMPLE DESCRIPTIVE STATISTICS FOR THE EXPLANATORY VARIABLES IN THE LOGISTIC MODEL

Conditions represented by Variables	Percent who met Condition
VSI and SSB Program "Takers"	9.1%
SSGT (E-6s)	35.9%
Afro-American Marines	31.4%
Other Minority Marines	9.6%
Marines with two years or less of college	5.4%
Marine with more than two years of college	1.3%
Marines not married	9.1%
Marines not married, who have dependents	5.9%
Marines married, no children	14.9%
Marines in Administrative MOSs	22.2%
Marines in Combat MOSs	14.9%
Marines in non-technical MOSs	31.8%
Marine recruiters/drill instructors	21.3%
Marines working in their MOS	77.3%
Marines assigned to FMFLANT	18.9%
Marines assigned to FMFPAC	32.7%
Marines who score first class on PFT	78.9%
Marines who are promoted at a slow rate	15.1%
Marines who are promoted at a fast rate	13.6%
Marines with less than three years TIG	16.5%
Number of months until EAS (mean) (standard deviation)	21.43 (15.05)
GCT score (mean) (standard deviation)	104.95 (13.57)
Marine's Age (mean) (standard deviation)	29.67 (2.10)

Source: Author

existence of a racial relationship in selection procedures used for promotion and determination of occupational specialty. However, conclusions regarding the promotion rate and the determination of occupational specialty for different races should not be drawn at this time for several reasons. First, the data set in this research is not representative of the entire Marine Corps due to the limitations placed on it. Second, a multivariate analysis is necessary to account properly for the relationships of race and VSI and SSB program selection.

B. RESULTS OF THE LOGISTIC MODEL

Table 12 contains the results of the logistic regression model constructed to forecast the acceptance rates of the VSI and SSB programs among Marine, male, E-5s and E-6s. The probability that a member separates under either of these programs is obtained from the equation:

$$P = 1/(1 + e^{-Z}), \text{ where}$$

$$\begin{aligned} Z = -0.98 \; & + (0.06)\mathbf{SSGT} - (0.44)\mathbf{BLACK} - (0.21)\mathbf{OTHMIN} \\ & + (0.19)\mathbf{SOMCOLL} + (0.26)\mathbf{YR3COLL} - (0.74)\mathbf{ADMIN} \\ & - (0.25)\mathbf{FIGHTER} - (0.43)\mathbf{NONTECH} - (0.46)\mathbf{BONUSMOS} \\ & + (0.01)\mathbf{INTEL} + (0.04)\mathbf{INMOS} - (0.14)\mathbf{FMFLANT} \\ & - (0.27)\mathbf{FMFPAC} - (0.21)\mathbf{PFT\_1ST} - (0.009)\mathbf{SLOWTRK} \\ & - (0.33)\mathbf{FASTTRK} - (0.02)\mathbf{MOSEAS} - (0.28)\mathbf{JUNIOR} \\ & + (0.28)\mathbf{SINGLE} + (0.35)\mathbf{SINWDEP} - (0.07)\mathbf{MARRIED} \\ & - (0.04)\mathbf{AAGE} \end{aligned}$$

Note, the coefficients have been rounded off to two significant digits for equation presentation purposes.

Statistics to measure the overall fit of the equation to the data include the -2 LOG Likelihood statistic. This

TABLE 12.
LOGISTIC REGRESSION MODEL RESULTS IN FORECASTING THE TAKE-RATES OF THE VSI AND SSB PROGRAMS

Variable Name	Sign Hypothesis	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square
INTERCEPT	NA	-0.977	0.867	1.27	0.260
SSGT	-	0.065	0.119	0.29	0.588
BLACK***	-	-0.437	0.112	15.32	0.001
OTHMIN	-	-0.207	0.156	1.77	0.184
SOMCOLL	+	0.194	0.181	1.15	0.283
YR3COLL	+	0.265	0.366	0.52	0.470
SINGLE**	+	0.283	0.135	4.38	0.036
SINWDEP**	+	0.353	0.163	4.66	0.031
MARRIED	+	-0.066	0.123	0.29	0.592
ADMIN***	-	-0.741	0.128	33.36	0.001
FIGHTER*	-	-0.249	0.131	3.61	0.057
NONTECH***	-	-0.431	0.106	16.60	0.001
BONUSMOS***	-	-0.462	0.127	13.23	0.003
INTEL***	+	0.010	0.004	8.67	0.003
INMOS	-	0.038	0.115	0.11	0.736
FMFLANT	+	-0.135	0.113	1.44	0.250
FMFPAC***	+	-0.272	0.101	7.31	0.007
PFT_1ST**	-	-0.212	0.097	4.80	0.029
SLOWTRK	+	-0.009	0.136	0.01	0.945
FASTTRK**	-	-0.334	0.134	6.18	0.013
MOSEAS***	-	-0.021	0.003	39.68	0.001
JUNIOR**	-	-0.288	0.144	4.02	0.045
AAGE*	+	-0.042	0.026	2.63	0.100

* Significant at the 0.1 level.
** Significant at the 0.05 level.
*** Significant at the 0.01 level.
Source: Author

statistic has a chi-square distribution with the null hypothesis that the coefficient for each explanatory variable of the equation is zero. For this model, the -2 LOG Likelihood statistic equals 209.030 with twenty-two degrees of freedom. The probability that the null hypothesis is true for this equation is less than 0.0001. Therefore, one may be at least 99.99 percent confident that at least one of the parameters of this logistic equation is not equal to zero.

In Chapter IV, hypotheses for the sign of each variable are provided along with a potential explanation. Table 12 also contains the hypothesized signs. In all but seven of the cases, the signs are as expected. In five of these seven cases, the parameter estimate is not significantly different from zero at the 0.10 confidence level. Therefore, one can conclude the parameter estimate has no effect, positive or negative, on the probability that the service member separates under the VSI and SSB programs.

The other two variables, **AAGE** and **FMFPAC**, are significant and have opposite signs from those hypothesized. As a Marine becomes older, he may be more concerned with the benefits of retirement. Noticing that the military's retirement plan is one of the more enticing, older Marines may decide to remain in the Corps to receive the benefits associated with a military career. This may explain why the sign for the variable **AAGE** is negative. Evidently, a Marine's age as a proxy for the value placed upon the military's retirement plan outweighs the effect argued in Chapter IV that age is a good proxy for work experience.

The variable **FMFPAC** represents Marines assigned to units under the operational command of the Commander, Fleet Marine Forces Pacific and may capture other information causing the unexpected sign. For instance, most units assigned to the FMFPAC Commander are located in Hawaii and California, locations with a high cost of living. In these areas, service members may receive higher indirect pecuniary benefits from military "perks" such as base housing, military exchange and commissary facilities, and morale, welfare, and recreation facilities. There is potential for Marines on the West Coast to place more value on these "perks" than Marines who are located on the East Coast and elsewhere. A Marine may feel the military allows him to maintain a better lifestyle than he could maintain in the same area if he worked in the civilian sector. This would tend to lower the probability that he would separate from the service under the VSI and SSB programs and explains the unexpected sign.

C. AN INTERPRETATION OF THE MODEL PARAMETERS

Since the functional form of the model yields the log of the odds of a Marine separating under the VSI and SSB programs, it is not useful to directly interpret the variable parameters as it is with a linear probability model. One method to evaluate the partial effects of the parameters is to develop a "reference" case and then change the value of a single explanatory variable while holding all other variables constant. The resulting delta in the probability is the effect of a change in the explanatory variable on the probability that the observation is a program taker.

Here, the reference case has all the dummy variables set to zero, with one exception, and the continuous variables take on their mean values. The exception is the variable **NONTECH** which is set equal to one. Therefore, this reference case is a male Marine Sergeant (E-5) who is white, has a high school education, possesses a non-technical military occupational specialty (see Appendix B for a breakdown of the different MOS groupings), is not a drill instructor or recruiter, has a GCT score of 105, is not working in his MOS, is not assigned to a unit in the Fleet Marine Force, does not score a first class on latest physical fitness test, is promoted at the average of his peers, is 21 months from his End-of-Active-Service, has more than two years time-in-grade, is married and has children, and is 30 years old. A person meeting this description stands a probability of 0.043 of separating under the VSI and SSB programs.

The remainder of this section outlines the effects of changes in each explanatory variable for the reference or base case Marine. These changes would have different numerical values for a different reference Marine. In a situation where the variable has no significant effect, there is no significant difference from the reference case. Table 13 summarizes the partial effects of changes in the statistically significant variables.

1. Military Compensation

a. Rank

SSGT, the variable representing the rank of the individual, has an unexpected sign but it is not significantly different from zero at the 0.10 level. Thus, at the 0.10 level of significance, there is no difference in being an E-6 versus an E-5 on the probability of separating from the Marine Corps under the VSI and SSB programs.

b. Marital Status

Two of the three variables representing marital status, **SINGLE** and **SINWDEP** have the expected, negative hypothesized sign. The variable **MARRIED** has an unexpected sign but it is not significantly different from zero at the 0.10 significance level. Both **SINGLE** and **SINWDEP** are significant to the 0.05 level. If a person is single, he stands a 1.3 percent higher probability of separating from the Marine Corps under the VSI and SSB programs than a person who is married and has children. For a Marine who is single and has dependents, this change increases to 1.7 percent. There is no significant difference for taking the VSI and SSB programs between a married Marine who has children and a married Marine who does not have any children.

TABLE 13.

CHANGES IN PROBABILITIES WHEN ALTERING THE VALUES OF THE EXPLANATORY VARIABLES

Variable Name	Sign Hypothesis	Probability of Separation Under Programs	Change from Reference Case
Reference Case	NA	0.0430	NA
BLACK	-	0.0282	-0.0148
SINGLE	+	0.0562	0.0132
SINWDEP	+	0.0601	0.0171
ADMIN	-	0.0210	-0.0220
FIGHTER	-	0.0338	-0.0091
CSSTECH	-	0.0646	0.0216
BONUSMOS	-	0.0275	-0.0155
INTEL (Increment = ten points)	+	0.1280	0.0850
FMFPAC	+	0.0331	-0.0099
PFT_1ST	-	0.0351	-0.0079
FASTTRK	-	0.0312	-0.0118
MOSEAS (Increment = one month)	-	0.0278	-0.0152
JUNIOR	-	0.0326	-0.0104
AAGE (Increment = one year)	+	0.0120	-0.0310

Footnote: Those variables not listed are not statistically significantly different from zero at the 0.10 level and, therefore, have no significant impact on changes in the reference case.

Source: Author

2. Civilian Compensation

a. *Race*

The two variables representing ethnic origin, **BLACK** and **OTHMIN** possess the expected signs but only **BLACK** is significant at the 0.10 level. Marines of Afro-American decent are 1.5 percent less likely to separate under the VSI and SSB programs than their Caucasian counterparts.

b. *Education Level*

Both variables symbolizing education level have the expected positive signs. However, neither variable is significantly different from zero at the 0.10 level.

c. *Military Occupational Specialty*

The three variables depicting occupational specialty, **ADMIN**, **FIGHTER**, and **NONTECH** have the expected negative signs and are significant at the 0.05 level. Marines who possess an administrative MOS have an 2.2 percent lower probability of leaving the service than those in a non-technical MOS. For Marines in a combat or technical MOS, the change is a 0.9 percent decrease and a 2.2 percent increase, respectively.

Individuals possessing the additional MOSs of Recruiter or Drill Instructor face a lower probability of leaving the service under the VSI and SSB programs. This is illustrated by the negative sign on the variable **BONUSMOS**. This variable is significant to the 0.01 level. Marines with these MOSs stand a 1.6 percent lower probability of leaving the Corps than those who do not fill these billets.

d. Intelligence Level

The variable **INTEL** has the expected sign and is significant at the 0.05 level. Marines with higher GCT scores tend to have a higher probability of leaving the Marine Corps than those with lower scores. For a ten point increase in score, the probability increases by 8.5 percent.

e. Age

As a Marine's age increases, he tends not to participate in the VSI and SSB programs. The variable, **AAGE**, is significant at the 0.1 level. For a year older than thirty, a Marine stands a 3.1 percent lower probability of leaving the Marine Corps than a person who is thirty years old and in the reference case.

3. Non-Pecuniary Aspects Of Military Life

a. Job Satisfaction

Six variables captured different aspects of job satisfaction. First, the variable INMOS indicates whether a Marine works in his primary MOS. The sign of this variable is opposite from the hypothesized sign but the coefficient is not significantly different from zero. An individual working in his primary MOS does not stand a significantly different probability of leaving under the programs than a person not working in his MOS.

The two variables, **SLOWTRK** and **FASTTRK** capture promotion rates. **FASTTRK** has the expected sign but **SLOWTRK** does not. **SLOWTRK's** opposite sign causes little concern because the coefficient is not significantly different from zero. A Marine on the "fast track" stands a 1.2 percent lower probability of separating than a Marine who is promoted at an average rate.

The variable, **MOSEAS**, represents the amount of time until the Marine reaches his end of contract. The negative sign for the coefficient is the same as for the original hypothesis. The variable is significant at the .01 level. For a month further from his EAS than, a Marine is 1.5 percent less likely to leave under the VSI and SSB programs.

The variable **PFT_1ST** has the expected negative sign and is significant at the 0.05 level. A Marine who scores a first class on his physical fitness test stands a 0.8 percent lower probability of leaving the Marine Corps under the VSI and SSB programs than a Marine who does not score as well on the physical fitness test.

The last variable representing job satisfaction captures the amount of time since a Marine's last promotion. **JUNIOR** has the expected sign and is significant at the 0.05 level. A Marine with less than two years time-in-grade tends to have a 1.0 percent lower likelihood of separating under either of the programs than a reference case Marine with more than two years time-in-grade.

b. Job Type and Location

The two variables, **FMFPAC** and **FMFLANT**, represent job type and location. Each variable has a sign different from the hypothesis. One variable, **FMFLANT**, is not significantly different from zero but the other, **FMFPAC**, is significant at the 0.01 level. A Marine who is assigned to the operating forces of the Pacific stands a 1.0 percent lower probability of separating under the programs while a Marine in the Atlantic Forces does not act significantly differently from a Marine not in the operating forces.

D. PREDICTIVE CAPABILITIES OF THE CLASSICAL REGRESSION MODEL

To determine the predictive capability of the model, it is applied to a test data subset withheld from the data used to estimate the model. For each observation, the model estimates the ratio of the log of the odds of participating in the VSI and SSB programs. After converting this value to the probability of separating from the Marine Corps under either program, the usefulness of the model may be determined.

The estimated probabilities range from 0.0081 to 0.3388. To determine whether each test observation is a taker or a non-taker, a cut-off point of 0.0913 is used. This number is chosen because it is the percentage of program takers in the data set. If the model returns a probability higher than 0.0912, the person is labeled as a program taker. If the probability is lower, then the person is a non-taker. Using this criterion, the model correctly predicted 60.9 percent, 58 takers and 441 non-takers, of the 819 observations. Table 14 provides a breakdown of the outcomes when the model is applied to the test set. As currently calibrated, the model tends to over-predict the likelihood of separating under the VSI and SSB programs.

TABLE 14.

PREDICTIVE CAPABILITIES OF THE LOGISTIC MODEL ON THE TEST DATA SET

	QUANTITY	PERCENTAGE
CORRECT TAKERS	58	7.1
CORRECT NON-TAKERS	441	53.8
INCORRECT TAKERS	298	36.4
INCORRECT NON-TAKERS	22	2.7
TOTAL	819	100.0

Source: Author

VI. NEURAL NETWORK MODEL RESULTS

A. INTRODUCTION

This chapter provides the results of the ANN models. It begins with a brief discussion of the decision process involved in constructing the networks. Descriptive statistics for the additional variables used in the ANNs follow this discussion. The chapter concludes with an evaluation of the predictive capabilities of both models.

Prior to examining the design of each model is it useful to review the neural network learning process. The goal of the process is for the network to learn patterns in the training data set and then be able to generalize these patterns to data sets not previously seen. A network learns the training data by estimating outcomes and matching these outcomes to the actual outcomes of the training data set. In other words, during learning, the network attempts to correctly predict the outputs of the training data set. Differences in the actual and estimated outputs cause the network to continually adjust neuron connection weights until the error is minimized. It is important to note that it is possible for a network to over-learn the training data. Over-learning causes the network to lose the ability to accurately forecast outcomes for data never seen before.

B. ARTIFICIAL NEURAL NETWORK MODEL DESIGN

This research employs two neural network models. The major difference between the two centers on the network topology. Each network has thirty-seven input neurons and one output neuron. One network contains one hidden layer with 30 hidden neurons. This network is referred to as Network #1. The other network contains two hidden layers, the first having twenty hidden neurons and the second having ten hidden neurons. This is referred to as Network #2. This section describes the design of these two networks. Table 15 provides an overview of the important design factors of the two neural networks developed during this thesis.

1. Network Architecture

Network architecture refers to the learning algorithm applied to the network. Back-propagation is chosen as the

TABLE 15.

DESIGN FACTORS OF THE NEURAL NETWORKS

Design Factor	Network #1	Network #2
Network Architecture	Back-Propagation	Back-Propagation
Hidden Layers	One	Two
Hidden Neurons	30	30
Pruned Connections	389	448
Transfer Function	Sigmoid	Sigmoid
Learning Rule (Delta Rule Type)	Normalized Cumulative	Normalized Cumulative
Momentum	0	0

Source: Author

architecture for both neural networks. Current literature suggests this architecture provides the best results when attempting to forecast the outcomes of an event. While this is the only architecture investigated in this thesis, further research could be directed towards the usefulness of other architectures. Klimasauskas and Rathburn (1993) make a case for the use of the Extended Delta-Bar-Delta architecture.

2. Network Topology

Network topology encompasses two factors, the number of layers in the network and the number of neurons in each layer. As discussed earlier, Network #1 and Network #2 differ in network topology. Current literature suggests that most forecasting problems may be solved with a network design of one hidden layer.[Ref. 12] Lapedes [Ref. 13] furnishes a calculus-based argument that no more than two hidden layers are required to train any neural network employing a back-propagation learning algorithm. This thesis examines the impacts of changing the number of layers from one to two.

In determining the number of hidden neurons within each layer, two general rules of thumb are utilized. First, the total number of connections are restricted to be less than the number of observations in the training set. If the total number of connections exceeds the number of observations, the network tends to lose its ability to generalize to data sets never seen before.

In this research, the training data set has 1,400 observations. Network #1 has 1,171 connections. Network #2 has 981 connections. Table 16 contains a summary of the number of connections in each network.

The bias neuron is connected to each neuron in the hidden and output layers. Its primary function is to counteract random noise that appears in the network. The bias element is synonymous with the intercept coefficient in the classical regression model.

Table 16.

NUMBER OF CONNECTIONS WITHIN EACH NETWORK

Location of Connections		Number of Connections	
From Neurons In:	To Neurons In:	Network #1	Network #2
Input Layer	Hidden Layer #1	1,110	740
Hidden Layer #1	Hidden Layer #2	0	200
Hidden Layer #1	Output Layer	30	0
Hidden Layer #2	Output Layer	0	10
Bias Element	Hidden Layers	30	30
Bias Element	Output Layer	1	1
Total Connections:		1,171	981

Source: Author

The other rule of thumb concerns the placement of neurons when using more than one hidden layer. The ratio of neurons between connecting hidden layers is kept at a ratio of two to one. Therefore, in Network #2, there are twenty hidden neurons in the first hidden layer and ten in the second hidden layer. Network #1 has all the neurons in the first, and only, hidden layer. When there are many neurons within a hidden layer, it becomes harder for the network to generalize to data sets it has never seen before. To counter this problem, the technique of **pruning** is used. **Pruning** is a method of eliminating connections between neurons which have negligible connection weights.

NeuralWorks Professional II/Plus (hereafter referred to as NeuralWorks) contains a pruning technique which allows the user to determine the minimal weight required for a connection to remain active. This technique eliminates connections which have a weight with a magnitude that is less than some specified percentage of the average of all of the absolute weights. NeuralWorks allows the network designer to specify the percentage value. This research uses a value of forty percent for two reasons. First, it is large enough to prune many negligible weights. Second, it is not so large that it causes excessive pruning. The forty percent value resulted in the pruning of 389 connections in Network #1 and 448 in Network #2.

3. Transfer Functions

The ba -propagation algorithm within NeuralWorks supports three transfer functions. They are the summation, sigmoid, and hyperbolic tangent transfer functions. This research experiments with both the sigmoid and hyperbolic tangent transfer functions. The best results are achieved with the sigmoid transfer function. Using the sigmoid transfer function results in a normal distribution of the weights and a small number of weights possessing large absolute values; two criteria recommended by NeuralWorks to determine if the network is an optimal solution.

The hyperbolic tangent transfer function results in a "split" distribution with many weights having large absolute values. The summation transfer function is not investigated.

4. Learning Rules And Rates

NeuralWorks supports the Delta Rule, Cumulative Delta Rule, and the Normalized Cumulative Delta Rule when using the back-propagation algorithm. This research chooses the Normalized Cumulative Delta Rule based upon the recommendations of the software manufacturers. The advantage of this rule over the others is that it accelerates the training time required by the network.

The epoch size[1] is set as a balance between two recommendations from NeuralWorks. First, NeuralWorks recommends the epoch size should be set to a multiple of the training data set size. This ensures that connection weight changes are made upon complete passes through the data. The

[1] Epoch size is the number of observations presented to the network before the connection weights are adjusted to account for error difference between the estimated and actual outcomes.

training data set contains 1,400 observations. NeuralWorks also recommends the epoch size should be restricted to the range sixty-five to one hundred. This prevents the network from learning for long periods of time without updating the connection weights. Thus, the epoch size is set at seventy.

As discussed in Chapter III, the learning rate is a value determined by the neural network designer which is used in calculating the change in the connection weights. This change is determined by multiplying the error of a connection times the learning rate times the output from the connection. Error is measured as the difference between the estimated value (a value between zero and one) and the actual value (a value of zero or one). Within NeuralWorks, the learning rate is adjustable during the learning of the network. Table 17 contains the values of the learning rates and the intervals during which they became effective.

Momentum is a factor added to the weight changes which allows the network to avoid becoming trapped in local error minimums. The momentum term for this thesis is set to zero. NeuralWorks possess the capability to vary the learning rates and momentum values between layers. For simplicity, the learning rates are held constant across each layer.

5. Stopping Criteria

Once a network is built and begins to learn, stopping criteria are needed to determine the point to stop learning.

TABLE 17.

LEARNING RATE SCHEDULE FOR THE ARTIFICIAL NEURAL NETWORKS

Number of passes through data set	Learning Rate Value
400-800	2.0
801-1,200	1.0
1,201-1,600	0.5
1,601-2,000	0.25
2,001-2,500	0.01

Source: Author.

This point may be the number of passes through the data set, the elapsed computer time, or functions of comparisons of the actual versus estimated outcomes such as root mean square (RMS) error or the correlation value between the predicted and actual outcome. Researchers may choose any combination which is appropriate for their research. It is important that stopping criteria be develcped prior to model construction. This allows researchers to establish ground rules for the research before it begins instead of developing new rules as the research progresses.

In this research, the criterion used is the number of passes through the data set. The network is presented with the training data set 2,500 times. This number is chosen because the network views the training data set a sufficient number of times for the network to learn patterns which exists in the training data set.

When the network passes through the training data set 2,500 times, the RMS error ranges from 0.20 to 0.35 and averages roughly 0.275. The correlation between the predicted and actual outcomes is 0.8 and above. The probability distribution for the training data set is 0.5. One half of the observations are takers and the other half are non-takers. If a random assessment model is used to predict the outcomes of the training data set, the correlation between actual and predicted would be 0.5. Thus, the neural networks provide a substantial improvement over the predictive capabilities of a random assessment model.[2]

When building ANN models, researchers must decide what criteria to use in determining the learning stop point. The longer a network runs, the greater the possibility it may not be able to generalize to the test data. When this occurs, the ANN model has essentially memorized the training data set.

This research experiments with both a smaller and larger number of passes through the data set. In both cases, smaller and larger numbers of presentations of the data set to the network lead to higher RMS error and lower correlation between the predicted and the actual outcomes.

---

[2] The processing time required to complete 2,500 passes through the data with a personal computer that possesses an INTEL 486 chip working at 25 megahertz is approximately eighteen hours.

C. DESCRIPTIVE STATISTICS OF AUGMENTED VARIABLES

Table 18 contains the descriptive statistics for the seven variables added to the ANN models which are not contained in the classical regression model. The mean and standard deviation of each variable is provided. Marines in the data set have an average of 11 years of service, have been at their present jobs on the average about a year and a half, and average about two months of deployed training time. On the average, it has been about three and a half years since they were assigned to an overseas duty station without their families. The average score on the ASVAB test for these Marines is in the forty-fifth percentile.

D. PREDICTIVE CAPABILITIES OF THE ANN MODELS ON THE TEST DATA

To determine the predictive capability of the models, they are applied to a test data subset withheld from the data used to estimate the model. For each observation, the model estimates a continuous outcome value between zero and one. If this value is over 0.5 then the observation is labeled a taker of the VSI and SSB programs. If the value is less than 0.5, the observation is labeled a non-taker. This value is picked as the cutoff point because it represent the probability distribution of program takers and non-takers within the training data set.

Using this criterion, Network #1 correctly predicted 58.1 percent, 49 takers and 427 non-takers, of the 819 decisions.

TABLE 18.

SIMPLE DESCRIPTIVE STATISTICS FOR THE EXPLANATORY VARIABLES ADDED TO THE ANN MODELS

Conditions represented by Variables	Mean	Standard Deviation
Years of Service	11.24	1.80
Elapsed Time At Present Job (In Years)	1.47	0.83
Accumulated Deployed Time (In months)	1.99	3.15
Years since Last Deployed Overseas Without Family	3.66	4.09
Percentile Ranking on ASVAB	45.18	25.92
Raw Score on Arithmetic Reasoning (ASVAB Test)	17.61	6.47
Raw Score on Paragraph Comprehension (ASVAB Test)	7.49	5.80
Raw Score on Word Knowledge (ASVAB Test)	24.97	6.90
Raw Score on Mathematical Knowledge (ASVAB Test)	13.09	5.11

Source: Author

T ͻle 19 provides a breakdown of the outcomes when the Network #1 is applied to the test set. As currently calibrated, the network tends to over-predict the likelihood of separating under the VSI and SSB programs. Table 20 displays the predictive abilities of Network #2. This network correctly predicted 66.8 percent, 47 takers and 500 non-takers, of the 819 observations.

Although it appears the prediction rates are better for Network #2, there is no way to determine if this difference is statistically different from zero. One of the drawbacks to using neural networks is the lack of statistical measures of model fit and Wald Chi-square values for parameter estimates in determining significance levels when assessing individual models and making model comparisons.

TABLE 19.

PREDICTIVE CAPABILITIES OF NEURAL NETWORK #1 IN PREDICTING TAKERS OF THE VSI AND SSB PROGRAMS ON THE TEST DATA SET

	QUANTITY	PERCENTAGE
CORRECT TAKERS	49	6.0
CORRECT NON-TAKERS	427	52.1
INCORRECT TAKERS	312	38.1
INCORRECT NON-TAKERS	31	3.8
TOTAL TEST DATA	819	100.0

Source: Author

TABLE 20.

PREDICTIVE CAPABILITIES OF NEURAL NETWORK #2
IN PREDICTING TAKERS OF THE VSI AND SSB PROGRAMS
ON THE TEST DATA SET

	QUANTITY	PERCENTAGE
CORRECT TAKERS	47	5.7
CORRECT NON-TAKERS	500	61.1
INCORRECT TAKERS	239	29.2
INCORRECT NON-TAKERS	33	4.0
TOTAL TEST DATA	819	100.0

Source: Author

VII. MODEL COMPARISONS

A. INTRODUCTION

This chapter compares the effectiveness of the three models estimated for this research. Two model evaluation criteria are examined. First, the predictive capabilities of each model are assessed using two previously defined metrics. Second, the partial effects of each explanatory variable within the three models are examined.

B. Model Forecasting Abilities

Table 21 contains the metrics employed to compare the models. The metrics are generated by applying the models to the test data set (n = 819). There are two comparison metrics. The first, M_1, was developed by Sharda and Wilson (1993) and is reviewed in Chapter II. M_1 has two components, M_{1t} and M_{1n}. The first, M_{1t}, represents the probability that the model correctly predicts a program taker. The second, M_{1n}, represents the probability that the model correctly predicts a program non-taker. The other metric, a simulated R^2, was developed by Wiggins et al. and is also described in Chapter II. The goal is to maximize each of these values.

Applying the metrics, the logistic regression model is a better forecaster than the neural network model with one hidden layer (Network #1). The probability it correctly

TABLE 21.

COMPARISON OF THE FORECASTING ABILITIES OF THE NEURAL NETWORK MODELS AND THE LOGISTIC REGRESSION MODEL ON THE TEST DATA SET (N = 819)

	Model Types		
	Logistic Regression	Neural Network #1	Neural Network #2
Number Correctly Predicted	499 (60.9)	478 (58.1)	547 (66.8)
Correctly Predicted Program Takers	58	49	47
Correctly Predicted Program Non-Takers	441	427	500
Incorrectly Predicted Program Takers	298	312	239
Incorrectly Predicted Program Non-Takers	22	31	33
M_{1t} (Prob of correct taker prediction)	0.16	0.14	0.16
M_{1n} (Prob of correct non-taker prediction)	0.95	0.93	0.94
Simulated R^2	0.0981	0.0721	0.0979

Source: Author
Footnote: The number in parenthesis is the percentage value.

predicts a program taker, M_{1t}, equals 0.16, and is higher than Network #1's probability of correctly predicting a program taker (0.14). The logistic model is also better at correctly predicting program non-takers; 0.95 versus 0.93. However, the neural network with two hidden layers (Network #2) is as good as the logistic model. Both models have the same probability of predicting a taker and the logistic model is slightly better at predicting non-takers.

The simulated R^2 is described in detail in Chapter II. If a model produces outputs equal to the actual outputs in every case, then the simulated R^2 is equal to one. The simulated R^2 may generate a negative number if the model produces a predicted take-rate lower than the take-rate obtained by simply using just the mean take-rate of the sample.[Ref. 11:p.15] The logistic model's simulated R^2 (0.0981) is small. Logistic regressions normally provide smaller R^2 values than other regression models. [Ref. 7:p. 55] Network #1 produces a lower simulated R^2 (0.0721) than the logistic model. However, Network #2 results in a simulated R^2 (0.0979) which is essentially the same as that which is provided by the logistic model.

C. Partial Effects of the Explanatory Variables

In this research, neural networks provide forecasting results which are almost as good as those obtained from logistic regression models. However, military manpower and personnel analysts also concern themselves with policy implications suggested by models. That is, they are interested in the partial effects of the variables in the model as well as the model's prediction capabilities.

Interpreting partial effects for logistic models is fairly straightforward. However, with neural networks the task is more difficult for two reasons. First, neural networks lack the statistical tests required to determine the significance levels associated with both the explanatory variables of the neural network and the overall fit of the network to the data. Therefore, researchers are unable to determine if the connection weights associated with an explanatory variable are significantly different from zero. Without these statistical tests, interpretation of the significance of the explanatory variables is difficult, at best.

The other drawback to interpreting the partial effects of the explanatory variables for neural networks evolves from the lack of a probability level associated with the network output. In this research, the neural networks are trained on data which possess output values of zero or one. A neural network produces probabilities associated with a specific set of input variables only when the network is trained on a data set which contains output values that are probabilities associated with a specific set of input variables.

Obtaining a data set with this characteristic is extremely difficult. Most manpower and personnel data sets contain information on the decision a person has made. Does the person reenlist or separate? Does he accept VSI or SSB? It does not contain the probability that the individual chooses one path or another.

Therefore, neural networks do not lend themselves to interpreting the partial effects of explanatory variables because they lack statistical tests to determine variable significance and the output of a neural network is not a probability level.

These two drawbacks hinder the ability of researchers employing neural network modeling techniques in analyzing the impacts of policy changes. In an attempt to overcome these deficiencies, this research compares the impact on the classification of an observation when changing one explanatory variable for a reference case while holding the other conditions constant.

As in Chapter V, the reference case is defined as a male Marine Sergeant (E-5) who is white, has a high school education, possesses a non-technical military occupational specialty, is not a drill instructor or recruiter, has a GCT of 105, is not working in his primary MOS, is not assigned to a unit in the Fleet Marine Force, does not score first class on his most recent physical fitness test, is promoted at the average of his peers, is twenty-one months from his end-of-active-service date, has more than two years time-in-grade, is married and has children, and is thirty years old.

In the logistic model, if the model output is greater than 0.0913 (the proportion of program takers in the test data set), the observation is categorized as a program taker. In the neural network models, if the output is greater than 0.5

(the midpoint of all possible outcomes), the observation is categorized as a program taker.

Table 22 contains classifications of the reference case as a taker or non-taker of the VSI and SSB programs and the resulting classification when each explanatory variable is changed while holding the other variables constant. The results are disturbing for two reasons. First, there is a lack of consistency between the classification of the reference case. If neural networks provided the same results as classical regression models, the base case should receive the same classification regardless of the model which is used. Here, Neural Network #1 (one hidden layer) classifies the reference case as a program taker while Neural Network #2 (two hidden layers) classifies the reference case the same as the logistic model does, as a program non-taker.

The other undesired result is a lack of consistency in the impact on the observation's classification of changing the explanatory variables one at a time and holding the others constant.[3] When changing the twenty-two explanatory variables in the logistic model one at a time, there is no switch in the classification of the reference case observation from a

---

[3] If more than one change to the reference case is desired simultaneously, a new reference case must be developed and incremental changes to it are required. Interpreting partial effects of explanatory variables requires changing one explanatory variable at a time while holding all others constant. The method applied in this research to examine partial effects of explanatory variables does not lend itself to an analysis of multiple changes in the explanatory variables.

TABLE 22.

CLASSIFICATION OF OUTCOMES FOR A REFERENCE CASE AND CHANGES IN THE EXPLANATORY VARIABLES FOR EACH OF THE THREE MODELS

Changed Condition	Classification		
	Logistic	Net #1	Net #2
Reference Case	Non-Taker	Taker	Non-Taker
Rank of E-6	Non-Taker	Non-Taker	Non-Taker
Afro-American Decent	Non-taker	Non-Taker	Non-Taker
Other Minority	Non-Taker	Non-Taker	Non-Taker
2 Yrs and less College	Non-Taker	Taker	Non-Taker
3 Yrs and More College	Non-Taker	Non-Taker	Non-Taker
Single W/Dependents	Non-Taker	Taker	Taker
Married No Dependents	Non-Taker	Non-Taker	Non-Taker
Single, No Dependents	Non-Taker	Taker	Taker
In an Admin Job	Non-Taker	Taker	Non-Taker
In a Combat Job	Non-Taker	Taker	Non-Taker
In a Technical Job	Non-Taker	Taker	Taker
Recruiter/Drill Instructor	Non-Taker	Non-Taker	Non-Taker
GCT Score	Taker	Non-Taker	Non-Taker
Working in Primary MOS	Non-Taker	Taker	Non-Taker
Assigned to FMFLANT Unit	Non-Taker	Non-Taker	Non-Taker
Assigned to FMFPAC Unit	Non-Taker	Non-Taker	Non-Taker
Scores 1st Class on PFT	Non-Taker	Taker	Taker
Months Until EAS	Non-Taker	Taker	Non-Taker
Less Than Two Years TIG	Non-Taker	Non-Taker	Non-Taker
Age	Non-Taker	Taker	Non-Taker
Slow Promotion Track	Non-Taker	Taker	Taker
Fast Promotion Track	Non-Taker	Non-Taker	Non-Taker

Source: Author

program taker to a non-taker, with one exception. If the observation's GCT score is increased by ten points, the observation becomes classified as a program taker. All other changes to the reference case result in observations which are classified as program non-takers.

Neural Network #1 switches the classification of the reference observation from a program taker to a program non-taker in eleven instances. Neural Network #2 switches the observation's classification from a program non-taker to a program taker in five instances. Even more confusing, none of the switches in classifications occur when changing the same explanatory variable.

In order to use artificial neural networks to obtain guidance for policy variables, it is important to have some degree of consistency of results among models. This is not the case with the networks developed in this research when the partial effects of the explanatory variables are explored.

D. CONCLUSIONS

To determine the usefulness of a modeling technique to military manpower and personnel analysts, two aspects of the model must be examined; its forecasting ability and its capacity to indicate the partial effects of the explanatory variables. When comparing the logistic and neural network models which are developed in this research, two aspects of neural networks are obvious. First, the forecasting ability of the network with two hidden layers is equal to that of the logistic model. Second, neural networks are not effective tools for analyzing the partial effects of explanatory variables. The networks lack consistency when compared to each other and compared to the results provided by the logistic model. Since military manpower and personnel analysts concern themselves with the impacts of policy, extreme caution should be used when solely employing artificial neural networks in analyzing the partial effects of explanatory variables.

VIII. CONCLUSIONS AND RECOMMENDATIONS

A. INTRODUCTION

This thesis compares the forecasting abilities of two neural network models with the forecasting abilities of a logistic regression model. The models attempt to forecast the take-rates for the VSI and SSB programs for male, Marine Corps E-5s and E-6s in fiscal year 1993. The research builds three models, one classical regression model and two neural network models. The logistic regression model contains twenty-two variables and is based upon the theory of occupational choice. The neural networks are constructed using all potential explanatory variables available from the data set (37 variables). Three sub-samples are created to build and test each model; a training data set consisting of 7,223 observations for the logistic regression model, a training data set of 1,400 observations for the neural networks, and a test data set of 816 observations for all three models.

The neural network models employ a back-propagation architecture with a normalized cumulative delta learning rule, a sigmoid transfer function, and are trained for 2,500 passes through the data set. The overriding difference between the two models is the number of hidden layers. One network has one hidden layer while the other has two. Both networks have thirty hidden neurons in the hidden layers. The logistic regression model is based on previous research of the VSI and SSB programs conducted by Noblit (1993).

This chapter presents several conclusions from the comparisons of the three networks. It also provides policy implications and concludes with recommendations for future research.

B. CONCLUSIONS

Three conclusions may be drawn from this research. First, the neural networks designed with two hidden layers possess similar forecasting abilities to the logistic regression model built in this research. Even though the logistic model is a slightly better forecaster of program non-takers, artificial neural networks provide forecasting results as good as those obtained from classical regression models.

The second conclusion concerns which topology provides better results. In this research, a neural network designed with two hidden layers (Network #2) supplies slightly better forecasting results than a neural network with one hidden layer (Network #1). However, Network #2 generated slightly more false negatives[4] than Network #1. This is a cause for some concern since false negatives carry a relatively high cost in the execution of the VSI and SSB programs. In this

[4] A false negative occurs when the model predicts a program non-taker and in reality, the observation is a program taker.

research, increasing the number of hidden layers increases the forecasting capabilities of the neural networks.

The third conclusion concerns the usefulness of neural networks in evaluating policy decisions. Military manpower and personnel analysts concern themselves with policy implications suggested by models. That is, they are interested in the partial effects of the variables in the model as well as the model's prediction capabilities.

Interpreting partial effects for neural networks is a difficult task because 1) neural networks lack the statistical tests required to determine the significance levels of explanatory variables and the overall model fit to the data and 2) neural networks lack the ability to produce a probability level associated with a set of explanatory variables. In addition to these difficulties, there is a lack of consistent results when evaluating the partial effects of explanatory variables across neural network models constructed on identical data sets. These inconsistencies and difficulties do not allow researchers to interpret easily the partial effects of the explanatory variables contained in neural network models. Therefore, neural networks should be used in conjunction with modeling techniques that support analyzing the partial effects of explanatory variables.

C. POLICY IMPLICATIONS

Models developed in this research have several implications about the VSI and SSB programs. First, if a Marine E-5 or E-6 has a technical MOS, he stands a higher probability of taking the program than a Marine in another MOS. This indicates that Marines with specialized training are leaving the Marine Corps at a greater rate than their non skilled counterparts. If the Marine Corps is attempting to reduce the number of its technicians, then this is not problem. However, decision makers must realize that the training cost associated with these individuals is high. It is important for the Marine Corps to receive a return on its investment in these individuals. Further evaluation of the VSI and SSB programs is required to determine if these programs are causing the Marine Corps to receive a rate of return below acceptable levels on the technical training of Marines in highly specialized MOSs.

Three variables in the logistic model **PFT_1ST, BONUSMOS**, and **FASTTRK**, capture characteristics of the type of individual the Marine Corps desires to retain. This research reveals that male E-5s and E-6s with these characteristics are not separating from the Marine Corps under the VSI and SSB programs. Those male E-5s and E-6s who score first class on their PFT (represented by **PFT_1ST**) stand a 2.8 percent lower probability of separating from the Marine Corps under the VSI and SSB programs than their non-first class PFT counterparts.

Drill Instructors and Recruiters (represented by **BONUSMOS**) stand a 5.5 percent lower probability of separating than their counterparts who are not Drill Instructors and Recruiters. Those Marine E-5s and E-6s promoted faster than their peers (represented by **FASTTRK**) stand a 4.2 percent lower probability of separating than their counterparts who are promoted at an average rate.

Artificial neural networks have potential to provide excellent forecasting results. However, before they may be used independently of other modeling techniques in the research of manpower and personnel related problems, the algorithms within the neural network "black box" must be advanced to a level which provides more information on the importance of individual explanatory variables. This impacts on military policy in two ways. First, policy makers must ensure sound statistical methods are employed when analyzing military manpower and personnel related data. Second, further research into neural networks is required to examine the appropriate domain of application for this emerging technology.

D. RECOMMENDATIONS FOR FURTHER STUDY

One area for further research centers on the different architectures available within the context of NeuralWorks. This thesis examined only the back-propagation learning algorithm. NeuralWorks supports fifteen other learning algorithms. The application of these algorithms to manpower related data sets is rare. Two algorithms which have produced useful results in other types of applications are the Delta-Bar-Delta and Extended-Delta-Bar-Delta algorithms. An analysis of the impacts of these architectures on the forecasting capabilities of a neural network built with a manpower related data set would be valuable.

A second area of research is related to the choice of the transfer function. This research experiments with both the hyperbolic tangent transfer function and the sigmoid transfer function. The transfer function is held constant across all layers. NeuralWorks supports the ability for changing the transfer function from one layer to the next. Researchers could experiment with adjusting the transfer functions between layers and evaluate any change in network performance. One possibility is to use a threshold function instead of a saturation function as applied in this thesis. A threshold function forces a neuron to excite and output a value of one if it receives inputs above some level. Otherwise, it remains dormant and outputs a zero. Further research into the implications of employing a threshold function within the design of neural networks may provide some useful insights into further understanding this developing field.

A third area of potential research concerns the modeling techniques applied to the VSI and SSB take-rate problem. Marines who choose not to accept these programs the first time they become eligible are also eligible in the following year. Sequential modeling techniques are the best suited to model this type of problem. Single period logistic modeling techniques provide excellent exploration results but are inadequate at controlling many of the features needed in situations where individuals are faced with a recurring decision.

This research builds a model for male, Marine Enlisted between the ranks of E-5 and E-6. Models constructed for other subgroups (i.e., females, Marine Officers, etc.) may reveal very different decision patterns. This would assist policy analysts in examining the impacts of policy changes for other Marines eligible for the VSI and SSB programs.

This research did not distinguish between those who separate under the VSI program and those who separate under the SSB program. Each program differs in the way that the cash incentive is paid out. Research on the differences between program selection and the characteristics of individuals choosing the different programs would provide policy makers with information to determine which program is best suited to meet desired goals. If one program is more productive than the other, future offerings could be constructed to reduce financial strain on available resources and eliminate program redundancies.

E. CONCLUDING COMMENTS

This thesis examined the ability of neural networks to assist manpower and personnel policy analysts in the execution of their duties. Neural networks provide forecasting results which are comparable to those provided by classical regression models. However, neural networks should not be used independently in the analysis of manpower and personnel related problems. Neural networks present too many inconsistencies and lack the statistical tests required to examine the partial effects of the explanatory variables. However, those unfamiliar with the theory underlying a particular problem can present all data elements to the network and it will sort out variables and variable combinations which are important in determining the actual outcomes.

Neural networks can add to the policy analyst's tool box. Combining neural networks with conventional and traditional statistical modeling methods enhances the capabilities of policy analysts and therefore, increases the knowledge base of decision makers.

APPENDIX A

VARIABLES USED IN BYRNE'S AND NOBLIT'S RESEARCH

VARIABLE NAME	VARIABLE DESCRIPTION	BYRNE	NOBLIT
	PROXIES FOR MILITARY COMPENSATION		
E4	= 1 IF AN E4 = 0 IF NOT AN E4	NO	YES
E5	= 1 IF AN E5 = 0 IF NOT AN E5	NO	YES
E6 (Noblit) SSGT (Byrne)	= 1 IF AN E6 = 0 IF NOT AN E6	YES	NO
E7	= 1 IF AN E7 = 0 IF AN E7	NO	YES
PROPAY	= 1 IF RECEIVING PRO PAY = 0 OTHERWISE	NO	YES
BONUS	= 1 IF RECEIVING AN SRB = 0 OTHERWISE	NO	YES
SINGLE	= 1 IF SINGLE/NO DEPEND = 0 OTHERWISE	YES	YES
MARRIED	= 1 IF MARRIED/NO CHILD = 0 OTHERWISE	YES	YES
S_DEP (Noblit) SINWDEP (Byrne)	= 1 IF SINGLE/NO DEPEND = 0 OTHERWISE	YES	YES
	PROXIES FOR CIVILIAN COMPENSATION		
BLACK	= 1 IF AFRO-AMERICAN = 0 OTHERWISE	YES	YES
OTH_MIN(Noblit) OTHMIN (Byrne)	=1 IF NOT WHITE OR BLACK = 0 OTHERWISE	YES	YES
ADMINSU(Noblit) ADMIN (Byrne)	= 1 IF IN ADMIN JOB = 0 OTHERWISE	YES	YES
CMBT (Noblit) FIGHTER (Byrne)	= 1 IF IN COMBAT JOB = 0 OTHERWISE	YES	YES
CSS_NT (Noblit) NOT_TECH(Byrne)	= 1 IF IN NON-TECH JOB = 0 OTHERWISE	YES	YES

VARIABLE NAME	VARIABLE DESCRIPTION	BYRNE	NOBLIT
PROXIES FOR CIVILIAN COMPENSATION			
FEMALE	= 1 IF FEMALE = 0 OTHERWISE	NO	YES
INTEL	= OBSERVATION'S GCT SCORE	YES	NO
AAGE	= OBSERVATION'S AGE	YES	NO
BONUSMOS	= 1 IF RECRUITER/DRILL INST = 0 OTHERWISE	YES	NO
COLLEGE	= 1 IF FINISHED ANY COLLEGE = 0 OTHERWISE	NO	YES
SOMCOLL	= 1 IF COMPLETED 1 OR 2 YRS = 0 OTHERWISE	YES	NO
YR3COLL	= 1 IF FINISHED 3 OR MORE = 0 OTHERWISE	YES	NO
NON-PECUNIARY FACTORS INVOLVED IN STAY OR LEAVE DECISION			
TIG	TIME IN GRADE (IN YEARS)	YES	YES
PRO_IND	A CONTINUOUS VARIABLE REPRESENTING PROMOTION RATE OF OBS COMPARED TO PEERS	NO	YES
SLOWTRK	= 1 IF PROMOTED SLOWER = 0 OTHERWISE	YES	NO
FASTTRK	= 1 IF PROMOTED FASTER = 0 OTHERWISE	YES	NO
INMOS	= 1 IF IN PRIMARY MOS = 0 OTHERWISE	YES	YES
NON_FMF	= 1 IF NOT IN THE FMF = 0 OTHERWISE	NO	YES
FMFLANT	= 1 IF ASSIGN TO FMFLANT = 0 OTHERWISE	YES	YES
FMFPAC	= 1 IF ASSIGN TO FMFPAC = 0 OTHERWISE	YES	NO
YRS_EAS	NUMBER OF YEARS TO EAS	NO	YES
PFT_1ST	= 1 IF 1ST CLASS ON PFT = 0 OTHERWISE	YES	NO
MOS_EAS	THE NUMBER OF MONTHS TO EAS	YES	NO

APPENDIX B

MILITARY OCCUPATIONAL SPECIALTY CATEGORIES

Category (Variable Name)	Military Occupational Specialty	
	Numerical	Description
Combat (FIGHTER)	03XX	Infantry
	08XX	Field Artillery
	18XX	Tank and Amphibious Assault
Administrative (ADMIN)	01XX	Personnel and Administration
	30XX	Supply Admin and Operations
	31XX	Traffic Management
	34XX	Auditing, Finance, Accounting
	40XX	Data Systems
	41XX	Marine Corps Exchange
	43XX	Public Affairs
	44XX	Legal Affairs
	46XX	Train and Visual Equipment
	55XX/98XX	Music
	66XX	Aviation Supply
	84XX	Recruiting
Tech Support (CSS_TECH)	11XX	Utilities
	13XX	Engineer, Construction Equip
	21XX	Ordnance
	28XX	Data/Communications Maint.
	59XX	Electronics Maintenance
	60XX/61XX	Aircraft Maintenance
	63XX/64XX	Avionics

Source: Navy Marine Corps Form 1008-A (April, 1992)

Category (Variable Name)	Military Occupational Specialty	
	Numerical	Description
Service Support Non-Technical (NON_TECH)	02XX	Intelligence
	04XX	Logistics
	15XX	Printing and Reproduction
	23XX	Ammunition and Explosive Ordnance Disposal
	25XX	Operational Communications
	26XX	Signals Intelligence/Ground Electronic Warfare
	33XX	Food Service
	35XX	Motor Transport
	57XX	Nuclear, Biological, and Chemical
	58XX	Military Police and Corrections
	65XX	Aviation Ordnance
	70XX	Airfield Services
	72XX	Air Control/Air Support/Anti-Air Warfare
	73XX	Air Traffic Controller Enlisted Flight Crews
	68XX	Weather Service

Source: Navy Marine Corps Form 1008-A (April, 1992)

APPENDIX C

ASSORTED TABLES CONTAINING FREQUENCIES FROM DATA SET

PROMOTION RATE BY RACE/ETHNIC GROUP

Promotion Rate	Caucasian	Afro-American	Hispanic	Asian	Other
Slow Rate	689 (14.52)	403 (15.93)	80 (14.57)	20 (20.48)	34 (28.18)
Average Rate	3,341 (70.43)	1,849 (73.15)	401 (72.65)	58 (60.24)	72 (59.09)
Fast Rate	714 (15.05)	276 (10.92)	70 (12.77)	19 (19.28)	16 (12.73)
Totals	4,744 (100.0)	2,528 (100.0)	551 (100.0)	97 (100.0)	122 (100.0)

MILITARY OCCUPATIONAL SPECIALTY CATEGORIES
BY RACE/ETHNIC GROUP

Military Occupational Special	Caucasian	Afro-American	Hispanic	Asian	Other
Admin	888 (18.72)	709 (28.04)	133 (24.15)	27 (27.71)	35 (29.09)
Fighter	690 (14.54)	397 (15.71)	96 (17.37)	14 (14.46)	19 (59.09)
Non-Tech	1,407 (29.67)	911 (36.05)	176 (31.94)	27 (27.71)	41 (12.73)
Tech-nical	1,759 (37.07)	511 (20.20)	146 (26.54)	29 (30.12)	27 (21.82)
Totals	4,744 (100.0)	2,528 (100.0)	551 (100.0)	97 (100.0)	122 (100.0)

LIST OF REFERENCES

1. Weinstein, John N. et al., "Neural Computing in Cancer Drug Development," Science, 16 October, 1992, pp. 447-451.

2. Yoon, Y. and G. Swales, "Predicting Stock Market Price Performance," Proceedings of the 24th Hawaii International Conference on System Sciences, Vol. 4, 156-162, 1991.

3. Klimasauskas, Casimir C., "Applying Neural Networks; Part II: A Walk Through The Application Process," PC AI, March/April 1991, 27-34.

4. Marsan, Carol Duffy, "Neural Networks Pays Off," Federal Computer Week, 15 February 1993, p. 16.

5. Klimasauskas, Casimir C., and Rathburn, Thomas A., Basics of Back-Propagation, (Pittsburgh, Pennsylvania: NeuralWare Inc, 1993).

6. Stanley, J., Introduction to Neural Networks, ed Luedeking, S., 2nd ed., (Sierra Madre, California: California Scientific Software, 1989).

7. Russell, Bradly R., "Neural Networks," Master's Thesis, Naval Postgraduate School, 1993.

8. ______. Neural Computing, (Pennsylvania: NeuralWare, Inc.), 1991.

9. Wiggins, Vince L., Looper, Larry T., and Engquist, Sheree K., "Neural Networks: A Primer," (AL-TP-1991-0011, AD-A235 920), Brooks AFB, Texas: Human Resources Directorate, Manpower and Personnel Division, Armstrong Laboratory.

10. Raghupathi, W., Schkade, Lawrence L., and Raju, Babi S., "A Neural Network Application for Bankruptcy Prediction," in Proceeding of the 24th Hawaii International Conference on System Sciences, Vol. 4 (1991), 147-155.

11. Wiggins, Vince L., Looper, Larry T., and Engquist, Sheree K., (1992), "Applying Neural Networks To Air Force Personnel Analysis," (AL-TR-1991-0118, AD-A248 762), Brooks AFB, Texas: Human Resources Directorate, Manpower and Personnel Division, Armstrong Laboratory.

12. Klimasauskas, Casimir C., "Applying Neural Networks; Part III: Training A Neural Network," PC AI, May/June 1991, 20-24.

13. Lapedes, A., and Farber, R., "How Neural Networks Work," Neural Information Processing Systems, American Institute of Physics, New York, 1998, pp. 442-456.

14. Caudill, M. (1990), "Neural Network Primer," Reprinted from AI Expert, San Francisco, California: Miller Freedman Publications.

15. Wassermann, Philip D., Neural Computing: Theory and Practice, (New York: Van Nostrand Reinhold, 1989)

16. Beale, R. and Jackson, T., Neural Computing, An Introduction, 2nd ed., (New York: Adam Hilger, 1991).

17. Hill, T., O'Connor, M., and Remus, W., "Neural Network Models for Time Series Forecasts," A working paper presented at the 24th Hawaii International Conference on System Sciences, 1991.

18. Sharda, Ramesh and Wilson, Rick L., "Performance Comparison Issues in Neural Network Experiments for Classification Problems," Proceedings of the IEEE Conference on Neural Networks, 1993, p. 649-656.

19. Commandant Of the Marine Corps, Washington D. C., UNCLASSIFIED ALMAR 007/92, Subject: Fiscal Year 92 Voluntary Incentive (VSI) and Special Separations Benefit (SSB) Implementing Guidance, 10 January, 1992.

20. Commandant Of the Marine Corps, Washington D. C., UNCLASSIFIED ALMAR 055/93, Subject: Phase III of Fiscal Year 93 Voluntary Incentive (VSI) and Special Separations Benefit (SSB), 13 February, 1993.

21. Noblit, Mark L., "Forecasting VSI/SSB Take-Rates For Enlisted Marine Corps Personnel," Master's Thesis, Naval Postgraduate School, 1993.

22. Headquarters United States Marine Corps, UNCLASSIFIED Memorandum 1000 MPP-25, Subject: Marine Corps Targeting for the FY92 VSI/SSB Program, 15 July 1992.

23. Commandant Of the Marine Corps, Washington D. C., UNCLASSIFIED ALMAR 233/92, Subject: Phase I of Fiscal Year 93 Voluntary Incentive (VSI) and Special Separations Benefit (SSB), 21 August, 1992.

24. Ehberg, Ronald G. and Smith, Robert S., Modern Labor Economics, Fourth ed., (New York: Harper Collins, 1990).

25. Tinney, Robert, The Effects of Selective Re-Enlistment Bonuses, Part I; Background and Theoretical Issues, Research Note 91-83, United States Army Research Institute for the Behavioral Sciences, Alexandria, Virginia, June 1991.

26. Studenmund, A. H., Using Econometrics: A Practical Guide, 2nd ed. (New York: Harper Collins Publishers Inc., 1992).

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Technical Information Center Cameron Station Alexandria, VA 22304-6145	2
2.	Library, Code 052 Naval Postgraduate School Monterey, CA 93943-5002	2
3.	Director, Training and Education MCCDC, Code C46 1019 Elliot Road Quantico, Virginia 22134-5027	2
4.	George Thomas, Code AS/Te Department of Administrative Sciences Naval Postgraduate School Monterey, Ca 93943-5001	2
5.	Tim Hill E601J College Of Business Administration University of Hawaii 2404 Maile Way Honolulu, HI 96822	1
6.	Tony Rathburn Penn Center West Building IV, Suite 227 Pittsburgh, PA 15276-8220	1
7.	Captain Brian J Byrne 1106 Milky Way Colorado Springs, CO 80906	1