INFORMATION TO USERS

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

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

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

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
FINANCIAL APPLICATIONS OF GENERALIZED NONLINEAR NONPARAMETRIC ECONOMETRIC METHODS (ARTIFICIAL NEURAL NETWORKS)

DISSertation

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate
School of The Ohio State University

By

Min Qi

The Ohio State University
1996

Dissertation Committee:
Professor G. S. Maddala, Adviser
Professor Stephen R. Cosslett
Professor Nelson C. Mark

Approved by

Adviser
Department of Economics
ABSTRACT

Artificial neural networks, a generalized nonlinear nonparametric econometric method has been shown to perform very well in the two applications outlined in this thesis.

In the first application, the ANN has been applied to price S&P 500 index call options. Using current asset price, interest rate, time to maturity, exercise price, and open interest as inputs and call price as output, the ANN with five middle layer units outperforms the Black-Scholes formula based on all four performance measures. The residual analysis shows that the ANN outperforms the Black-Scholes formula in reducing the residual autocorrelations and the extent of underpricing, and the residual heteroscedasticity is less severe in the ANN than in the Black-Scholes formula. The ANN also outperforms the Black-Scholes formula in reducing the moneyness bias and the timeliness bias which are often found in the option pricing literature. The economic implications of the ANN are analyzed through pseudo weights, sum of input weights and sensitivity analysis. The pseudo weights indicate that the economic implications of the ANN model are consistent with the properties of call option prices, i.e., the lower the exercise price, or the higher the current asset price, or the longer the time to maturity, the higher the call price. Moreover, ANN can also capture the ambiguous effect of risk-free...
interest rate on option prices. Open interest has long been ignored in the option pricing literature, and is found here to be an important factor in determining option prices. The sensitivity analysis shows both the sign and magnitude of the effect of a pricing variable on option price, and is consistent with the results of pseudo weights. This application provides an alternative to the traditional Black-Scholes formula that does not require the many unrealistic assumptions required by the Black-Scholes formula and gives much superior performance. The result is very important for investors and portfolio managers who always have to judge whether a given option is worth buying or selling.

In the second application, ANN has been applied to forecast returns on S&P 500 index using five economic variables, inflation rate, unemployment insurance claim rate, price-to-earning ratio, real interest rate and money growth rate. The linear regression gives better fit than those reported in the existing return forecasting literature. It is then shown that the goodness of fit can be largely improved using an ANN by not imposing linear restrictions. The first order residual autocorrelation is smaller and the residual heteroscedasticity is less severe for the ANN than for the linear regression model. The substantial nonlinearity shown in the sensitivity analysis invalidates the economic implications from the pseudo weights and sum of input weights. The recursively computed neural network nonlinear forecasts can improve the average annual gains generated by a simple profit trading rule, though they do not statistically significantly improve the linear or the random walk forecasts in terms of conventional criteria based on squared forecasting errors. Moreover, the ANN outperforms the linear regression and random walk in predicting the direction change of returns on S&P 500 index, and more
importantly, the ANN is much better than the linear regression and the random walk models in predicting the down turns of the S&P 500 index.
Dedicated to my mother
ACKNOWLEDGMENTS

I thank my adviser, G. S. Maddala, for intellectual support, encouragement, and enthusiasm which made this dissertation possible, and for his patience in correcting both my stylistic and scientific errors.

I thank Stephen Cosselett for meaningful questions on the dissertation proposal and for providing some of the literature on the dissertation topic.

I thank Nelson Mark for stimulating discussions and helpful comments on the stock market prediction.

I also thank the participants of the 9th Graduate Research Forum at the Ohio State University, the 3rd International Conference on Neural Networks in the Capital Markets, the seminars at San Francisco State University and Kent State University for useful feedback.

I also wish to thank those who helped me to handle various problems while I was a Ph.D. student, a TA, and a job candidate, especially Chris Colflesh, Jo Ducey, Lucia Dunn, Eric Fisher, Belton Fleisher, Masanori Hashimoto, Linda Lewis, Roy McCarthy, Lisa Nock, Gail Nowosadko, Andrew Paine, Patricia Reagan, Margaret Renshler, Lars Sandberg, René Stulz and Chunjun Zhao.
I am grateful to my husband and best friend, Mingjie Wang, for constant understanding and help, my two sweet daughters, Nancy and Michelle, for giving me endless joy, my parents and parents-in-law for helping me take care of the kids.
VITA

August 12, 1966 ................................... Born in Sichuan, P. R. China.

1988 ...................................................... M.E. Management Information System,
Tsinghua University, Beijing.

1991 ....................................................... M.B.A. International Trade and Finance,
Tsinghua University, Beijing.

1991 - present ....................................... Fellow, Graduate Teaching and Research
Associate, The Ohio State University.

PUBLICATIONS

Research Publication


500 index call options. Proceedings of the 3rd International Conference on Neural

FIELD OF STUDY

Major: Economics
Fields: Econometrics, Finance, International Trade and Development
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dedication</td>
<td>v</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>vi</td>
</tr>
<tr>
<td>Vita</td>
<td>viii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xi</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xii</td>
</tr>
<tr>
<td>Chapters:</td>
<td></td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>2. ANN methods</td>
<td>3</td>
</tr>
<tr>
<td>2.1. ANN structure</td>
<td>4</td>
</tr>
<tr>
<td>2.2. ANN learning</td>
<td>5</td>
</tr>
<tr>
<td>2.3. Universal approximation</td>
<td>7</td>
</tr>
<tr>
<td>2.4. Relationship between ANNs and traditional statistical models</td>
<td>8</td>
</tr>
<tr>
<td>2.4.1. Linear regression</td>
<td>8</td>
</tr>
<tr>
<td>2.4.2. Logit and probit models</td>
<td>9</td>
</tr>
<tr>
<td>2.4.3. Principal component analysis</td>
<td>11</td>
</tr>
<tr>
<td>2.4.4. Latent variables model with multiple indicators and multiple causes (MIMIC model)</td>
<td>13</td>
</tr>
<tr>
<td>2.5. ANN implementation and interpretation</td>
<td>14</td>
</tr>
<tr>
<td>2.5.1. Model selection</td>
<td>15</td>
</tr>
<tr>
<td>2.5.1.1. ANN inputs and outputs</td>
<td>15</td>
</tr>
<tr>
<td>2.5.1.2. ANN architecture</td>
<td>17</td>
</tr>
<tr>
<td>2.5.1.3. ANN evaluation criteria</td>
<td>19</td>
</tr>
<tr>
<td>2.5.2. Statistical inference in ANN</td>
<td>22</td>
</tr>
<tr>
<td>2.5.3. Model implications</td>
<td>23</td>
</tr>
<tr>
<td>2.6. Applications in Finance and Economics</td>
<td>26</td>
</tr>
<tr>
<td>2.7. Summary</td>
<td>28</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>In-sample RMSE, MAE, $R^2$, and $\rho$ of alternative option pricing models</td>
<td>73</td>
</tr>
<tr>
<td>2</td>
<td>Out-of-sample RMSE, MAE, $R^2$, and $\rho$ of alternative option pricing models</td>
<td>74</td>
</tr>
<tr>
<td>3</td>
<td>Performance of an ANN with volatility $\sigma$ as an additional input variable</td>
<td>75</td>
</tr>
<tr>
<td>4</td>
<td>Descriptive statistics of residuals of the ANN and the Black-Scholes option pricing models</td>
<td>76</td>
</tr>
<tr>
<td>5</td>
<td>White’s heteroscedasticity test for the ANN and the Black-Scholes option pricing models</td>
<td>77</td>
</tr>
<tr>
<td>6</td>
<td>Moneyness and timeliness bias of alternative option pricing models</td>
<td>78</td>
</tr>
<tr>
<td>7</td>
<td>Pseudo weights of ANN option pricing for all five subsamples</td>
<td>79</td>
</tr>
<tr>
<td>8</td>
<td>Sum of absolute input weights for ANN option pricing across all five subsamples</td>
<td>79</td>
</tr>
<tr>
<td>9</td>
<td>Pseudo weights, sum of input weights and in-sample RMSE, $R^2$, $\rho$, and Theil’s U for ANN stock prediction</td>
<td>80</td>
</tr>
<tr>
<td>10</td>
<td>Out-of-sample RMSE, $\rho$, and Theil’s U of alternative forecasting models</td>
<td>81</td>
</tr>
<tr>
<td>11</td>
<td>The t-statistics and p-values for testing the null hypothesis of no difference in the squared prediction errors of two competing forecasts</td>
<td>80</td>
</tr>
<tr>
<td>12</td>
<td>Average annual gains (%) using naive strategy and trading rule based on recursive linear regression and ANN forecasts</td>
<td>81</td>
</tr>
<tr>
<td>Figures</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>A three-layer feedforward neural network</td>
<td>82</td>
</tr>
<tr>
<td>2</td>
<td>ADALINE network</td>
<td>83</td>
</tr>
<tr>
<td>3</td>
<td>MADALINE network</td>
<td>83</td>
</tr>
<tr>
<td>4</td>
<td>A three-layer feedforward linear ANN</td>
<td>83</td>
</tr>
<tr>
<td>5</td>
<td>Stability of alternative structures by RMSE</td>
<td>84</td>
</tr>
<tr>
<td>6</td>
<td>Stability of alternative structures by MAE</td>
<td>84</td>
</tr>
<tr>
<td>7</td>
<td>Stability of alternative structures by ρ</td>
<td>85</td>
</tr>
<tr>
<td>8</td>
<td>Stability of alternative structures by $R^2$</td>
<td>85</td>
</tr>
<tr>
<td>9</td>
<td>Economic implications of open interest V</td>
<td>86</td>
</tr>
<tr>
<td>10</td>
<td>Sensitivity analysis of the ANN option pricing model</td>
<td>87</td>
</tr>
<tr>
<td>11</td>
<td>Actual and fitted returns on S&amp;P 500 index by alternative models (in sample)</td>
<td>90</td>
</tr>
<tr>
<td>12</td>
<td>Sensitivity analysis for the ANN stock prediction model</td>
<td>91</td>
</tr>
<tr>
<td>13</td>
<td>Actual and forecasted returns on S&amp;P 500 index by alternative models (out of sample)</td>
<td>94</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

Most econometric work is on linear parametric models. Though they are relatively well developed and easy to use, there are two major limitations of this approach: the assumption of linear functional forms and the assumption of a particular parametric specification. The poor fit of linear models motivates research and applications of nonlinear regressions. Furthermore, nonparametric regression allows researchers to avoid the restrictions of parametric modeling assumptions. These two elements are incorporated in the method of artificial neural networks. Though artificial neural networks (ANNs) are unorthodox to many econometricians and statisticians, they can be considered as a class of generalized nonlinear nonparametric methods. ANNs allow one to fully utilize the data and let the data determine the structure and parameters of a model. They are appealing to applications where there are abundant high quality data and no good models. As the speed of computers increases and the cost of computing declines exponentially, this computer intensive method becomes attractive.

In the present thesis I show with some illustrations that ANNs perform very well in option pricing and stock market prediction. The plan of the thesis is as follows: Chapter
2 outlines ANN methods and reviews previous work in this area. In Chapter 3, I consider an application of ANN methods to option pricing. In Chapter 4, I apply ANN methods to prediction of stock returns. Chapter 5 presents the conclusions.
CHAPTER 2

ANN METHODS

The past decade has seen an explosive growth in studies of neural networks after three consecutive cycles of enthusiasm and skepticism since the 1940’s. This has been brought about largely by the realization that ANNs have powerful pattern recognition properties that may outperform other existing modeling techniques in many applications. ANNs have attracted attention of researchers from a diverse field of applications including signal processing, medical imaging, economic and financial modeling (to name only a few). Meanwhile researchers from cognitive science, neuroscience, psychology, biology, computer science, mathematics, physics and statistics have contributed to the structural and methodological developments of ANNs. Many different networks, such as multilayer feedforward networks, recurrent and statistical networks, associative memory networks and self-organization networks, etc., thus have been developed for different purposes. A variety of supervised or unsupervised learning rules are now available to train a network from data. Among these, multilayer feedforward backpropagation network is the most popular one in economic and financial applications and is the focus of the present thesis.
Most of the materials presented in this chapter are drawn from Qi (1996). Wide-ranging introductions to neural network theory can be found in Hecht-Nielsen (1990), Hertz, Grogh and Palmer (1991), Wasserman (1993) and Bose and Liang (1996).

2.1. **ANN structure**

Inspired by studies of the brain and nerve system, neural networks simulate a highly interconnected, parallel computational structure with many relatively simple individual units. Individual units are organized in layers: the input, middle and output layers. Feedforward networks map inputs into outputs with signals flowing in one direction only, from the input layer to the middle layer and then the output layer. Each unit in the middle and output layers has a transfer function which transfers the signal it receives. The input layer units do not have a transfer function, but they are used to distribute input signals to the network. Each connection has a numerical weight, which modifies the signals that pass through it.

Consider a three-layer feedforward network with a single output unit, k middle layer units and n input units (see Figure 1). The input layer can be represented by a vector $X = (x_1, x_2, ..., x_n)'$, the middle layer can be represented by a vector $M = (m_1, m_2, ..., m_k)'$, and $y$ is the output. Any middle layer unit receives the weighted sum of all inputs and a bias term (denoted by $x_0$, $x_0$ always equals one), and produces an output signal

$$m_j = F(\sum \beta_{ij} x_i) = F(X' \beta_j), \quad j = 1, 2, ..., k, \quad i = 0, 1, 2, ..., n,$$  \hspace{1cm} (2.1)
where $F$ is the transfer function, $x_i$ is the $i$th input signal, and $\beta_{ij}$ is the weight of the connection from the $i$th input unit to the $j$th middle layer unit. In the same way, the output unit receives the weighted sum of the output signals of the middle layer units and produces a signal

$$y = G(\sum_{j} \alpha_j m_j), \quad j = 0, 1, 2, ..., k,$$

(2.2)

where $G$ is the transfer function, $\alpha_j$ is the weight of the connection from the $j$th middle layer unit to the output unit, and $j = 0$ indexes a bias unit with $m_0$ always equals one. Substituting (2.1) into (2.2), we get

$$y = G(\alpha_0 + \sum_{j=1}^{k} \alpha_j F(\sum_{i} \beta_{ij} x_i)) = f(X, \theta),$$

(2.3)

where $X$ is the vector of inputs, and $\theta = (\alpha_0, \alpha_1, \alpha_2, ..., \alpha_k, \beta_{01}, \beta_{11}, ..., \beta_{n1}, \beta_{02}, \beta_{12}, ..., \beta_{n2}, \beta_{0k}, \beta_{1k}, ..., \beta_{sk})'$ is the vector of network weights. $F$ and $G$ can take several functional forms, such as the threshold function which produces binary ($\pm 1$) or (0/1) output, the sigmoid (or logistic) function which produces output between 0 and 1,

$$F(a) = G(a) = \frac{1}{1 + \exp(-a)},$$

or $F(a) = a$ (identity) and $G(a) = \frac{1}{1 + \exp(-a)}$.

(2.3) can be interpreted as a nonlinear function which represents the described three-layer feedforward neural network. As will be shown in Section 2.4, this representation nests many familiar statistical models, such as regression (linear and nonlinear), classification (logit, probit), structural equation models (MIMIC, PCA), and time series analysis (ARMA, GARCH).
2.2. **ANN learning**

The most widely used estimation method (or so called learning rule) of the ANN described in the previous section is error backpropagation (Rumelhart, Hinton and Williams, 1986a, b), which is considered to be a major reason of the explosive reemergence of interest in multilayer neural networks in the mid-1980's. Backpropagation is a recursive gradient descent method that minimizes the sum of the squared error of the system by moving down the gradient of the error curve. More specifically, network weight vector $\theta$ is chosen to minimize the loss function,

$$\min_{\theta} L = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2,$$

where $N$ is the sample size, $y_i$ is desired (or target, actual) output value and $\hat{y}_i$ is the calculated output value, and

$$\hat{y}_i = f(X_i, \theta) = G(\alpha_0 + \sum_{j=1}^{k} \alpha_j F(\sum_{i} \beta_{ij} x_{ij})).$$

Then the iterative step of the gradient descent algorithm takes $\theta$ to $\theta + \Delta \theta$, and

$$\Delta \theta = -\eta \nabla f(X_t, \theta)(y_t - f(X_t, \theta)),$$

where $\eta > 0$ is the step size, or learning rate, $\nabla f(X_t, \theta)$ is the gradient of $f(X_t, \theta)$ with respect to $\theta$ (a column vector), and the chain rule is used to calculate $\nabla f(X_t, \theta)$.

The error surface is multi-dimensional and may contain many local minima. As a result, training the network often requires experimentation with different starting weights, adjusting the learning rate, or adding a momentum term to avoid getting stuck in local optima or slow convergence. For most studies that aim at comparing the ANN with some
alternative models, as long as the ANN performs significantly better than its counterpart, it is not necessary to search for global minima. For studies that try to search for global minima, a grid search method is often used (see Gorr, Nagin and Szczypula, 1994, for example). Other methods have also been proposed, for example, Baldi and Hornik (1989) find that the error surface has a unique minimum which corresponds to the projection onto the subspace generated by the first principal component vector of the covariance matrix of the data. White, Gallant, Hornik, Stinchcombe and Wooldridge (1992) has more discussion about the global optimization. The iteration stops when either the prespecified maximum number of iterations or the error goal has been reached.

2.3. Universal approximation

A major advantage of ANNs is their ability to provide a flexible mapping between inputs and outputs. Based on a series of studies by Kolmogorov (1957), Sprecher (1965), Lorentz (1976), and Hecht-Nielsen (1987, 1990), any continuous function can be computed using linear summations and a single properly chosen nonlinear function. Therefore, the arrangement of the simple units into a multilayer framework produces a mapping between inputs and outputs that is consistent with any underlying functional relationship regardless of its “true” functional form. Having a general mapping between the input and output vectors eliminates the need for unjustified \textit{a priori} restrictions which are needed in common statistical and econometric modeling.

However, to implement a perfectly general mapping between inputs and outputs, correct transfer functions are needed. Sigmoid middle layer transfer function has been
shown to serve the purpose by studies like Cybenko (1989), Funahashi (1989), Hecht-Nielsen (1989), Hornik, Stinchcombe and White (1989). Stinchcombe and White (1989) show that some non-sigmoid functions can also be used. Thus, an ANN can be viewed as a "universal approximator", i.e., a flexible functional form that can approximate an arbitrary function arbitrarily well, given sufficiently many middle layer units and properly adjusted weights.

2.4. Relationship between ANNs and traditional statistical models

Most of the development in neural networks has been achieved primarily by non-statisticians. Consequently, few statistical concepts and methods have been applied in this development. Nevertheless, some familiar statistical models can be represented in a general ANN framework, and many concepts and constructs can be expressed in a neural network notation (Cheng and Titterington, 1994). On the other hand, ANNs can be considered as a particular class of nonlinear parametric models, and "learning" corresponds to statistical estimation of the model parameters. As a result, modern theory of estimation and inference for nonlinear models can be applied to neural network learning (White, 1989a; Kuan and White, 1994). This section briefly outlines the relationship between ANNs and some of the traditional statistical methods.
2.4.1. Linear regression

Multiple linear regression models can be represented by a simple two-layer feedforward network with a linear transfer function \( F(a) = a \), an ADALINE network of Widrow and Hoff (1960) (see Figure 2),

\[
y = \sum_{i=0}^{n} \beta_i x_i = X'\beta,
\]

where \( y \) is the output value, \( X = (x_0, x_1, \ldots, x_n)' \) is the input vector, and \( \beta = (\beta_0, \beta_1, \ldots, \beta_n)' \) is the weight vector. While such a network has been proved useful in a variety of applications, it cannot generalize or perform well on patterns that have never been presented. It is also less efficient than linear regression. However, it does not assume any thing (such as linearity and orthogonality in linear regression) about the true data generating process, and thus is more robust than classical linear regression.

A multiple adaptive linear network, MADALINE of Widrow and Hoff (1960), can be used to represent the standard systems of seemingly unrelated regressions (Figure 3):

\[
y_1 = \sum_{i=0}^{n} \beta_{i1} x_i = X'\beta_1, \\
y_2 = \sum_{i=0}^{n} \beta_{i2} x_i = X'\beta_2, \\
\vdots \\
y_k = \sum_{i=0}^{n} \beta_{ik} x_i = X'\beta_k.
\]

If lagged outputs are used as network inputs in an ADALINE network, we get linear AR(d) time series equation:
\[ y_t = \sum_{i=1}^{d} \beta_i y_{t-1} . \]  

(2.9)

2.4.2. Logit and probit models

In the two-layer ADALINE network with a linear discriminant transfer function, units are not activated until some threshold level is reached, i.e.,

\[ y = F\left(\sum_{i=0}^{n} \beta_i x_i \right), \]  

(2.10)

where the transfer function \( F(a) = 1 \) if \( a > 0 \) and \( F(a) = 0 \) if \( a \leq 0 \). The output unit is thus a threshold unit.

Networks with a threshold output unit are suited for classification and pattern recognition problems. Since the transfer function \( F \) can be any continuous, non-decreasing function, \( F \) can represent a cumulative distribution function (cdf).

When \( F \) is the logistic cumulative distribution function, \( F\left(\sum_{i=0}^{n} \beta_i x_i \right) \) is the conditional expectation of the familiar binary logit model. When \( F \) is the normal cumulative distribution function, \( F\left(\sum_{i=0}^{n} \beta_i x_i \right) \) is the conditional expectation of a binary random variable generated by a probit model. For a more detailed introduction of logit and probit models, see Maddala (1983).

Therefore, a two-layer neural network can represent the familiar logit and probit regression models, which are very popular in financial applications where binary classifications or decisions are involved. However, due to the limitations of a two-layer
neural network, most of the classification applications of ANNs use one or more middle layers. It has been shown by Tam and Kiang (1992) that a two-layer ANN has a performance similar to that of linear discriminant analysis, but the incorporation of a hidden layer considerably improves the predictive accuracy. More work on ANNs and related methods for classification is discussed in Ripley (1994).

2.4.3. Principal component analysis

Principal component analysis (PCA) is a common statistical method of data analysis often used for reduction in the dimension of data matrix. The purpose is to find a set of m orthogonal vectors in data space that account for as much as possible of the data variance. Typically m is smaller than the dimension of the original data, thus, PCA performs a dimension reduction that retains most of the intrinsic information in the data and makes the reduced data much easier to handle. For a more detailed discussion, see Rao (1964) who examines the issue in what sense principal components provide a reduction of the data without much loss of information we are seeking from the data.

Specifically, the first principal component is taken to be along the direction with maximum variance. The second principal component is constrained to lie in the subspace perpendicular to the first, within which it is taken along the direction with the maximum variance. Then the third principal component is taken in the maximum variance direction in the subspace perpendicular to the first two, and so on. In general the kth principal component direction is along an eigenvector direction belonging to the kth largest eigenvalue of the full covariance matrix.
Several ANNs can perform PCA (Hertz, Grogh and Palmer, 1991). Let's first consider a two-layer linear feedforward network (see Figure 3),

$$y_j = \sum_{i=1}^{n} \beta_{ij} x_i = X' \beta_j,$$  \hspace{1cm} (2.11)

where the input vector $X = (x_1, x_2, ..., x_n)'$ is n-dimensional, and $\beta_j$ is the weight vector for the jth output. Under either of the following learning rules,

$$\Delta \beta_{ij} = \eta y_j (x_i - \sum_{k=1}^{j} y_k \beta_{ki}), \hspace{1cm} \text{(Sanger, 1989), or}$$  \hspace{1cm} (2.12)

$$\Delta \beta_{ij} = \eta y_j (x_i - \sum_{k=1}^{n} y_k \beta_{ki}), \hspace{1cm} \text{(Oja, 1989),}$$  \hspace{1cm} (2.13)

when an equilibrium has been reached, the average weight change is expected to be zero. It can be shown that

$$\text{mean}(\Delta \beta_j) = C \beta_j - (\beta_j' C \beta_j) \beta_j = 0,$$  \hspace{1cm} (2.14)

where $C$ is the correlation matrix. An equilibrium weight vector thus must satisfy

$$C \beta_j = \lambda_j \beta_j,$$  \hspace{1cm} (2.15)

with

$$\lambda_j = \beta_j' C \beta_j = \beta_j' \lambda_j \beta_j = \lambda_j \beta_j' \beta_j.$$  \hspace{1cm} (2.16)

(2.15) shows clearly that an equilibrium $\beta_j$ must be an eigenvector of the correlation matrix $C$, and (2.16) proves that $|\beta| = 1$. It can also be shown that $\lambda_j$ is the jth largest eigenvalue.

PCA can also be performed by a three layer linear ANN with n inputs, n outputs, and m < n middle layer units, using a self-supervised backpropagation approach (Sanger, 1989).
1989). The idea is to make the target outputs equal to the inputs. As the outputs become arbitrarily close to the inputs in the training set, the m middle layer units end up projecting onto the subspace of the first m principal components. Various generalizations of neural PCA-type learning algorithms containing nonlinearities have been derived and discussed in Karhunen and Joutsensalo (1995).

2.4.4. Latent variable model with multiple indicators and multiple causes (MIMIC model)

Causal models which contain latent variables have been extensively applied in several areas of social science, such as psychology, economics, education. They are potentially useful in financial applications. The latent variables are hypothetical and not directly observable, but have implications for relationships among observable variables. The observable variables may be effects ("indicators"), or causes of the latent variables, or both. Causal models with multiple indicators and multiple causes of latent variables are sometimes called MIMIC models. Such a MIMIC model can be easily represented in a three-layer feedforward linear ANN (see Figure 4):

\[ M = X' \beta \]  \hspace{1cm} (2.17)

\[ Y = M' \alpha \]  \hspace{1cm} (2.18)

where \( \beta \) is the weight matrix of connections between the input and middle layers, and \( \alpha \) is the weight matrix of connections between the middle and output layers.

In (2.17), the middle layer units of the ANN, \( M = (m_1, m_2, \ldots, m_k)' \) (comparable to the latent variables in a MIMIC model), is linearly determined by the input vector of the
ANN, \( X = (x_1, x_2, \ldots, x_n)' \) (corresponding to a set of observable exogenous causes). In (2.18), the middle layer units of the ANN linearly determine the output units of the ANN, \( Y = (y_1, y_2, \ldots, y_m)' \), which correspond to a set of observable endogenous indicators.

Under some assumptions about the disturbances added to (2.17) and (2.18), and some restrictions on the reduced form, the MIMIC model can be estimated by maximum-likelihood or some limited information approach (Jöreskog and Goldberger, 1975). While no additional restrictions are needed to train an ANN MIMIC model, such a multilayer linear network has the same limitations as a two-layer one. It can only work if the input patterns are linearly independent (Hertz, Grogh and Palmer, 1991). A multilayer nonlinear network which can represent nonlinear MIMIC models will be more interesting.

### 2.5. ANN implementation and interpretation

It is well known that there are several limitations that may restrict the use of neural networks. First, there is no formal theory for determining optimal network structure, and the appropriate number of layers and middle layer units must be determined by experimentation. Second, there is no optimal algorithm to ensure the global minimum because of the multi-minima error surface. Third, ANN does not have parametric statistical properties, thus no statistical inference can be carried out. Fourth, it is difficult to interpret a trained ANN model.

These limitations call for further studies in three broad areas outlined in Cheng and Titterington (1994): (1) mathematical modeling of real cognitive process; (2) theoretical investigations of networks and neurocomputing; (3) development of useful tools for
practical prediction and pattern recognition. While the first two areas are certainly important, they are not the focus of the present thesis. In this section, I outline some of the useful techniques and procedures that aim to overcome the aforementioned limitations.

2.5.1. *Model selection*

Though ANNs can be universal approximators, the optimal network structure is not determined automatically. Failures in applications are sometimes due to a suboptimal ANN structure.

To develop the optimal network in any financial application, one needs to (1) identify the most relevant inputs and outputs; (2) choose an appropriate network structure including the necessary number of hidden layers and hidden layer units; (3) use proper model evaluation criteria. I now clarify these points one by one.

2.5.1.1. *ANN inputs and outputs*

The choice of network input and output variables and the quality of data are critical to the success of ANN applications. The choice depends heavily on the type of task that an ANN is expected to perform and is more or less subjective to the modeler’s discretion on the model and the scope of the study. It is common practice to use independent variables as network inputs and use dependent variables as network outputs in a model.

For example, in a seminal study aimed at extracting nonlinear regularities from economic time series, White (1988) uses the lagged one day returns on IBM stock, \( r_{t-1}, r_t \).
as the network inputs and the one day return on day \( t \), \( r_t \), as the network output.

The goodness of fit of such an ANN provides evidence for or against the efficient markets hypothesis and the presence of nonlinear regularities in the case of IBM daily stock returns. However, as the author points out, in order to expand the scope of the search for evidence against the efficient markets hypothesis, the network needs to be elaborated by allowing additional inputs, such as volume, other stock prices and volume, leading indicators, macroeconomic data, etc.

In another study by Grudnitski and Osburn (1993), 24 input units and one output unit are used in their ANN based on the belief that general economic conditions and traders' expectation about the futures market are related to price movements of futures. The input units represent six input variables per month (i.e., price change, price volatility, money growth rate, three percentage commitments of large speculators, large hedgers, and small traders) presented four months at a time. The output is the change of the monthly centered price mean for the forecast month.

Sometimes, if there are more independent variables than one desires to include in the network input, dimension reduction techniques can be used. One can choose a smaller group of statistically significant variables from a regression of the dependent variable on a large group of independent variables. Principal component analysis and stepwise regression can also be used. For example, Salchenberger, Cinar and Lash (1992) perform a stepwise regression on 29 financial ratios, which results in the identification of five variables. Then the five financial variables are used as inputs of a neural network to forecast the probability of failure of thrift institutions.
In order to minimize the effect of magnitude among the inputs and outputs and increase the effectiveness of the learning algorithm, the data set is often normalized (or scaled) to be within a specific range depending on the transfer function. For example, if an ANN has sigmoid or logistic transfer function in the output unit, output needs to be scaled to fall in the range of $[0, 1]$. Otherwise, a target output which falls outside that range will constantly create large backpropagation errors, and the network will be unable to learn the input-output relationship that is implied by the particular training pattern. Typically, variables will be normalized to have zero mean and unit standard deviation.

The quality of data and the degree to which data sets properly represent the population are very important, as is the case in any econometric and statistical modeling. To train and test an ANN, it is also important to have enough data.

2.5.1.2. ANN architecture

After specifying the network input and output layers, the ANN architecture remains undetermined unless the necessary number of hidden layers and hidden layer units are determined. Consider layered networks of continuous-valued units with logistic transfer functions for hidden units and linear transfer functions for output units. Overall such a network implies a function, $y = f(X)$, from input variables, $X = (x_1, x_2, \ldots, x_n)'$, to output value, $y$.

Due to the limitation of the capacity of a two-layer ANN (Hertz, Grogh and Palmer 1991), networks with at least one middle layer are often used. Cybenko (1988) proves that ANN with at most two hidden layers can approximate a particular set of
function $y = f(X)$ with arbitrary accuracy given enough units per layer. It has also been proved that only one hidden layer is enough to approximate any continuous function (Cybenko, 1989; Hornik, Stinchcombe and White, 1989). Many empirical studies, such as Collins, Ghosh and Scofield (1988), Dutta and Shekhar (1988), Salchenberger, Cinar and Lash (1992) (to name a few) have confirmed this. The correctness of these results, however, hinge on the appropriate number of hidden units.

The choice of $k$, the number of hidden units, represents a compromise. If $k$ is too small, an ANN may not approximate $y = f(X)$ at the desired accuracy. However, if $k$ is too large, an ANN may overfit and can not generalize (or forecast) out of sample. A useful method is cross-validation, by which the number of middle layer units is selected to optimize out-of-sample performance (White, 1990). Another related model selection criterion, predictive stochastic complexity (PSC; defined as Equation (2.28)) can also be used (Kuan and Liu, 1995).

Other common methods for optimal network design have been reviewed by Refenes (1995b). These methods fall into three groups. The first is analytic techniques in which algebraic or statistical analysis is used to determine \textit{a priori} hidden unit size. Several rules of thumb have been cited, such as the number of connections should be less than $0.1T$ and the number of hidden units is of the order of $(T-1)$ or $\log_2 T$, where $T$ is the sample size. The main problem with these techniques is that they perform static analysis and can only provide a very rough estimate for hidden unit size. However, they compare well with current experimental methods for network design.
The second type is constructive techniques, such as cascade correlation (Fahlman and Lebiere, 1990), tiling algorithm (Mezard and Nadal, 1989), neural decision tree (Gallant, 1986), upstart algorithm (Frean, 1989) and the CLS procedure (Refenes and Vithlani, 1991). These methods construct the hidden units in layers one by one as they are needed. Though these techniques guarantee the network convergence, generalization and stability are not guaranteed.

The last type, network pruning, operates in the opposite direction by pruning the network and removing "redundant" or least sensitive connections. These include network pruning (Sietsma and Dow, 1991) and artificial selection (Hergert, Finnoff and Zimmermann, 1992). However, optimal pruning is not always possible.

2.5.1.3. ANN evaluation criteria

A criterion is always needed to compare the performance of alternative models and select the best one. Let \((\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_N)\) denote the predicted values and \((y_1, y_2, \ldots, y_N)\) be the actual values, where \(N\) is the sample size. Some of the commonly used criteria are listed below.

(1) Mean square error (MSE) and root mean square error (RMSE):

\[
MSE = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2, \tag{2.19}
\]

\[
RMSE = \sqrt{MSE}. \tag{2.20}
\]

(2) Mean absolute error (MAE) and mean absolute percentage error (MAPE):
\[
\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i|, \quad (2.21)
\]

\[
\text{MAPE} = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{y_i - \hat{y}_i}{y_i} \right|. \quad (2.22)
\]

(MAPE is not available for samples which contain zero actual values).

(3) Coefficient of determination (\(R^2\)):

\[
R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}, \quad (2.23)
\]

where \(\bar{y} = \frac{1}{N} \sum y_i\).

(4) Pearson correlation coefficient (\(\rho\)): \(\rho\) measures the linear correlation between predicted values and actual values, which quantifies the extent to which the fitted or forecasted returns coincide with the actual turning points.

\[
\rho = \frac{\sum (y_i - \bar{y})(\hat{y}_i - \bar{y})}{\sqrt{\sum (y_i - \bar{y})^2 \sum (\hat{y}_i - \bar{y})^2}}. \quad (2.24)
\]

(5) Theil's coefficient of inequality (\(U\)): Theil's U gives prediction performance relative to the random walk prediction, a smaller than one Theil's U indicates that the forecast is better than the random walk forecast in RMSE sense.

\[
U = \frac{\text{RMSE}}{\sqrt{\frac{1}{N-1} \sum (y_i - y_{i-1})^2}}. \quad (2.25)
\]

(6) Akaike information criterion (AIC): AIC adjusts MSE to account for the model complexity,
$$AIC = \text{MSE} \left( \frac{N+k}{N-k} \right),$$  \hspace{1cm} (2.26)

where $k$ is the number of free parameters in the model, or the number of free weights in an ANN.

(7) Schwarz information criterion (SIC), or Bayesian information criterion (BIC): SIC or BIC is another way of adjusting MSE to account for model complexity,

$$\text{SIC} = \text{BIC} = \ln(\text{MSE}) + \frac{\ln(N)}{N} k.$$  \hspace{1cm} (2.27)

(8) Predictive stochastic complexity (PSC):

$$\text{PSC} = \frac{1}{N-k} \sum_{i=k+1}^{N} (y_i - \hat{y}_{ii})^2,$$  \hspace{1cm} (2.28)

where $\hat{y}_{ii}$ is the predicted value based on parameters obtained from the data up to the $i-1$ observation.

(9) Direction accuracy (DA) and confusion rate (CR):

$$DA = \frac{1}{N} \sum a_i,$$  \hspace{1cm} (2.29)

where $a_i = \begin{cases} 1 & \text{if } (y_{i+1} - y_i)(\hat{y}_{i+1} - y_i) > 0, \\ 0 & \text{otherwise}. \end{cases}$

$$CR = 1 - DA.$$  \hspace{1cm} (2.30)

Sometimes, the significance of the difference in the performance of alternative models needs to be tested. T-test or Diebold-Mariano test (Diebold and Mariano, 1995) are often used to test the null hypotheses that there is no difference in the square errors of two alternative models. The hypothesis of independence between the actual and predicted
directions can be tested by the HM test (Henriksson and Merton, 1981; Pesaran and Timmerman, 1994).

It is worth noting that the in-sample performance of any properly designed and well trained ANNs, evaluated by the above measures, is usually much better than from their traditional statistical counterparts. This is not surprising given the universal approximation property of ANNs. To avoid spurious fit or overfit, it is important to test the resulting ANN using hold-out sample, i.e., to evaluate the trained ANN using data not been used in training the ANN. Whether the selected ANN model is useful or not depends primarily on the out-of-sample performance.

Swanson and White (1995a, b) show that compared to a variety of out-of-sample forecast-based model selection criteria, such as forecast mean squared error, forecast direction accuracy, or forecast-based trading system profitability, an in-sample Schwarz information criterion (SIC) does not appear to be a reliable guide to out-of-sample performance. In cases where out-of-sample performance measures are used as model selection criteria, it is important to test the model using strictly untouched data set, i.e., data not used in training and validation. Otherwise, an upward bias in out-of-sample forecasting accuracy is likely to occur.

2.5.2. Statistical inference in ANN

Very few empirical studies of ANN applications report confidence intervals or conduct hypothesis testing, because the classical statistical properties are generally not available. However, if we view (2.4) as a nonlinear least squares regression, then the
estimator of $\theta$ will have the statistical properties of a nonlinear least square estimator. Thus, statistical inference can be carried out. Under some general conditions, such as i.i.d., the nonlinear least squares estimator $\hat{\theta}_n$ that solves (2.4) converges almost surely to $\theta^*$ that solves $\min_{\theta} \mathbb{E}(\mathbb{E}(y_t|X_t) - f(X_t, \theta))^2$. Thus $\theta^*$ is a parameter vector of a minimum mean squared error approximation $f(X_t, \theta^*)$ to $\mathbb{E}(y_t|X_t)$. One can therefore test hypothesis about the parameters of the best approximation. For details, see White (1989a, b), Kuan and White (1994).

A bootstrap method has been proposed by Lebaron and Weigend (1994) to determine the quality and reliability of a neural network predictor. Though the method is extremely computationally intensive and requires i.i.d. assumption, it does provide more robust forecasting along with the probability distribution of the forecast results. In their multivariate time series prediction of daily total trading volume on the New York Stock Exchange, the bootstrapping results show that the performance variation due to different splits between training, cross-validation, and testing samples is significantly larger than the variance due to different network architecture and initial weights.

2.5.3. Model implications

Artificial neural networks are often viewed as "black boxes", because the estimated models are difficult to explain due to their complex functional forms. However, the relationship between weights, inputs and outputs is clearly defined, which allows us to look into the "black boxes" and find the economic implications of ANN models. Following
the notation in Section 2.1 for a three-layer ANN as shown in Figure 1, several practical methods have been proposed to interpret the relative significance of each input variable on the output.

(1) Pseudo weights

In an application of ANN to price call options using five input variables, Qi and Maddala (1995) use the weighted average of the input weights, or so called pseudo weights, to approximate the marginal contribution of an input variable to the output. The idea is that since the training data set have been normalized by units so that all input values are between -3 and 3, the higher the magnitude of the weight of an input, the more the input can affect the output. However, this is not necessarily true due to the existence of middle layer units. For example, an input with a large weight may have a small effect on output if the weight from the middle layer unit to the output is very small, yet an input with a small weight to a middle layer unit may have a large effect on output if the weight from the middle layer unit to the output is quite large. The pseudo weight for the $i$th input variable is defined as:

$$PW_i = \sum_{j=1}^{k} \alpha_{ij} \beta_{ij} = \alpha \beta_i.$$  \hspace{1cm} (2.31)

It is reported by Qi and Maddala that the economic implications of PW are consistent with the call option properties.

(2) Sum of input weights

Sen, Oliver and Sen (1995) proposed, and Refenes, Zapranis and Francis (1995) adopted the idea of summing the absolute values of the input weights for each input
variable to approximate the degree of impact that an input variable has on the outcome. The sum of input weights (SW) for the $i$th input variable is calculated as:

$$SW_i = \sum_{j=1}^{k} |\beta_{ij}|.$$  \hspace{1cm} (2.32)

Sen, Oliver and Sen (1995) find that all the variables found significant in the logit analysis to predict corporate mergers are included in the set of five variables with the highest sum of input weights.

Notice the difference between PW and SW. SW loses information about the negative effect of an input variable on the output by taking the absolute values. If the weights were all positive, PW and SW should end up with the same rank order for the different input variables. More importantly, as will be shown in Section 4.6 that in the presence of substantial nonlinearity, both PW and SW are no longer relevant, and a useful tool of model interpretation is sensitivity analysis.

(3) Sensitivity analysis

Sensitivity analysis shows the sensitivity of the network output to changes in the input variables. To perform the sensitivity analysis, the minimum, maximum and the mean (or median) of each input variable are first determined. The value of each input variable is varied one at a time, holding the values of other input variables fixed at the their mean (or median). For each predictor being varied, the values are spread over certain number of equal intervals over its whole range. The neural network model is then used to compute the output. The plot of the neural network outcome against the value of the input variable indicates how the network output changes with particular input variable with other input
variables being fixed. This sensitivity analysis has been utilized by Sen, Oliver and Sen (1995) and Refenes, Zapranis and Francis (1995) to gain the insights into their models.

(4) Sensitivity index

Sen, Oliver and Sen (1995) use a sensitivity index to find out the relative strength of the influence of an input variable on the output. The index for the $i$th input variable is computed by averaging the changes of output for certain number ($M$) of equal interval changes over the whole range of that input variable:

$$SI_i = \frac{1}{M} \sum_i (\hat{y}_{i+1} - \hat{y}_i). \quad (2.33)$$

The sensitivity index provides a measure of “significance” of the input variables in predicting the output. The results in Sen et al. (1995) agree, in part, with the logistic regression.

2.6. Applications in Finance and Economics

Neural networks are gaining their popularity in Finance and Economics. Major applications are in the following areas:

Refenes (1995a). Most of these studies find that ANN forecasts can generate larger profits than a benchmark linear model, though the forecast errors may not significantly different from those of a benchmark linear model.

(ii) Option pricing: Studies by Bailey, Thompson and Feinstein (1988), Hutchinson, Lo and Poggio (1994), Qi and Maddala (1995), and White (1995) find that the ANN is a good alternative to the benchmark Black-Scholes option pricing formula. Section 3.1 will review the literature in greater detail.

(iii) Prediction of bond rating: Dutta and Shekhar (1988), Surkan and Singleton (1990), Utans and Moody (1991), Moody and Utans (1995), Singleton and Surkan (1995) report that the ANN is superior to the traditional methods such as linear regression and discriminant analysis in predicting bond ratings.


(v) Bankruptcy prediction: Odom and Sharda (1990), Tam and Kiang (1990), Raghupathi, Schkade and Raju (1991), Tam and Kiang (1992), Coats and Fant (1992), Salchenberger, Cinar and Lash (1992), Huang (1993), Altman, Marco and Varetto (1994) and Poddig (1995), are some of the many applications of the ANN in this area. There is strong evidence that ANNs are suitable for the task of bankruptcy prediction and that they are able to outperform traditional discriminant analysis and logit model.
(vi) Corporate merger prediction: Sen, Oliver and Sen (1995) show that neural networks have better modeling capabilities than logistic regression. However, neural networks are still not capable of predicting mergers well.

(vii) Market response models: Dasgupta, Dispensa and Ghose (1994) report that the ANN performs better than two statistical market response models: logit model and discriminant analysis. However, the improvement is not significantly higher.

(viii) Model specification: White (1989c) tests for neglected nonlinearity using a multilayer feedforward networks. Lee, White and Granger (1993) compare the neural network test for neglected nonlinearity with the Keenan test, the Tsay test, the White dynamic information matrix test, the McLeod-Li test, the Ramsey RESET test, the Brock-Dechert-Scheinkman test, and the Bispectrum test. The relative performance of the ANN test is found to be encouraging and many of the economic series tested exhibit potential nonlinearities. Donaldson, Kamstra and Kim (1993) find the ANN does outperform the specialized finance models such as ARCH in removing the leptokurtosis and symmetric and asymmetric heteroskedasticity from the data.

(ix) Macroeconomic forecasting: Swanson and White (1995a) show that the premium of the forward rate over the spot rate helps to predict the sign of future changes in the interest rate and ANN models appear to be promising for use in this forecasting context. Swanson and White (1995b) report that the real-time macroeconomic forecasts from a model selection approach using linear models and ANNs outperform those in the Survey of Professional Forecasters.
2.7. Summary

In this chapter, I briefly outline the ANN methods. The ANN structure and learning methods are introduced in Section 2.1 and Section 2.2. The universal approximation property of ANNs is explained in Section 2.3. I also points out the relationship between ANNs and some traditional statistical models in Section 2.4. Some practical issues such as model section, statistical inference and model interpretation are reviewed in Section 2.5. Several major areas of applications in Finance and Economics are briefly reviewed in Section 2.6.
CHAPTER 3

OPTION PRICING USING ANN

3.1. Introduction

Options trading started on a national exchange in 1973 when the Chicago Board Options Exchange began listing call options. Now option contracts are traded on several exchanges, written on a wide variety of financial instruments including common stocks, stock indices, foreign exchange rates, agricultural commodities, precious metals, and various futures. Popular and potent tools to increase an investor's exposure to asset prices, or to provide insurance against volatility of asset prices, options have become essential tools that a portfolio manager must understand.

There are two basic types of options: calls and puts. A call (or a put) option gives its holder the right to buy (or sell) the underlying asset by a certain date for a certain price. The price specified in the contract is the exercise price or strike price; the date specified in the contract is the expiration date, exercise date or time to maturity. American options can be exercised at any time up to the expiration date. European options can only be exercised on the expiration date. The holder of an option has the right to but does not have to exercise the right to buy or sell the underlying. An
An investor must pay a price to purchase an option contract, and the price paid by an investor to hold an option contract is known as the option price.

Financial economists searched for years for a workable option-pricing model before Black and Scholes (1973) and Merton (1973) derived the famous Black-Scholes formula (BSF). Based on the principle that “if options are correctly priced in the market, it should not be possible to make sure profits by creating portfolios of long and short positions in options and their underlying stocks” (Black and Scholes 1973, p637), the authors set up a perfect hedge position with a return equal to the short term risk-free rate. Using stochastic calculus and solving the derived partial differential equation, they find the closed form solution for call option prices:

\[ q = SN(d_1) - Xe^{-rT}N(d_2), \]

where \( d_1 = \frac{\ln(S/X) + rT}{\sigma \sqrt{T} + \frac{\sigma \sqrt{T}}{2}} \), \( d_2 = d_1 - \sigma \sqrt{T} \), \( S \) is current underlying asset price, \( X \) is exercise price, \( r \) is short-term risk-free interest rate, \( T \) is time to maturity, and \( \sigma \) is the volatility of the underlying asset.

Often considered as the most significant breakthrough in finance theory in the last three decades, the Black-Scholes model suffers from some unrealistic simplifying assumptions: (i) The option is exercised only at the expiration date (the so-called European option). While the American options which can be exercised before the expiration date are traded in the US, Canada, and Europe, the standard European options are rarely traded except for some foreign currency options. Moreover, when the underlying asset pays dividends, an American call option owner has the incentive to
exercise the option early to own the underlying asset and earn the dividend. Since American option owners have the choice of exercise early, the price of an American call option is in general higher than that of an otherwise the same European call option. Therefore, the Black-Scholes formula which gives European option prices tends to underestimate the American option prices. As can be seen in Section 3.6, the differences between the actual option prices and the option prices predicted by the Black-Scholes formula are positive for almost all cases. (ii) The underlying asset pays no dividends. While this is true for assets such as precious metals and interest rate futures, this is obviously not true for common stocks and foreign exchange rates. (iii) The short-term interest rate is a known constant over time, yet, recent studies have derived option pricing models of various stochastic interest rate processes. (iv) Stock prices are lognormally distributed with a constant variance rate, yet, studies such as Fama (1965) and Akgiray (1989), to name a few, have rejected this assumption. (v) There are no transaction costs. (vi) There is no restriction on borrowing at the short-term interest rate. (vii) There are no penalties to short selling. The last three assumptions are obviously not true in reality.

Therefore, it is not surprising that several studies including Black and Scholes (1972) have found that the actual option prices deviate in certain systematic ways from the values predicted by the Black-Scholes formula. Hence, a more accurate valuation of options is still in need.

Though neural networks have been extensively applied in financial areas such as bank failure assessment, bond rating prediction, commodities trading, credit rating, investment screening, loan underwriting, and return forecasting as reviewed in Section 2.6,
very few applications discuss the pricing of options. Since ANN can approximate various functions of input vector from the data, although might not be a substitute for the Black-Scholes formula, it should be promising to discover a more accurate and efficient option pricing formula when the price dynamics of the underlying asset are unknown, or when the pricing equation associated with the no-arbitrage condition cannot be solved analytically.

The first well known experiment of option pricing using neural networks is by Hutchinson, Lo and Poggio (1994). First, the potential value of neural network pricing formula has been shown by the fact that neural networks can discover the Black-Scholes formula from a two-year training set of simulated daily option prices. The option prices are simulated based on all the Black-Scholes assumptions, such as geometric Brownian motion with constant mean and volatility, constant interest rate, etc. The resulting network formula has been shown to be successful in pricing and delta-hedging options out-of-sample. Then the network is applied to the pricing and delta-hedging of S&P 500 futures options from 1987 to 1991. The results show that neural networks outperform the Black-Scholes formula.

However, Hutchinson et al. (1994) assume constant risk-free interest rate and constant volatility of the underlying asset. They further assume that the return of the underlying asset is independent of the level of the stock price, so that the option pricing formula is homogeneous of degree one in both S, the asset price and X, the exercise price. Thus, their networks have only two inputs, S/X and time to maturity, one output, the ratio of the call price to X. It is reasonable to doubt whether such a network can capture all the option price variations.
Another research on option pricing using ANN has been done by Qi and Maddala (1995). Unlike Hutchinson, Lo and Poggio (1994), Qi and Maddala use variables that are believed to be important in determining option prices as network inputs, and use option prices as network output. The input variables are the underlying asset price $S$, exercise price $X$, risk-free rate $r$, time-to-maturity $T$, and open interest $V$. Such a network provides superior performance to the Black-Scholes formula both in- and out-of-sample for S&P 500 index call options, and the results are better than those reported by Hutchinson, Lo and Poggio (1994). Moreover, by analyzing the network weights, Qi and Maddala find that the economic implications of the neural network model are consistent with the option price properties and the open interest is found to be important in determining option prices.

However, it is still not clear whether the improvement reported by Qi and Maddala (1995) is due to the flexible functional relationship that is superior to the Black-Scholes formula, or due to the more appropriate set of pricing variables which exclude the volatility and include the open interest. In the present thesis, I investigate the first issue by considering an ANN that uses exactly the same input variable as the Black-Scholes formula and address the second issue by removing the volatility and adding the open interest to the previous ANN. Residuals of both the ANN and the Black-Scholes formula are also analyzed to gain some insight into the source of improvement. Moreover, I analyze the model implications by some additional methods other than pseudo weights, such as sum of input weights, and sensitivity analysis.
3.2. Data

I use the same data as in Qi and Maddala (1995). Standard & Poor's 500 index is chosen to be the underlying security. On average, there are about thirty S&P 500 index call options of different strike prices and expiration dates on each trading day. The whole data set consists of 1107 S&P 500 index calls spanning Dec. 1, 1994 through Jan. 19, 1995. The S&P 500 index closing prices on each day are used as the current security prices. Time to maturity, exercise price, and open interest are obtained directly from Index Options. The yield of the three month treasury bill with an expiration date closest to that of the index option is used as the short-term interest rate. All data are collected from the Wall Street Journal.

The Black-Scholes option price is calculated from underlying asset price, exercise price, interest rate, time to maturity and volatility using the Black-Scholes formula. The volatility, $\sigma$, is calculated from historical data, specifically, for each trading day, $\sigma$ is calculated as the annualized standard deviation of the recent 106 daily continuously compounded S&P 500 returns. Ceteris paribus, more data generally lead to more accuracy. However, $\sigma$ does change over time and data that are too old may not be relevant for predicting the future. The choice of the recent 106 trading days is a compromise, which is in the range suggested by Hull (1993).

To yield sufficient training and testing patterns to train the network and to evaluate the performance of an ANN effectively, it is important to have enough data. The amount of data required for training a network depends heavily on the network structure, the training method, and the problem being addressed. There are no rigid rules. I use roughly
four-fifth of the whole sample (or 886 patterns) as the training set, and use the remaining data (or 223 patterns) as the test set.

3.3. ANN paradigm

Due to the limitation of the capacity of a two-layer ANN (Hertz, Grogh and Palmer, 1991), networks with at least one middle layer are often used. Cybenko (1988) proves that ANN with at most two hidden layers can approximate a particular set of function $y = f(X)$ with arbitrary accuracy given enough units per layer. It has also been proved that only one hidden layer is enough to approximate any continuous function (Cybenko, 1989; Hornik, Stinchcombe and White, 1989). Many empirical studies, such as Collins, Ghosh and Scofield (1988), Dutta and Shekhar (1988), Salchenberger, Cinar and Lash (1992) (to name a few) have confirmed this. The correctness of these results, however, hinge on the appropriate number of hidden units.

The choice of $k$, the number of hidden units, represents a compromise. If $k$ is too small, an ANN may not approximate $y = f(X)$ at the desired accuracy. However, if $k$ is too large, an ANN may overfit and cannot generalize (or forecast) out of sample. A useful method is cross-validation, by which the number of middle layer units is selected to optimize out-of-sample performance (White, 1990).

In the present thesis, I choose the number of middle layer units based on both the in- and out-of-sample performance of alternative ANN structures. The four performance measures are: root mean squared error (RMSE), mean absolute error (MAE), Pearson correlation coefficient ($\rho$), and coefficient of determination ($R^2$).
Figure 5 shows how the in- and out-of-sample RMSE changes as the number of iterations increases for ANNs with three, five, eight and ten units in the middle layer. Though the ANN with ten middle layer units has the smallest RMSE within sample, its out-of-sample RMSE is the largest and is getting larger as the number of iterations increases, which suggests the overfit of the ANN with ten middle layer units. Similar result holds for the ANN with eight middle layer units.

Compared to the ANN with three middle layer units, the ANN with five middle layer units has similar RMSE in sample, but has smaller RMSE out of sample, thus is the best among the alternative structures. The ANN with five middle layer units is also the best based on the other three performance measures, MAE (Figure 6), ρ (Figure 7), and $R^2$ (Figure 8).

Therefore, a three-layer feedforward ANN with five middle layer units has been used in this application. The input layer consists of important factors in determining option prices, such as, current asset price ($S$), interest rate ($r$), time to maturity ($T$), exercise price ($X$), and open interest ($V$). The output layer has only one unit, the call price ($C$).

3.4. Model fitting

To ensure the reliability of the empirical findings, I use a five-fold cross-validation resampling design suggested by Efron and Tibshirani (1993). The data are divided into five roughly equal-sized parts, and for each of the five parts, the model is fitted to the other four parts of the data, and the pricing error is calculated for the chosen part of the data. The results of all five subsamples are reported.
In order to find out whether the superior performance reported in Qi and Maddala (1995) is due to the flexible functional relationship of ANN that is superior to the Black-Scholes formula, or due to the choice of more appropriate factors, I report three sets of results in Table 1. The first set is the results of the Black-Scholes formula. The second set is the results of an ANN with exactly the same input variables as in the Black-Scholes formula, i.e., S, r, T, X, and σ. The third set is the results of an ANN with S, r, T, X, and V as the input variables. The results are reported for all four performance measures and for all five subsamples.

The performance of the ANN with the same input variables as in the Black-Scholes formula is unanimously better than that of the Black-Scholes formula based on all four performance measures and for all five subsamples. Across all five subsamples, RMSE and MAE of the ANN are smaller than those of the Black-Scholes formula, and the coefficient of determination R^2 and the Pearson correlation coefficient ρ are larger for the ANN than for the Black-Scholes formula. The results suggest that the option pricing accuracy might be improved purely due to the flexible functional form that can be achieved by an ANN, and that the functional form of the Black-Scholes formula might be suboptimal.

Moreover, as I substitute volatility σ with open interest V in the previous ANN, the performance has been improved unanimously again based on all four performance measures and for all five subsamples. Across all five subsamples, RMSE and MAE of the ANN with the open interest input are smaller than those of the ANN with the volatility input, and the R^2 and ρ are larger for the former. The evidence thus suggests that the option pricing accuracy might be improved further by using a better set of input variables.
than the Black-Scholes option pricing variables. More specifically, open interest is more useful in pricing S&P 500 index call options than volatility.

3.5. Model testing

Given the universal approximation property of ANNs and the many unrealistic restrictions of the Black-Scholes model, the improvement in model fitting by the ANN is not surprising. However, whether the ANN option pricing model is useful or not, depends on its out-of-sample performance. What model testing does is to feed patterns of each retained test subsample into the corresponding fitted ANN option pricing model and to evaluate its pricing accuracy. The results of alternative models based on all four performance measures and for all five test subsamples are reported in Table 2.

Unlike the in-sample results, the out-of-sample results are a little mixed, at least at the first glance. Consider first the ANN with the same input variables as the Black-Scholes formula. For test subsample 1, the performance of the ANN is better than the Black-Scholes formula based on all four performance measures, which suggests that the functional form implied in the ANN is better than that of the Black-Scholes formula. However, the results are the opposite for test subsamples 4 and 5. For test subsamples 2 and 3, though the ANN has smaller MAE, it has larger RMSE and smaller ρ than the Black-Scholes formula. Moreover, the average MAE across five test subsamples is better than that of the Black-Scholes formula, but the average RMSE, R², and ρ across five test subsamples are worse than those of the Black-Scholes formula. Therefore, it is unclear out
of sample whether the improvement in pricing accuracy by the ANN is purely due to the improvement in the functional relationship among Black-Scholes option pricing variables.

Now consider the ANN with input variables $S$, $r$, $T$, $X$ and $V$. There is unanimous improvement in the performance based on all four performance measures for test subsamples 1, 2 and 3. Though the results for test subsamples 4 and 5 are a little mixed, the average, minimum and maximum performance of the ANN are all better than those of the Black-Scholes formula for all four performance measures. Moreover, the performance of the ANN with open interest as one input variable is better than that of the ANN with volatility as one input variable based on the average, minimum and maximum of all four performance measures across five test subsamples. The evidence thus clearly suggests that open interest is more relevant in option pricing than volatility.

This is a surprising result. Since options are sometimes purchased as a hedge against the risk, the higher the volatility of the underlying asset, the more valuable are the options. In order to find out whether volatility can marginally improve the pricing accuracy, consider an ANN with six input variables, $S$, $r$, $T$, $X$, $V$ and $\sigma$. The results for subsample 5 are shown in Table 3.

Based on all four performance measures, both the in- and out-of-sample performance of the ANN with six input variables are worse than those of the ANN without volatility $\sigma$, and the out-of-sample performance is even worse. The evidence suggests the overfit of the ANN with six input variables. Moreover, as the number of iterations increases from 2000 to 4000, the in-sample performance of such an ANN has been improved unanimously, but the out-of-sample performance becomes dramatically
worse based on all four performance measures, which provides further evidence of the overfit of the ANN with six input variables. The results for other subsamples are similar and thus are not reported here. Therefore, to our surprise, volatility $\sigma$ does not improve the pricing accuracy even marginally.

3.6. Residual analysis

To gain some insights into the difference in the performance of the Black-Scholes option pricing formula and its ANN counterpart, the residuals of both models are analyzed. Table 4 reports the first order residual autocorrelation, the percentage of positive residuals and the mean of residuals for all subsamples. Table 5 reports the results of White's heteroskedasticity test for all subsamples.

As can be seen in Table 4, residuals are all positively autocorrelated, and the Durbin-Watson test statistics are much smaller for the Black-Scholes formula than for ANN, thus the residual autocorrelation is much higher for the Black-Scholes formula than for the ANN both in and out of sample for all five subsamples. Therefore, the ANN outperforms the Black-Scholes formula in reducing the residual autocorrelations.

The Black-Scholes model has much higher percentages of positive residuals than the ANN both in and out of sample, and across all five subsamples, and the percentages are almost all well over 50%, which provides evidence that the Black-Scholes formula systematically underprices call options. This is consistent with the analysis in Section 3.1, where it is pointed out that the Black-Scholes formula which gives European option prices tends to underestimate the American option prices. For the ANN, however, the
percentages are below 50% both in-sample and out-of-sample, which suggests that the
ANN may slightly overprice call options. Moreover, the means of the ANN residuals are
mostly small and negative while the means of the residuals of the Black-Scholes formula
are mostly large and positive, which suggests that while the ANN slightly overprices S&P
500 call options, the Black-Scholes formula significantly underprices S&P 500 index call
options. Therefore, the ANN outperforms the Black-Scholes formula in reducing the
extent of underpricing.

Table 5 reports the results of White’s heteroscedasticity test. Most of the
regression coefficients are significant for all five subsamples, and there is clear evidence of
heteroskedasticity from White’s test for both the ANN and the Black-Scholes formula.
However, the average $R^2$ is 0.1804 for the ANN, but is 0.3217 for the Black-Scholes
formula, thus heteroscedasticity is less severe in the ANN than in the Black-Scholes
model. Therefore, the ANN outperforms the Black-Scholes formula in reducing the
residual heteroscedasticity.

In option pricing literature, different pricing errors are often found for options that
are in, at, and out of the money (the so called moneyness bias) and for options that have
different time to expire (the so called timeliness bias). Since no options in my sample are at
the money, the means of the residuals for in and out of the money options are calculated.
The correlations between the residuals and the time to expire are also computed. The
results of the ANN and the Black-Scholes formula for all subsamples are listed in Table 6.

For the Black-Scholes formula, the MAEs are much larger for the in-the-money
options than for the out-of-the-money options for all five subsamples and for all five test
subsamples, which is consistent with the moneyness bias found in the option pricing literature. The average in-sample MAE of the ANN across five subsamples is 0.6797 for in-the-money options and 0.2831 for out-of-the-money options, while the numbers for the Black-Scholes formula are 1.1092 and 0.5660 respectively. The average out-of-sample MAE of the ANN across five test subsamples is 0.9703 for in-the-money options and 0.3462 for out-of-the-money options, whereas the numbers for the Black-Scholes formula are 1.1823 and 0.5511 respectively. Therefore, the ANN outperforms the Black-Scholes formula in reducing the moneyness bias.

As to the timeliness bias, the correlations between the absolute values of the residuals and the time to expire are all positive for the Black-Scholes formula both in and out of sample, which is consistent with those found in the option pricing literature. The average in-sample correlation is 0.0501 for the ANN and 0.3314 for the Black-Scholes formula, and the average out-of-sample correlation is 0.0323 for the ANN and 0.2898 for the Black-Scholes formula. Therefore, the ANN outperforms the Black-Scholes formula in reducing the timeliness bias.

3.7. Economic implications

In the present thesis, pseudo weights, sum of input weights, and sensitivity analysis have been used to investigate the significance of each input variable on the output.
3.7.1. Factor significance through network weights

Table 7 reports the pseudo weights for each of the five training subsamples. Since ANNs are unidentifiable in general, the error surfaces are multimodal, and the initial weights are randomly generated, the network weights may differ from subsample to subsample. However, for all subsamples, X has large negative pseudo weights which is consistent with the property of call options: the higher the exercise price, the lower the value of a call option. S has positive weights for all five subsamples, which is also consistent with the property of call prices: the higher the current asset price, the higher the call price. Furthermore, T has positive weights across all five subsamples, which is consistent with the property of call options: the longer the time to expire, the higher the call value.

For r, the short-term interest rate, the signs are mixed across the five subsamples, which may reflect the mixed effects of the short-term interest rate on call option prices. Theoretically, the higher the r in the economy, the higher the expected growth rate of the asset price, thus the higher the call price. However, the higher the r, the lower the present value of any future cash flows to be received by a call option holder, the lower the call price. Ceteris paribus, the first effect always dominates the second, so the higher the r, the higher the call price. In real world, when r rises, stock prices tend to fall, the net effect of r might therefore be the opposite. So the mixed signs of pseudo weights have actually reflected the ambiguous effect of r on call option prices.

Open interest also has mixed signs. Moreover, among five input variables, V has the second largest weight in magnitude (only next to the exercise price) for four out of the
five subsamples, which shows that open interest may indeed be an important factor that affects option prices. Since $V$ is the sum of all the long positions or all the short positions and represents the equilibrium quantity supplied or demanded, the negative pseudo weights in subsamples 1 and 2 are presumably indications of shifts in supply as shown in Figure 9(a), while the positive pseudo weights in subsamples 3, 4 and 5 are presumably indications of shifts in demand as shown in Figure 9(b). Since supply and demand might shift one at a time, or simultaneously, signs may be mixed.

The absolute values of the pseudo weights indicate the magnitudes of the effect of an input variable on call option prices. The rank order of the average magnitudes per absolute pseudo weights is $X > T > S > V > r$.

Table 8 reports the sum of absolute input weights for each of the five subsamples. Among the five input variables, exercise price $X$ has the largest sum of weights, thus is the most significant factor in determining option prices by sum of input weights, the same as by pseudo weights. However, the rank order of magnitudes according to sum of input weights is $X > V > T > r > S$, which is different from that by pseudo weights. Whether the pseudo weights or the sum of input weights gives a more accurate rank order will be shown in Section 3.7.2.

3.7.2. Sensitivity analysis

The results of sensitivity analysis for subsample 3 are shown in Figure 10. Each plot shows how the call option price changes with two input variables with the rest of the three input variables being held at their mean values. In each plot, the call prices are shown
along the direction of the Z-axis, and the two inputs are shown along the direction of the X- and Y-axis respectively.

The positive or negative effect of an input variable on the call option price is reflected in the positive or negative slope of the surface along the direction of the input variable. Plots (a), (b), (c), and (d) in Figure 10 clearly show that stock price $S$ is positively related to call option price. Plots (b), (e), (h), and (i) in Figure 10 clearly show that time to maturity $T$ is positively related to call option price. The negative effect of $X$ on call option price can be clearly seen in plots (c), (f), (h), and (j) in Figure 10. The negative relationship between $r$ and call option prices is shown in plots (a), (e), (f), and (g) of Figure 10. Finally, $V$ is positively related to call option price from plots (d), (g), (i), and (j) in Figure 10. Thus the signs of the effect in sensitivity analysis are consistent with the signs of pseudo weights in Table 7.

Moreover, the magnitude of the effect of an input variable on the call option price is also consistent with that of the pseudo weights. The magnitude can be visualized by the steepness of the surface along the direction of an input variable. For all plots that contains exercise price $X$ in Figure 10, the slope of the surface is steeper in the direction of $X$ than in the other direction, thus exercise price has the largest effect on call option prices; For all plots that contain time to maturity $T$ (except plot (h) which also contains $X$), the surface is steeper in the direction of $T$ than in the other direction, thus time to expire has the second largest effect on call option prices; For all plots that contain current stock price $S$ (except plot (b) which contains $X$ and plot (c) which contains $T$), the slope is larger in the direction of $S$ than in the other direction, thus current stock price has the third largest
effect on call option prices. Finally, for the two remaining input variables interest rate $r$ and open interest $V$, the surface in plot (g) is steeper along the direction of $V$ than along the direction of $r$, thus short term interest rate has the smallest effect on call option price. The rank order of magnitudes from sensitivity analysis, $X > T > S > V > r$, is consistent with the rank order of the absolute pseudo weight in Table 7. Therefore, pseudo weights are better indicators of factor significance both in magnitudes and signs than sum of input weights.

3.8. Summary

In this chapter, the ANN has been applied to pricing options and the performance has been compared to the traditional Black-Scholes formula. Section 3.1 introduces the background of option pricing and reviews the literature regarding ANNs and option pricing. Section 3.2 describes the data. In Section 3.3, the ANN structure is chosen based on both in- and out-of-sample performance. The following two sections reports the in- and out-of-sample performance of the ANN which uses five input variables $S$, $r$, $X$, $T$ and $V$ to price S&P 500 index call options, the results are also compared to those of the Black-Scholes formula and an ANN which uses Black-Scholes option pricing variables, $S$, $r$, $T$, $X$, $\sigma$ as inputs. Section 3.6 analyzes the residuals of the ANN and the Black-Scholes formula. Finally, Section 3.7 investigates the economic implications of the ANN option pricing model using both network weights and sensitivity analysis.
CHAPTER 4

STOCK MARKET PREDICTION USING ANN

4.1. Introduction

Accurate stock market prediction is problematic and many of the models that have been developed perform poorly. Such models are not justified in a truly efficient market where prices instantaneously reflect all available information. Any predictable profit opportunity will be exploited long before an analyst has completed his or her calculation. The question remains why so many individuals and financial institutions continue to make predictions and to take positions against the market.

Traditional finance theory based on the Capital Asset Pricing Model and its generalization, the Arbitrage Pricing Theory, supplies part of the answer. There may be a compensation for the systematic risk associated with a position in the stock market. Apart from a risk free return, which is the same for all asset positions, the theory states that the higher the risk, the higher the expected return on an investment. Analysis of the stock price behavior and the underlying economic factors certainly provides insight in the systematics of the price movements and the associated risk. In this way part of the return,
namely the risk premium, may be predictable. To predict stock returns, two things need to be done. First, the underlying economic factors need to be identified. Second, the underlying functional relationship between stock return and the economic factors needs to be discovered.

In finance literature, several economic factors, such as expected inflation, industrial production, term spread, default spread, dividend yield have been found to be significantly linearly related to stock returns, see Chen, Roll and Ross (1986), Keim and Stambaugh (1986), Fama and French (1989) for example. Practitioners also try to use various macroeconomic variables to forecast stock market. For instance, Leeb and Conrad (1995) use inflation rate, unemployment insurance claim rate, real interest rate, money growth rate, and price-to-earning ratio to forecast stock market movements. It is reported that the trading strategy of buying stocks whenever their linear regression predicts at least 10% gain and selling when the regression forecasts at least 1.5% loss has reaped an average annual gain of 19.6% since 1954, nearly three times that of a simple buy-and-hold strategy. Yet, the book reports no significance tests on the regression coefficients, nor the goodness of in-sample fit and the out-of-sample forecasting accuracy.

All above research uses linear models. Recent studies into the time series properties of various financial variables by Hinich and Patterson (1985), Larrain (1991) and Peters (1991), to name a few, however, have indicated the presence of structural nonlinear dynamics. The underperformance of the traditional linear models of stock price may be due to misspecification, rather than market efficiency. Generalized nonlinear nonparametric econometric methods, such as artificial neural networks, do not force a
linear model onto a possibly nonlinear relationship, thus enable one to investigate the relationship between economic factors and the stock market from a new perspective.

4.2. Data

The inflation rate, INF, is calculated as the percentage change in Consumer Price Index, CPI, for the previous twelve months. The rate of change in unemployment insurance claims, UIC, is calculated as the percentage change in initial unemployment insurance claims in the last twelve months. The price-to-earning ratio, P/E, is calculated from S&P 500 stock composite, and is based on seasonally adjusted annual rate of earnings before 1971 and on four quarter total of earnings available at the time of computation since 1971. Following Leeb and Conrad (1995), the real interest rate, r, is calculated as the difference between the annual yield on twenty-year AAA-rated corporate bonds and the previous twelve-month’s rate of change in the all commodity producer price index. The money growth rate, m, is measured as the difference between the average monthly growth rate of m1 over the last twelve months and the average monthly growth rate of m1 over the whole sample period, which goes from January 1959 to June 1995.

The five macroeconomic variables described above are used to predict the returns on S&P 500 index at a horizon of twelve months. All data are obtained from CITIBASE.
4.3. Linear regression

The relationship between the returns on S&P 500 index and the above five macroeconomic variables is first investigated by a linear regression. Four benchmarks, RMSE, Pearson correlation coefficient (ρ), Theil's U statistics, and the coefficient of determination R², are used in comparing the performance of alternative models.

The estimation is performed over the period January 1960-June 1994. The multiple regression of stock returns (R) on inflation rate (INF), initial unemployment insurance claim rate (UIC), price-to-earning ratio (P/E), real interest rate (r) and money growth rate (m) gives:

\[ R = 14.37 - 0.29 \text{INF} + 0.22 \text{UIC} - 0.66 \text{P/E} + 0.85r - 0.20m, R^2 = 0.39 \]

\[ (6.16) \ (-4.88) \ (7.17) \ (-5.28) \ (5.00) \ (-1.08) \]  

(4.1)

All the regression coefficients except for m are highly significant and the signs are consistent with those given by Leeb and Conrad, which indicates that these forecasting variables indeed contain information about the future stock returns. Pearson's ρ equals 0.6279, designating that the fitted returns coincide with the actual returns to certain extent. Theil's U of 0.77 indicates that the RMSE of the linear regression, 10.85, is about 25% less than that of a random walk.

R² of 0.39 is much higher than the values of 0.00 to 0.18 reported in Fama and French (1989), where stock returns are regressed on term spread, dividend yield or default spread. It is well above the adjusted R² of -0.003 to 0.020 in Keim and Stambaugh (1986), where stock returns are regressed on their three ex ante variables. It is also much higher
than the $R^2$ of -0.03 to 0.22 in Fama and French (1988), where dividend yields are used to predict stock returns.

The linear regression shows that the macroeconomic variables used by practitioners can also explain a significant portion of stock return variations, in addition to those found in the finance literature. Next it is shown that using the same macroeconomic variables as in the linear regression, an artificial neural network method can improve the goodness of fit by relaxing the linear restriction and allowing non-linear relationships among variables.

4.4. Artificial neural networks

A multilayer feed-forward artificial neural network with five inputs, five middle layer processing elements and one output is used. The five inputs are the explanatory variables used in the linear regression: INF, UIC, P/E, r and m, and the output is the dependent variable used in the linear regression: the return on S&P 500 index in the succeeding twelve months. Table 9 reports the pseudo weights, sum of input weights and the four performance measures.

RMSE is now only 7.90, dropping by about 27%, compared to the linear regression. Pearson’s $\rho$ rises from 0.63 to 0.83, Theil’s $U$ drops from 0.77 to 0.56, and $R^2$ increases from 0.39 to 0.68. The unanimous improvement in all four performance measures suggests that the functional relationship between the stock returns and the five macroeconomic variables is unlikely to be linear.
Figure 11 plots the actual and fitted returns by the linear regression, the ANN and the random walk. Consistent with the four performance measures, although both the linear regression and the ANN forecasts look generally close to the actual returns, the ANN forecast is closer.

4.5. Residual analysis

The residuals of both the linear regression and the ANN are analyzed. The first order autocorrelation is 0.8812 for the linear regression, 0.7532 for the ANN, and the Durbin-Watson test statistic is 0.2370 for the linear regression and 0.4896 for the ANN. The lower autocorrelation of the ANN may be due to the avoidance of the linear restrictions among variables. However, the high autocorrelation of both the linear regression and the ANN residuals may indicate either the omission of important predictors or the non-stationarity of the data. In the later case, a linear regression in first difference is necessary.

The percentages of positive residuals are very close, 54.59% for the linear regression and 54.11% for the ANN. The mean of residuals is 0 for the linear regression and 0.5791 for the ANN.

White's tests are conducted to detect the heteroscedasticity for both the linear regression and the ANN models. The squared residuals are regressed on the five predicting variables and their square values. The coefficient of determination $R^2$ is 0.0651 for the ANN and 0.1005 for the linear regression, indicating that heteroscedasticity is not a big problem in both models.
4.6. Economic implications

From Table 9, the signs of pseudo weights for UIC, P/E and r are consistent with those in the linear regression. The money growth rate, m, is not significant at 90% confidence interval in the linear regression and the sign is opposite here, which may suggest that the effect of money growth on the stock market is either ambiguous or smeared by other variables, because m is likely to be correlated with other independent variable, such as the inflation rate and the real interest rate. Surprisingly, though inflation rate is shown to be significantly negatively related to stock return in the linear regression, it has a positive pseudo weight in the ANN. The puzzle of different signs in inflation rate and money growth rate will be explained later in sensitivity analysis.

According to the pseudo weights, the rank order of the magnitudes of factor significance is \( r > UIC > P/E > m > INF \). The sum of input weights are also reported in Table 9. However, it indicates a different rank order, \( r > INF > m > P/E > UIC \). Interest rate still bears the largest significance by sum of input weights, the same as by pseudo weights. But for the other four variables, the rank order is reversed. Further evidence of factor significance is provided in the following sensitivity analysis.

Figure 12 reports the sensitivity analysis for the ANN stock prediction. Each plot shows how returns on S&P 500 index over the subsequent twelve month change with two predicting variables with the rest of the three predicting variables being held at their means.
The immediate conclusion one can draw from the ten pair-wise plots in Figure 12 is that there is substantial nonlinearity between the stock return and the economic factors. For example, in plot (h), at lower level of interest rate, stock return monotonically decreases with the P/E ratio. However, at higher interest rate level, stock return first increases then decreases with the P/E ratio. Similarly, in plot (b), stock return monotonically decreases with P/E ratio at higher level of inflation rate INF, but increases up to certain level then decreases with P/E ratio at lower level of inflation rate. Therefore the improvement in the performance of ANN over linear regression is in fact due to its ability to handle the complex nonlinear relationships among variables.

The complex nonlinear relationship between stock return and economic factors invalidates the economics implications from pseudo weights and sum of input weights. Now the stock return may be positively related to an economic factor within a particular range, but may be negative related to the same economic factor in another range. Similarly, the stock return may be very responsive to the change in an economic variable in certain range, but less responsive to the same economic factor in other ranges. For example, in plot (f), at lower level of real interest rate, unemployment seems to have little effect on stock returns. However, as r increases, stock return becomes positively related to unemployment insurance claim rate, but this positive effect diminishes as r increases further toward its maximum. The pseudo weights, sum of input weights thus are no longer relevant in the interpretation of economic implications for a complex model with substantial nonlinear. Nevertheless, they do help capturing some of the dominant factors, such as the interest rate (r). In all four plots that contain r, there is larger variation in stock
return along the direction of $r$ than in the other direction, and in most ranges, $r$ is positively related to future stock return. For inflation rate (INF) and money growth rate (m), however, their relationships to stock return are not monotone in most ranges, and this may partly explain why the signs of their pseudo weights are different from the signs of their linear regression coefficients.

Similarly, in the ANN option pricing model, the contours in plot (d) and (i) of Figure 10 show that $V$ is not monotonically related to the option price. This may explain why the signs of the pseudo weights for subsamples 1 and 2 are different from that for subsamples 3, 4 and 5.

4.7. Out-of-sample forecasts

Very few of the previous studies have conducted out-of-sample forecasts. However, a model of stock returns may have no practical use if it cannot provide out-of-sample forecast of certain accuracy, no matter how well the model fits the data in sample. In this thesis, returns are recursively forecast for one year horizon. Each forecast is from a linear regression or ANN estimated with overlapping one-year returns that begin in January 1960 and end in December of the preceding year. For example, to forecast the first twelve overlapping one-year returns (January 1980 to December 1980, February 1980 to January 1981, ..... December 1980 to November 1981), the coefficients or weights estimated with the 240 overlapping one-year returns from January 1960 to December 1979 are used. The out-of-sample forecasts are evaluated based on three categories of measures: RMSE based measures, profit measures, and direction accuracy.
4.7.1. RMSE based measures

Table 10 compares the out-of-sample forecasts of the linear regression, ANN and random walk models for the forecasting period January 1980-June 1994. Each random walk forecast is just the average monthly overlapping one-year return during the period beginning January 1960 and ending in the month preceding the out-of-sample forecast. For example, the random walk one-year return forecast for January 1980 to December 1980 is the average monthly overlapping one-year return from January 1960 to December 1979.

For both the linear regression and the ANN, all three out-of-sample performance measures tend to be worse than those in sample. However, Theil's U is less than one for both the linear regression and the ANN, which suggests that both the linear regression and the ANN show some improvement in forecasting accuracy as compared to the random walk model, especially for the ANN, the improvement is more significant. As can be seen in Figure 13, the ANN forecasts are generally more close to the actual returns than linear regression forecasts. This is also reflected in the Pearson correlation coefficients. $p$ is 0.58 for the ANN, 0.32 for the linear regression, and -0.08 for the random walk model. Therefore, the out-of-sample forecasts seem to provide further evidence of the nonlinear relationship between the macroeconomics factors and the stock market returns.

To ascertain whether the linear regression significantly improves upon the random walk, whether the ANN significantly improves upon the linear regression, and whether the ANN significantly improves upon the random walk, t test is conducted to test the null hypothesis that there is no difference in the squared forecasting errors of two competing
forecasts for the random walk and the linear regression, the linear regression and the ANN, the random walk and the ANN. The results are listed in Table 11. The t-statistics show that although all three differences of the squared forecasting errors are positive, none of them are statistically significant. Thus, it is not clear whether the linear regression or the artificial neural network can statistically significantly improve the forecasting accuracy of a random walk in terms of squared forecasting errors.

However, Leitch and Tanner (1991) find that the conventional criteria based upon some measure of the size of the forecast error, such as RMSE, Theil’s U statistics, have no systematic relationship to profits. Furthermore, Satchell and Timmermann (1992) show that the RMSE is not particularly suited for assessment of the economic value of predictions of nonlinear processes where the predicted value and the prediction error typically are not independent. Therefore, given the evidence of the nonlinear relationship between the economic factors and the stock returns, a more appropriate measure of forecast accuracy is profitability, not the size of the forecast error or its squared value.

4.7.2. Profit measures

A simple trading rule is used to generate profit measures for alternative forecasting systems the linear regression and the ANN. Without loss of generality, holding periods are assumed to be one year and new positions, based upon the updated forecasts for each year, are assumed to be taken at the January and closed at the December. Since the average rate of return on S&P 500 index is used as the profit measure, the size of the position does not matter. The trading strategy is to go “long” whenever the model projects
a gain of at least 10%, go “short” when the model forecasts losses of at least 10%, and do not take a position if the model predicts a return greater than -10% but less than 10%. For the 1980-1994 period, the strategy has realized an average annual gain of 17.26% using ANN forecasts and 8.77% based on linear regression forecasts. In comparison, a naive trading strategy of buying each year without considering any forecasts has an average annual gain of 8.86%.

The average annual gains are also calculated with new positions taken at the rest of the eleven months and closed in one year. The results are reported in Table 12. For seven out of twelve holding period starting months, the average annual gain based on the linear forecasts is higher than those generated by the naive trading strategy and the average annual gain across all starting months is about 1.77% higher on average. More importantly, the trading profits based on the ANN forecasts are higher than those based on the linear regression forecasts for nine out of twelve starting months, and the average annual returns across all starting months is about 3% higher than the linear regression forecasts. Therefore, the artificial neural network is superior to the linear regression and the naive buying strategy in generating trading profits.

4.7.3. Direction accuracy

In investment analysis, sometimes it is more important to forecast the direction change or the turning point than to forecast the exact values of the future price or the future return. To evaluate the ability to forecast turning point of alternative models, I calculate the direction accuracy in two different ways: in levels and in returns.
First I consider whether the models are able to forecast the direction change of S&P 500 index levels. If a model can accurately forecast whether the S&P 500 index will go up or down in the future, then the market timing ability of the model can be easily turned into profits. Without loss of generality, I still use one year forecasting horizon. Let $y_i$ be the actual value of future S&P 500 index, $\hat{y}_i$ be the forecasted value of future S&P 500 index, for a sample of size $N$, the direction accuracy is calculated as:

$$DA = \frac{1}{N} \sum a_i,$$

(4.2)

where $a_i = \begin{cases} 1 & \text{if } (y_{i+1} - y_i)(\hat{y}_{i+1} - y_i) > 0, \\ 0 & \text{otherwise.} \end{cases}$

The results show that while the linear regression can correctly forecast 72.41% of the direction changes in the level of S&P 500 index, the direction accuracy of the ANN and random walk are very close, they are 75.29% and 75.86% respectively. Therefore, ANN and random walk slightly outperform linear regression in forecasting the direction changes of S&P 500 index levels.

However, since there is a general upward trend in the levels of S&P 500 index, it is very easy to get a higher than 50% direction accuracy in levels. In fact, from January 1980 to June 1994, 75.86% of the monthly one-year changes in S&P 500 index levels are positive. Therefore a naive strategy of upward prediction at the beginning of each investment horizon, such as the random walk, will yield an accuracy of 75.86%, which is just the percentage of positive monthly one-year changes in S&P 500 index. Therefore it would be more interesting and challenging to forecast the down turn of the S&P 500
index, or to forecast the direction changes in returns on S&P 500 index rather than the direction changes in levels.

Out of the 42 down turns of the actual monthly one-year changes in the S&P 500 index across the forecasting period January 1980 - June 1994, ANN can correctly predict 25, linear regression can predict 20, and random walk model can predict none. Therefore ANN is superior to linear regression and random walk models in that it can accurately predict 11.9% more of the down turns in the S&P 500 index than the linear regression, and 59.52% more than the random walk model.

As to the monthly one-year returns on the S&P 500 index, the direction accuracy is 53.76% for ANN, 49.13% for the linear regression, and 52.02% for the random walk model. Therefore, ANN is also superior to the linear regression and the random walk models in forecasting the direction changes of returns.

4.8. Summary

In this chapter, ANN has been applied to predicting stock returns, and the performance has been compared to the linear regression and random walk models. Section 4.1 introduces the background of stock return prediction and reviews the literature. Section 4.2 describes the data. Section 4.3 reports the results of linear regression. Section 4.4 reports the model fitting result of ANN. Residuals of linear regression and ANN models are analyzed in Section 4.5. Section 4.6 investigates the economic implications of ANN option pricing model using both network weights and sensitivity analysis. Section
4.7 compares out-of-sample performance of linear regression, ANN and random walk based on RMSE, profit measure and direction accuracy.
CHAPTER 5

SUMMARY AND CONCLUSIONS

Artificial neural networks, a generalized nonlinear nonparametric econometric method has been shown to perform very well in the two applications outlined in this thesis.

In the first application, the ANN has been applied to price S&P 500 index call options. Using current asset price, interest rate, time to maturity, exercise price, and open interest as inputs and call price as output, the ANN with five middle layer units outperforms the Black-Scholes formula. The ANN has smaller average squared errors, smaller mean absolute errors, larger $R^2$ and larger $\rho$ than the Black-Scholes formula both in and out of sample. The residual analysis shows that the ANN outperforms the Black-Scholes formula in reducing the residual autocorrelations and the extent of underpricing, and the residual heteroscedasticity is less severe in the ANN than in the Black-Scholes formula. The ANN also outperforms the Black-Scholes formula in reducing the moneyness bias and the timeliness bias which are often found in the option pricing literature. The economic implications of the ANN are analyzed through pseudo weights, sum of input weights and sensitivity analysis. The pseudo weights indicate that the
economic implications of the ANN model are consistent with the properties of call option prices, i.e., the lower the exercise price, or the higher the current asset price, or the longer the time to maturity, the higher the call price. Moreover, ANN can also capture the ambiguous effect of risk-free interest rate on option prices. Open interest has long been ignored in the option pricing literature, and is found here to be an important factor in determining option prices. The sensitivity analysis shows both the sign and magnitude of the effect of a pricing variable on option price, and is consistent with the results of pseudo weights. This application provides an alternative to the traditional Black-Scholes formula that does not require the many unrealistic assumptions required by the Black-Scholes formula and gives much superior performance. The result is very important for investors and portfolio managers who always have to judge whether a given option is worth buying or selling.

In the second application, ANN has been applied to forecast returns on S&P 500 index using five economic variables, the inflation rate, unemployment insurance claim rate, price-to-earning ratio, real interest rate and money growth rate. The linear regression gives better fit than those reported in the existing return forecasting literature. It is then shown that the goodness of fit can be largely improved using an ANN by not imposing linear restrictions. The first order residual autocorrelation is smaller and the residual heteroscedasticity is less severe for the ANN than for the linear regression model. The substantial nonlinearity shown in the sensitivity analysis invalidates the economic implications from the pseudo weights and sum of input weights. The recursively computed neural network nonlinear forecasts can improve the average annual gains generated by a
simple profit trading rule, though they do not statistically significantly improve the linear or the random walk forecasts in terms of conventional criteria based on squared forecasting errors. Moreover, the ANN outperforms the linear regression and random walk in predicting the direction change of returns on S&P 500 index, and more importantly, the ANN is much better than the linear regression and the random walk models in predicting the down turns of the S&P 500 index.
LIST OF REFERENCES


### APPENDIX A

### TABLES

<table>
<thead>
<tr>
<th>subsample</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Black-Scholes formula</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSE</td>
<td>.8861</td>
<td>1.1185</td>
<td>1.2007</td>
<td>1.2372</td>
<td>1.2356</td>
</tr>
<tr>
<td>MAE</td>
<td>.6034</td>
<td>.7277</td>
<td>.7901</td>
<td>.8346</td>
<td>.8268</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.9965</td>
<td>.9944</td>
<td>.9936</td>
<td>.9935</td>
<td>.9938</td>
</tr>
<tr>
<td>$\rho$</td>
<td>.9984</td>
<td>.9974</td>
<td>.9972</td>
<td>.9972</td>
<td>.9976</td>
</tr>
<tr>
<td><strong>ANN input = (S, r, T, X, \sigma)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSE</td>
<td>.6229</td>
<td>.7862</td>
<td>1.0464</td>
<td>.8224</td>
<td>.8825</td>
</tr>
<tr>
<td>MAE</td>
<td>.3934</td>
<td>.4856</td>
<td>.6474</td>
<td>.5315</td>
<td>.5539</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.9983</td>
<td>.9972</td>
<td>.9951</td>
<td>.9971</td>
<td>.9968</td>
</tr>
<tr>
<td>$\rho$</td>
<td>.9992</td>
<td>.9986</td>
<td>.9978</td>
<td>.9986</td>
<td>.9985</td>
</tr>
<tr>
<td><strong>ANN input = (S, r, T, X, V)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSE</td>
<td>.5471</td>
<td>.6228</td>
<td>.7183</td>
<td>.6820</td>
<td>.6613</td>
</tr>
<tr>
<td>MAE</td>
<td>.3627</td>
<td>.4060</td>
<td>.4681</td>
<td>.4581</td>
<td>.4159</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.9987</td>
<td>.9983</td>
<td>.9977</td>
<td>.9980</td>
<td>.9982</td>
</tr>
<tr>
<td>$\rho$</td>
<td>.9993</td>
<td>.9991</td>
<td>.9989</td>
<td>.9990</td>
<td>.9991</td>
</tr>
</tbody>
</table>

Table 1: In-sample RMSE, MAE, $R^2$, and $\rho$ of alternative option pricing models.
<table>
<thead>
<tr>
<th>test subsample</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>average</th>
<th>minimum</th>
<th>maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Black-Scholes formula</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSE</td>
<td>1.8436</td>
<td>1.2374</td>
<td>.8758</td>
<td>.6403</td>
<td>.6601</td>
<td>1.0514</td>
<td>.6403</td>
<td>1.8436</td>
</tr>
<tr>
<td>MAE</td>
<td>1.3703</td>
<td>.8720</td>
<td>.6219</td>
<td>.4431</td>
<td>.4776</td>
<td>.7570</td>
<td>.4431</td>
<td>1.3703</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.9868</td>
<td>.9941</td>
<td>.9970</td>
<td>.9981</td>
<td>.9974</td>
<td>.9947</td>
<td>.9868</td>
<td>.9981</td>
</tr>
<tr>
<td>$\rho$</td>
<td>.9962</td>
<td>.9981</td>
<td>.9987</td>
<td>.9991</td>
<td>.9991</td>
<td>.9982</td>
<td>.9962</td>
<td>.9991</td>
</tr>
<tr>
<td><strong>ANN input = (S, r, T, X, $\sigma$)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSE</td>
<td>1.7622</td>
<td>1.3452</td>
<td>1.0409</td>
<td>.8689</td>
<td>2.3885</td>
<td>1.4811</td>
<td>.8689</td>
<td>2.3885</td>
</tr>
<tr>
<td>MAE</td>
<td>.8263</td>
<td>.7841</td>
<td>.5473</td>
<td>.4668</td>
<td>.9373</td>
<td>.7124</td>
<td>.4668</td>
<td>.9373</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.9932</td>
<td>.9948</td>
<td>.9959</td>
<td>.9959</td>
<td>.9855</td>
<td>.9931</td>
<td>.9855</td>
<td>.9959</td>
</tr>
<tr>
<td>$\rho$</td>
<td>.9976</td>
<td>.9978</td>
<td>.9980</td>
<td>.9983</td>
<td>.9961</td>
<td>.9976</td>
<td>.9961</td>
<td>.9980</td>
</tr>
<tr>
<td><strong>ANN input = (S, r, T, X, $V$)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSE</td>
<td>1.4085</td>
<td>.9702</td>
<td>.5262</td>
<td>.7768</td>
<td>.9543</td>
<td>.9272</td>
<td>.5262</td>
<td>1.4085</td>
</tr>
<tr>
<td>MAE</td>
<td>.9091</td>
<td>.5847</td>
<td>.3655</td>
<td>.3982</td>
<td>.5747</td>
<td>.5664</td>
<td>.3655</td>
<td>.9091</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.9923</td>
<td>.9963</td>
<td>.9989</td>
<td>.9972</td>
<td>.9945</td>
<td>.9958</td>
<td>.9923</td>
<td>.9989</td>
</tr>
<tr>
<td>$\rho$</td>
<td>.9969</td>
<td>.9988</td>
<td>.9995</td>
<td>.9987</td>
<td>.9973</td>
<td>.9982</td>
<td>.9969</td>
<td>.9995</td>
</tr>
</tbody>
</table>

Table 2: Out-of-sample RMSE, MAE, $R^2$, and $\rho$ of alternative option pricing models.
<table>
<thead>
<tr>
<th>RMSE</th>
<th>MAE</th>
<th>$R^2$</th>
<th>$\rho$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4000 iterations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>in-sample</td>
<td>.8286</td>
<td>.5654</td>
<td>.9972</td>
</tr>
<tr>
<td>out-of-sample</td>
<td>601.72</td>
<td>22.64</td>
<td>-2.6438</td>
</tr>
<tr>
<td>2000 iterations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>in-sample</td>
<td>.9808</td>
<td>.7039</td>
<td>.9961</td>
</tr>
<tr>
<td>out-of-sample</td>
<td>102.91</td>
<td>9.1729</td>
<td>.3768</td>
</tr>
</tbody>
</table>

Table 3: Performance of an ANN with volatility $\sigma$ as an additional input variable.
Table 4: Descriptive statistics of residuals of the ANN and the Black-Scholes option pricing models.

<table>
<thead>
<tr>
<th>subsample</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First order residual autocorrelation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>in-sample</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANN</td>
<td>.2527</td>
<td>.2523</td>
<td>.2349</td>
<td>.2727</td>
<td>.2944</td>
</tr>
<tr>
<td></td>
<td>(1.49)</td>
<td>(1.48)</td>
<td>(1.49)</td>
<td>(1.43)</td>
<td>(1.37)</td>
</tr>
<tr>
<td>BSF</td>
<td>.5411</td>
<td>.6165</td>
<td>.5775</td>
<td>.5874</td>
<td>.5722</td>
</tr>
<tr>
<td></td>
<td>(.88)</td>
<td>(.70)</td>
<td>(.74)</td>
<td>(.70)</td>
<td>(.67)</td>
</tr>
<tr>
<td><strong>out-of-sample</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANN</td>
<td>.1205</td>
<td>.2865</td>
<td>.5438</td>
<td>.1781</td>
<td>-.1967</td>
</tr>
<tr>
<td></td>
<td>(.9323)</td>
<td>(.5571)</td>
<td>(.1899)</td>
<td>(.6051)</td>
<td>(1.1133)</td>
</tr>
<tr>
<td>BSF</td>
<td>.4974</td>
<td>.4092</td>
<td>.6493</td>
<td>.4585</td>
<td>.0648</td>
</tr>
<tr>
<td></td>
<td>(.3217)</td>
<td>(.2820)</td>
<td>(.1330)</td>
<td>(.1404)</td>
<td>(.1993)</td>
</tr>
<tr>
<td><strong>Percentage of positive residuals</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>in-sample</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANN</td>
<td>43.34</td>
<td>46.16</td>
<td>38.04</td>
<td>41.87</td>
<td>48.76</td>
</tr>
<tr>
<td>BSF</td>
<td>56.66</td>
<td>59.14</td>
<td>66.37</td>
<td>67.38</td>
<td>71.61</td>
</tr>
<tr>
<td><strong>out-of-sample</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANN</td>
<td>79.64</td>
<td>31.22</td>
<td>42.99</td>
<td>36.20</td>
<td>34.98</td>
</tr>
<tr>
<td>BSF</td>
<td>94.57</td>
<td>84.62</td>
<td>55.66</td>
<td>51.58</td>
<td>34.98</td>
</tr>
<tr>
<td><strong>Mean of residuals</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>in-sample</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANN</td>
<td>-.0298</td>
<td>.0217</td>
<td>-.1007</td>
<td>-.0596</td>
<td>.0813</td>
</tr>
<tr>
<td>BSF</td>
<td>.1880</td>
<td>.3136</td>
<td>.4046</td>
<td>.4670</td>
<td>.5639</td>
</tr>
<tr>
<td><strong>out-of-sample</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANN</td>
<td>.6012</td>
<td>-.4022</td>
<td>-.0115</td>
<td>.0195</td>
<td>-.0854</td>
</tr>
<tr>
<td>BSF</td>
<td>1.1864</td>
<td>.6829</td>
<td>.3183</td>
<td>.0680</td>
<td>-.3125</td>
</tr>
</tbody>
</table>

Note: the numbers in parenthesis are the Durbin-Watson statistics.
<table>
<thead>
<tr>
<th></th>
<th>constant</th>
<th>S</th>
<th>r</th>
<th>T</th>
<th>X</th>
<th>V</th>
<th>σ</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subsample 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANN</td>
<td>5.3370</td>
<td>-0.0030</td>
<td>20.7803</td>
<td>-1.6409</td>
<td>-0.0095</td>
<td>0.0000</td>
<td>0.1361</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.6448)</td>
<td>(-0.7161)</td>
<td>(3.1448)</td>
<td>(-4.1450)</td>
<td>(-9.6377)</td>
<td>(-2.9444)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BSF</td>
<td>57.4655</td>
<td>-0.0516</td>
<td>-20.2460</td>
<td>6.1282</td>
<td>-0.0305</td>
<td>-190.0394</td>
<td>0.3021</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.2374)</td>
<td>(-5.8747)</td>
<td>(-1.3417)</td>
<td>(7.6449)</td>
<td>(-15.5772)</td>
<td>(-2.0991)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Subsample 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANN</td>
<td>11.4810</td>
<td>-0.0115</td>
<td>17.7891</td>
<td>-1.1033</td>
<td>-0.0138</td>
<td>0.0000</td>
<td>0.1421</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.3357)</td>
<td>(-2.0397)</td>
<td>(1.9027)</td>
<td>(-1.9261)</td>
<td>(-10.0605)</td>
<td>(-3.1793)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BSF</td>
<td>123.8821</td>
<td>-0.1072</td>
<td>-42.9477</td>
<td>10.0384</td>
<td>-0.0432</td>
<td>-539.6371</td>
<td>0.3130</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(14.2656)</td>
<td>(-6.1033)</td>
<td>(-1.6086)</td>
<td>(6.7360)</td>
<td>(-11.9325)</td>
<td>(-6.1133)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Subsample 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANN</td>
<td>12.6998</td>
<td>-0.0040</td>
<td>18.0434</td>
<td>-0.5625</td>
<td>-0.0238</td>
<td>0.0000</td>
<td>0.2284</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.3048)</td>
<td>(-4.3084)</td>
<td>(1.1409)</td>
<td>(-0.7811)</td>
<td>(-14.2288)</td>
<td>(-3.3155)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BSF</td>
<td>119.8200</td>
<td>-0.0826</td>
<td>8.1622</td>
<td>10.2648</td>
<td>-0.0480</td>
<td>-539.6371</td>
<td>0.3279</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(12.2616)</td>
<td>(-5.8424)</td>
<td>(2.2725)</td>
<td>(6.9732)</td>
<td>(-13.1033)</td>
<td>(-6.1133)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Subsample 4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANN</td>
<td>11.8630</td>
<td>-0.0085</td>
<td>21.3035</td>
<td>-0.9172</td>
<td>-0.0180</td>
<td>0.0000</td>
<td>0.1791</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.4514)</td>
<td>(-1.5688)</td>
<td>(1.8626)</td>
<td>(-1.4162)</td>
<td>(-11.8862)</td>
<td>(-2.5053)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BSF</td>
<td>126.4748</td>
<td>-0.0794</td>
<td>-65.5375</td>
<td>12.7069</td>
<td>-0.0510</td>
<td>-652.6742</td>
<td>0.3277</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(13.3062)</td>
<td>(-5.4990)</td>
<td>(-2.1980)</td>
<td>(8.3419)</td>
<td>(-13.6528)</td>
<td>(-7.7992)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Subsample 5</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANN</td>
<td>17.0972</td>
<td>-0.0159</td>
<td>9.9811</td>
<td>-0.9457</td>
<td>-0.0208</td>
<td>0.0000</td>
<td>0.2164</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.6437)</td>
<td>(-2.0628)</td>
<td>(0.8499)</td>
<td>(-1.4063)</td>
<td>(-13.3380)</td>
<td>(-3.7190)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BSF</td>
<td>145.0784</td>
<td>-0.1412</td>
<td>-81.7844</td>
<td>13.0301</td>
<td>-0.0500</td>
<td>-551.9775</td>
<td>0.3378</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(14.6167)</td>
<td>(-6.8122)</td>
<td>(-2.8488)</td>
<td>(8.9785)</td>
<td>(-13.7949)</td>
<td>(-6.2700)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: The numbers are the regression coefficients of square pricing errors on five pricing variables and a constant term, and the numbers in parenthesis are t-ratios.

Table 5: White’s heteroscedasticity test for the ANN and the Black-Scholes option pricing models.
Table 6: Moneyness and timeliness bias of alternative option pricing models.

<table>
<thead>
<tr>
<th>subsample</th>
<th>in-the-money MAE</th>
<th>out-of-the-money MAE</th>
<th>correlation with T</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ANN</td>
<td>BSF</td>
<td>ANN</td>
</tr>
<tr>
<td>1</td>
<td>.5674</td>
<td>.9101</td>
<td>.2413</td>
</tr>
<tr>
<td>2</td>
<td>.6526</td>
<td>1.0562</td>
<td>.2713</td>
</tr>
<tr>
<td>3</td>
<td>.7523</td>
<td>1.1094</td>
<td>.3136</td>
</tr>
<tr>
<td>4</td>
<td>.6975</td>
<td>1.1794</td>
<td>.3260</td>
</tr>
<tr>
<td>5</td>
<td>.7286</td>
<td>1.2909</td>
<td>.2633</td>
</tr>
<tr>
<td>average</td>
<td>0.6797</td>
<td>1.1092</td>
<td>0.2831</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>test subsample</th>
<th>out of sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ANN</td>
</tr>
<tr>
<td>1</td>
<td>1.4483</td>
</tr>
<tr>
<td>2</td>
<td>.9781</td>
</tr>
<tr>
<td>3</td>
<td>.5801</td>
</tr>
<tr>
<td>4</td>
<td>.8538</td>
</tr>
<tr>
<td>5</td>
<td>.9913</td>
</tr>
<tr>
<td>average</td>
<td>0.9703</td>
</tr>
</tbody>
</table>

Table 6: Moneyness and timeliness bias of alternative option pricing models.
<table>
<thead>
<tr>
<th>subsample</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>average</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>1.9689</td>
<td>2.3459</td>
<td>3.3734</td>
<td>3.3011</td>
<td>2.6389</td>
<td>2.7256</td>
</tr>
<tr>
<td>r</td>
<td>.5392</td>
<td>.5493</td>
<td>-2.0784</td>
<td>-3.7803</td>
<td>-3.1693</td>
<td>-1.5879</td>
</tr>
<tr>
<td>T</td>
<td>3.6322</td>
<td>3.3162</td>
<td>5.3696</td>
<td>6.7213</td>
<td>6.1061</td>
<td>5.0291</td>
</tr>
<tr>
<td>V</td>
<td>-4.7383</td>
<td>-4.3987</td>
<td>4.7083</td>
<td>7.6396</td>
<td>7.0717</td>
<td>2.0565</td>
</tr>
</tbody>
</table>

Table 7: Pseudo weights of ANN option pricing for all five subsamples.

<table>
<thead>
<tr>
<th>subsample</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>average</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>.8571</td>
<td>.6857</td>
<td>.9975</td>
<td>1.0201</td>
<td>1.1222</td>
<td>.9365</td>
</tr>
<tr>
<td>r</td>
<td>1.6126</td>
<td>1.3320</td>
<td>.6139</td>
<td>.9958</td>
<td>1.4070</td>
<td>1.1923</td>
</tr>
<tr>
<td>T</td>
<td>1.5820</td>
<td>1.5756</td>
<td>1.9494</td>
<td>2.2002</td>
<td>2.3897</td>
<td>1.9394</td>
</tr>
<tr>
<td>V</td>
<td>5.9485</td>
<td>5.8578</td>
<td>2.0234</td>
<td>3.0114</td>
<td>2.0656</td>
<td>3.7813</td>
</tr>
</tbody>
</table>

Table 8: Sum of absolute input weights for ANN option pricing across all five subsamples.
<table>
<thead>
<tr>
<th>INF</th>
<th>UIC</th>
<th>P/E</th>
<th>r</th>
<th>m</th>
</tr>
</thead>
<tbody>
<tr>
<td>pseudo weights</td>
<td>0.3859</td>
<td>1.9421</td>
<td>-0.7585</td>
<td>6.0863</td>
</tr>
<tr>
<td>sum of input weights</td>
<td>5.2794</td>
<td>2.4055</td>
<td>4.1891</td>
<td>6.2138</td>
</tr>
<tr>
<td>RMSE: 7.8985</td>
<td>R²: 0.68</td>
<td>ρ: 0.8250</td>
<td>Theil's U: 0.56</td>
<td></td>
</tr>
</tbody>
</table>

Table 9: Pseudo weights, sum of input weights and in-sample RMSE, R², ρ, and Theil’s U for ANN stock prediction.

<table>
<thead>
<tr>
<th></th>
<th>LR</th>
<th>ANN</th>
<th>RW</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>14.20</td>
<td>13.69</td>
<td>14.56</td>
</tr>
<tr>
<td>Pearson ρ</td>
<td>0.32</td>
<td>0.58</td>
<td>-0.08</td>
</tr>
<tr>
<td>Theil’s U</td>
<td>0.98</td>
<td>0.94</td>
<td></td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th></th>
<th>RW vs LR</th>
<th>LR vs ANN</th>
<th>RW vs ANN</th>
</tr>
</thead>
<tbody>
<tr>
<td>t-statistics</td>
<td>0.7196</td>
<td>0.7007</td>
<td>0.9669</td>
</tr>
<tr>
<td>p-value (one-tailed)</td>
<td>0.2364</td>
<td>0.2422</td>
<td>0.1675</td>
</tr>
</tbody>
</table>

Table 11: The t-statistics and p-values for testing the null hypothesis of no difference in the squared prediction errors of two competing forecasts.
<table>
<thead>
<tr>
<th>holding period starting</th>
<th>holding period ending</th>
<th>LR</th>
<th>ANN</th>
<th>Naive</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>December</td>
<td>8.77</td>
<td>17.26</td>
<td>8.86</td>
</tr>
<tr>
<td>February</td>
<td>January</td>
<td>6.71</td>
<td>9.16</td>
<td>8.72</td>
</tr>
<tr>
<td>March</td>
<td>February</td>
<td>10.43</td>
<td>15.85</td>
<td>9.96</td>
</tr>
<tr>
<td>April</td>
<td>March</td>
<td>19.65</td>
<td>15.29</td>
<td>11.05</td>
</tr>
<tr>
<td>May</td>
<td>April</td>
<td>21.05</td>
<td>18.07</td>
<td>10.85</td>
</tr>
<tr>
<td>June</td>
<td>May</td>
<td>12.79</td>
<td>18.83</td>
<td>10.67</td>
</tr>
<tr>
<td>July</td>
<td>June</td>
<td>18.48</td>
<td>14.42</td>
<td>11.09</td>
</tr>
<tr>
<td>August</td>
<td>July</td>
<td>13.28</td>
<td>16.05</td>
<td>10.35</td>
</tr>
<tr>
<td>September</td>
<td>August</td>
<td>9.68</td>
<td>13.02</td>
<td>11.37</td>
</tr>
<tr>
<td>October</td>
<td>September</td>
<td>4.38</td>
<td>11.69</td>
<td>10.71</td>
</tr>
<tr>
<td>November</td>
<td>October</td>
<td>7.84</td>
<td>13.05</td>
<td>9.54</td>
</tr>
<tr>
<td>December</td>
<td>November</td>
<td>9.86</td>
<td>15.89</td>
<td>8.55</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td></td>
<td><strong>11.91</strong></td>
<td><strong>14.88</strong></td>
<td><strong>10.14</strong></td>
</tr>
</tbody>
</table>

Table 12: Average annual gains (%) using naive strategy and trading rule based on recursive linear regression and ANN forecasts: January 1980-June 1994.
APPENDIX B

FIGURES

Figure 1: A three-layer feedforward neural network.
output layer: $y$

weight vector: $\beta = (\beta_0, \beta_1, \ldots, \beta_n)'$

input layer: $X = (x_0, x_1, \ldots, x_n)'$

Figure 2: ADALINE network

output layer: $Y = (y_1, y_2, \ldots, y_k)'$

weight matrix: $\beta$

input layer: $X = (x_1, x_2, \ldots, x_n)'$

Figure 3: MADALINE network

output layer: $Y$

weight matrix: $\alpha$

middle layer: $M$

weight matrix: $\beta$

input layer: $X$

Figure 4: A three-layer feedforward linear ANN
Figure 5: Stability of alternative structures by RMSE.

Figure 6: Stability of alternative structures by MAE.
Figure 7: Stability of alternative structures by $\rho$.

Figure 8: Stability of alternative structures by $R^2$. 
(a) supply shifts: 
C and V are negatively related

(b) demand shifts: 
C and V are positively related

Figure 9: Economic implications of open interest V.
Figure 10: Sensitivity analysis of the ANN option pricing model.
Figure 10: Sensitivity analysis of the ANN option pricing model (continued).
Figure 10: Sensitivity analysis of the ANN option pricing model (continued).
Figure 11: Actual and fitted returns on S&P 500 index by alternative models (in sample).
Figure 12: Sensitivity analysis for the ANN stock prediction model.
Figure 12: Sensitivity analysis for the ANN stock prediction model (continued).
(Figure 12 continued)

(i) $R = f(p/e, m)$

(ii) $R = f(r, m)$

Figure 12: Sensitivity analysis for the ANN stock prediction model (continued).
Figure 13: Actual and forecasted returns on S&P 500 index by alternative models (out of sample).