

ARTIFICIAL NEURAL NETWORK MODELS  
FOR FORECASTING AND DECISION MAKING

Tim Hill  
University of Hawaii  
2404 Maile Way  
Honolulu, HI USA 96822  
(01) 808-956-7430  
THILL@UHUNIX.UHCC.HAWAII.EDU

Leorey Marquez  
CSIRO  
Melbourne, Australia

Marcus O'Connor  
University of New South Wales  
P.O. Box 1  
Kensington 2033, Australia  
(61) 2-6974640  
75B1162@CSDVAX.CSD.UNSW.OZ.AU

William Remus  
University of Hawaii  
2404 Maile Way  
Honolulu, HI USA 96822  
(01) 808-956-7430  
REMUS@UHUNIX.UHCC.HAWAII.EDU

September 30, 1993

ARTIFICIAL NEURAL NETWORK MODELS  
FOR FORECASTING AND DECISION MAKING

ABSTRACT

Some authors advocate artificial neural networks as a replacement for statistical forecasting and decision models; other authors are concerned that artificial neural networks might be oversold or just a fad. In this paper we review the literature comparing artificial neural networks and statistical models, particularly in regression-based forecasting, time series forecasting, and decision making. Our intention is to give a balanced assessment of the potential of artificial neural networks for forecasting and decision making models.

We survey the literature and summarize several studies we have performed. Overall, the empirical studies find artificial neural networks comparable to their statistical counterparts. We note the need for using the many mathematical proofs underlying artificial neural networks to determine the best conditions for using artificial neural networks in the forecasting and decision making.

Key Words: Artificial Neural Networks, Regression, Forecasting, Decision Making, Time Series.

ARTIFICIAL NEURAL NETWORK MODELS  
FOR FORECASTING AND DECISION MAKING  
INTRODUCTION

Over the last few decades, there has been much research directed at predicting the future and making better decisions. This research has led to many developments in forecasting methods.

Most of these methodological advances have been based on statistical techniques. Currently, there is a new challenger for these methodologies - artificial neural networks.

Artificial neural networks (ANN) have been widely touted as solving many forecasting and decision modeling problems (e.g., Hiew and Green, 1992). For example, they are argued to be able to model easily any type of parametric or non-parametric process and automatically and optimally transform the input data. These sorts of claims have led to much interest in artificial neural networks.

On the other hand, Chatfield (1993) has queried whether artificial neural networks have been oversold or are just a fad.

In this paper, we will attempt to give a balanced review of the literature comparing artificial neural networks and statistical techniques. Our review will be segmented into three different application areas: time series forecasting, regression-based forecasting, and regression-based decision models. Additionally, we will note the literature comparing artificial neural networks and other models such as discriminant analysis. But before that review, we will first examine the general claims made for artificial neural networks that are relevant to forecasting and decision making.

THE POTENTIAL OF ARTIFICIAL NEURAL NETWORKS

Artificial neural networks are mathematical models inspired by the organization and functioning of biological neurons. There are numerous artificial neural network variations that are related to the nature of the task assigned to the network. There are also numerous variations in how the neuron is modeled. In some cases, these models correspond closely to biological neurons (e.g., Gluck and Bower, 1988; Granger et al., 1989) and in other cases the models depart from biological functioning in significant ways. See the appendix for a more detailed explanation of the artificial neural network paradigm.

The literature suggests several potential advantages that artificial neural networks have over statistical methods. Artificial neural networks can be universal function approximators for even non-linear functions. Artificial neural networks can also estimate piece-wise approximations of functions. In the following paragraphs we will elaborate these possibilities.

Artificial neural networks can be mathematically shown to be universal function approximators (Hornik et al., 1989). This

means that artificial neural networks can automatically approximate whatever functional form best characterizes the data.

While this property gives little value if the functional form is simple (e.g., linear), this property allows artificial neural networks to extract more signal from complex underlying functional forms. Also artificial neural networks can at least partially transform the input data automatically (Connor, 1988; Donaldson, Kamstra and Kim, 1993).

Artificial neural networks also are inherently non-linear (Rumelhart and McClelland, 1986; Wasserman, 1989). That means not only do they estimate non-linear functions well but also they can extract any residual non-linear elements from the data after linear terms are removed. De Gooijer and Kumar (1992) recently argued the benefits of using non-linear forecasting models.

With artificial neural networks using one or more hidden layers, the networks can partition the sample space automatically and build different functions in different portions of that space.

This means that artificial neural networks have a modest capability for building piece-wise non-linear models. The artificial neural network model for the exclusive OR function is a good example of such a model (Wasserman, 1989, pp. 30-33). Collopy and Armstrong (1992) surveyed forecasting experts and found that the experts considered it important to select extrapolation techniques that identified and treated abrupt changes in historical data patterns, suggesting the utility of piece-wise models.

Some statistical time series methods have inherent limitations due to the way in which the models are estimated. The estimation of many kinds of statistical time series models require human interaction and evaluation. The estimation of artificial neural networks, however, can be automated (Hoptroff, 1993). Also, many statistical models must be re-estimated periodically when new data arrive. Many artificial neural network algorithms learn incrementally (Widrow and Sterns, 1985).

There are also problems with artificial neural networks. First, artificial neural network methodology and modeling techniques are rapidly changing, whereas many statistical modeling techniques are stable and well developed. Second, software is readily available for statistical techniques but commercial artificial neural network software, although of good quality, often lags developments in the field. Third, artificial neural network models are harder to interpret and to give physical meaning than are many forecasting models. Fourth, artificial neural networks contain more parameters to estimate than do most statistical forecasting models; this can result in overfitting problems. Also, artificial neural networks require more computer time than statistical models. These points are elaborated in the appendix.

Although several classes of forecasting models share some of

the advantages discussed earlier in this section, artificial neural networks provide all the latter advantages. In fact if needed, artificial neural networks can even nest ARIMA forecasting models (e.g., Donaldson, Kamstra, and Kim, 1993). Thus, the use of artificial neural networks for forecasting and decision making models is worthy of evaluation.

#### ARTIFICIAL NEURAL NETWORKS AND TIME SERIES FORECASTING MODELS

Artificial neural networks and traditional time series techniques have been compared in several studies. The best of these studies have used the data from the well-known "M-competition" (Makridakis et al., 1982). Makridakis et al. gathered 1001 real time series; this and the soon-to-be described studies use a systematic sample of 111 series from the original database. In the original competition, various groups of forecasters were given all but the most recent data points in each of the series and asked to make forecasts for those most recent points. Each competitor's forecasts were then compared to the actual values in the holdout data set. The results of this competition were reported in Makridakis et al. (1982). Since the competition, all 1001 series including the holdout data have been made available.

Sharda and Patil (1990) used 75 series from a systematic sample of 111 series and found that artificial neural network models performed as well as the automatic Box-Jenkins (Autobox) procedure. In the 36 deleted series, however, neither the artificial neural network nor Autobox models had enough data to estimate the models. Foster et al. (1991) also used the M-competition data. They found artificial neural networks to be inferior to Holt's, Brown's, and the least squares statistical models for yearly data but comparable with quarterly data; they did not compare the models on monthly data. Sharda and Patil (1992) and Tang et al. (1991) found that for time series with a long memory, artificial neural network models and Box-Jenkins models produce comparable results. However, for time series with short memory, Tang et al. (1991) found artificial neural networks to be superior to Box-Jenkins.

Kang (1991) compared artificial neural networks and Box-Jenkins (Autobox) on the 50 M-competition series designated by Pack and Downing (1983) to be most appropriate for the Box-Jenkins technique. Kang found Autobox to be superior or equivalent to the average of eighteen different artificial neural network architectures in terms of MAPE (Mean Absolute Percentage Error). Kang also compared the eighteen artificial neural network architectures and the Autobox model on seven sets of simulated time series patterns. Kang found the MAPE for the average of the eighteen artificial neural network architectures only superior when trend and seasonal patterns were in the data. In each case,

at least one artificial neural network was better than the Autobox model. Thus, architecture is crucial in designing a successful forecasting model. Interestingly, both Kang (1991) and Tang et al. (1991) found that artificial neural networks often performed better when predicting beyond the first few periods.

The above studies are encouraging but are equivocal; thus, we (Hill, O'Connor, and Remus, 1993) were inspired to attempt a more definitive comparison of artificial neural networks and statistical models. In our analysis we included the models used by both Sharda and Patil (1990) and Foster et al. (1991); also we included three important time series forecasting models used in the M-competition<sup>1</sup>, one method based on combining the results of six forecasting models<sup>2</sup>, a naive forecasting model (next period's forecast is just whatever happened in the prior period), and a method based on human judgment (Lawrence et al., 1985). This method performed as well as the best statistical method in the M-competition (Lawrence et al., 1985) and appears to be the most widely used method in business (Dalrymple, 1987).

We compared the models across yearly, quarterly, and monthly data from a systematic sample of 111 M-competition time series; the models also were designed to cover the cases that Sharda and Patil (1990) were unable to forecast. Additionally, we standardized many other procedural differences (like artificial neural network model structure) among the earlier discussed studies. All the above gave us a sound comparison of statistical and artificial neural network time series models. We found the artificial neural networks were significantly better than

---

<sup>1</sup>The models used included the deseasonalized single exponential smoothing, Box-Jenkins, and deseasonalized Holt exponential smoothing method. These were chosen both because they represented numerous well-used methods and because they performed relatively well in the M-competition described earlier. All of the above models were estimated as part of the M-competition by experts in the technique.

<sup>2</sup>There is a large literature in forecasting that suggests that the best forecasts can be made by combining the results of several forecasting models (e.g., Makridakis and Winkler, 1985); we included such a model from the M-competition. This model is the average of the forecasts of six statistical methods (deseasonalized single exponential smoothing, deseasonalized adaptive response rate exponential smoothing, deseasonalized Holt's exponential smoothing, deseasonalized Brown's linear exponential smoothing, Holt-Winter's linear and exponential smoothing, and Carboni-Longini filter method). It should be noted that Makridakis and Winkler's conclusion that the simple averaging is the best way to combine forecasts is not universally supported; see Donaldson, Kamstra, and Kim (1993) for a contrasting view of this literature.

statistical and human judgment methods by about 5% MAPE in the quarterly and about 2% MAPE in the monthly time series.

Why did our work outperform Sharda and Patil (1990), Foster et al. (1991), and Kang (1991)? Part of the answer is suggested when we compare the results on our two architectures. The first architecture forecast all periods in the forecast horizon simultaneously. This model is similar to that used by Foster et al. (1991), Kang (1991), and Sharda and Patil (1990 and 1992). In all four studies (including ours) this ANN architecture performed roughly as well as the statistical models.

In the second ANN architecture, a forecast for the first period of the forecast horizon was generated. It was fed back into the model to forecast the second period of the forecast horizon. This process was continued until there were forecasts for the entire horizon. This second ANN architecture outperformed the first ANN architecture, Sharda and Patil (1990) and Foster et al. (1991). The bootstrapping in the second architecture may have been the crucial difference since bootstrapping also improves judgmental forecasting models (see Armstrong, 1985, pp. 271-293 for a good review of this literature).

We also compared the second ANN architecture with the statistical models across all periods in the forecast horizon; the superiority of the second ANN architecture was in the later periods of the forecast horizon (confirming the findings of Kang (1991) and Tang et al. (1991)). Both Sharda and Patil (1990) and Foster et al. (1991) made comparisons based on a one period ahead forecast.

Two of the results above deserve additional discussion. First, both Kang (1991) and our work found artificial neural networks better in forecasting monthly and quarterly series than in forecasting annual series. Second, Kang (1991), Tang et al. (1991), and our work found artificial neural networks to be superior in the later periods of the forecast horizon. Since artificial neural networks are most advantageous when non-linear patterns are present, we favor that explanation for our results. In the first case, non-linear patterns could result from the relatively more non-linearities in deseasonalized quarterly and monthly series than in annual series. The second result could also be explained if the series contained non-linearities. If so, the non-linear artificial neural networks should make better forecasts than linear models; the improvement would be increasingly apparent as the forecast horizon increased.

Overall, the early studies on artificial neural networks are equivocal. In our work, however, we found artificial neural networks to perform as well as classical statistical models in forecasting annual time series and they may actually perform better than statistical models for monthly and quarterly data. The case for artificial neural network models may be strengthened by other model attributes that are important in a particular

application (for example, the artificial neural network's ability to automate the forecasting process).

#### ARTIFICIAL NEURAL NETWORKS AND REGRESSION-BASED MODELS

Regression models are widely used to make forecasts and to model human decision tasks. In this section we will examine the comparative studies of artificial neural networks and regression models that are independent of the task to be modeled; in the next section we will examine comparative studies of artificial neural networks and regression models when used to model human decision tasks. Interestingly, there are no studies that compare artificial neural networks and regression using real multivariate forecasting data.

##### Comparative Studies Independent of Task

In Marquez et al. (1992) we generated data representing common bivariate functional forms used in forecasting (linear, log-linear, squared, square-root, and reciprocal). For each functional form, 100 sets of  $n$  points (15, 30, or 60) each with three noise levels ( $R^2 = .30, .60, \text{ and } .90$ ) were created. Artificial neural networks were estimated for each of the above data sets. The fit of the artificial neural networks was compared to the fit of the true functional forms using holdout data sets of 100 points each. The fit of the artificial neural networks to the data was almost as good as the fit of the true functional forms (i.e., within 2% of the MAPE). Also, for all practical purposes the artificial neural networks fit the data as well as a correctly specified regression model with the appropriate independent variable transformation. In this study artificial neural networks performed comparatively well with high noise and low sample sizes.

This work continued as Marquez (1992) expanded the simulation study to other functional forms. The Marquez et al. (1992) results generalized to other functional forms including second degree polynomials, forms with multiple independent variables, the Michaelis-Menten equation, the exponential rise equation, and the logistic model. The artificial neural networks again were within 2% of the MAPE of the true functional form; the artificial neural networks fit the data about as well as a correctly specified regression model with appropriate independent variable transformations. Artificial neural networks were also noted to be as vulnerable as regression to outliers.

The work in this section shows the comparable performance of artificial neural networks and regression. It also shows that artificial neural networks are effective in estimating true functional forms; their MAPE is less than 2% of that of the true functional form. This ability to work well when the functional form is unknown makes artificial neural networks an attractive choice in many applications.

##### Comparative Studies for Decision Tasks



A decision model tries to predict a human's judgment based on the factors that a person uses to make the judgment. Generally, these models have linear functional forms and regression is used to estimate them. There are several interesting business decisions that have been examined using these models.

Dutta and Shekhar (1988) used 10 factors to predict the ratings of corporate bonds. They estimated the artificial neural network and regression models on 30 randomly selected bonds rated in Standard and Poor's and Valueline. They found that the artificial neural networks outperformed the regression models in accurately predicting the ratings of 17 randomly selected bonds in a holdout sample. Duliba (1991) compared artificial neural network models with four types of regression models in predicting the financial performance of a single group of transportation companies. She found that the artificial neural network model outperformed the random effects regression model but not the fixed effects model for this decision.

One of the classic human judgment problems has been modeling the graduate school admission decision. Gorr, Nagin, and Szczypula (1993) undertook modeling this decision with artificial neural networks, linear regression, stepwise polynomial regression, and index used by a graduate admissions committee. While they found the artificial neural network to identify model structure beyond that of the regression model, the three empirically-estimated models performed equally well but none of the three models outperformed the graduate admissions committee's index.

Remus and Hill (1990) compared the production scheduling decisions as modeled by artificial neural networks and regression-based decision rules. The data used were from 62 decision makers who each made 24 decisions (Remus, 1987); thus, artificial neural network and regression models were developed and compared for each of the 62 decision makers. The model structure used was the same structure used in the comparable regression models. Artificial neural network models performed as well but not better than those using the linear regression models. Both models outperformed the actual decision makers.

In a second study, Hill and Remus (1993) continued the above research and aggregated the data from all 62 decision makers to estimate a composite artificial neural network model. The resulting artificial neural network model performed better than both the statistical models and artificial neural networks from the earlier study. The performance of the composite artificial neural network was not significantly different than the performance of the statistical composite models in intermediate levels of environmental variability; however, in low levels of variance the statistical composite model performed better.

Overall, we believe the research shows that artificial neural networks can perform as well as but not necessarily better than

regression in modeling human decision making. However, artificial neural network models may be particularly valuable if the decision contains important non-linear elements or some of the other advantages of artificial neural networks are crucial in a given application (for example, the artificial neural network's ability to automate the process).

#### ARTIFICIAL NEURAL NETWORKS AND OTHER MODELS

There have been numerous other studies where models (other than ordinary least squares regression and time series models) have been compared to artificial neural networks in common business tasks. For example, logistic regression is commonly used in classification problems where the response variable has a binary value. Bell et al. (1989) compared backpropagation networks against logistic regression models in predicting commercial bank failures. The artificial neural network model performed well in failure prediction and was a better predictor of bank failure than the logistic model.

Roy and Cosset (1990) also used artificial neural network and logistic regression models in predicting country risk ratings using economic and political indicators. The artificial neural network models had lower mean absolute error in their predictions of country risk ratings and were more sensitive to changes in risk indicators than their logistic counterparts.

Artificial neural networks are an alternative to discriminant analysis. Practical problems where the comparison between artificial neural networks and discriminant analysis has been applied include the prediction of stock price performance (Yoon and Swales, 1990), the prediction of company bankruptcy (Odom and Sharda, 1990; Raghupathi et al., 1991; Koster et al., 1990), and the assignment of ratings to bonds (Surkan and Singleton, 1990). In all of these studies, the artificial neural network model outperformed discriminant analysis. The above studies suffer from many technical problems; in particular, the number of data sets is limited and the data sets are small.

Artificial neural networks are also an alternative to discriminant analysis in the prediction of bank failures in Texas (Tam, 1991; Tam and Kiang, 1992). In this work care was taken to carefully estimate several models (including artificial neural networks, discriminant analysis, logistic regression, and ID3) and rigorously test them on holdout samples. The artificial neural networks had better predictive accuracy than the other models. It is not clear how well this result generalizes given the uniqueness of the problem modeled.

Artificial neural networks are also an alternative to specialized non-linear models such as those used in finance. In a well performed study, Donaldson, Kamstra, and Kim (1993) used stock index data from the London, New York, Tokyo, and Toronto exchanges to evaluate the ability of several popular conditional

volatility models to account for the fat-tailed and heteroskedastic nature of stock returns<sup>3</sup>; the models used included artificial neural networks and specialized finance models such as the autoregressive conditioned heteroskedasticity (ARCH). They found that these models were able only to partially remove the leptokurtosis and symmetric and asymmetric heteroskedasticity from the data. However, the artificial neural networks did outperform the specialized finance models in this task. Incidentally, this is one of the few studies that exploit the strengths of artificial neural networks suggested by the formal theory described earlier.

In general, the combination of numerous studies which are only suggestive because of assorted methodological weaknesses and several well conducted studies (Donaldson, Kamstra, and Kim, 1993; Tam, 1991; Tam and Kiang, 1992) gives us guarded optimism for artificial neural networks as viable alternatives to the many specialized models in business.

#### SUMMARY

In this review, we found artificial neural networks did as well as (and occasionally better than) statistical models; however, we suspect that there are only certain conditions where this is true. This borderline needs to be defined and honored or artificial neural networks could be oversold; already there are warnings from Chatfield (1993) that authors such as Hiew and Green (1992) are overselling artificial neural networks.

The theory discussed in the second section should give us an idea of where this borderline is. The theory should also define the needed empirical studies to confirm the borderline. For example, artificial neural networks can be mathematically shown to be universal function approximators (Hornik et al., 1989). The theory-based research should identify problem characteristics (like functional form and sample size) that predict when artificial neural networks will forecast better than statistical models. Similarly Connor (1988) and Donaldson, Kamstra, and Kim (1993) noted the ability of artificial neural networks to at least partially automatically transform input variables. Theory-based research should identify with what input variable characteristics predict when artificial neural networks will improve model estimation. Artificial neural networks can be piece-wise linear models (Wasserman, 1989, pp. 30-33); theory-based research should identify when this advantage would give substantially improved forecasting performance. Such theory-based predictions should set the agenda for the next round of research.

The research suggested above should be of great interest to

---

<sup>3</sup> This is an important issue given the problems the conditional volatility introduces into the capital asset pricing model (CAPM) in finance. This results from the heteroskedasticity and fat-tails associated with the stock price data used in CAPM.

the forecasting community since Collopy and Armstrong's (1992) experts wanted to select extrapolation techniques that handle nonlinearities and also handle discontinuities. Theoretically, artificial neural networks can do both well. However, as noted earlier well conducted evaluations of these capabilities have not yet been made.

The artificial neural network community is composed largely of information and computer scientists, electrical engineers, psychologists, and physicists; the background of this community is reflected in the type of research reported. Much of this research is in case study form. Formal theory-based evaluations of these emerging artificial neural network applications in forecasting and elsewhere are also needed.

Most of the studies reviewed use backpropagation for artificial neural network estimation. Recently, improvements have been proposed for the backpropagation algorithm and alternative artificial neural network models have been proposed. These improvements and alternatives will also need the same rigorous evaluation in managerial tasks such a forecasting and decision making.

Although we believe the future looks bright for artificial neural network applications in forecasting and decision making, it is still necessary to rigorously evaluate these applications. Given the weaknesses in much of the current research, however, rigorous theory-based research needs to be done before these new models become an accepted part of our modeling tools.

Appendix  
An Introduction to Artificial neural Networks

Artificial neural Networks (ANN) are mathematical, algorithmic, software models inspired by biological artificial neural networks. The ANN consists of basic units, termed neurons, whose design is suggested by their biological counterparts. These artificial neurons have input paths just as biological neurons have dendrites; they have output paths just as biological neurons have axons. Both artificial and biological neurons also have predispositions that affect the strength of their output. The neuron combines the inputs, incorporates the effect of the predisposition (bias), and outputs signals. In both real and artificial neurons, learning occurs and alters the strength of connections between the neurons and the biases. In the following paragraphs, we will detail the structure, mathematics, and learning algorithm found in the most common artificial neural network.

In biological neurons, nature sets the way that the input signals on dendrites are processed and the way in which the latter is translated into an axon activation. There are a great variety of methods used in nature. With artificial neurons, the modeler sets both.

In ANN's, the neuron input path  $i$  has a signal on it ( $X_i$ ) and the strength of the path is characterized by a weight ( $w_i$ ). The neuron is modeled as summing the path weight times the input signal over all paths and adding the node bias ( $\hat{E}$ ). The output ( $Y$ ) is usually a sigmoid shaped logistic function of the latter sum. Mathematically, the sum is expressed as:

$$\text{sum} = \sum w_i x_i + \hat{E}$$

and it is transformed into the output  $Y$  with the sigmoid shaped logistic function showed mathematically below and depicted in Figure 1:

$$Y = 1/(1 + e^{-\text{sum}})$$

Note that this S-shaped function reduces the effect of extreme input values on the performance of the network.  $x$

\*\*\*\*\*

Insert Figure 1 About Here

\*\*\*\*\*

Learning occurs through the adjustment of the path weights and node biases. The most common method used for the adjustment is backpropagation. In this method, the weights are adjusted to minimize the squared difference between the model output and the desired output for an observation in the data set. The squared error is then propagated backward through the network and used to adjust the weights and biases. The error is:

$$E = (1/2) \sum \sum (y_{j,c} - d_{j,c})^2$$

where  $c$  is an index over the data set used to estimate the network,  $j$  is an index over the output units of the network,  $y$  is the actual state of the output unit for a given set of inputs, and  $d$  is the desired state of the output unit for that set of inputs.

This adjustment process leads to a gradient descent to a minimum point on the error surface like that depicted in Figure 2.

This process is not without problems since there is no assurance that the minimum is not a local minimum; this problem and its solution is discussed later in this section.

\*\*\*\*\*

Insert Figure 2 About Here

\*\*\*\*\*

The simplest version of the gradient descent is to change each weight by an amount proportional to the accumulated error associated with that weight; that is, for any given weight:

$$\Delta w(t) = - \alpha \partial E / \partial w$$

Often an additional term is added to reflect previous adjustments made to that weight. The latter term is called momentum; the above equation becomes:

$$\Delta w(t) = - \alpha \partial E / \partial w + \beta \Delta w(t-1)$$

Rumelhart et al. (1986) solve this problem for the sigmoid transfer function to get the gradient descent formula used in most standard backpropagation software. Note that  $\alpha$  and  $\beta$  must be hand selected for the task or alternatively the Rumelhart and McClelland (1986) defaults are used.

As in nature, many neurons combine to form a artificial neural network as is shown in figure 3. The network consists of an input layer, an output layer, and perhaps one or more intervening layers; the latter are termed hidden layers. Each layer consists of multiple artificial neurons; these artificial neurons are connected to other neurons in adjacent layers. Since these networks contain many interacting non-linear neurons in multiple layers, the networks can capture relatively complex phenomenon. The learning occurs in the same way as described above for an individual neuron; that is, the error is propagated backward and used to adjust the weights and biases in the network using the earlier describe algorithm. The Rumelhart et al. (1986) solution to the gradient descent problem accommodates this multi-layer backpropagation.

\*\*\*\*\*

Insert Figure 3 About Here

\*\*\*\*\*

Usually the modeling begins by postulating an initial network based on general modeling practices. One general practice is to include as many reasonable predictor variables and dummy variables as possible; hence, the size of the input layer is determined by the number of these variables. In the time series models, the general practice is to try to capture the seasonality by having as

many input variables as there are periods in one seasonal cycle. To get more concise models, the number of input variables is often reduced; a good way to do this is with the pruning method discussed later in the appendix.

Only one hidden layer is needed for a artificial neural network to be a universal function approximator to a continuous function (Hornik et al., 1989) so often only one hidden layer is used; however, more than one hidden layer maybe used since overall fewer neurons will be required. If the function to be approximated is discontinuous, the model will require at most two hidden layers (Cybenko, 1988).

The number of nodes in the output layer corresponds to the number of variables to be predicted; thus, artificial neural networks can simultaneously forecast several variables. In the time series case, the output layer can be structured to simultaneous forecast several periods into the future. However, we found in the current study that the best performing artificial neural networks for time series use relatively simple networks and bootstrap the one period forecasts to forecast additional periods.

To begin estimating the network, the weights and biases are usually initialized with random values. The observations are then input to the network and parameters adjusted using one of the following two methods. In the first method, an observation is presented to the input layer and an output generated. The difference between the network's output and the desired output provides the error that is backpropagated to adjust the weights. Then the next observation of the data is presented and more adjustments made. In the second method, only after the entire data set has been presented is the adjustment made; the adjustment is based on the overall fit between the network outputs and the desired output values calculated across the entire data set. Usually the first method is used to start the gradient descent process (since it leads to a fast descent of the error surface) and the second is used to close in on the minimum point.

The adjustment process is repeated until the error converges on a minimum point. However, it is not quite that simple since the error surface may have both local minimums and the global minimum. To avoid the local minimums, a good starting point is required; thus, a downhill simplex algorithm may be used to find such a starting point. Also, a downhill simplex algorithm may be used to hone in on the global minimum after backpropagation has ended. Marquez (1992) used the methods of Nelder and Mead (1965) to accomplish both of the above.

In general, artificial neural networks use more parameters than their classical counterparts and, thus, are more prone to overfitting problems. There are two general approaches to the overfitting problem. The first and simplest is fit the model using only part of the data and to evaluate the model's performance on the other part of the data; the latter is usually

termed a "holdout sample."

The second approach to this problem is to use one of the many network pruning algorithms reduce the network size (and hence parameters to estimate). Marquez (1992) used a pruning method based on the work of Sietsma and Dow (1991) and Weigend et al. (1990). This technique includes a term based on the number of artificial neural network parameters in the error function; when the augmented function is minimized, both the error and the number of parameters are reduced. We might note that it is best to use both methods as did Marquez (1992).

There are three practical problems with artificial neural networks that are worthy of elaboration in the following paragraphs. These problems are the difficulty in tracking the rapidly changing literature on artificial neural networks, the computer intensiveness of these models, and the difficulty in interpreting the model structure.

Artificial neural network methodology and modeling techniques are rapidly changing and improving so researchers must keep current in the many artificial neural network journals and conference proceedings. The algorithm improvements are often presented in the mathematical dialect of its authors; the dialect used can differ widely among the members of the artificial neural network community and thus make challenging reading.

Commercial software is readily available for statistical techniques but commercial artificial neural network software, although of good quality, often lags developments such as those described above and must be revised periodically. For example, the software we used was locally written, augmented version of Rumelhart and McClelland's code (1986).

Although convergence on either a local or global minimum is guaranteed with backpropagation, convergence may require extensive CPU and elapsed time. This means that models often must be developed in batch mode rather than interactive mode. In our work, the model estimation took many hours on a 486 personal computer or a SUN SPARC server and, thus, were run overnight.

A last point we should make about artificial neural networks is that they are much less interpretable than are traditional times series and regression models. As can be inferred from figure 1, knowledge of the value of the weights and biases in the network gives, at best, a general idea of the functional relationships embedded in this complex, non-linear network. Thus, even if a artificial neural network is based on causally related data, the resulting model may not give great insight into the strength and nature of the relationships.



## References

- Armstrong, J.S., Long-Range Forecasting, Second Edition, New York: Wiley, 1985.
- Bell, T., G. Ribar, and J. Verchio, "Neural Nets vs. Logistic Regression," presented at the University of Southern California Expert Systems Symposium, November, 1989.
- Chatfield, C., "Neural Networks: Forecasting Breakthrough or Passing Fad?," International Journal of Forecasting, 1993, 9, 1-3.
- Collopy, F. and J.S. Armstrong, "Expert Opinions about Extrapolation and the Mystery of the Overlooked Discontinuities," International Journal of Forecasting, 1992, 8, 575-582.
- Connor, D., "Data Transformation Explains the Basics of Neural Networks," EDN, May 12, 1988, 138-144.
- Cybenko, G., "Continuous Valued Neural Networks with Two Hidden Layers Are Sufficient," Technical Report, Department of Computer Science, Tufts University, 1988.
- Dalrymple, D.J., "Sales Forecasting Practices: Results of a United States Survey," International Journal of Forecasting, 1987, 3, 379-392.
- De Gooijer, I.G. and K. Kumar, "Some Recent Developments in Non-Linear Modelling, Testing, and Forecasting," International Journal of Forecasting, 1992, 8, 135-156.
- Donaldson, R.G., M. Kamstra, and H.Y. Kim, "Evaluating Alternative Models for Conditional Stock Volatility: Evidence from International Data," Working Paper, University of British Columbia, 1993.
- Duliba, K., "Contrasting Neural Nets with Regression in Predicting Performance," Proceedings of the 24th Hawaii International Conference on System Sciences, 1991, Vol. 4, 163-170.
- Dutta, S. and S. Shekhar, "Bond Rating: A Non-Conservative Application of Neural Networks," Proceedings of the 1988 International Conference on Neural Networks, 1988, Vol. 2, 443-450.
- Foster, B., F. Collopy, and L. Ungar, "Neural Network Forecasting of Short, Noisy Time Series," presented at the ORSA TIMS National Meeting, May, 1991.

Gluck, M.A. and G.H. Bower, "Evaluating an Adaptive Model of Human Learning," Journal of Memory and Language, 1989, 27, 166-195.

Gorr, W.I., D. Nagin, and J. Szczypula, "Comparative Study of Artificial Neural Networks and Statistical Methods for Predicting Student Grade Point Averages," Working Paper, Carnegie Mellon University, 1993.

Granger, R., J. Ambros-Ingerson, U. Staubli, and G. Lynch, "Memorial Operation of Multiple, Interacting Simulated Brain Structures," in Neuroscience and Connectionist Models, M. Gluck and D. Rumelhart, eds., Hillsdale: Erlbaum Associates, 1989.

Hiew, M. and G. Green, "Beyond Statistics. A Forecasting System That Learns," The Forum, 1992, Vol. 5, pp. 1 and 6.

Hill, T., M. O'Connor, and W. Remus, "Neural Networks for Time Series Forecasting," Working Paper, University of Hawaii, 1993, Third round at Management Science.

Hill, T. and W. Remus, "Neural Network Models for Intelligent Support of Managerial Decision Making," Decision Support Systems, 1993, Forthcoming.

Hoptroff, R.G., "The Principles and Practice of Time Series Forecasting and Business Modeling Using Neural Nets," Neural Computing & Applications, 1993, 1, 59-66.

Hornik, K., M. Stinchcombe, and H. White, "Multilayer Feedforward Networks are Universal Approximators," Neural Networks, 1989, 2(5), 359-366.

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

Koster, A., N. Sondak, and W. Bourbia, "A Business Application of Artificial Neural Network Systems," The Journal of Computer Information Systems, 1990, Vol. XI, 3-10.

Lawrence, M.J., R.H. Edmundson, and M.J. O'Connor, "An Examination of the Accuracy of Judgemental Extrapolation of Time Series," International Journal of Forecasting, 1985, 1, 25-35.

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

Makridakis, S. and R. L. Winkler, "Averages of Forecasts: Some

Empirical Results," Management Science, 1985, 29, 987-996.

Marquez, L., Function Approximation Using Neural Networks: A Simulation Study, Ph.D. Dissertation, University of Hawaii, 1992.

Marquez, L., T. Hill, W. Remus, and R. Worthley, "Neural Network Models as an Alternative to Regression," in Neural Network Applications in Finance and Investing edited by R. Trippi and E. Turban, Chicago: Probus Publishing, 1992, 435-450.

Nelder, J. and R. Mead, "The Downhill Simplex Method," Computer Journal, 1965, 7, 308-310.

Odom, M. and R. Sharda, "A Neural Network Model for Bankruptcy Prediction," Proceedings of the 1990 International Joint Conference on Neural Networks, 1990, Vol. 2, 163-168.

Pack, D.J. and D.J. Downing, "Why Didn't Box-Jenkins Win (Again)?" presented at the Third International Symposium on Forecasting, 1983.

Raghupathi, W., L. Schkade, and R. Bapi, "A Neural Network Application for Bankruptcy Prediction," Proceedings of the 24th Hawaii International Conference on System Sciences, 1991, Vol. 4, 147-155.

Remus, W., "A Study of the Impact of Graphical and Tabular Displays and Their Interaction with Environmental Complexity," Management Science, 1987, Vol. 33, 1200-1205.

Remus, W. and T. Hill, "Neural Network Models of Managerial Judgment," Proceedings of the 23rd Hawaii International Conference on System Sciences, 1990, Vol. 4, 340-344. Forthcoming in Advances in Artificial Intelligence in Business and Finance.

Roy, J. and J. Cosset, "Forecasting Country Risk Ratings Using a Neural Network," Proceedings of the 23rd Hawaii International Conference on System Sciences, 1990, Vol. 4, 327-334.

Rumelhart, D. and J. McClelland, Parallel Distributed Processing, Cambridge: MIT Press, 1986.

Rumelhart, D., G.E. Hinton, and R.J. Williams, "Learning Representations by Back-Propagating Errors," Nature, 1986, 323, 533-536.

Sharda, R. and R. Patil, "Neural Networks as Forecasting Experts: An Empirical Test," Proceedings of the 1990 International Joint Conference on Neural Networks Meeting, 1990, Vol. 2, 491-494.

Sharda, R. and R. Patil, "Connectionist Approach to Time Series Prediction: An Empirical Test," Journal of Intelligent Manufacturing, 1992, Forthcoming.

Sietsma, J. and R. Dow, "Creating Artificial Neural Networks That Generalize," Neural Networks, 1991, 4, 67-79.

Surkan, A. and J. Singleton, "Neural Networks for Bond Rating Improved by Multiple Hidden Layers," Proceedings of the 1990 International Joint Conference on Neural Networks, 1990, Vol. 2, 157-162.

Tam, K.Y., "Neural Network Models and the Prediction of Bank Bankruptcy," Omega, The International Journal of Management Science, 1991, 19, 429-445.

Tam, K.Y. and M.Y. Kiang, "Managerial Applications of Neural Networks: The Case of Bank Failure Predictions," Management Science, 1992, 38, 926-947.

Tang, Z., C. de Almeida, and P. Fishwick, "Time Series Forecasting Using Neural Networks vs. Box-Jenkins Methodology," Simulation, 1991, Vol. 57, 303-310.

Wasserman, P.D., Neural Computing: Theory and Practice, Van Nostrand Reinhold: New York, 1989.

Weigend, A., B. Huberman, and D. Rumelhart, "Predicting the Future: A Connectionist Approach," International Journal of Neural Systems, 1990, 1, 193-209.

Widrow, B. and S.D. Sterns, Adaptive Signal Processing, Englewood Cliffs, NJ: Prentice-Hall, 1985.

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

Figure 1  
The Sigmoid Transfer Function  
(Adapted from Wasserman, 1989)

Figure 2  
Descent Down an Error Surface for Two Weights ( $W_0$  and  $W_1$ )  
(Adapted from Rumelhart and McClelland, 1986)

Figure 3  
An Artificial Neural Network