NEURAL NETWORK MODELS USING SAS ENTERPRISE MINER

DIANA BINTI ABDUL WAHAB
Faculty of Economics and Administration
Universiti Malaya
diana.abdwahab@um.edu.my

Neural network is advocated to outperform traditional methods in time series analysis. This paper intends to investigate the dominance of neural network to methods such as exponential smoothing, regression and Box-Jenkins using the same ground rules. The models are developed using large number of historical data with dynamic patterns to predict the future by dividing the data set into training and validating set. Models are then compared based on their forecast accuracy in the validation sample, and model with the least forecast error is selected. Significant emphasis in this paper is on the use of one of the powerful data-mining method in SAS module called Enterprise Miner. The modeling process incorporates the Enterprise Miner SEMMA methodology which stands for Sampling, Exploring, Modifying, Modeling, and Assessing data. Lastly, this paper will conclude with a brief summary of the advantages and disadvantages of using neural network model in time series forecasting.

Keywords: SAS Enterprise Miner, neural network, time series.

1. Introduction

Forecasting for time series data is crucial in operations and future planning. The advancement in computerization and the emergence of online data collection have lead to data deluge, which is resulted in damped data containing uncovered pattern and information. In forecasting time series data especially in economics and financial sector, involves high fluctuated data which is not easy to model. Traditional methods such as the M-competition model use historical data to explain and predict the future. In this paper, we are examining the advantages of neural network over the traditional method in capturing the irregular pattern in time series. The data used in this paper is collected over the period of 10 years.

2. Forecasting Using Classical Decomposition

Time series is a sequence of data points where the study of time series involves the process of understanding the series and its underlying contexts in order to make accurate forecast. In theory, time series realization begins with infinite past and continues to infinite future (Diebold, 2006). If the structure changes over time, there is no way to predict the future. In classical methodologies, the ground rule is first to identify the pattern of a time series data.

Time series has significant impact on decision making in diverse areas of application: operations planning, financial management, demographic, marketing, business and economics. Time series has many different forecasting method but the principles underlying the forecasts are identical (Diebold, 2006). It has dynamic pattern need to be understood; consists of trend, seasonal, cycle and irregular pattern in additive or multiplicative algorithm which need to be identified then extrapolate these components to make forecast. The selection process of a model includes the identification of patterns, evaluation of fit and finally evaluation of forecast.

Many methodologies are designed to explain the patterns. The best method to forecast should be the simplest yet the error term follows white noise using KISS methodology (Diebold), Keep It Sophisticatedly Simple. 'Classical decomposition's performance as a forecasting technique, however, rates as very disappointing' (Hansen, 2003). The inadequacies of traditional forecasting methods, as explained by Hansen (2003) are: the traditional procedure may yield biased components and inadequately extrapolate the components into the future. Third, the irregular component may contain important residual information. Finally, the simple recombination of components implied by additive and multiplicative classical decomposition may not be optimal (Hansen, 2003).

The data used in this experiment is a series of spot prices for foreign currency in Australian dollars and the exchange rates of the currency relative to US dollar over the period of 10 years. Changes in the daily spot price over a period of 1500 daily returns begin on 10/9/86. Figure 1 represents the time series plot for the daily return for Australian dollar in relative to US dollar.

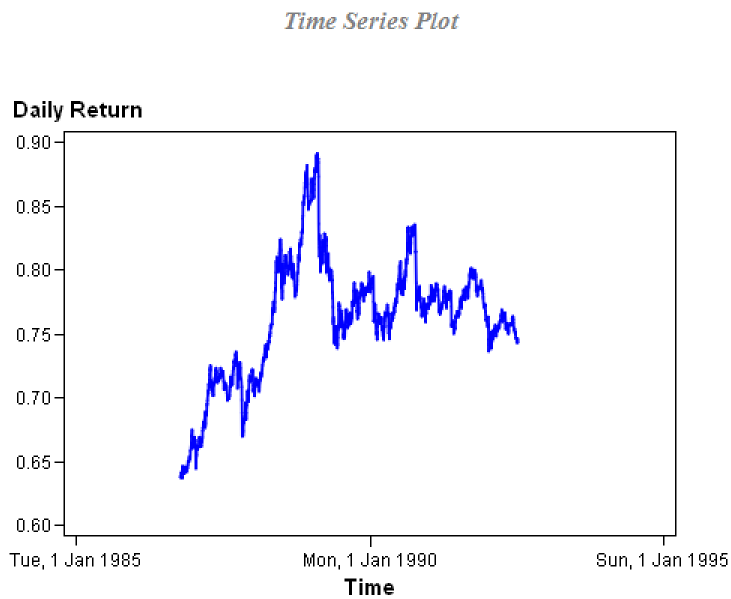

Figure 1: Time series plot of Australian dollar.

The time series plot for the daily returns of Australian dollar shows a slow evolution of trend from Jan 1985 until Jan 1990 but it becomes more stable after then. The plot suggests some cyclical pattern with high fluctuation and irregular fashion. High returns can be seen in early 1989. The mean of the returns is 0.7617 in a range of 0.2546.

Exponential smoothing method is used to capture the pattern and forecast the future prediction. It captures the irregular pattern in time series data. In exponential smoothing method, the most recent data are weighted more heavily than data in the early part of the series to make forecast in the future. The weight function is given by

$$w_\tau = w(1-1)^{t-\tau}$$

Table 1 shows the sum squares of error and mean square error for four types of exponential smoothing method. From the output given, Holts-Winter method is the best method suggested because it heeds the seasonal effect of the time series data.

Table 1: Sum of squares of error and mean square error

Exponential Smoothing methods	Sum of Squares Error	Mean Square Error
Single exponential	0.0409	0.00272484
Double exponential	0.0355	0.00002304
Holts-Winter method	0.0289	0.00001936
Holts-Winter with multiplicative	0.0289	0.00001936

Since the time series plot of data shows dynamic pattern over large amount of data, we can not just use exponential smoothing method because of the lack of its ability to capture the irregular component of the pattern. ARMA model suggest that the series might be affected by its past value, the error past value and also the current and past values of other time series. Basically it involves three phases: identification, estimation and forecast.

For an initial modeling, we try to predict the time series using a simple linear regression. Durbin-Watson statistics shows that the model is highly autocorrelated. The forecast model selection criterion Akaike Information Criterion and also Schwarz Information Criterion show negative values. However, the explanatory variable time does not give a significant impact on the model. Using the log data to smooth out the cyclical and irregular component, the estimation result shows no difference. The errors are still highly correlated with Durbin-Watson statistics value is 0.0103. Next we are going to find a parsimonious ARMA(p,q) model that fits the data well and to gauge its adequacy. The MSE for each model developed is shown in the following table:

Table 2: ARMA(p,q) and its mean square error

	MA		
AR	**0**	**1**	**2**
0	-	0.03	0.029
1	0.03	0.004	0.004
2	0.03	0.004	0.004

The results shown in Table 2 suggest that the Box-Jenkins methodology can capture the fluctuations in the data. The errors follow white noise and the coefficient of determination is 99% at ARMA(2,2).

3. Neural Network In Time Series

Neural network is one of the new generation method that has advantage in analysis of trend and patterns, besides knowledge based system and genetic algorithm (Lin, 1995). Neural network gains its superiority in capturing nonlinear trend, which is most useful in forecasting highly fluctuated economics and financial data.

Neural network gains its popularity because of the adaptation from biological neural system (Walsh, 2005). The function *f(x)* is a composition of another function *g(x)* which is also a composition of other functions (Wikipedia, 2008). Artificial neurons have input paths just as biological neurons have dendrites; they have output paths just as biological neurons have axons (Hill, 1996). Basic architecture is called hidden units - the input variables are fully connected to the first hidden layer, and the hidden layers are then fully connected until the last hidden layer and the last hidden layer is fully connected to the output (Walsh, 2005).

Lin (1995) discussed scheme for time series forecasting at three phases. It starts with detecting the input pattern. The autocorrelation function graph is used to identify the right pattern. The pattern suggested by the autocorrelation function gives information about the dynamic structure of a time series model. Then, it is followed by determining number of neurons in hidden layer(s) using Baum Haussler rules. Finally, construct the neural network forecaster. Using a simulation approach, Lin (1995) concluded the three phases as the effective scheme.

Sarle (1994) likened the process if training a neural network model to a kangaroo searching for the top of Mt Everest, which is denoted for the global optimum. Reaching at the top of other really tall mountain is also acceptable. However, the top of small hill, which is called as bad local optimum would not be acceptable. The optimization process can also be applied to minimization by multiplying the error measure by -1 (Sarle, 1994).

Learning occurs and alter the neurons (Hill, 1996), and this process is called back propagation. Back propagation is 'a method for computing the error gradient for multilayer perceptron, a straightforward application of the chain rule of elementary calculus' (Sarle, 1994). The process adjusts the weights so that the model will have a least square error.

Synapses of neurons are modeled as weights. The strength of the connection is the value of the weight. In the kangaroo case, 'the compass direction represents the values of synaptic weights in the network' (Sarle, 1994). Training the network with more than two weights will results in multidimensional landscape, which is more complicated to visualize. According to Sarle, initial weights play an important role in determining the level of complication of searching the global optimum. Imagine putting an initial weight as dropping a kangaroo on any land on earth. It will be much better to drop the kangaroo on Himalaya rather than let it 'drowned in the Indian ocean' (Sarle, 1994).

Neural network calculates the arithmetic mean differently. As the mean is calculated by dividing a sum of a number to its count, neural network calculate the mean by setting an initial value and then calculate the mean through its learning process over many iterations (Sarle, 1994). If the number of the sample is large enough, the estimate can reach near the true mean.

The time series plot is represented through a histogram using Enterprise Miner, which is shown in Figure 2:

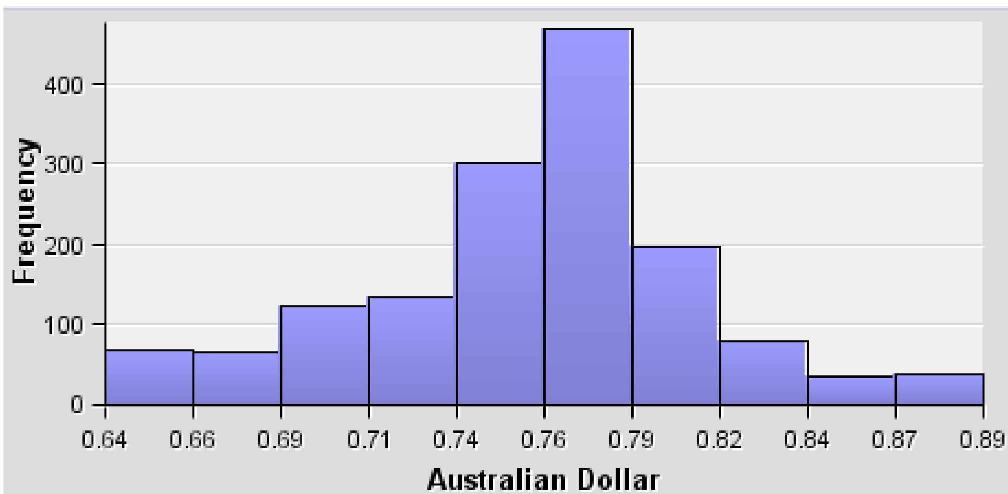

Figure 2: Histogram for Australian Dollar

The sample is then partitioned into two groups; the training set and the validation set. Using 75 per cent of the total sample as training set, neural network estimates the weight of its synaptic coefficient through the learning process. The remaining 25% will be used as validation set which will test the accuracy of the prediction.

Figure 3 shows the modeling process of neural network forecast model. The data node is connected to statExplore to investigate the descriptive statistics of the variable at hand. Then, the data is partitioned into two groups of training and validation set. Finally, the node is connected to the Neural Network node with MLP architecture. The model selection criterion used is misclassification.

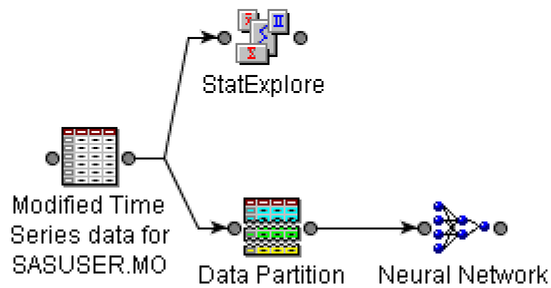

Figure 3: Neural Network modeling for time series forecasting.

The MSE from neural network model is 0.00082, much smaller than the traditional methods shown in the earlier part of this paper. The mean for predicted score between the training and validation set is also very near, indicates that the forecast accuracy is good.

4. Evaluation And Comparison

In universal application, most criteria find the model with the smallest out-of-sample 1-step-ahead mean squared prediction error to be the best forecast model. However as we include more variables, the MSE wont rise. It typically falls continuously as more variables are added. Thus, MSE is not consistent and is very unlikely we get coefficient of nearly 0 on a newly added variable though the coefficient is actually 0 in population (Diebold, 2006). So we use more consistent method: AIC and SIC, to reduce bias associated with MSE. We also look at Durbin Watson statistics to see if the regression disturbance is serially correlated. This is done in order to examine pattern in forecast error.

Hill et al (1996) compared neural network model to six time series method generated in major forecasting competition and found out that neural network did a significant better than the traditional method. Some of the limitations in the traditional method which neural network can tackle are misspecification of functional form of independent and dependent variables, failure to make necessary data transformation, outliers lead to biased estimate of model parameters, unable to capture nonlinear pattern and do not learn incrementally as new cases added into the model (Hill et al, 1996).

According to Hill et al (1996), neural network performs better because it is a universal approximator of each models suggested by hidden layers. It can also capture the nonlinearity of relationship. Hansen (2003) examines performance of neural network in forecasting time series and assess the use of stacked generalization as a way of refining this process. The methodology is applied to four economic and business time series. The main focus is on combining tools from the statistical community with neural network to gain better forecast accuracy.

The ability to generalize and its ability in fault tolerance give neural network a competitive edge in dealing with incomplete data and missing values. In fault tolerance, ‘when some of the neuron malfunction, neural network can still produce approximately correct output’ (Sarle, 1994). When a neural network fails, it can continue without the need to reprogramming. Neural network generates its own rules through training examples.

The advancement in computing also helps to make the neural network works faster on its parallel processing. With the adaptive system, neural network change its structure based on the information that go through network during the learning phase.

The mean square error (MSE) is used to judge the performance of the models. From the outputs given in the earlier part of this paper, we can see that neural network outperform the traditional methods with a much smaller value of MSE. The forecast model selection criterion AIC and SIC also suggest that neural network is the best method.

5. Conclusion

Neural network has been advocated to be the best prediction model for data that is highly fluctuated and has dynamic pattern. The investigation over the advantages of neural network shows that the model is adaptive enough to capture the dynamic pattern in the data at hand. However, behind the powerful ability of neural network, it actually requires high processing time to reach to the global optimum. Neural network also requires training in order to operate.

6. References

Hansen J. V. and Nelson R. D. (2003), Forecasting and recombining time-series components by using neural networks, Journal of the Operational Research Society, Vol. 54, No. 3, pp. 307 - 317

Hill T., O'Connor M. and Remus W. (1996), Neural Network Models for Time Series Forecasts, Management Science, Vol. 42, No. 7, pp. 1082-1092

Koskivaara E. (2000), Artificial Neural Network Models for Predicting Patterns in Auditing Monthly Balances, The Journal of the Operational Research Society, Vol. 51, No. 9, (Sep., 2000), pp. 1060-1069

Diebold F. (2006), Element of Forecasting, 4th Ed.