RESEARCH ARTICLE

Time series regression and artificial neural network approaches for forecasting unit price of tea at Colombo auction

H.A.C.K. Hettiarachchi [*] **and B.M.S.G. Banneheka**
Department of Statistics and Computer Science, Faculty of Applied Sciences, University of Sri Jayewardenepura, Gangodawila, Nugegoda.

Abstract: Tea export plays a vital role in the Sri Lankan economy. It is of immense importance to forecast the prices in the Colombo Tea Auction Center (CTAC) at which a majority of the Sri Lankan tea is marketed. There was no evidence of former studies on forecasting prices of tea at CTAC. The most familiar and the standard practice in the conventional context for forecasting a series varying with time is the building of time series models based on the stationarity and the characteristics of the relevant series, which are autoregressive (AR) terms and moving average (MA) terms. But the auction prices of tea are inherently noisy, non-stationary and chaotic in nature and therefore, the conventional methods cannot be applied. Alternatively, time series regression with generalized least squares and artificial neural network (ANN) were identified as two suitable methods for forecasting the price for a unit of Sri Lankan tea at the CTAC one month ahead. Models were fitted using the prices in 160 months at seven tea auction centers worldwide and assessed and compared using the mean absolute percentage error (MAPE), mean squared error (MSE), coefficient of determination and correlation coefficient between observed and fitted values. Both methods were found to perform well, ANN performing slightly better.

Keywords: Artificial neural network, chaotic series, Colombo tea auction, forecasting prices, time series regression

INTRODUCTION

Tea industry is crucial to the Sri Lankan economy, contributing to a noteworthy amount to the gross domestic product (GDP). Ceylon tea from Sri Lanka is reputed as the best tea for more than a century. Sri Lanka produces a diversity of tea products with varying flavours and aromas, synonymous with quality. As the

4th biggest tea producing country globally, Sri Lanka has a production share of 8.5 % in the international sphere. It is the world's 2nd largest tea exporter with a share of around 18.3 % of the global demand (Sri Lanka Tea Board Statistical Bulletin, 2007). Sri Lankan tea is exported to the global market in many forms such as in bulk form, tea packets, tea bags, instant tea and green tea. The major export destinations of Sri Lankan tea are Russia, UAE, Syria, Iran and Turkey. Sri Lankan tea is marketed mainly through the weekly auctions held at the Colombo Tea Auction Center (CTAC). There are eight internationally renowned tea auction centers currently operating around the world, namely, Colombo - Sri Lanka; Kolkata - India; Cochin - India; Guwahati - India; Chittagong - Bangladesh; Mombassa - Kenya; Jakarta - Indonesia and Malawi - South Africa. (Sri Lanka Tea Board Statistical Bulletin, 2007).

It is very useful for the tea producers to have an idea on how much they are going to earn in the forthcoming month for their crops. However, previous studies on methods for forecasting the auction price of tea are sparse. Dharmasena (2003) has tested the ability of forecasting prices at some of the above auction centers using vector autoregression (VAR) models and found that for most markets a random walk forecast outperforms the VAR generated forecasts. Apart from this, no other study for forecasting the tea auction prices could be found. Artificial neural network (ANN) method has been intensively used for forecasting many series such as the short-term electricity prices (Catalão *et al.*, 2006), next-day price of electricity in the energy market (Pino *et al.*, 2008), precipitation (Mar & Naing, 2008), and

commodity prices (Kohzadia *et al.*, 1996). However, no application of ANN was found in the area of forecasting the prices in tea auctions.

This study was performed with the primary goal of finding suitable methods to obtain a one-month ahead forecast for the price of a unit of Sri Lankan tea at the CTAC. The most familiar and the standard practice in the conventional context for forecasting a series varying with time is the building of time series models based on the stationarity and the characteristics of the relevant series, which are autoregressive terms (AR) and moving average terms (MA). But auction prices are inherently noisy, non-stationary and chaotic in nature. Therefore alternative methods are required for the above purpose. The auction prices at the CTAC could not solely compensate efficiently for forecasting its price in the subsequent month due to its extremely chaotic nature. Therefore the prices at other auction centers were also used for forecasting the auction prices at CTAC more rigorously.

The remainder of this paper is organized as follows. Section 2 presents the methods and material employed in the study. This has two sub sections, which portray the two different approaches explored for forecasting. Section 3 recapitulates the entire study stating the conclusions, which have been drawn from the analysis and also a discussion about the study. Finally the conclusion and the recommendations are presented in Section 4.

METHODS AND MATERIALS

Seven auction centers for which a considerable amount of data existed namely, Colombo, Kolkata, Cochin, Guwahati, Chittagong, Mombassa and Jakarta were employed in the study. Auction prices at these centers were available from January, 1997 to May, 2010. Although the auctions are held weekly, only the monthly prices (which are the averages of the weekly prices during a month) were available for centers other than Colombo. Therefore, the monthly prices were used for this study. All the prices were considered in US dollars.

Powers and lags of the prices at other auction centers having the highest correlations with the Colombo price were determined through a computer programme written in R (R Development Core Team, 2010). The prices at the above selected lags, raised to the above selected powers, were used as explanatory variables to predict the price at Colombo. Two models using different approaches were developed and compared.

Time series regression approach

The dataset was divided into two subsets, one consisting of the first 136 records for building the model (train set) and the other consisting of the last 24 records (test set) for validating it. Initially in this study an ordinary least square regression model was fitted for data from CTAC using the explanatory variables described above. The residuals of this model were non-normal and serially correlated, violating the basic underlying assumptions of a linear regression model.

Normality of the residuals could be achieved through Box-Cox transformation.

$$Y_t^{\lambda} = \beta_0 + \beta_1 X_{1,t-k_1} + \beta_2 X_{2,t-k_2} + \beta_3 X_{3,t-k_3}$$

$$+ \beta_4 X_{4,t-k_4} + \beta_5 X_{5,t-k_5} + \beta_6 X_{6,t-k_6} + N_t$$

$$...(1)$$

Here Y_t^{λ} is the auction price at CTAC at month t, transformed using the Box-Cox method and $X_{i,t-k_i}$'s ($i = 1,2,...,6$ and k_i's are the selected lags resulting in highest correlations between CTAC and other auction centers) are the auction prices at other auction centers mentioned above at month t-k_i raised to selected powers. The value of λ, which results in normal residuals was determined to be equal to -2 using the method of maximum likelihood according to the Box-Cox transformation technique.

Here N_t exhibited a ARIMA(1,1,0) pattern. In such cases the parameter estimates yielded by ordinary least squares method have low reliance (Tsay, 1984). The problem of autocorrelations could be remedied by transforming the explanatory variables using the Hildreth-Lu procedure (Neter *et al*., 1990). However, this dramatically reduced the coefficient of determination. Therefore as an alternative method, a regression model was fitted using the method of generalized least squares applying a ARIMA(1,1,0) model for the errors. For this model also the initial explanatory variables were same as above and the response variable was Box-Cox transformation of Colombo prices. The final model was fitted dropping the insignificant explanatory variables.

Artificial neural network (ANN) approach

ANN is a popular method used for forecasting non-stationary time series in the modern world. One of the major advantages of neural networks is that,
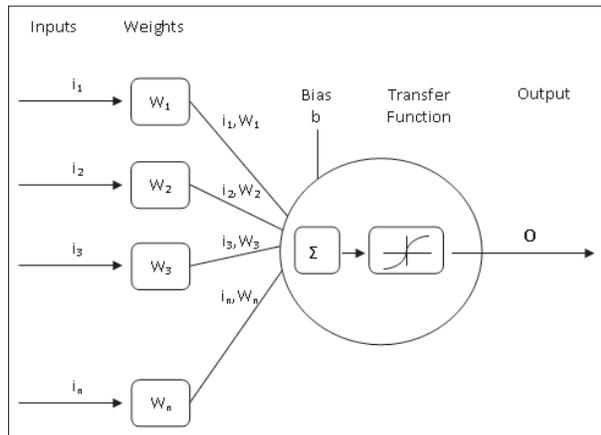
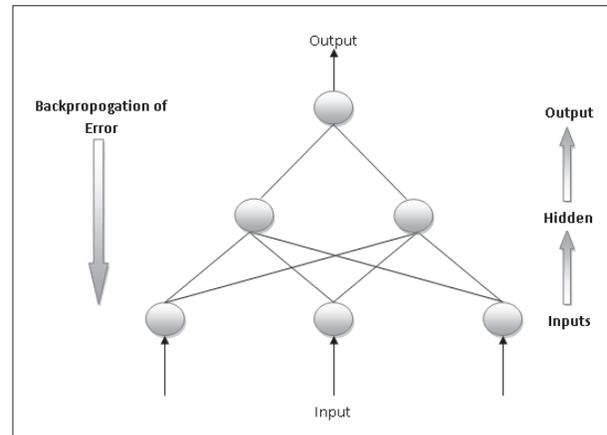**Figure 1:** Graphical representation of a neuron

Figure 2: Backpropogation algorithm

theoretically, they are capable of approximating any continuous function, and thus it is not required to have any hypotheses about the underlying model. On the other hand, the final neural network solution depends on the initial conditions of the network, and it is not easy to interpret the solution in traditional, analytic terms, such as those used to build theories that explain phenomena. Neural networks should not, however, be heralded as a substitute for statistical modeling, but rather as a complementary effort or an alternative approach to fitting non-linear data.

ANN is made up of simple processing units (neurons), which have the ability to learn functional dependencies from data. Each neuron is a simple processing unit which receives some weighted data, sums them with a bias and calculates an output to be passed on (Figure 2). The function that the neuron uses to calculate the output is called the activation function. The manner in which the neurons of a neural network are structured is intimately linked with the learning algorithm, which is used to train the network. The most common architecture is the multilayer perceptron (MLP). These networks are a feed forward network where the neurons are structured in one or more hidden layers. Each perceptron in one layer is connected to every perceptron on the next layer; hence information is constantly "fed forward" from one layer to the next.

As in the first method the dataset was divided into two, such that the 'train set' consists of 134 records and the 'test set' consists of 24 records. Here the records were assigned to each set randomly since this method does not require them to be in a timely order as in the first method.

The normalized values of the six explanatory variables of the previous model along with the one lagged behind Colombo price series were the inputs to the ANN. The inputs varied in a wide range, hence they were normalized to allow a faster training and reduce the chances of being misled by local optima. Forecast price at the CTAC in the forthcoming month was the output. Thus, in this context, the ANN consisted of seven input nodes and a single output node. A feedforward neural network was build and trained for the 'train set' varying the number of hidden layers, number of neurons in each hidden layer and transfer functions, etc. The complexity of an ANN increases and convergence slows down as the number of hidden layers in the network increases. Therefore, the number of hidden layers was varied from one to three while keeping the other factors constant and thereby selected the optimum number of layers. The number of layers that resulted in forecasts, which had least mean square error and highest correlation with the target values was determined as the optimum number. The number of neurons in each hidden layer was also determined in the same procedure.

The resulting network was trained for the train dataset using the Levenberg-Marquardt (LM) backpropagation algorithm. In the backpropagation algorithm, the combination of weights, which minimizes the error function is considered to be a solution of the learning problem. From many algorithms for training ANN, LM algorithm was selected in our study as it has been found that it is advantageous for training ANN owing to its shorter training duration and more satisfactory performance criteria (Cigizoglu & Kisi, 2005).

ANNs for forecasting the price of Sri Lankan tea at the CTAC two months ahead and three months ahead was also created using the same procedure.

RESULTS

The determined autocorrelation coefficients of CTAC and cross correlation coefficients with other auction centers at different lags are presented in Table 1. Accordingly there is a significant correlation in the auction prices between CTAC and other auction centers.

The final model resulted from the time series regression approach is

$$\widehat{Colombo_t} =$$

$$\{-0.1241140 \left((Cochin^{0.5})_{t-1} - 1.2665 \, (Cochin^{0.5})_{t-2} + 0.2665 \, (Cochin^{0.5})_{t-3}\right) +$$

$$0.0045484 \left((Guwahati^{2.5})_{t-1} - 1.2665 \, (Guwahati^{2.5})_{t-2} + 0.2665 \, (Guwahati^{2.5})_{t-3}\right) + 0.2665205 \, (Colombo^{-2})_{t-1} -$$

$$0.2665205 \, (Colombo^{-2})_{t-2}\}^{-1/2}$$

$$...(4)$$

Table 1: Autocorrelation coefficients of CTAC and cross correlation coefficients with other auction centers at different lags

| Lag	Colombo	Kolkata	Cochin	Guwahati	Chittagong	Jakarta	Mombassa
0	1	0.634379	0.681698	0.63378	0.709015	0.757966	0.75142
1	0.961869	0.619478	0.651245	0.613466	0.663064	0.729918	0.723532
2	0.911345	0.623816	0.610814	0.604512	0.625507	0.693588	0.66626
3	0.852035	0.634519	0.578287	0.61605	0.586266	0.656243	0.603773
4	0.80096	0.601148	0.559321	0.589839	0.153037	0.625485	0.550811
5	0.752635	0.552673	0.541648	0.545048	0.483104	0.599792	0.513702
6	0.712905	0.493042	0.518047	0.48299	0.44893	0.561854	0.475381

(a)

(b)

(c)
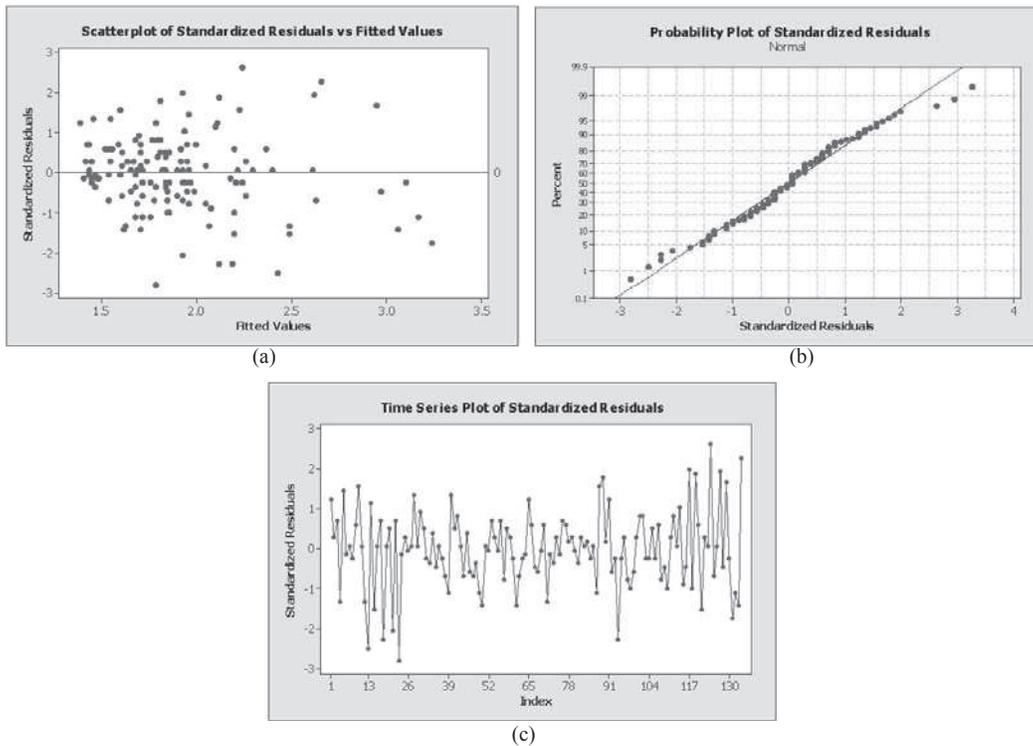
Figure 3: Plots of model adequacy of the time series regression model (a) plot of standardized residuals against fitted values; (b) normal probability plot of standardized residuals; (c) time series plot of standardized residuals
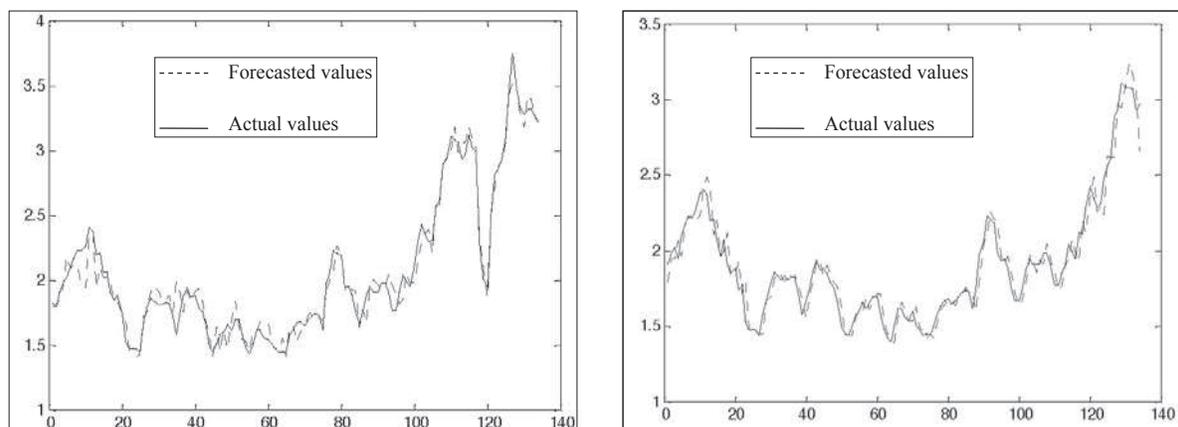
Figure 4: Plots of forecasted and actual values against the index for (a) artificial neural network (ANN) approach and (b) time series regression approach

Table 2: Comparison of the models

| | Train set | | Test set | |
	Time series regression	Artificial neural network	Time series regression	Artificial neural network
MAPE	0.03529	0.0351	0.08524	0.0610
MSE	0.0088	0.0098	0.100727	0.0368
ρ	0.9706	0.9841	0.8121	0.9466
R^2	92.89 %	96.84 %	74.09 %	90.25 %

MAPE – Mean absolute percentage error

MSE – Mean square error

r – Correlation coefficient between the observed values and the fitted values

R^2 – Coefficient of determination

Table 3: Prediction performances of ANNs for forecasting one month, two months and three months ahead auction price

| | One month | | Two months | | Three months | |
	Train set	Test set	Train set	Test set	Train set	Test set
MAPE	0.0351	0.0610	0.0499	0.1229	0.0499	0.1856
MSE	0.0098	0.0368	0.1190	0.1750	0.1190	0.4315
ρ	0.9841	0.9466	0.9687	0.8945	0.9682	0.7972
R^2	96.84 %	90.25 %	93.83 %	83.2 %	93.83 %	73.13 %

MAPE – Mean absolute percentage error

MSE – Mean square error

r – Correlation coefficient between the observed values and the fitted values

R^2 – Coefficient of determination

Coefficient of determination of this model is 92.89 %. This model satisfied the underlying assumptions of regression analysis (Figure 1). The normality of the standardized residuals was significant at 0.046 and Durbin Watson statistic (equal to 2.10) verified that residuals are uncorrelated.

An ANN with two hidden layers having six and four neurons respectively, which produced output values that were highly correlated with the target values, was selected as the best model. The coefficient of determination of the model was 96.84 %. The plot of forecasted values and actual values against the index is presented in Figure 4.

A comparison between the two models developed for forecasting the auction price in the subsequent month was performed (Table 2).

ANNs for forecasting the price of Sri Lankan tea at the CTAC two and three months ahead was also created using the same procedure and the results are presented in Table 3. According to Table 3 the correlation coefficients between outputs and targets of above two ANNs respectively, are 0.9687 and 0.9682.

DISCUSSION AND CONCLUSION

The prices prevailing at the Jakarta auction center mostly correlated to the prices prevailing at the Colombo auction center where the correlation coefficient is 0.758. Mombassa and Chittagong has the second and third largest correlation coefficients, respectively. Models developed using both time series regression and artificial neural network approaches performed well with coefficient of determinations of 92.89 % and 96.84 %, respectively. However, based on the coefficient of determination, MAPE and correlation coefficient between the target values and the output values for both train set and test set, it can be concluded that the ANN approach performs slightly better than the time series regression approach.

The quality characteristic of tea are not incorporated into this study. To extend this study, one can incorporate the price structure based on different quality of tea traded in each market.

Further, due to restrictions on the availability of data the ANN was not validated in this study. In future studies, if more data can be collected, the ANN could be validated making it much more efficient. It is recommended that future researchers conduct a sensitivity analysis for improvement of the ANN. The researchers could optimize the ANN especially through effective use of pruning algorithms and can incorporate magnitudebased pruning, which eliminates unwanted links and skeletonization, which eliminates unwanted nodes.

REFERENCES

1. Catalão J.P.S., Mariano S.J.P.S., Mendesb V.M.F. & Ferreira L.A.F.M. (2006). Short-term electricity prices forecasting in a competitive market: a neural network approach. *Mendeley* **77**(10): 1297 − 1304.

2. Cigizoglu H.K. & Kisi Ö. (2005). Flow prediction by three back propagation techniques using k-fold partitioning of neural network training data. *Nordic Hydrology* **36**: 49 − 64.

3. Dharmasena K.S.D.B. (2003). *International Black Tea Market Integration and Price Discovery*. Texas A&M University, Texas, USA.

4. Kohzadia N., Boyd M.S., Kermanshahib B. & Kaastrac I. (1996). A comparison of artificial neural network and time series models for forecasting commodity prices. *Neurocomputing* **10**(2):169 − 181.

5. Mar K.W. & Naing T.T. (2008). Optimum neural network architecture for precipitation prediction of Myanmar. *World Academy of Science, Engineering and Technology* **48**: 130 − 134.

6. Market Intelligence and Resource Division of Sri Lanka Tea Board (2007). *Sri Lanka Tea Board Statistical Bulletin*. Sri Lanka Tea Board, Colombo.

7. Neter J., Wasserman W. & Kutner M.H. (1990). *Applied Linear Statistical Models*, 3rd edition. Erwin, Boston, USA.

8. Pino R., Parreno J., Gomez A. & Priore P. (2008). Forecasting next-day price of electricity in the Spanish energy market using artificial neural networks. *Engineering Applications of Artificial Intelligence* **21**(1): 53 − 62.

9. Tsay R.S. (1984). Regression models with time series errors. *Journal of the American Statistical Association* **79**(385): 84 − 96.

10. R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. *http://www. Rproject. org*