www.irmbrjournal.com December 2017
International Review of Management and Business Research Vol. 6 Issue.4

Forecasting of Pakistan's Import Prices of Black Tea Using ANN and SARIMA Model

MUHAMMAD NOUMAN HUSSAIN
Department of Statistics, University of Lahore, Sargodha, Pakistan
Email: noumanhussain472@gmail.com
Tel: +923206622341

ASAD ALI
Department of Quantitative Methods, SBE, University Of Management & Technology, Lahore, Pakistan
Email: F2017204002@umt.edu.pk
Tel: +923041611166

Abstract

Forecasting is very helpful tool for making decisions and plans for upcoming time periods. In this study Seasonal Auto Regressive Integrated Moving Average (SARIMA) and Artificial Neural Networks (ANN) models are used to forecast the Pakistan's import prices of black tea by using data for the time period Jan 2004 to Dec 2014. The performance of SARIMA and ANN models are evaluated on the basis of root mean square error (RMSE), mean square error (MSE) and mean absolute error (MAE). The selected model under Box-Jenkins is SARIMA (0, 1, 1)(0, 1, 1)$_{12}$. ANN models with different combination of input, hidden and output layers were tested with four activation functions (semi linear, sigmoid, bipolar sigmoid and the hyperbolic tangent function). This study reveals that ANN model performed well as compare to SARIMA model because RMSE, MAE and MSE of ANN model are less as compare to RMSE, MAE and MSE of SARIMA model. Therefore, ANN can be effective for forecasting the Pakistan's import prices of black tea.*

Key Words: *SARIMA, ANN, Black Tea, RMSE, MSE, MAE.*

Introduction

Tea is a most common beverage in almost all over the world. It is 2nd most common beverage drunk after the water. It is used and loved by the persons having different ages. More than 3 billion cup of tea are daily used in the world (Tea board Kenya). Pakistan does not produce tea as much as it is consumed. Among the top ten tea importing countries, Pakistan is the 3rd largest importer of tea in the world. The per capita consumption of tea in Pakistan is around one kilogram and is constantly increasing due to increase in demand. So, in terms of tea consumption, Pakistan is the seventh largest country.

Pakistan has a long tradition of tea drinking that has now become an integral part on the country's social life. In Pakistan, a morning and evening cup of tea are an essential part of everyday culture. Even in summer when the weather is extremely hot, tea is drunk twice a day. This tradition has existed from many years. Every guest in a home in Pakistan is most likely to ask for a cup of tea to relax after a journey. This trend has always created a strong demand for tea in Pakistan. Drinking of green tea in Pakistan is very minor, it is imported from four different countries such as Bangladesh, China, Indonesia, and Vietnam.

Forecasting is a decision making tool of predicting or estimating the future based on the past and present observations of the same variable. Forecasting provides information about the possible future events and their cost for the organization. Forecasting may not reduce the future uncertainty but it increases the confidence of management to make an important decision.

Literature Review

Many researchers have used time series analysis to forecast the different variables. Review of some past studies have been done in this section. Abeysinghe (1994) showed that using seasonal dummies in removing seasonality will likely produce spurious regression. Therefore, it was important to identify the nature of seasonality exhibited in the series prior to making any treatment to the seasonality and before deciding on a correct forecast model that can control for the seasonality. Qu and Zhang (1996) applied an autoregressive model to forecast tourist arrivals in 12 tourist destinations. Kanungo, D. P., et al. (2006) discussed that ANN modeling approach had numerous advantages over conventional phenomological, as ANN structure just require data set and don't need to follow any assumption about the underlined data set. Rabenja et al. (2009) utilized seasonal ARIMA and non-seasonal ARIMA for forecasting of monthly rainfall as well as discharge of the Namorona River in the Vohiparara River Basin of Madagascar. Study revealed that seasonal ARIMA was more efficient for forecasting these variables as compare to non-seasonal ARIMA. Gijo (2011) used time series data of demand of tea for India. The underline series was modeled by Box-Jenkins seasonal auto regressive integrated moving average (SARIMA) model. Adequacy of the fitted model was tested by using Lung-Box test criteria followed by residual analysis. Hamidreza and Leila (2012) utilized SARFIMA model to study and predict the Iran's oil supply. The results showed that the best model was SARFIMA $(0,1,1)(0,-0.199,0)_{12}$ which was used to predict the quantity of oil supply in Iran till the end of 2020.

Leila and Masoud (2012) carried out an empirical study of the usefulness of SARFIMA models in energy science. The results revealed that the appropriate model was SARFIMA $(2,1,0)(0,0.473,0)_{12}$. Which was utilized to predict the consumption rates of petroleum products till the end of 2013. Etuk (2013) used ARIMA model under Box-Jenkins methodology to model and forecast Nigeria's monthly inflation rates for the period November 2003 to October 2013 with a total of 120 data points. Hossain & Abdulla. (2015) considered yearly data of tea production in Bangladesh over the period 1972 to 2013. On the basis of AIC and BIC, the most suitable model to forecast the tea productions in Bangladesh was ARIMA (0, 2, 1). Adequacy of the fitted model has been tested using run test and Jarque-Bera (JB) test criteria followed by residual analysis. The comparison between the original series and forecasted series shows the same manner indicating the fitted model behaved statistically well and suitable to forecast the tea productions in Bangladesh.

Ali et al. (2016 a) utilized ARIMA model and compare the mortality rate under five years of Pakistan, Bangladesh and India. Results revealed that mortality rate has decreasing trend in all countries under consideration. By utilizing Box-Jenkins methodology, Ali et al. (2016 b) forecasted the daily gold price. On the basis of model selection criterion and forecasting checks (Mean Absolute Error, Mean Absolute Percentage Error, and Root Mean Square Error) they concluded that ARIMA (0,1,1) was most appropriate model for forecasting of daily gold price. Iqbal et al. (2016) found that for forecasting of wheat production for Pakistan and India, box Jenkins methodology was appropriate. They suggest ARIMA (1,1,1) and ARIMA (1,1,0) for forecasting wheat production of Pakistan and India respectively. Ali and Zubair (2017) compared the ANN and ARIMA model for forecasting of Pakistan's mortality rate. By utilizing mean absolute error and root mean square error they found that ANN was more appropriate for forecasting of Pakistan's mortality rate.

After comprehensive review of some forecasting techniques, we select SARIMA and ANN for forecasting of Pakistan's Import Prices of Black Tea and formulate following objectives for the study:

(1) To make model of import prices of black tea in Pakistan using SARIMA and ANN model.
(2) To Forecast the Pakistan's import prices of black tea with the help of these models.
(3) To compare the performance of SARIMA and ANN model base on their respective RMSE, MSE and MAE.
(4) To identify the best and reliable model for forecasting the future prices of import prices of black tea.

Methodology

The data set under consideration is monthly import tea prices of Pakistan from January 2004 to December 2014. We have taken tea price data from Tea Association of Pakistan.

SARIMA Methodology

In time series analysis, an autoregressive integrated moving average (ARIMA) model is a generalization of an ARMA model. Both of these models are fitted to time series data either to better understand the data or to predict future points in the series (forecasting). ARIMA models are applied in some cases where data show evidence of non-stationary, where an initial differencing step (corresponding to the "integrated" part of the model) can be applied to reduce the non-stationary. The AR part of ARIMA indicates that the evolving variable of interest is regressed on its own lagged (i.e., prior) values. The MA part indicates that the regression error is actually a linear regression of error terms whose values occurred contemporaneously and at various times in the past. The integrated indicate that the data values have been replaced with the difference between their values and the previous values (and this differencing process may have been performed more than once). The purpose of each of these features is to make the model fit the data as well as possible. If in ARIMA model we attach seasonal part than it become seasonal ARIMA (SARIMA) model. The seasonal part of an ARIMA model has the same structure as the non-seasonal part: it may have an AR factor, an MA factor, and/or an order of differencing. In the seasonal part of the model, all of these factors operate across multiples of lag s (the number of periods in a season). A seasonal ARIMA model is classified as an SARIMA (p, d, q)*(P, D, Q)$_s$ model, where P = number of seasonal autoregressive (SAR) terms, D=number of seasonal differences, Q = number of seasonal moving average (SMA) terms. In identifying a seasonal model, the first step is to determine whether or not a seasonal difference is needed, in addition to or perhaps instead of a non-seasonal difference. You should look at time series plots and ACF and PACF plots for all possible combinations of 0 or 1 non-seasonal difference. In general SARIMA can be represented as

$$(B)\emptyset_p(B^s)w_t = \theta_q(B)\tau_q(B^s)z_t \quad (1)$$

$$\emptyset(B^s)\varphi(B)(1-B)^d(1-B^S)^D w_t = \tau(B^s)\theta(B)z_t \quad (2)$$

Where, B denotes the back shift operator, $\varphi_p, \emptyset_p, \theta_q, \tau_q$ are the polynomial functions of p, P, q, Q, Zt denotes the random process. The SARIMA model is generally represent in the form ARIMA (p, d, q)*(P, D, Q)$_s$.

For testing the stationary, we utilized Augmented Dickey Fuller (ADF) test. It is named after the statisticians David Dickey and Wayne Fuller, who developed the test in 1979. Null and alternative hypothesis of ADF test are

Ho: $\delta = 0$ i.e. there is unit root in the series or series is non-stationary
H1:$\delta \neq 0$ i.e. there is no unit root or series is stationary.
Test Statistic of ADF is

$$DF_\tau = \frac{\hat{\theta}}{SE(\hat{\theta})}$$

Decision rule

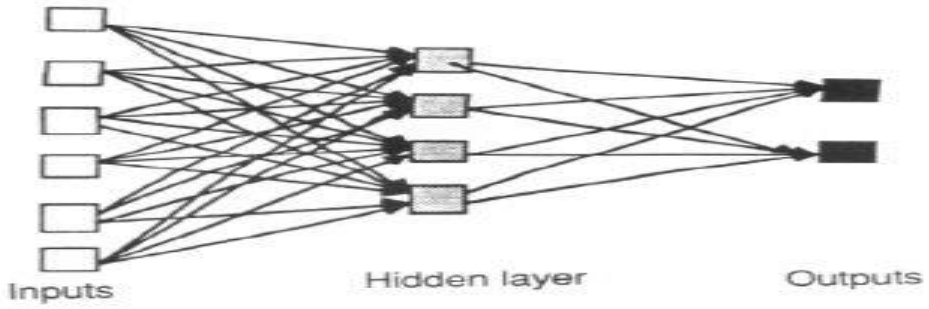
If t* > $ADF_{critical\ value}$ then do not reject Ho i.e. series is not stationary.
If t* < $ADF_{critical\ value}$ then reject Ho i.e. series is stationary.

We find out the values of p and q on the basis of correlogram and then we move toward the estimation step and estimate the unknown coefficients. In this we use maximum likelihood estimation outlined in Box-Jenkins (1976). The maximum likelihood equations are solved by nonlinear function maximization. Back casting is used to obtain estimates of the initial residuals. The estimation process has intensive and iterative calculation, so it often takes a few seconds to obtain a solution. Once a model has been fitted, the final step is the diagnostic checking of the model. The checking is carried out by studying the autocorrelation plots of the residuals to see if further structure (large correlation values) can be found. If all the autocorrelations and partial autocorrelations are small, the model is considered adequate and forecasts are generated. If some of the autocorrelations are large, the values of p and/or q are adjusted and the model is re-estimated.

This process of checking the residuals and adjusting the values of p and q continues until the resulting residuals are random.

Artificial Neural Networks Methodology

Multilayer perception (MLP) model belongs to a general class structure ANN's called feed forward neural network. A feed forward neural network is a basic type of neural network that is capable to approximate both continue and enterable functions. Network architecture of MLP consists of neurons that grouped in layers. The model is characterized by a network which contains three layers that is input, hidden and output layer, connected by acyclic links. Hidden layers are may be more than one. The nodes in these layers are also known as processing elements. There are many neural network models, but the basic structure of ANN involves a system of layered, interconnected nodes and neurons are presented.



For verification of forecasted model, the residuals from the resulting ANN structure are tested and plotted to judge whether the series is uncorrelated or correlated. When the residuals revealed to be uncorrelated, the selected model is then used to forecast drought indices with the lowest criterion value of mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE). While bipolar sigmoid as an activation function is chosen for each tea import price with one month time scale.

Analysis and Discussion

SARIMA

First of all we plot the data and check its behavior, whether there exist a deterministic trend (increasing or decreasing) or a seasonal deterministic pattern.

Figure 1: Pakistan's Import Prices of Black Tea (Tons)

Figure 1 shows the time series plot of Pakistan's monthly import prices of black tea in tons from January 2004 to December 2014. The higher values display more variation than the lower values and there is upward trend in a data.

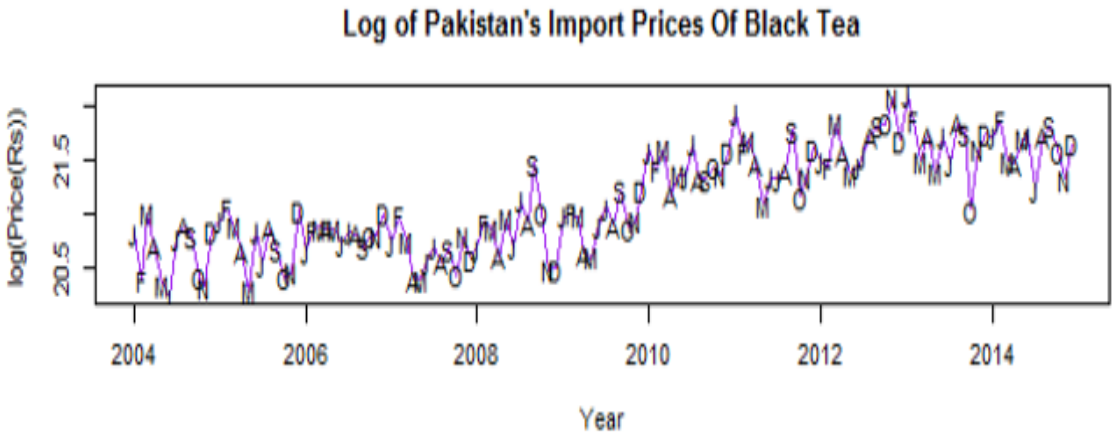


Figure 2: Plot of log Import prices of Black Tea (Ton)

We report the plot of log of black tea prices as well in Figure 2. We notice the amount of variations around the upward trend is much more uniform across high and low values of the series. There is an upward trend but also seasonality can be seen as January is followed by January and similarly for all other months.

Unit Root Test

We used Augmented Dickey Fuller (ADF) test to check the stationary of the series. The results of the ADF test are given in a table 1.

Table 1: Augmented Dickey Fuller (ADF) Test

t-statistics	Prob.	Level of Significance
-2.0684	0.5481	5%

Table 1 shows that data is still non-stationary as P- value is greater than 0.05, so we take integrated difference to make the data Stationary.

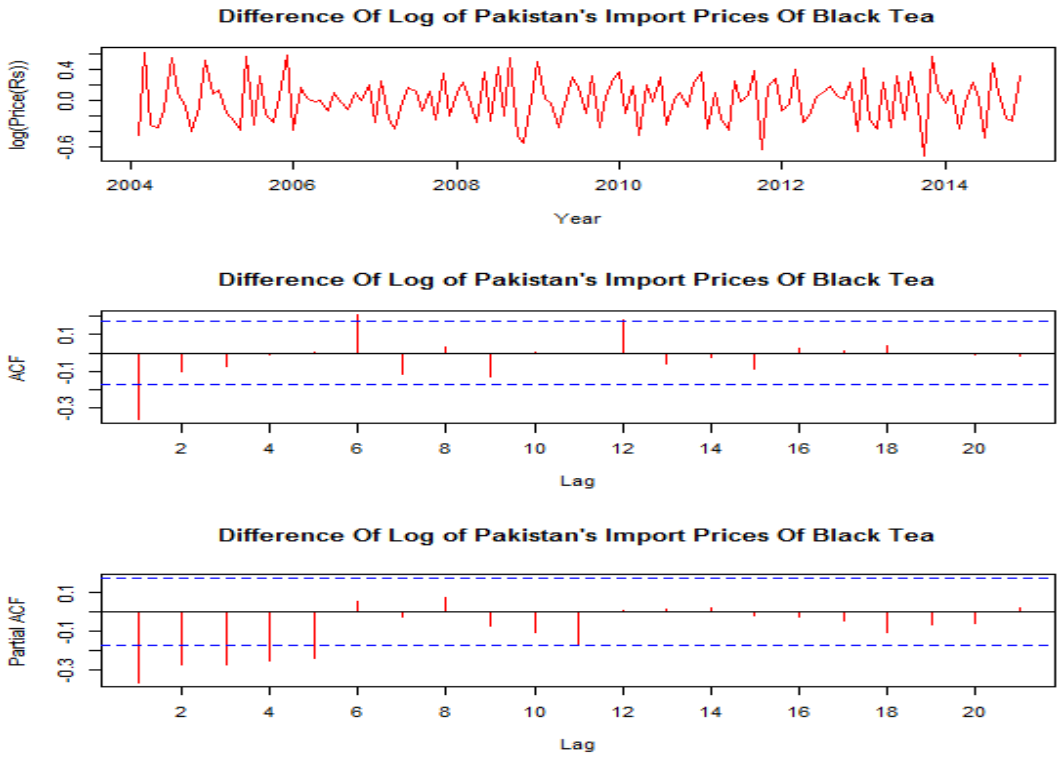


Figure 3: Plot of ACF and PACF of difference of log Pakistan's Import prices of black tea

Figure 3 shows the time series plot of first differenced log data of import prices of black tea. The general upward trend has now disappeared but strong seasonality is still present in a data. Plot of ACF shows significant spikes at lag 1, 6, 12. It strongly suggests that seasonal difference is essential to remove seasonality from data. Figure 4 shows the time series plot after first and seasonal difference has been taken. Now the seasonality is removed and series is stationary. Plot of ACF and PACF suggests that a simple model which incorporates at lag 1 and lag 12 and autocorrelations might be adequate. We report Augmented Dickey Fuller test in table 2.

Table 2: Augmented Dickey Fuller (ADF) Test

t-statistics	Prob.	Level of Significance
-6.6618	0.01	5%

Table 2 shows that the series is now stationary as trend and seasonality is removed from the series. Now we can use this series for forecasting the Pakistan's import prices of black tea.

Model Identification and Estimation

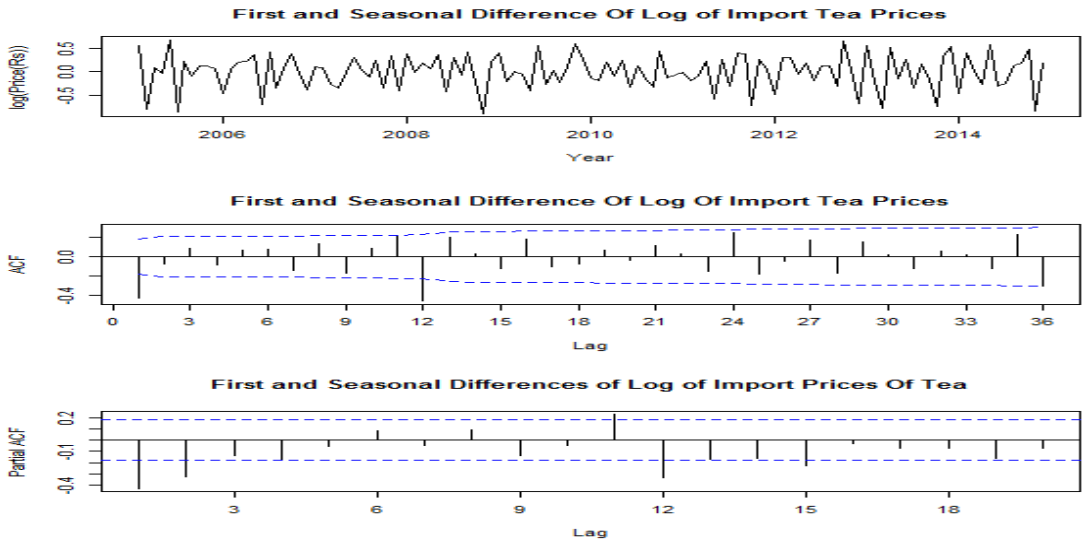


Figure 4: First and Seasonal Difference of log Import Prices of Black Tea

From Figure 4 the plot of ACF has significant spikes at lag 12 therefore seasonal MA model has of order 1 and plot of Partial Autocorrelation function has significant spikes at lags 12 and 24 , this indicate that seasonal AR model has of order 2. So, we proposed tentative models along with their AIC in Table 3.

Table 3: Tentative Models and their corresponding AIC's

SARIMA (p, d, q)*(P, D, Q)	AIC
$(0,1,1)*(0,1,1)_{[12]}$	**180.31**
$(1,1,0)*(1,1,0)_{[12]}$	245.47
$(2,1,0)*(2,1,0)_{[12]}$	214.77
$(0,1,2)*(0,1,2)_{[12]}$	183.74

From Table 3, we can observe that the best fitted Model is ARIMA $(0, 1, 1)*(0, 1, 1)_{[12]}$ as this model possess minimum AIC i.e180.31.

Table 4: Summary Of best Model Fitted

Coefficient	Ma 1	Sma1
Estimate	-0.8640	-1.0000
Standard error (S.E)	0.0521	0.2583
RMSE		**MAE**
0.456		0.275

Table 4 contains the maximum likelihood estimates of best fitted model with corresponding standard errors.

Diagnostic Checking

Correlogram of the residuals of best fitted model is check to see whether they are white noise or not. In diagnostic checking, we check the residuals of the model from ACF plot of residuals. If all spikes are within a limits than our model is adequate.

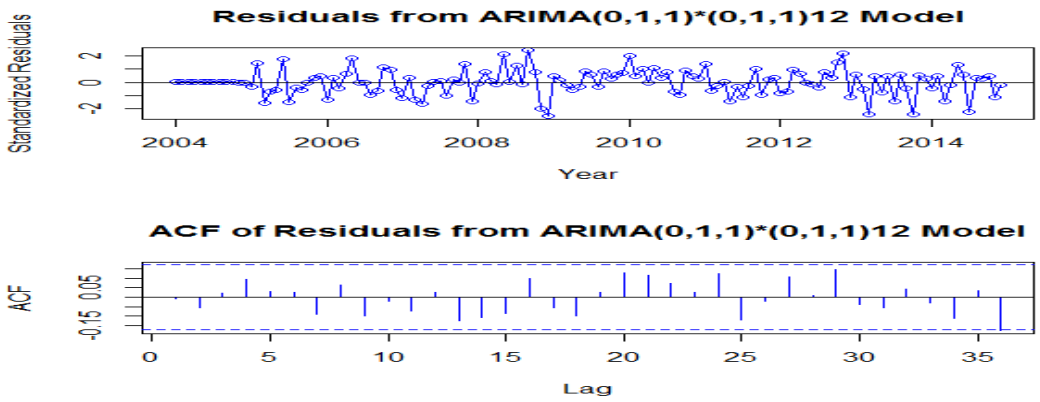


Figure 5: Simple and ACF plots of Residuals of Model SARIMA (0, 1, 1) (0, 1, 1)$_{[12]}$

Figure 5 shows the time series and ACF plot of standardized residuals for seasonal ARIMA model. The plots do not show any major irregularities. So, this model is suitable for forecasting purpose.

ANN

We have taken tea import prices data of Pakistan tea association from January 2004 to December 2014. On the basis of the observed data, we forecast future prices of import for 2015 to 2018. After selection of the appropriate parameters of ANN based on hit and trial method, the forecasting model for each index is then validated.

Table: 5: Summary of ANN Model

Input Layers	Hidden Layers	Output Layers	MSE	MAE
10	10	1	0.105	0.217

Validation of forecasting model is done by using the values of mean absolute error (MAE) and mean square error (MSE) in Table 5.

Diagnostic Checking

Figure 6 shows the residual of the ANN model. The residuals of the ANN model have zero mean and cosntant variance lies between -2.5 to 1.

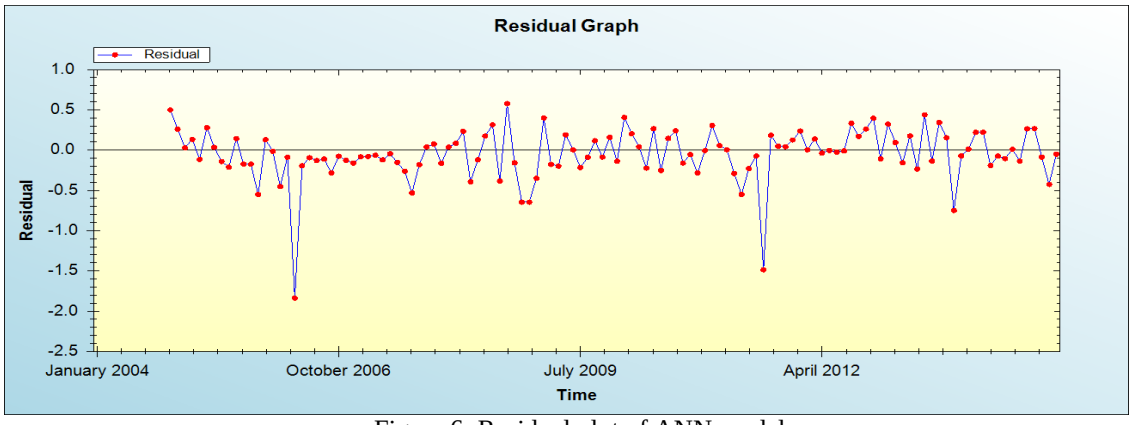


Figure 6: Residual plot of ANN model

Forecasting

Table 6 shows the forecasted values of both models (SARIMA and ANN) for import price of black tea in Pakistan from January 2015 to December 2018.

Table 6: Forecasted Results of Import Price of Black Tea from 2015 to 2018.

Time	SARIMA	ANN	Time	SARIMA	ANN
Jan-2015	21.909	21.475	Jan-2017	22.113	21.590
Feb-2015	21.882	21.705	Feb-2017	22.086	21.597
Mar-2015	21.894	21.627	Mar-2017	22.098	21.598
Apr-2015	21.419	21.495	Apr-2017	21.623	21.602
May-2015	21.582	21.595	May-2017	21.785	21.596
June-2015	21.713	21.777	June-2017	21.917	21.607
July-2015	21.756	21.662	July-2017	21.960	21.610
Aug-2015	21.617	21.674	Aug-2017	21.821	21.618
Sep-2015	21.723	21.645	Sep-2017	21.926	21.634
Oct-2015	21.638	21.648	Oct-2017	21.841	21.615
Nov-2015	21.689	21.695	Nov-2017	21.893	21.609
Dec-2015	21.892	21.699	Dec-2017	22.096	21.611
Jan-2016	22.011	21.579	Jan-2018	22.214	21.603
Feb-2016	21.984	21.625	Feb-2018	22.188	21.601
Mar-2016	21.996	21.666	Mar-2018	22.199	21.598
Apr-2016	21.521	21.662	Apr-2018	21.725	21.602
May-2016	21.684	21.658	May-2018	21.887	21.603
June-2016	21.815	21.659	June-2018	22.019	21.606
July-2016	21.858	21.647	July-2018	22.061	21.608
Aug-2016	21.719	21.695	Aug-2018	21.923	21.632
Sep-2016	21.824	21.694	Sep-2018	22.028	21.612
Oct-2016	21.740	21.661	Oct-2018	21.943	21.612
Nov-2016	21.791	21.655	Nov-2018	21.995	21.609
Dec-2016	21.994	21.671	Dec-2018	22.197	21.608

Comparison

Table 7 shows the comparison between SARIMA model and ANN. It clearly indicates that ANN is outperform the SARIMA as its root mean square error, mean square error, and mean absolute error are small.

Table 7: Forecasted Checks on SARIMA and ANN

Model	RMSE	MSE	MAE
SARIMA	0.456	0.207	0.275
ANN	0.324	0.105	0.217

Conclusion

Forecasting is very helpful tool in making decisions and plans for upcoming time periods. In this study we use Seasonal Auto Regressive Integrated moving average (SARIMA) and Artificial Neural Networks (ANN) structure to forecast the Pakistan's import prices of black tea by using data for the time period January 2004 to December 2014.

Simple time series plot is use to check whether the series is stationary or not. As the underlined data set exhibit a trend as well as seasonality therefore we move towards the Box Jenkins methodology. Under Box Jenkins methodology, first of all we check the stationary of data. To check stationary Augmented Ducky Fuller is utilize. After making the data stationary (remove trend and seasonality), we use ACF and PACF plots to examine the order of model. We have selected the best model SARIMA $(0, 1, 1)*(0, 1, 1)_{12}$ among the tentative models on the basis of minimum AIC. Maximum likelihood method of estimation is utilize to find out the estimates of unknown coefficients. To check autocorrelation among the residuals of best fitted model Correlogram is utilize. On the basis of the selected model (under Box Jenkins methodology), we forecast the future prices of imports.

ANN structure is also use in this study to forecast future prices of black tea. In ANN structure we use bipolar sigmoid function as activation function and appropriate 10 input, 10 hidden and 1 output layer neurons. The residuals under ANN structure show the random pattern with zero mean and unit variance. For the comparison purpose, we have utilized RMSE, MSE and MAE for SARIMA $(0, 1, 1)*(0, 1, 1)_{12}$ model and ANN. Results reveal that all these three values are smaller for ANN 0.324, 0.105 and 0.217 respectively. Therefore, we conclude that ANN model performed well as compare to SARIMA model. Thus, ANN can be effectively used for forecasting the Pakistan's import prices of black tea in future.

Recommendations

For further study we suggest the following recommendations.

- To compare the performance of SARIMA and ANN models; we can further compare the performance of ANN model to any other model like exponential smoothing method, ARAR, ARMA or ARIMA model.
- We have considered time series data of Pakistan's Import prices of black tea and forecast future prices of black tea, similar study can be performed for any other country or any particular area in the world.
- These results may be applicable for drought forecasting and municipal decision making and planning.

The empirical results of ANN and SARIMA models clearly suggest that ANN model can perform better for forecast the Pakistan's import prices of black tea data. So ANN model can be used for modeling other time series phenomena.

References

A. Ali, et al. 2016a. A Comparative Study of Pakistan, Bangladesh and India on The Basis of Mortality Rate Under Five Years. *Journal of ISOSS, 2(1):* 79-84.

A. Ali, et al. 2016b. Forecasting of Daily Gold Price by Using Box-Jenkins Methodology. *International Journal of Asian Social Science, 6(11):* 614-624.

A. Ali, and M. Zubair, 2017. The Comparison between Artificial Neural Networking and Box-Jenkins Methodology for Forecasting of Mortality Rate. *Int.J. Adv. Res, 5(5):* 1090-1095.

A. T. Rabenja, et al. 2009. Forecasting of the rainfall and the discharge of the Namorana River in Vohiparara and FFT analysis of these data. *Proceedings, 4th International Conference in High-energy Physics, Antananarivo, Madagascar,* 1-12.

Box, G. E. P. and Jenkins, G. M. 1976. Time Series Analysis: Forecasting and Control. Rev. Ed. San Francisco. Holden-Day.

D. A. Dickey, and W. A. Fuller, 1979. Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *Journal of the American Statistical Association, 74(366):* 427-431.

D. P. Kanungo, et al. 2006. A comparative study of conventional, ANN black box, fuzzy and combined neural and fuzzy weighting procedures for landslide susceptibility zonation in Darjeeling Himalayas. *Engineering Geology, 85(3):* 347-366.

E. H. Etuk, 2013. The fitting of a SARIMA model to monthly Naira-Euro Exchange Rates. *Mathematical Theory and Modeling, 3(1):* 17-26.

E. V. Gijo, 2011. Demand forecasting of tea by seasonal ARIMA model. *International Journal of Business Excellence, 4(1):* 111-124.

H. Qu, and H. Q. Zhang, 1996. Projecting International Tourist Arrivals in East Asia and The Pacific to The Year 2005. *Journal of Travel Research, 35(1)*: 24-34.

Jan, M. S. 2016. *PTA.* Retrieved from Pakistan Tea Assosiation: http://www.pakistanteaassociation.com/statistics.html

M. Hamidreza, and S. Leila, 2012. Using SARFIMA model to study and predict the Iran's oil supply. *International Journal of Energy Economics and Policy, 12(1)*: 41-49.

M. M. Hossain, and F. Abdulla, 2015. Forecasting the Tea production of Bangladesh: Application of ARIMA model. *Jordan Journal of Mathematics and Statistics (JJMS), 8(3):* 257-270.

M. Iqbal, et al. 2016. Forecasting of Wheat Production: A Comparative Study of Pakistan and India. *Int. J. Adv. Res. 4(12):* 698-709.

S. Leila, and Y. Masoud, 2012. An Empiricle Study of the Usefulness of SARFIMA models in Energy Science. *International Journal of Energy Science, 2(2)*.

T. Abeysinghe, 1994. Deterministic seasonal models and spurious regression. *Journal of Econometrics, 61*: 259-272.