Proceedings of 12th Agricultural Research Symposium (2013) 351-355

Forecasting of Tea Production Using Time Series Models

W.H.E.B.P. WEERAPURA and N.R. ABEYNAYAKE

Department of Agribusiness Management, Faculty of Agriculture and Plantation Management, Wayamba University of Sri Lanka Makandura, Gonawila (NWP)

ABSTRACT

Tea industry has a tremendous impact on the Sri Lankan economy. Total tea production in 2011 was 329 million kilograms and the total tea exports earned 164.9 billion rupees which accounts for 58.9% of total agricultural exports. Forecasting of tea production is one of the major important requirements to individual producers, agribusiness firms and policymakers for various purposes. There is a significant research need to update and develop models with present data for accurate forecasting in near future. With this background this study was undertaken to identify the trend and appropriate time series models to forecast elevation wise black tea production. Trend analysis revealed that, high grown and medium grown tea showed the decreasing trend during the early period while low grown tea showed increasing trend. But recent period production of all elevations showed a declining trend, which is a problem that should be addressed strategically. Exponential smoothing techniques and ARIMA methodology were employed to identify appropriate models for elevation wise tea production. MAPE was used as model selection criteria with residual analysis. Among the exponential smoothing models tested, Single exponential models were selected as good models and among the ARIMA models tested, ARIMA (1,0,1), ARIMA (1,1,1) and ARIMA (1,1,0) were selected for high, medium and low grown tea respectively.

KEY WORDS: ARIMA, Forecasting, MAPE, Residual analysis, Trend

INTRODUCTION

Tea as a crop was first introduced to Sri Lanka in the middle of 19th century. JamesTaylor started the first commercial tea plantation at Loolkandura estate in 1867. The tea produced in this country is popularly known as "Ceylon tea" and ranks among best available teas in the international trade. Over the years, the word Ceylon has become synonymous with quality tea.

There are three agro climatic tea growing regions according to the different elevation zones; high grown, medium grown and low grown. Teas that are grown in Badulla and NuwaraEliya are high grown teas where the elevation is above 1200 m from sea level. Teas grown in Kandy and Matale are medium grown teas, where the elevation is 600 m-1200 m and teas grown in Galle, Matara, Kalutara and Ratnapura are low grown teas, where the elevation is below 600 m. Tea industry is a strong pillar in Sri Lankan economy in terms of foreign exchange earnings and employment. Tea exports earnings reached 164.9billion rupees out of total agricultural exports of 279.5 billion rupees in 2011 (Anon, 2011a).Two millions of people are employed directly and indirectly on the industry (Anon, 2011b).

However forecasting of production of tea enable policymakers and planners to estimate the production requirement of tea in future and formulate appropriate strategies to meet the future demand.

There is a significant research need to update and develop models with present data for accurate forecasting in the near future.With this background this study wascarried out with the objective of identifying appropriate time series models: Auto Regressive Integrated Moving Averages (ARIMA) and Smoothing techniques, in order to forecast elevation wise tea production and assess the trend of elevation wise tea production.

METHODOLOGY

Data Collection

Time series data on elevation wise annual black tea production in kg from 1963 to 2011 were collected from statistical bulletin published by Sri Lankan tea board, which provided a total of 49 years production observations.

Analysis

Different trend models and time series models were tested for the data. Based on the Mean Absolute Percentage Error (MAPE) value, the best models were selected.

Trend Models

Linear, Exponential and Quadratic models were tested to find out the most suitable trend.

Exponential Smoothing Models

At the first phase of the analysis, Single exponential models were tested with different constant values.

The methods of single exponential forecasting take the forecast for the previous period and adjust it using the forecast error. [Forecast error = $(Y_t - F_t)$]

$$F_{t+1} = F_t + \alpha (Y_t - F_t)$$

Where,
Y_t= observed value for time period t
F_t= fitted value for time period t
α= weighting factor, which ranges from 0 to 1
t = current time period

At the second phase of analysis Holt's Linear Exponential Smoothing model (Double Exponential model) were fitted.

Holt, (1957) extended single exponential smoothing to linear exponential smoothing to allow forecasting of data with trends. The forecast for Holt's linear exponential smoothing is found using two smoothing constants, α and β (with values between 0 and 1), and three equations:

$$L_t = \alpha Y_t + (1-\alpha)(L_{t-1} + b_{t-1}) \ldots \textbf{(a)}$$
$$b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1} \ldots \textbf{(b)}$$
$$F_{t+m} = L_t + b_t m \qquad \ldots \textbf{(c)}$$

Here, L_t denotes an estimate of the level of the series at time t and b_t denotes an estimate of the slope of the series at time t. Equation (a) adjusts (L_t) directly for the trend of the previous period, b_t-1, by adding it to the last smoothed value, L_t-1. This helps to eliminate the lag and brings L_t to the approximate level of the current data value. Equation (b) then updates the trend, which is expressed as the difference between the last two smoothed values. This is appropriate difference between the last two smoothed values. This is appropriate because if there is a trend in the data, new values should be higher or lower than previous ones. Since there may be some randomness remaining, the trend is modified by smoothing with β the trend in the last period (L_t-L_{t-1}), and adding that to the previous estimate of the trend multiplied by (1-β). Thus, equation (b) is similar to the basic form of single smoothing but applied to the updating of the trend. Finally equation (c) is used to forecast ahead. The trend, b_t, is multiplied by the number of periods ahead to be forecast, m, and added to the base value, L_t.

ARIMA Models

The general model introduced by Box and Jenkins (1976) includes autoregressive as well as moving average parameters, and explicitly includes differencing in the formulation of the model. Specifically, the three types of parameters in the model are: the autoregressive parameters (p), the number of differencing passes (d), and moving average parameters (q). In the notation introduced by Box and Jenkins, models are summarized as ARIMA (p, d, q); so, for example, a model described as (0, 1, 2) means that it contains 0 (zero) autoregressive and 2 moving average which were computed for the series after it was differenced once.

ARIMA (p,d,q) :

$$\nabla^d Y_t = (1-B)^d \varepsilon_t$$

Where, p= non seasonal autoregressive parameters; d= non seasonal number of differencing passes;q = non seasonal moving average parameters.

ARIMA (p,1,q):

$$(1-\phi_1 B)(1-B)Y_t = C + (1-\theta_1 B)\varepsilon_t,$$

Where, B = Back shift notation ($BY_t = Y_{t}$-1).

Model Selection and Validation

As the model selection criteria, Mean Absolute Percentage Error (MAPE) which is illustrated in following equation was used to select the best fitted model.

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}|\text{PE}_t|$$

Where, $PE_t = 100*(Y_t - F_t)/Y_t$

RESULTS AND DISCUSSION

Time series plots of tea production for each tea growing areas are given in Figure 1, 2 and 3. It is very clear that, production behavior of the tea can be separated into two phases: 1963-1992 (Phase I) and 1993-2011 (Phase II).

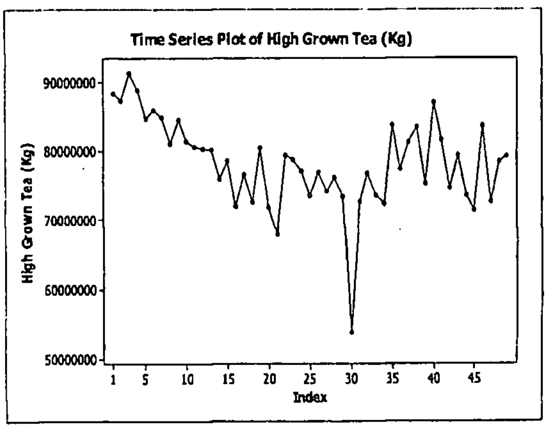

Figure 1. Time series plot for high grown production

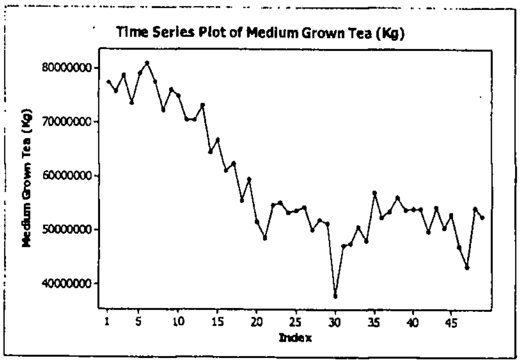

Figure 2. Time series plot for medium grown production

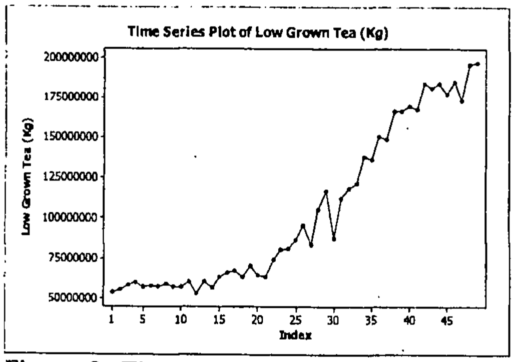

Figure 3. Time series plot for low grown production

Trend Analysis

Due to the clear phases of the time series plot, trend analysis was performed separately for the two phases.

Trend analysis was carried out for selected three methods; Linear, Exponential and Quadratic.

The lowest MAPE was employed as the model selection criteria to select the best fitted general trend model. The MAPE values for the models tested are given in the Table 2. Selected models presented in the Table 1.

Table 1. Selected trend models

High Grown Tea (10^5)	
Phase I (1963-1992)	$Y_t = 923.9 - 14.3t + 0.2t^2$
Phase II (1993-2011)	$Y_t = 729.4 + 12.4t - 0.6t2$
Medium Grown Tea (10^5)	
Phase I (1963-1992)	$Y_t = 826.0 - 11.9t$
Phase II (1993-2011)	$Y_t = 469.8 + 12.7t - 0.6t^2$
Low Grown Tea (10^5)	
Phase I (1963-1992)	$Y_t = 625.6 - 18.5t + 1.1t^2$
Phase II (1993-2011)	$Y_t = 1020.3 + 88.9t - 2.3t^2$

Table 2. MAPE for fitted trend models

Tested Models	**MAPE Value**					
	High Grown Tea		**Medium Grown Tea**		**Low Grown Tea**	
	PhaseI	**Phase II**	**Phase I**	**Phase II**	**Phase I**	**Phase II**
Linear	3.45	4.99	**4.58**	5.89	9.72	4.44
Exponential	3.34	5.92	4.60	5.92	7.94	5.32
Quadratic	**2.96**	**4.89**	4.59	**4.72**	**5.63**	**2.82**

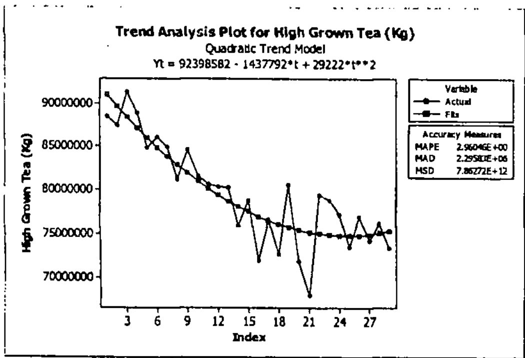

Figure 4a. Trend analysis plot for high grown tea-Phase I

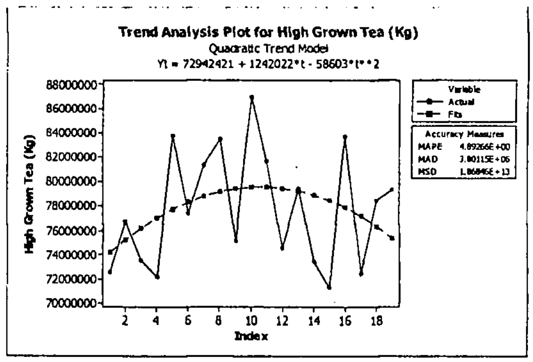

Figure 4b. Trend analysis plot for high grown tea - Phase II

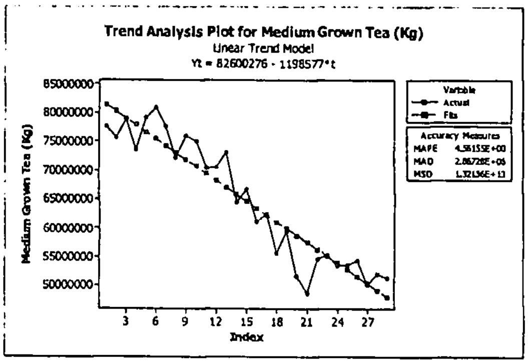

Figure 5a. Trend analysis plot for medium grown tea - Phase I

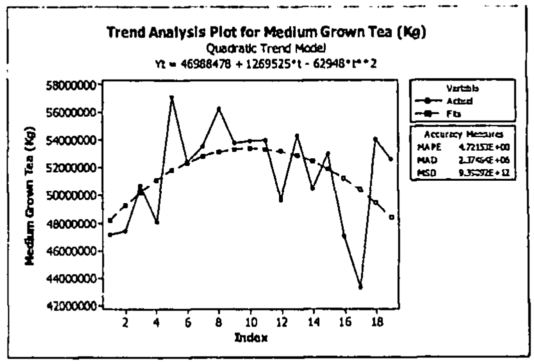

Figure 5b. Trend analysis plot for medium grown tea - Phase II

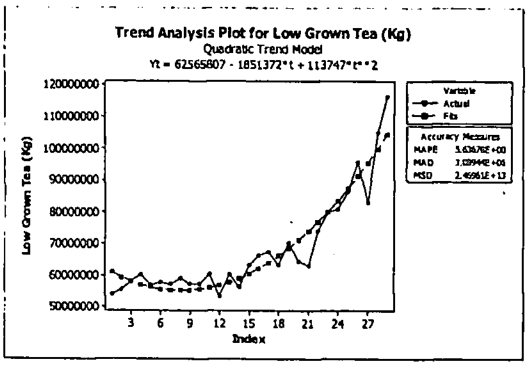

Figure 6a. Trend analysis plot for low grown tea - Phase I

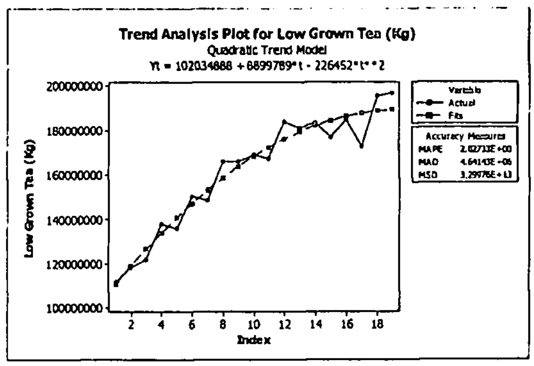

Figure 6b. Trend analysis plot for low grown tea - Phase II

Figure 4, 5 and 6 clearly show that the production trend of high, medium and low grown tea for the period 1963 to 1992 and 1994 to 2011separately. High grown and medium grown tea showed a declining trend during the early period while low grown tea showed increasing trend (Figure 4a, 5a and 6a). But recent period production of all elevations showed declining trend (Figure 4b, 5b, and 6b).

Exponential Smoothing Models

Table 3, 4 and 5 gives the MAPE values for Double and Single Exponential smoothing models and fitted values for the 2009, 2010, 2011and forecasted values for 2012, 2013 and 2014.

Table 3. High grown tea production (Kg)

		Fitted /forecasted values	
Year/ MAPE	**Observed**	**Single Exponential**	**Double Exponential**
2009	72479739	78064795	78860644
2010	78387908	76221770	75213792
2011	79300000	76936579	76880332
2012		72479739	78226720
2013		78387908	78257487
2014		79300000	78288253
MAPE		5.3	5.5

Table 4. Medium grown tea production (Kg)

		Fitted /forecasted values	
Year/ MAPE	**Observed**	**Single Exponential**	**Double Exponential**
2009	43313234	49289859	48844672
2010	54060029	45824315	45306371
2011	52600000	50599791	49607336
2012		51759612	50952012
2013		51759612	50712082
2014		51759612	50472153
MAPE		5.8	6.1

Table 5. Low grown tea production (Kg)

		Fitted /forecasted values	
Year/ MAPE	**Observed**	**Single Exponential**	**Double Exponential**
2009	173033501	183149779	189143355
2010	195668124	175582006	183230866
2011	196700000	190608005	188883154
2012		195165297	193859533
2013		195165297	196083532
2014		195165297	198307531
MAPE		6.0	6.9

According to the MAPE values, Single Exponential models were better for the forecasting purposes of all elevation with compare to Double Exponential models.

Table 6. Estimates of selected ARIMA models

Category	Selected Model	MAPE	Fitted values for 2009, 2010 and 2011 (in Mn kg)	Forecasted values for 2012,2013 and 2014 (in Mn kg)
High Grown	ARIMA (1,0,1)	5.1012	78.4629 76.6387 77.4041	78.1678 78.2763 78.3780
Medium Grown	ARIMA (1,1,1)	5.8842	43.3132 54.0600 52.6000	50.7370 50.4475 49.8618
Low Grown	ARIMA (1,1,0)	6.5508	185.372 183.136 189.192	200.648 203.189 206.408

ARIMA procedure was carried out for the same data set and three ARIMA models were identified for three elevations by following Box-Jenkins ARIMA methodology. Selected ARIMA models, relevant MAPE, fitted values and forecasted values were summarized in Table 6. Residual analysis was carried out separately for three selected models and it revealed that non randomness and non-autocorrelation between lags for residuals.

CONCLUSIONS

Trend analysis for tea production revealed that the latter period (1993 to 2011) has a decreasing trend for tea production in all elevation. Mostly it may be an adverse repercussion of increasing cost of production and recent climate changes.

According to the models fitted by Single Exponential models and Double Exponential models, Single Exponential models were better than double exponential models for forecasting of production of tea for the all elevations.

ARIMA models selected for high, medium and low grown tea are ARIMA (1, 0, 1), ARIMA (1, 1, 1) and ARIMA (1, 1, 0) respectively.

By increasing production, producer can increase the profitability and reduce the cost of production which is high in Sri Lanka. This high production cost may be a reason for the slipping of Ceylon tea from some competitive markets in the global tea arena where the major competitors like Kenya, who has a higher productivity and gain the competitive advantage in production and export performances. Adequate focus on replanting, fertilizing and other field operations will heighten the production and simultaneously strategic implementation of solutions for labor disputes will be another rigorous fact for enhancing the production of black tea in all elevations.

ACKNOWLEDGEMENTS

The authors wish to express their sincere gratitude to all the academic and non-academic staff members of Faculty of Agriculture and Plantation Management, for their support to the successful completion of the research.

REFERENCES

Anon. (2011a).Central Bank of Sri Lanka. Available From: www.cbsl.gov.lk (Accessed 5 April 2013).

Anon. (2011b). Annual report Thalavakelle. Available from: http://www.talawakelleteas.com (Accessed 6 March 2013).

Box, G.E.P. and Jenkins, G.M. (1970). Time series analysis: forecasting and control, San Francisco: Holden day.

Holt, C.C. (1957). Forecasting seasonal and trends by exponentially weighed moving average, Office of Navel Research, Research memorandum No. 52.