

Forecasting of Paddy Production in Sri Lanka: A Time Series Analysis using ARIMA Model

V. Sivapathasundaram* and C. Bogahawatte

Postgraduate Institute of Agriculture
University of Peradeniya
Sri Lanka

ABSTRACT. *Forecasting of paddy production is a need for planning purposes and import policy of rice should be based on such forecasts. Even though Sri Lanka has achieved self sufficiency in rice the expenditure on rice sector has increased continuously. The objectives of this study are to investigate the past, present and future trends of paddy production in Sri Lanka and to develop a time series model to detect the long term trend and prediction for future changes of paddy production for the three leading years. Autoregressive Integrated Moving Average (ARIMA) was used to fit the data set which is complementary to the trend regression approach and forecasting of the concerned variable to the near future. Time series forecasting analysis utilized the secondary data of the Department of Census and Statistics of Sri Lanka for the period of 1952 to 2010. Non-stationarity in mean was corrected through differencing of the data of order 1. ARIMA (2, 1, 0) was the most suitable model used as this model has the lowest AIC and BIC values. The Mean Absolute Percentage Error (MAPE) for paddy production was 10.5. The forecasts for paddy production during 2011 to 2013 were 4.07, 4.12 and 4.22 million Mt respectively, and the production for the year 2011 and 2012 was lower than in 2010. However in later year 2013 the production was higher. This model can be used by researchers for forecasting of paddy production in Sri Lanka. But, it should be updated continuously with incorporation of recent data.*

Keywords: *ARIMA model, forecasting, paddy production*

INTRODUCTION

Sri Lanka's rice sector has achieved spectacular results in the last 5 decades. The total paddy production increased rapidly and reached 4.18 million Mt in 2010 (CBSL, 2010). In 2006, the latest year for which detailed household expenditure data are available, the average per capita expenditure was Rs. 250 per month on rice (US\$ 2.50), equivalent to 12.8 percent of the total food expenditure. The per capita consumption of rice is 105 kg/year (average of 2004-08) or about 300 grams per day, which provides about 1,000 calories or almost half of the average total daily calorie intake. Reliable management information on production estimates of agriculture in each cropping system is a prerequisite for policy decision making on rice imports to Sri Lanka. Therefore, forecasting of paddy production and paddy price is a need and important for planning purposes, and import policy of rice should be based on such forecasts.

* Author for correspondence: sivavani@live.com

Sri Lanka has already reached self sufficiency level in rice (Department of Census and Statistics, 2010). Table 1 shows the self sufficiency rate of rice over last five years.

Table 1. Self-sufficiency rate of rice from 2005 to 2010

Year	Paddy production (million Mt)	Total rice availability for human consumption (million Mt)	Total rice requirement (million Mt)	Rate of self-sufficiency (%)
2005	3.24	2.06	2.05	100.7
2006	3.34	2.13	2.07	102.9
2007	3.13	2.01	2.08	96.6
2008	3.86	2.45	2.10	116.8
2009	3.65	2.42	2.25	107.4
2010	4.04	2.68	2.35	113.9

Source: Agriculture and Environment Statistics Division, Department of Census and Statistics, Sri Lanka

Even though the country reached self sufficiency in rice, government expenditure has been increasing in order to promote and support paddy production in Sri Lanka. Among the policies implemented to promote paddy cultivation in Sri Lanka fertilizer subsidy program is one of the long lasting, most expensive and most politically sensitive policies (Weerahewa *et al.*, 2010). The expenditure on the fertilizer subsidy in 2010 was Rs 26.9 billion which constituted 2.24 percent of total government expenditure (CBSL, 2010). If Sri Lanka is to become a rice exporting country, forecasting of paddy production is vital in preferring policy planning to achieve this goal.

Raghavender (2010) analyzed yearly rice production data for the period of 1955-56 to 2007-08 in Andhra Pradesh by time series methods. The forecasting power of autoregressive integrated moving average model was used to forecast rice production for three leading years. In Sri Lanka, Thattil and Walisinghe (2000) used ARIMA and regression models to create database for climate, price and production data on paddy, district-wise and each of the seasons, *maha* and *yala* and used this database to derive prediction models for forecasting paddy yields for following three years. A study done using time series forecasting analysis based on the secondary data revealed that cultivated and harvested land extents and yields have positive increasing trend while simple time series model explained the increasing trend of retail prices (De Silva and Yamao, 2009). According to Dayaratna Banda *et al.* (2008) forecasts, expected paddy prices should be about Rs 29 per kg in April 2008. This forecast value is quite similar to the actual price reported (Rs 30 per kg) by Agrarian Research and Training Institute for that year. These forecasts indicated that paddy price would also follow an increasing trend in the coming months as well. A research conducted by Partheepan and Jeyakumar (2005) used the ARIMA model as a tool of analysis in agriculture based research in order to forecast rainfall. The forecasted annual rainfall behaves on a 10 year cyclic pattern showed slightly increasing trend up to 2010. Hence, the objectives of the study are to determine the past, present and future trends of paddy production in Sri Lanka and develop a time series model to detect long term trend and prediction for future changes of paddy production for the three leading years in Sri Lanka.

METHODOLOGY

The econometric model of Autoregressive Integrated Moving Average (ARIMA) was used to fit the data set which is complementary to the trend regression approach and forecasting of the concerned variable to the near future (Box *et al.*, 1994). ARIMA (Box-Jenkins model) method is an extrapolation method for forecasting and, like any other such method, it requires only the historical time series data for the variable under forecasting. Among the extrapolation methods, this is one of the most sophisticated methods, as it incorporates the features of all such methods, does not require the investigator to choose the initial values of any variable and values of various parameters a priori and it is robust to handle any data pattern. As one would expect, this is quite a difficult model to develop and apply as it involves consecutive four stages such as Identification process, Estimation, Diagnostic testing and Forecasting (Janacek and Swift, 1993).

Theoretical basis of Time-series analysis

A time series is a set of values of a continuous variable Y (Y1, Y2, ..., Yn), ordered according to a discrete index variable t (1, 2, ..., n). The term time-series comes from econometric studies in which the index variable refers to intervals of time measured in a suitable scale. In general, in a given time series the following can be recognized and separated into the following components.

- 1) A regular, long-term component of variability, termed trend, that represents the whole evolution pattern of the series;
- 2) A regular, short-term component whose shape occurs periodically at intervals of lags of the index variable, currently known as seasonality, because this term is also derived by applications in economics;
- 3) An AR(p) autoregressive component of p order, which relates each value $Z_t = Y_t -$ (trend and seasonality) to the p previous Z values, according to the following linear relationship

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + \epsilon_t \dots \dots \dots (1)$$

where ϕ_i (i= 1, ..., p) is the parameter to be estimated and ϵ_t is a residual term; and

- 4) A MA(q) moving average component of q order, which relates each Z_t value to the q residuals of the q previous Z estimates

$$Z_t = \epsilon_t - \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q} \dots \dots \dots (2)$$

where θ_i (i= 1, ..., q) parameter to be estimated.

In mathematical terms, therefore, a seasonal ARIMA model is an ARIMA (p, d, q) model whose residuals ϵ_t can be further modeled by an ARIMA (P, D, Q)s structure with linear operators (P, D, Q) being functions of the Bs operator. Since annual value of the concerned variable was used, univariate, non-seasonal ARIMA (p, d, q) was used to fit the data sets.

Data and data analysis

Time series forecasting analysis utilized the secondary data of the Department of Census and Statistics of Sri Lanka for the time period of 1952 to 2010. Time series forecasting analysis

was carried out using statistical software of SPSS. The time series analysis was done to fulfill the objective of detecting the trend and predicting future changes of paddy production for the three leading years in Sri Lanka by developing a model according to the behavior of the data. The econometric model of ARIMA was used to fit the data set. Since annual value of the concerned variable was used, a univariate non-seasonal ARIMA (p, d, q) was estimated. Modeling was done by four stages such as Identification process, Estimation, Diagnostic testing and Forecasting.

Identification process

Annual values were plotted to identify the variation of the data (i.e. trend, seasonality, cyclical and extreme events). Autocorrelation function was used to identify the stationarity of the series ($\mu = 0$ and σ^2 constant throughout the observation), non-stationarity series was converted to stationary by taking the differences (i.e. 1st difference of $Y_{t+1} - Y_t$, 2nd difference of $Y_{t+2} - Y_t$ or first difference of the 1st differenced series). All data were checked for the pure randomness (White noise), hypothesis of randomness. The null hypothesis, H_0 : No autocorrelation up to a given lag significantly differs from zero. That type of series did not fit ARIMA. Tentative AR/MA/ARMA/ARIMA was selected using Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF).

Parameter estimation

Both AR and MA parameters were estimated. The Null Hypothesis for selection of the parameter was:

H_0 : All considered parameters should be elaborating the condition with the value is between -1 to +1. Various non-seasonal ARIMA models were tried and each of the parameters were estimated.

Diagnostic checking

Adequacy of the model was checked by using Box Pierce (Ljung – Box) Chi Square statistic residuals with null hypothesis of H_0 : Residuals follow white noise series and H_0 was rejected at $P < 0.05$. Sum of Squares residual of the best model was of the lower value. AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) also used to select the best model. Normal probability value was taken to check the residuals.

Forecasting

Selected best fitting model was used to forecast the future data for the three leading years. Simple linear regression model was fitted to auto correlated annual series to determine coefficients and trend.

RESULTS AND DISCUSSION

ARIMA model was estimated only after transforming the variable under forecasting into a stationary series whose values vary over time only around a constant mean and constant variance. The most common method is to check stationarity through examining the graph or time plot of the data. Fig. 1 revealed that the data was non-stationary. Non-stationarity in mean was corrected through appropriate differencing of the data. In this case difference of order 1 was sufficient to achieve stationarity in mean.

The newly constructed variable X was examined for stationarity. The graph of X was stationary in mean. The next step was to identify the values of p and q . For this, the autocorrelation and partial autocorrelation coefficients of various orders of X were computed (Table 2). The ACF and PACF (Fig. 2 and 3) showed that the order of p and q can at most be 1. Then entertained three tentative ARIMA models were used to choose the model which has the minimum AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion). The models were ARIMA (2, 1, 0), ARIMA (2, 1, 1) and ARIMA (2, 1, 2). Among the above models, ARIMA (2, 1, 0) was the most suitable model as this model has the lowest AIC and BIC values.

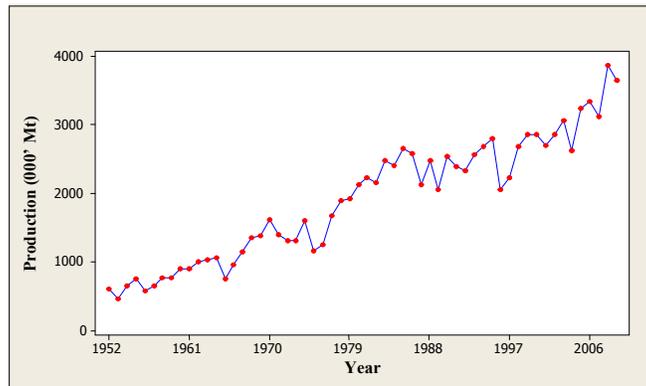


Fig. 1. Time plot of paddy production data

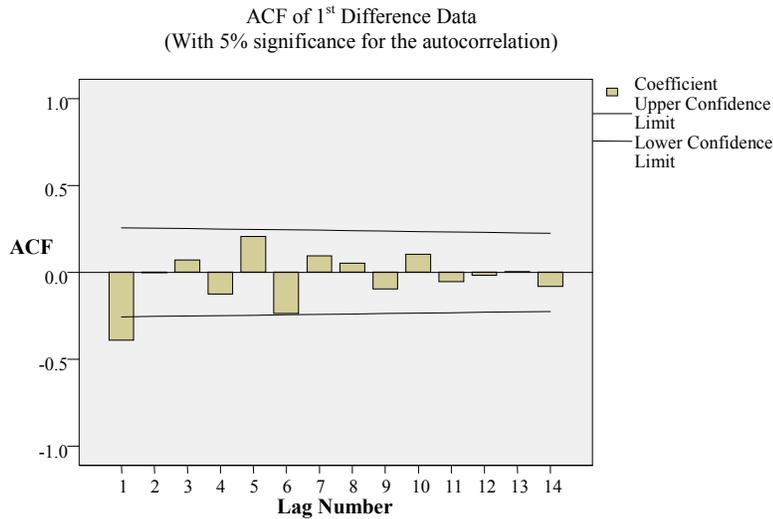


Fig. 2. ACF of differenced data

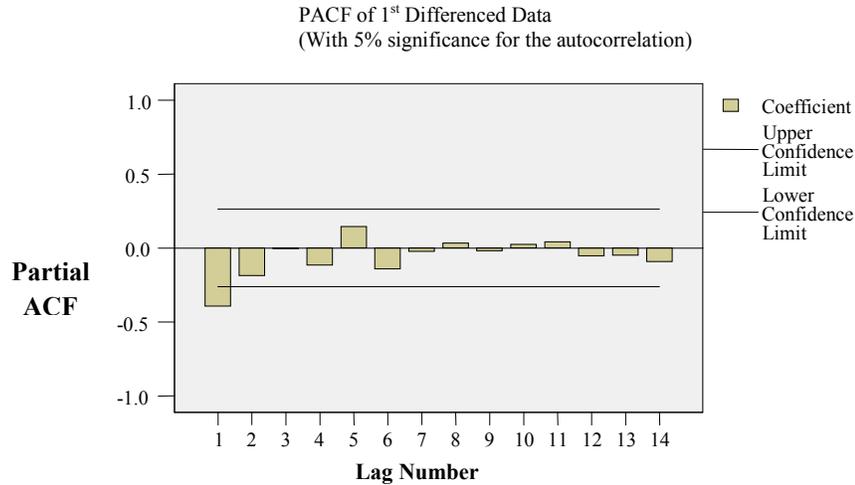


Fig. 3. PACF of differenced data

Table 2. Autocorrelations and partial autocorrelations

Lag	Autocorrelation	Std. Error	Lag	Partial Autocorrelation	Std. Error
1	-.392	.128	1	-.392	.131
2	-.004	.127	2	-.186	.131
3	.071	.126	3	-.004	.131
4	-.125	.125	4	-.116	.131
5	.206	.123	5	.145	.131
6	-.237	.122	6	-.141	.131
7	.094	.121	7	-.023	.131
8	.052	.120	8	.034	.131
9	-.096	.119	9	-.019	.131
10	.103	.117	10	.024	.131
11	-.054	.116	11	.043	.131
12	-.018	.115	12	-.052	.131
13	.002	.114	13	-.050	.131
14	-.082	.112	14	-.091	.131

Model estimation and verification

Model parameters were estimated using SPSS package. Results of estimation are reported in Table 3. The model verification is concerned with checking the residuals of the model to see if they contain any systematic pattern which still can be removed to improve on the chosen ARIMA. This is done through examining the autocorrelations and partial autocorrelations of the residuals of various orders. For this purpose, the various correlations up to 14 lags were computed and the same along with their significance was tested by using Box-Ljung test and the results are shown in Table 3. Results indicated that none of these correlations were significantly different from zero at a reasonable level. This proved that the selected ARIMA model is an appropriate model. The ACF and PACF of the residuals (Fig. 4 and 5) also

indicated ‘goodness of fit’ of the model. So the fitted ARIMA model for the production data (Equation 3),

$$Z_t = 60.139 + 1.482 Z_{t-1} - 0.669 Z_{t-2} + 0.187 Z_{t-3} + \varepsilon_t \dots\dots\dots(3)$$

Table 3. Estimates of the ARIMA model

	Estimates	Std Error	t	Approx Significance	
Nonseasonal Lags	AR1	.48	.135	-3.56	.001
	AR2	-.19	.136	-1.38	.174
Constant		60.14	19.897	3.02	.004
Number of Residuals	58				
Number of parameters	2				
Residual Df	55				
Adjusted Residual SS	3469459				
Residual SS	3470627				
Residual Variance	62808.51				
Model Std. Error	250.62				
Log-Likelihood	-401.31				
Akaike’s Information Criterion (AIC)	808.61				
Schwarz’s Bayesian Criterion (BIC)	814.79				

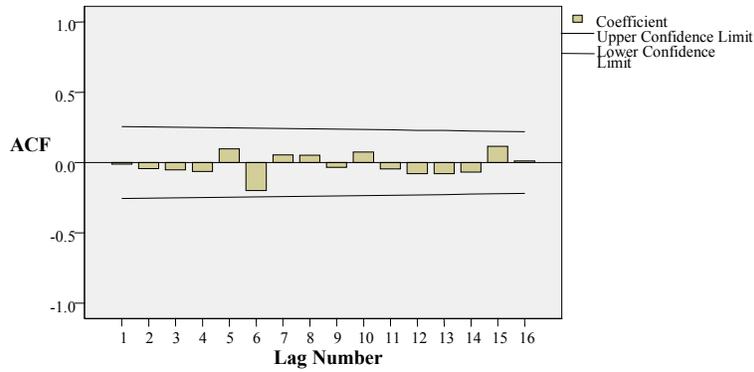


Fig. 4. ACF of residuals

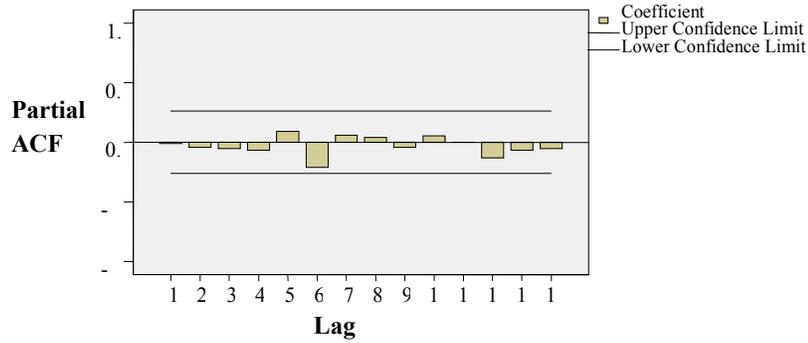


Fig. 5. PACF of residuals

Forecasting with ARIMA model

ARIMA models are developed basically to forecast the corresponding variable. There are two kinds of forecasts: sample period forecasts and post-sample period forecasts. The former are used to develop confidence in the model and the later to generate genuine forecasts for use in planning and other purposes. The ARIMA model can be used to yield both these kinds of forecasts.

Sample period forecasts:

The sample period forecasts were obtained simply by plugging the actual values of the explanatory variables in the estimated Equation 3. The explanatory variables here were the lagged values of Z_t and the estimated lagged errors. To judge the forecasting ability of the fitted ARIMA model, important measures of the sample period forecasts' accuracy were computed. The Mean Absolute Percentage Error (MAPE) for paddy production in Sri Lanka was 10.5. This measure indicated that the forecasting accuracy was high.

Post sample forecasts:

The principal objective of developing an ARIMA model for a variable is to generate post sample period forecasts for that variable. This was done through using Equation 3. The forecasts for paddy production during 2011 to 2013 are given in Table 4 where it has the increasing trend but it has decreased for the year 2011 and 2012 and increased for the year 2013 compared to 2010. Fig. 6 and 7 showed the time series plot and trend analysis plot of the forecasted paddy production respectively.

Table 4. Estimated values of paddy production and 95% confidence limits (CL)

Year	Estimated Production (000 Mt)	Lower CL	Upper CL
2011	4067.72621	3561.09154	4574.36088
2012	4123.41276	3548.76408	4698.06145
2013	4217.94313	3570.45428	4865.43198

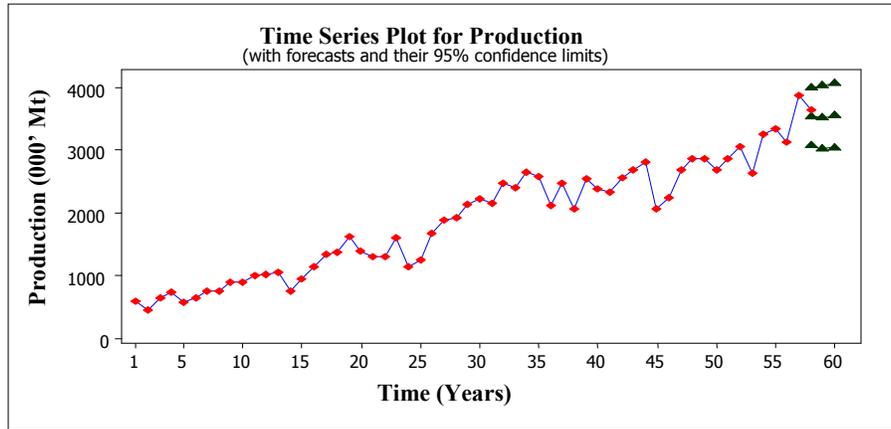


Fig. 6. Time series plot of the paddy production

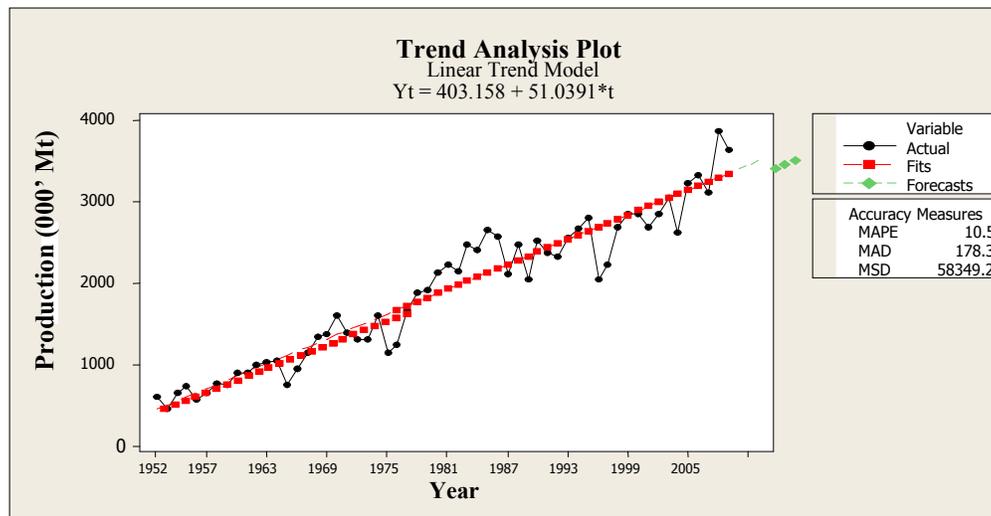


Fig. 7. Trend analysis plot of the paddy production

CONCLUSIONS

The developed model for paddy production was found to be ARIMA (2, 1, 0). From the forecast available by using the developed model, it can be seen that forecasted production for the years 2011 and 2012 was lower than in 2010 but in later year in 2013 the production was higher. Overall it has an increasing trend. The validity of the forecasted values can be checked when the data for the lead periods become available. The model can be used by researchers for forecasting of paddy production in Sri Lanka. However, it should be updated continuously with the incorporation of recent data.

REFERENCES

- Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (1994). Time Series Analysis: Forecasting and Control. (3rd ed.). Prentice Hall Press, Englewood cliffs, New Jersey.
- Central Bank of Sri Lanka. (2010). Annual Report, Colombo, Central Bank of Sri Lanka.
- Department of Census and Statistics, Agriculture Census, 2010 [on line]. [Accessed on 10.10.2011] Available at <http://www.statistics.gov.lk/agriculture/AGC2002/Pages/AGC2010>
- De Silva, D.A.M. and Yamao, M. (2009). Rice pinch to war thrown nation: an overview of the rice supply chain of Sri Lanka and the consumer attitudes on government rice risk management. *J. of Agric. Sci.* 4(2), 77-96.
- Dayaratna Banda, O.G., Jayawickrama, J.M.A. and Ranathilaka, M.B. (2008). Sense and nonsense of rice price controls in Sri Lanka. *The Peradeniya J. of Economics.* 2(1&2), 60-87.
- Janacek, G.J. and Swift, L. (1993). Time Series: Forecasting, Simulation, Applications. Ellis Horwood, New York, London,
- Partheepan, K. and Jeyakumar, P. (2005). Development of a time series model to forecast climate data in Batticaloa district, Sri Lanka. Proceedings of the Water Professional Day, Postgraduate Institute of Agriculture, Peradeniya. October, 2005. pp. 199-208.
- Raghavender, M. (2010). Forecasting paddy production in Andhra Pradesh with ARIMA model. *Int. J. of Agric. and Stat. Sci.* 6(1), 251-258.
- Thattil, R.O. and Walisinghe, W.M.P.K. (2003). Forecasting paddy yields. *In: Abey Siriwardena, D.S. de Z., Dissanayaka, D.M.N. and Nugaliyadde, L. (Ed.) Rice Congress 2000.* Department of Agriculture, Sri Lanka.
- Weerahewa, J., Kodithuwakku, K.A.S.S. and Ariyawardana, A. (2010). The fertilizer subsidy program in Sri Lanka. 12 pp. In: Andersen, P.P. (Ed.) Food Policy for Developing Countries: The Role of Government in the Global Food System. Ithaca, New York.