Prediction of Annual National Coconut Production - A Stochastic Approach

T.S.G. PEIRIS
Coconut Research Institute, Lunuwila,
Sri Lanka

T.U.S. PEIRIS
Coconut Research Institute, Lunuwila,
Sri Lanka

and

S. RAJAPAKSHA
Coconut Research Institute, Lunuwila,
Sri Lanka

ABSTRACT

Coconut is one of the most economically important crops and it comprises an important part of the diet in Sri Lanka. Prediction of annual national coconut yield is of vital importance to policy-makers, scientists and coconut growers for various reasons. The empirical statistical models based on drought indices via monthly rainfall values have failed to predict national yield due to many reasons. Properties of the auto correlation of the annual yield (1950-1988) confirmed that a stochastic model of type ARIMA (2,1,0) was the best model to forecast annual national yield. The model does not need any external variables. The adequacy of the model was tested based on various statistics. The percentage error of the predicted values from 1950 to 1998 varied from (1-12)%. Based on data 1950-1988, the percentage error for the estimated value for the year 1999 is 6.3% with respect to the actual value of 2830 million nuts. Using data (1950-1999) the estimated values for the years 2000 and 2001 are 2896 and 2705 million nuts.

Key Words: Stochastic approach, Coconut production, Time series analysis, Univariatic model, ARIMA

INTRODUCTION

Coconut is important to Sri Lanka, both as a food crop and as an industrial crop. The major part of the production (70%) is used for local consumption and the balance is used for export as kernel products (24%) and husk nuts (1%). The percentage contribution to GDP by the coconut sector was 2.6 in 1998, which was higher than the corresponding percentage by tea sector or rubber sector (Anonymous, 1998). Annual coconut production during 1995-1999 fluctuated between 2540 million nuts (in 1998) and 2830 million nuts (in 1999) with a mean of 2654 and standard error of 53 million nuts. The advanced knowledge of coconut production is an important tool for policy-makers mainly in utilization of nuts for various sectors and to search export markets.

In a given location the current crop condition, which can be estimated from the past weather during the growing period of the crop, determines its potential yielding ability (Peiris and Thattil, 1998). Thus the predictive models based on climatic variables are meaningful in estimating the production in a given location. However, as the climate variability in coconut grown areas in Sri Lanka is very complex and not been studied deeply, the crop-weather models developed to predict national yields were not successful in the past (Peiris and Mathes, 1997). The models used to predict national coconut yield are discussed by Peiris *et al.*, (1995). Presently no statistical methods are used to estimate the national coconut production in Sri Lanka.

In this paper a univariate time-series model based on the yield trend over the years was developed to predict the annual national yield and compared with regression models containing climatic variables.

METHODOLOGY

Secondary Data

National annual production data (1950 to 1999) were collected from the annual reports of the Coconut Development Authority, Colombo. To compute drought indices, monthly rainfall data from 1962 to 1997 of 18 locations in the coconut growing areas: Kurunegala, Horombawa, Lunuwila, Palugawewa, Rajakadaluwa, Ratmalagara, Kirama, Mediyawa, Nikaweratiya, Polonthalawa, Ridibendi-ela, Ratnapura, Gampaha, Galle, Kalutara and Matara were obtained from the Department of Meteorology, Colombo. These stations represent four agro-ecological regions of major coconut growing regions. The locations were selected as the authors are presently investigating the long-term rainfall variability of these locations.

Regression Analysis -Empirical Modelling

Abeywardena (1983) has developed an empirical statistical model using data from 1963 to 1976 to predict the national annual coconut yield a year ahead using 'drought indices' of eight locations. The locations were Tangalla, Kudawewa, Kuliyapitiya, Lunuwila, Rajakadaluwa, Kurunegala, Negambo and Polgahawela. The drought index for a given year is derived from the monthly total rainfall. The model explained 95% of yield variation around the mean of 1963-1976. The status of significance of the model (or the status of significance the variables in the model) was not mentioned. However, as the errors of the estimated values for some years were large, the model was not used. But no alternative climatological models have been attempted.

In our study, the concept of drought indices was further tested using more recent data. A basic computer program was developed to compute the drought indices. The drought indices were considered as independent variables in the regression approach. The correlation structure of the drought indices among the locations does not support strong multicollinearity. Of the 18 variables of drought indices the best model, which is statistically significant and explains about 60% of the yield variation from 1964 to 1998, is:

$$Y = 2996.2937_{(101)} - 1.0201_{(0.3931)}*X1 - 0.8524_{(0.2945)}*X2 + 0.5651_{(0.2993)}*X3 - 0.7737_{(0.2993)}*X4 + 0.8471_{(0.3068)}*X5$$

$(R^2 = 0.62; p < 0.001)$

Where, Y = National annual coconut production; X1 = drought index at Lunuwila; X2 = drought index at Rajakadaluwa; X3 = drought index at Kirama; X4 = drought index at Nikaweratiya and X5 = drought index at Polonthalawa. The standard errors of the parameters are given in parenthesis. All the variables in the model are significant at the 0.05 level.

Validity of the Model

The percentage error for the predicted values from 1964 to 1998 for the above model varied from 1% to 24% (as rainfall data were not available prior to 1964 for some stations, yield data were used for the period 1964 - 1998). The correlation between the observed and predicted values is significant ($r = 0.79^{***}$, n=36). The pattern of the residual was tested and found to be distributed randomly. These results confirmed the validity of the model for the selected data. The predicted value for 1999 is 2539 and the percentage error is 10.3% as against the actual value of 2830.

Using data from 1964 to 1999, the selected five variables of the model was tested for three twelve year periods: (1964 -1975, 1976 -1987 and 1988-1999) and it found that only the model for 1976 - 1987 was significant. Also none of the variables were significant in any of the models. When the parameters were further tested for two18-year periods (1964 - 1982 and 1983 - 1999) it was found that though the models were significant, none of the variables were significant. It suggests that the model (1) is dependent on period of the yield data used and thus a unique model cannot be recommended irrespective of duration of the data.

Nevertheless the drought indices may have biological importance because Y is significantly correlated negatively with X_i (i = 1,2,3,4,5). It indicates that the doughtiness of these places in a given year reduces the national yield on the following year. Thus the concept of drought index should not be under estimated. However, it suggests the necessity of thorough study on long-term rainfall variability in many coconuts growing areas.

Time Series Analysis - Univariate Stochastic Approach

A time series is a set of observation generated sequentially in time. The annual yield can be considered as a discrete time series observed at equal time intervals. An intrinsic feature of a time series is that the adjacent observations are dependent. It was found that the auto-correlations of yield at the lags 1, 3, and 4 are significant at the 5% level. It indicates that the yield in a given year is correlated to the prior yields and so time series approach is appropriate. In univariate time series the rainfall variability is not directly taken into consideration, but it assumed that yield fluctuation has occurred due to rainfall variability over the years in coconut growing areas.

It is assumed that an observed time series $Y_1, Y_2, \ldots, Y_n$ can be modeled by an auto-regressive-integrated moving average model of type ARIMA(p,d,q) (Box and Jenkins, 1976) and it can be represented as:

$$\phi(B)(1-B)^d Y_t = \theta(B)e_t \quad t = 1,2, \ldots, n$$

Where Y_t = national yield in the i^{th} year; n = the number of observations for the series; B = the back shift operator s.t. $B^j Z_j = Z_{t-j}$, $\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)$ and $\theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_p B^p)$ are two polynomials in B; p = the order of the auto-regressive part; q = the order of the moving average part and d = the order of non-seasonal differencing of the series. The ϕ_i 's and θ_i 's are parameters of auto-regressive part and moving average part of the model. The e_i 's are independently and identically distributed random variables with zero mean and constant variance σ^2_e. The ARIMA model (2) predicts a value in a time series at a given time as a linear combination of its own past values and past errors.

Identification of Time Series Models

The auto-correlation function of the annual crop production ('original series') from 1950 to 1998 indicated that the series is non-stationary and there is no seasonal or cyclic pattern in the national coconut production. The series was stationary after a single differencing. The simultaneous inspection of the sample ACF and PACF indicated that the annual yield variation can be represented by the models of ARIMA (2,1,0) and ARIMA (2,1,1). For each model the non-significant chi-square values for residuals were observed at lags 6,12,18, and 24. It indicated that residuals were uncorrelated in both models and therefore both models were considered as possible models to represent the annual coconut data.

Selection of The 'Best' Model

The maximum likelihood estimates of the parameters of ARIMA(2,1,0) and ARIMA(2,1,1) models along with their standard errors and correlation matrices are given in Table 1.

Table 1 indicates that the ARIMA (2,1,0) model has the lowest standard errors for parameters and also it has the lowest correlation between the parameters. The t-ratio of all the parameters in ARIMA (2,1,0) were significant at the 5% level. In ARIMA (2,1,1) only the second autoregressive parameter was significant. It indicates that both parameters are significantly important for the ARIMA (2,1,0) model. Further, the statistics such as Akaike's information criterion (AIC), and Schwartz' s Bayesian Criterion (SBC) were smaller in ARIMA (2,1,0) than in ARIMA (2,1,1). Thus it can be concluded that the ARIMA (2,1,0) model is the most appropriate, to model the change in annual coconut production. The model is represented as:

$$Y_t = (1 - \phi_1 B - \phi_2 B^2)(1 - B)Y_t = e_t$$

Where ϕ_1 and ϕ_2 are equal to - 0.26684 and - 0.66548 respectively.

Validation of the Predictions

Based on model (3) the percentage error of the predicted values for the period from 1950 to 1998 varied from 1% to 12% except on two occasions. The plot of actual and estimated values is given in Fig. 1.

Table 1: Estimates of the model parameters using maximum likelihood method.

Model	Parameters	Std. error of the parameters	Correlation matrix
ARIMA(2,1,0)	ϕ_1 = - 0.26684 ϕ_2 = - 0.66548	0.10701 0.10557	1.000 0.157 1.000
ARIMA(2,1,1)	θ_1 = 0.07177 ϕ_1 = - 0.22699 ϕ_2 = - 0.65982	0.22043 0.16167 0.11039	1.000 0.738 1.000 0.224 0.255 1.000

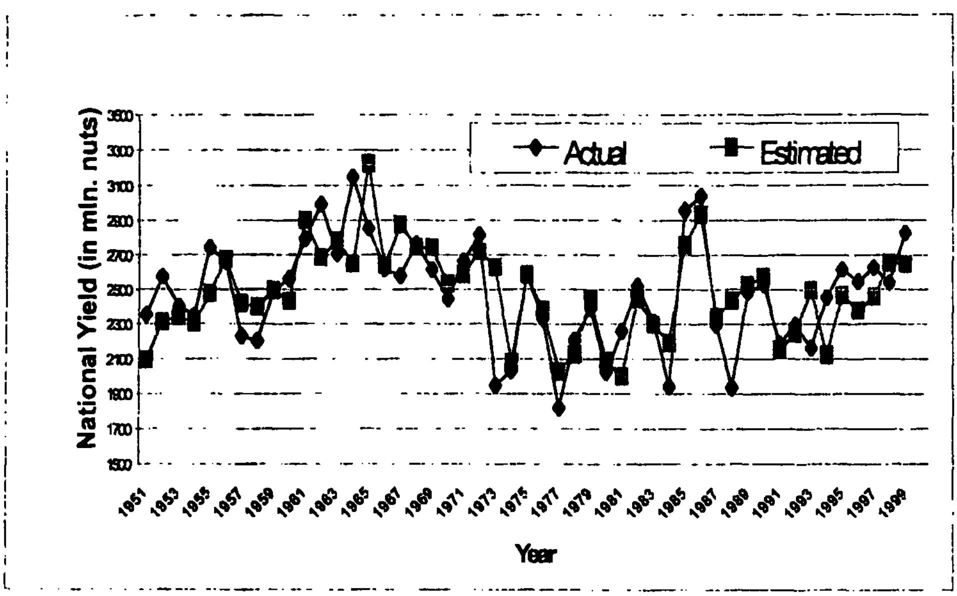

Figure 1: Plot of actual and estimated national annual coconut production.

The correlation between the actual and estimated values is significant ($r = 0.87^{***}$, n=50). The estimated value using model (3) for the year 1999 is 2650 million nuts as against the actual value of 2830. The percentage error is 6.3%, compared to 10.3% error with the regression model.

When two 30-year periods: 1950 - 1979 and 1968 - 1997 were considered it was found that the best fitted model is ARIMA (2,1,0). By including the most recent forecast value, prediction was done for one year ahead using ARIMA (2,1,0) type model. The estimated yield for the period 2000 to 2004 are 2896, 2705 , 2765, 2875 and 2815 million nuts, respectively.

Time Series Analysis - Bivariate Stochastic Approach

In addition to the pattern of climate ('rainfall')) in the previous year, other factors that could influence the national yield are the nut price of the previous years and the amount of fertilizer used in the previous years. It was found that the national coconut production has no significant cross correlation with the nut prices or the amount of annual fertilizer used in the previous years. Because of this reason Abeywaredena (1983) excluded the fertilizer as a dependant variable for his national production function. Therefore, for bi-variate time series those two factors need not to be considered.

DISCUSSION

The need for a national yield forecasting model has been felt by the coconut industry since the failure of Abeywardena's model during 1980 s; Therefore the model developed in this paper is of immense value for various activities of the industry. As the rate of national consumption for culinary use is 90 nuts per head per year, the number of nuts required for culinary purposes can be estimated based on the rate of annual population increase (1%). Therefore the balance nuts after culinary use can be computed using predicted values for future years. The advance knowledge of balance nuts will be useful in planning to share nuts among various activities in the coconut industry such as production of desiccated coconut, coconut oil, copra, coconut milk/cream/milk powder and fresh nut exports. Based on the past trend of nut utilization for the above activities, it can be predicted that there will not be a scarcity of coconut in future for the coconut industry.

The ARIMA class of time series models is a powerful tool for forecasting national coconut production than regression models. The advantage of time series model is that prediction of yield can be done even more than one year ahead without any external variables. Using the predicted values for 2000 and 2001, the predicted national coconut production for the years 2002 to 2004 are 2765, 2875 and 2815 million nuts Time series data are often contaminated with outliers due to the influence of unusual and non-repetitive events. Thus the incorporation of the method suggested by Chung and Liu (1993) for the time series packages will be beneficial to tackle such data.

CONCLUSION

A univariate stochastic model of type ARIMA (2,1,0) can be used to predict the annual coconut production in Sri Lanka for efficient planning of the coconut industry.

ACKNOWLEDGMENT

Part of the work was funded by the National Science Foundation, Colombo, Sri Lanka.

REFERENCES

Abeywardena, V. (1983). Effect of moisture stress and irrigation on yield of coconut in Sri Lanka. (In) *Coconut Research and Development*. N.M. Nayar (Ed.). pp. 98-106. Wiley Eastern Limited, New Delhi.

Anonymous (1998). *Annual Report of the Central Bank of Sri Lanka.* Colombo, Sri Lanka.

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

Chung, C. and L-M. Liu (1993). Forecasting time series with outliers. Journal of Forecasting, 12: 13-35.

Peiris, T.S.G., R.O. Thattil and R. Mahindapala (1995). An analysis of the effect of climate and weather on coconut (*Cocos nucifera*). *Experimental Agriculture* 31:(4) 451-460.

Peiris, T.S.G. and D.T. Mathes (1997). The analysis of long term rainfall variability at selected locations in Sri Lanka. *Proceedings of the International Congress on Modelling and Simulation* 06-11 December, 1997 Hobart, Australia. 69-74.

Peiris, T.S.G. and R.O. Thattil (1998). The study of climate effects on the nut yield of coconut using parsimonious models. *Experimental Agriculture* 34: 189-206.