

FORECASTING THE CROP YIELD OF A COCONUT ESTATE

by

T. S. G. Peiris*

ABSTRACT

Seasonal Autoregressive Integrated Moving Average (ARIMA) process of $(0,1,2) \times (0,1,1) \times 6$ that best fits a set of crop-wise coconut yield data, in Bandirippuwa, Lunuwila is identified without using variance stabilization transformation. In this process the present value of the series may be described as a linear function of the past observation of the series and past disturbances. The physical factors such as rainfall, temperature, day length etc. are not required for this method, however the past crop figures in the estate is needed. While such model is useful for short term forecasting, it also gives the upper and lower limits of the forecasts at a given probability. These intervals would provide the quantified information on the degree of duration of the forecasts.

INTRODUCTION

Coconut is a perennial crop harvested at bimonthly intervals (six harvests a year) in Sri Lanka. The first harvest generally refers to January/February while the sixth harvest is November/December. Variation in yields are observed both between harvests within a year as well as between years. It is believed that the yield fluctuation of coconut depend on physical parameters like rainfall, temperature, humidity and biotic factors like variety, fertility levels and other management conditions. However for a given estate it is reasonable to assume these biotic factors are not random variables. Further it is known that the annual distribution of these physical parameters is more important to the yield than the quantity and rainfall is the most important. But, as the distribution of these factors being highly erratic, forecasting yields has become a problem.

Several authors have attempted to develop different mathematical models for forecasting annual coconut yields of a given estate. To apply models it is necessary to maintain records of daily rainfall of the previous years. Such records are generally not maintained in many estates and therefore those models cannot be applied to them. In addition the models suggestions are not applicable to predict yield in a given harvest of a year. Therefore it would be more useful to identify a model which enables forecasting both harvest-wise and annual yields in an estate in the absence of any physical parameters.

In this paper a seasonal ARIMA model is identified to forecast the yield in the particular estate using Box Jenkins Methodology of time series analysis. The stationary of the series is achieved using one degree each from nonseasonal difference and seasonal difference with the length of seasonality as six. The autocorrelation function (ACF) and the partial autocorrelation function (PACF) are used to identify the model.

MATERIALS AND METHODS

The pick-wise coconut yield from 1955 to 1978 at Bandirippuwa estate of the Coconut Research Institute, Lunuwila was used in this study. The data were analysed using the SPSS package on the mainframe computer as the University of Canterbury, NewZealand.

*Biometry Division, Coconut Research Institute Lunuwila, Sri Lanka.

Let the series $(Z_t, Z_{t-1}, \dots, Z_{t-n+1})$ be observed at discrete equally spaced points in time, where Z_t be the present value of the series. Let X_t be the series of identically independently distributed random variables with zero mean and constant variance. The general nonseasonal ARIMA (p,d,q) can be written as

$$(1-B)^d (1-\phi_1 B - \dots - \phi_p B^p) Z_t = (1-Q_1 B - \dots - Q_q B^q) X_t$$

Where p and q are the respective orders of nonseasonal autoregressive part and moving average part of the model and B is a back shift operator on the index of time series such that, $B^m Z_t = Z_{t-m}$ and $B^m X_t = X_{t-m}$ (m is a positive integer). ϕ_i ($i=1,2,\dots,p$) and Q_i ($i=1,2,\dots,q$) are the non seasonal parameters of the model. The number of the non-seasonal difference is d . Seasonal series generally cannot make stationary only from the non seasonal differences Let D be the number of seasonal differences, S be the length of seasonality, and P and Q be the respective orders of the seasonal AR part and MA part.

Then the multiplicative seasonal ARIMA model is written as;

$$(1-B)^d (1-B^S)^D (1 - \phi_1 B - \dots - \phi_p B^p) (1-Q_1^s B^S - \dots - Q_p^s B^{pS}) Z_t \\ \phi_{is} B^{is} - \dots - \phi_{ps} B^{pS}) (1 - Q_1^s B^S - \dots - Q_q^s B^{qS}) X_t^t$$

Where ϕ_{is} is ($i = 1,2,\dots,p$) and ϕ_{is} ($j = 1,2,\dots,q$) are the seasonal parameters of AR part and MA part respectively. This model is denoted by ARIMA $(p,d,q) \times (P,D,Q) \times S$. The autocorrelation of time lag T ($T=1,2, \dots$) is the correlation between Z_t and Z_{t-T} and the partial autocorrelation of time T is the correlation between Z_t and Z_{t-T} with Z_{t+j} ($j = 1,2,\dots,T-1$) held fixed. The sample autocorrelation coefficient of the partial autocorrelation of lag k are denoted by r_k and P_k respectively. (Malridakis, S and Wheelwright S/C, 1978).

Results and Discussion

The plot of the 144 data points of the original data clearly exhibits a periodic components consisting of regular seasonal pattern. Different sizes of the amplitude of the series suggests that the original series is nonstationary. The fig. 1 shows the plot of autocorrelation coefficients for the data for time lags 1,2,3,.....24. The dashed horizontal lines indicate the 95% probability limits for a purely random process.

As the seasonal autocorrelation coefficients; etc. are highly significantly different from zero it confirms the existence of strong seasonal pattern with a length of six period. The need for seasonal and nonseasonal differencing is therefore obvious. The stationarity of the series was achieved by the combination of one degree of nonseasonal differencing and one degree of seasonal differencing. The plots of autocorrelation function and partial autocorrelation function of the stationary series are exhibited in of Fig. 2 and Fig. 3 respectively.

The seasonal ACF truncates to zero and the seasonal PACF drops to zero. Further the significance of the first seasonal autocorrelation coefficient is suggesting the presence of seasonal moving average part with order one and confirmed that the seasonal component could be represented by ARIMA $(0,1,1) \times 6$. Non seasonal autocorrelation coefficients of lag 1 and lag 2 significantly different from zero. Also the non seasonal autocorrelation coefficients at lag 4 is significant while at lag 8 and lag 10 are just on the boundary of the 95% confidence limits. The nonseasonal partial autocorrelation coefficients at lag 2 lies well outside the 95% significant interval and at lag 4 just touches the confidence limits. The rest of the non seasonal autocorrelation coefficients and partial autocorrelation coefficients oscillate around zero. The nonsignificance of the coefficient at lag 1 could be occurred due to inherent random variation of the series. However the simultaneous

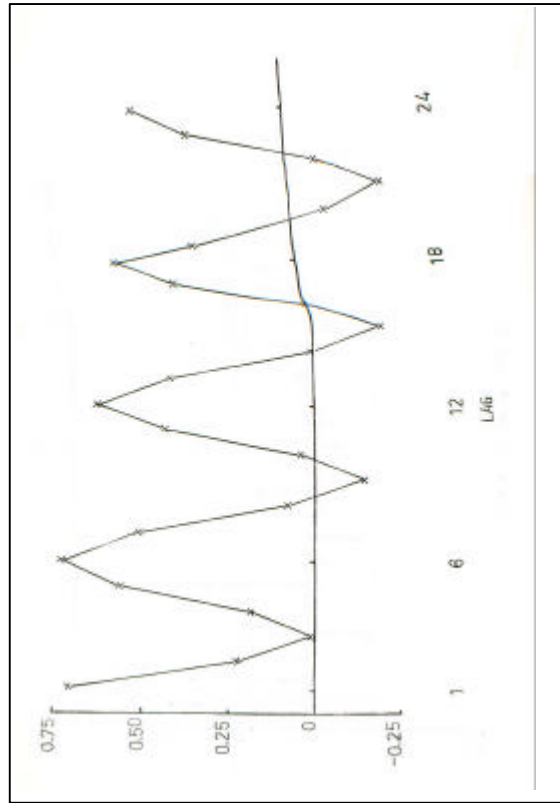
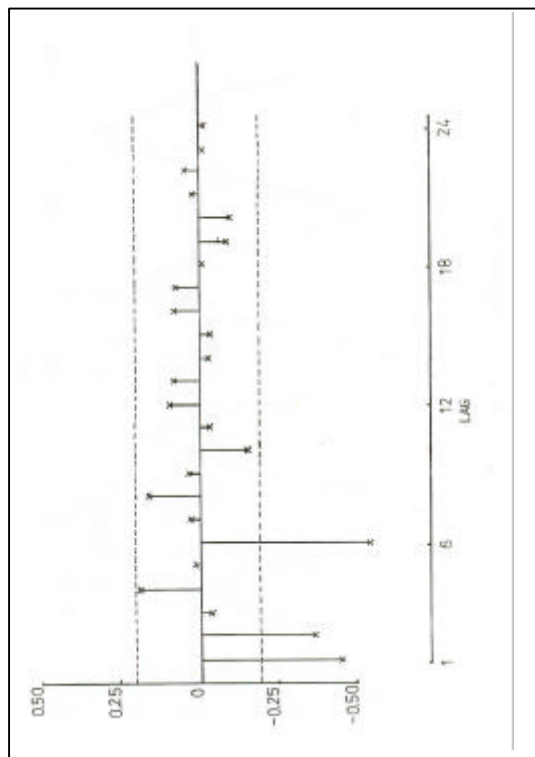


Fig. 1 ACF of the original data



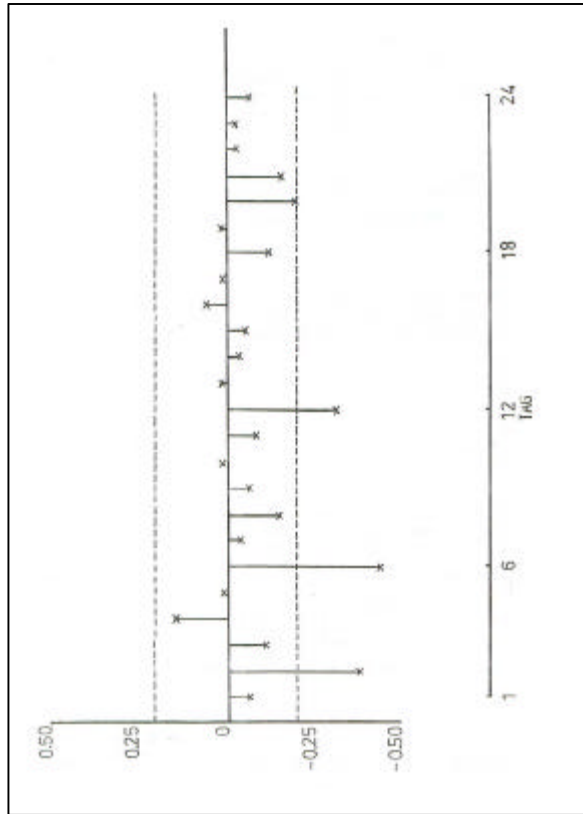


Fig. 3 PACF of 1st and 11t

inspection of both ACF and PACF would suggest that the nonseasonal component could be represented by the models of ARIMA (2,1,0), ARIMA (0,1,2) and ARIMA 1,1,1). Therefore the ARIMA models of (0,1,2,) x (0,1,1) x 6, (2,1,0) x 6, and (1,1,1) x (0,1,1) x 6 were considered as possible process to represent the progina; series. The three. models are denoted by the model 1,2 and 3 respectively.

The Table 1 lists estimates of the three parameters in all three models along with their errors and correlation matrices.

Table 1. Different estimates for the 3 process

Model	Parameters	Std. error of the Parameters	Correlation matrix parameters		
I	$Q_1 = 0.18077$	0.074	1.00		
	$Q_2 = 0.41823$	0.074	-0.032	1.00	
	$Q_3 = 0.91875$	0.025	0.014	-0.093	1.00
II	$?_1 = 0.00375$	0.077	1.00		
	$?_2 = -0.3540$	0.078	0.013	1.00	
	$?_{is} = 92656$	0.024	0.027	0.014	1.00
III	$?_1 = 0.48362$	0.496	1.00		
	$Q_1 = -0.58974$	0.458	0.989	1.00	
	$Q_{is} = 0.91150$	0.028	0.436	0.439	1.00

The model 1 has the lowest standard errors for the parameters and also lowest correlation coefficient between the nonseasonal parameters and the seasonal parameter. The Table 11 shows the diagnostic chi-square statistics for the residuals of the three models at lags of 6, 12, 18, 24 and 30.

Table 2. X^2 value of the 3 identified models at different lags.

Lag	Model I	Model II	Model III	DF
06	4.20ns	8.60ns	17.34***	3
12	12.93ns	17.01ns	25.71**	9
18	16.79ns	20.77ns	31.25**	15
24	20.36ns	24.26ns	33.34*	21
30	26.33ns	29.57ns	37.36	27
	*P=0.05	**P=0.01	***P=0.001	

The nonsignificant lowest chi-square statistics had been observed in the model I for all. lags. Thus, by comparison it can be confirm that the model, ARIMA (0,1,2) x (0,1,1) x 6 would be the most appropriate process to the set of data used in this paper. Plot of the residual autocorrelations of this model shows that the independence assumption of the residual is satisfied. All the values lie within the 95% significance interval which reveals that the residual has no particular pattern. The model I therefore adequately fits for the crop wise data in Bandirippuwa estate.

The multiplicative seasonal ARIMA model for the data is written as $(1-B)(1-B^6)Z_t = (1-0.01808B - 0.4182B^2)(1 - 0.9188B^6)X_t$. The forecasts values from the 145th observation (1st harvest of 1979) to 150th observation (6th harvest of 1979) and their 95% confidence limits along with the percentage values are listed in Table 3.

Table 3

Period	Forecast	95 percent limits		Actual	Percentage error
		Lower	Upper		
145	88,322	75,239	1,51,315	99,957	11.64
146	1,39,293	1,02,496	2,06,090	1,50,166	7.24
147	1,45,842	1,13,059	1,70,626	1,49,830	2.66
148	1,05,822	90,161	1,61,483	1,05,428	0.37
149	61,329	41,960	78,462	54,710	12.10
150	67,750	60,170	95,686	74,902	9.55

Conclusion

The main aim of this paper is to identify an appropriate time series process to a set of coconut data. Having applied the Box Jenkin methodology, it was shown that a multiplicative seasonal ARIMA process of (0,1,2) x (0,1,1) x 6 could fit to the particular set of data at Bandirippuwa estate. This model is useful in forecasting the both annual and crop wise yield. The model is also flexible because the forecast can be done without any physical variables. The percentage error in prediction is reasonably low. Though the confidence intervals obtained in this example are bit wider such intervals provide more information to the estate managers for their future planning programme.

The main difficulty in this method is the selection of an suitable model. Though Box Jenkins methodology provide detailed techniques for identification of the nonseasonal model, identification of seasonal ARIMA models have not been extensively studied.

ACKNOWLEDGEMENTS

The author wish to thank to Prof. J.J. Deely, of the Dept. of Mathematics, University of Canterbury for the valuable comments. Thanks are also to Mr. D.T. Mathes, Senior Biometrician, and to Mrs. D.M.T. Marasinghe for typing the script in the Coconut Research Institute.

REFERENCES

1. Abeywardena, V. (1969) "Forecasting coconut crops using rainfall data a preliminary study" Ceylon Coconut Quarterly 19, 161-176.
2. Abeywardena, V. (1971) "Yield variation in coconut" C.C.Q. 22 97-103.
3. Chatfield, C. (1979) "Inverse autocorrelations" J.R. Statist. Soc. A. 142, 363-372.
4. Godfrey, L. G. (1979) "Texting the adequacy of a time series model" Biometrika 66 67-72.
5. Hamilton, D.C. and Watts, D.G. (1978) "Interpreting partial autocorrelation function of seasonal time series models". Biometrika 65, 135-140.
6. Hull, C.H. and Nil, N.H. (1981) "SPSS update 7-9" Mcgrow Hill Book Company.
7. Makindatics, S. and Wheelwright, S.C. (1978) "Forecasting Methods and application" John Willey and on, New York.
8. Medhi, J. (1982) "Stochaxtic Process" John Willey and Sons, Inc., New York.
9. Newbold, P. (1981) "Some recent developments in Time Series Analysis" International Statistical Review 49, 53-6.
10. Smith, G.W. (1966) "The relation between rainfall, soilwater and yield of copra on a coconut estate in Trinidad" J. Appl. Ecol 3(1) 117-125.