

## Use of Unobserved Components Model for Forecasting Non-stationary Time Series: A Case of Annual National Coconut Production in Sri Lanka

N.K.K. Brintha\*, S. Samita<sup>1</sup>, N.R. Abeynayake<sup>2</sup>  
I.M.S.K. Idirisinghe<sup>3</sup> and A.M.D.P. Kumarathunga<sup>3</sup>

Postgraduate Institute of Agriculture  
University of Peradeniya  
Sri Lanka

**ABSTRACT:** *Forecasting a time series is generally done by using autoregressive integrated moving average (ARIMA) models. The main drawback of this technique is that the time series should be stationary. In reality, this assumption is rarely met. The Unobserved Component Model (UCM) is a promising alternative to ARIMA in overcoming this problem as it does not make use of the stationary assumption. In addition, it breaks down response series into components such as trends, cycles, and regression effects, which could be useful especially in forecasting the production of perennial crops. The present study was aimed at using UCM for annual national coconut production data from 1950 to 2012, which is non-stationary, and to forecast the coconut production in Sri Lanka. Results revealed that both the trend components, level and slope, have non-stochastic processes. Further, it revealed that the level was significant ( $p=0.0001$ ) and slope was non-significant ( $p>0.1$ ). The linear trend model zero variance slope was found to be the best fit for the data with 11.3 years of estimated period of the cycle. The forecasted error for 2011 and 2012 were 1.08% and 1.69%, respectively. From the fitted model, predicted annual coconut production for 2013 was 2739.1 million nuts and the 95% CI is 2048.7 to 3429.5 million nuts. Thus, the use of UCM is recommended for annual data series, too.*

**Keywords:** *Non stochastic process, predicted production, trend components*

## INTRODUCTION

Box Jenkins and to a limited extent the exponential smoothing techniques are commonly used in the analysis of time series in agriculture. Main drawbacks in these models are that, they are suitable only for the stationary series (Box *et al.*, 1994), empirical in nature and fail to explain the underlying mechanism. It is not always possible to create a time-series stationary by differencing or by some other means. Hence, this approach could be limited to few data sets. Also, correlogram and partial auto correlation function specifying the models are not always very informative, especially in small samples. This could lead to inappropriate models and predictions.

---

<sup>1</sup> Department of Crop Science, Faculty of Agriculture, University of Peradeniya, Sri Lanka

<sup>2</sup> Department of Agribusiness Management, Faculty of Agriculture and Plantation Management, Wayamba University of Sri Lanka

<sup>3</sup> Coconut Research Institute, Lunuwila, Sri Lanka

\* Corresponding author: ilbrintha@yahoo.com

Unobserved component model (UCM) is a promising alternative approach to overcome these problems (Harvey, 1996). It is also known as structural time series models and it is a flexible class of models which are useful for forecasting. It decomposes the response series into latent components such as trend, cycle and seasonal effect and linear and nonlinear regression effects. The salient feature of the UCM is latent components, which follow suitable stochastic models and it provides suitable set of patterns to capture the outstanding actions of the response series. UCM can also consist of explanatory variables. Apart from the forecast, structural modeling gives estimates of these unobserved components and it is of very useful in practical usage. UCM can handle intensive data irregularities too. It is very similar to dynamic models and also popular in the Bayesian time series (West & Harrison, 1999).

Perennial crop production is influenced by environmental and management factors. It is quite hard to assume that the underlying parameters are consistent and difficult to capture the latent components by univariate ARIMA models. Harvey & Todd (1983) reported in detail the advantages of UCM over seasonal ARIMA. Ravichandran & Prajneshu (2001) compared the efficiencies of ARIMA and State Space Modeling utilizing all-India Marine products export data and Kapombe & Colyer (1998) studied that structural time series model to estimate the supply response function for broiler production in the United States using quarterly data. Ravichandran & Muthurama, (2006) utilized UCM model to model and forecast the rice production of India. Even though the UCM has been used in actual scenario, there is hardly any use of UCM in forecasting perennial crop production. The aim of this study was to investigate the possibility of using UCM for modeling and forecasting annual national coconut production of Sri Lanka.

## METHODOLOGY

### Data used in the study

Annual coconut production from 1950 to 2012 collected from Coconut Research Institute (CRI) in Sri Lanka was used for the study.

### UCM

A UCM consists of trend, cycle, seasonal and irregular components, and specified of the form (Harvey and Stock, 1993).

$$Y_t = \mu_t + \varphi_t + \omega_t + \varepsilon_t \quad (1)$$

where  $\mu_t$ ,  $\varphi_t$  and  $\omega_t$  denotes the stochastic trend, stochastic cycle and seasonal component respectively. Here  $\varepsilon_t$  is the overall error (irregular component), which is assumed to be a Gaussian white noise with variance  $\sigma_\varepsilon^2$ . Since the data used is annual, seasonal effect cannot be identified and thus the UCM for the data can be formulated of the form

$$Y_t = \mu_t + \varphi_t + \varepsilon_t \quad (2)$$

### Estimating trend effect

There are two different ways to modeling the trend component in UCM. The first method is by mean of random walk (RW) model, (3). The RW model can be formulated of the form (Harvey & Koopman, 2009).

$$\mu_t = \mu_{t-1} + \delta_t, \delta_t \sim i.i.d N(0, \sigma_\delta^2) \quad (3)$$

The second method involves modeling the trend as a Locally Linear Time trend (LLT), which consist of both level and slope (Harvey, 2001). The trend,  $\mu_t$  is modeled as a stochastic component with varying level and slope and it can be formulated of the form,

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \delta_t, \delta_t \sim i.i.d N(0, \sigma_\delta^2) \quad (4a)$$

$$\beta_t = \beta_{t-1} + \tau_t, \tau_t \sim i.i.d N(0, \sigma_\tau^2) \quad (4b)$$

Where  $\beta_t$  is the slope of the local linear time trend. The disturbances  $\delta_t$  and  $\tau_t$  are assumed to be mutually independent. Special cases of this trend model is obtained by setting one or both of the disturbance variances,  $\sigma_\delta^2$  and  $\sigma_\tau^2$ , equal to zero. If  $\sigma_\tau^2$  is set equal to zero, then the trend becomes linear (fixed slope). If  $\sigma_\delta^2$  is set to zero, then the subsequent model generally has a smoother trend. If both the variances are set to zero, then the resulting model is the deterministic linear time trend,

$$\mu_t = \mu_0 + \beta_0 t \quad (5)$$

Thus the reduced form of a LLM is the ARIMA (0, 2, 2) model.

### Estimating cyclic effect

Cyclical function of time  $\varphi_t$  with frequency  $\lambda$  is usually measured in radians. The period of the cycle, which is the time taken to go through its complete sequence of values, is  $2\pi/\lambda$ . A cycle can be expressed as a mixture of sine and cosine waves, depending on two parameters,  $\alpha$  and  $\beta$  (Harvey & Stock, 1993).

Accordingly,

$$\varphi_t = \alpha \cos \lambda t + \beta \sin \lambda t \quad (6)$$

$$\varphi_t^* = -\alpha \sin \lambda t + \beta \cos \lambda t \quad (7)$$

where  $(\alpha^2 + \beta^2)^{1/2}$  is called the amplitude and  $\tan^{-1}(\beta/\alpha)$  is the phase. As with the linear trend, the cycle can be built up recursively, leading to the stochastic model.

$$\begin{bmatrix} \varphi_t \\ \varphi_t^* \end{bmatrix} = \begin{bmatrix} \cos \lambda & \sin \lambda \\ -\sin \lambda & \cos \lambda \end{bmatrix} \begin{bmatrix} \varphi_{t-1} \\ \varphi_{t-1}^* \end{bmatrix}, t = 1, 2, \dots, T$$

Then the cyclic component  $\varphi_t$  is modeled of the form.

$$\begin{bmatrix} \varphi_t \\ \varphi_t^* \end{bmatrix} = \rho \begin{bmatrix} \cos \lambda & \sin \lambda \\ -\sin \lambda & \cos \lambda \end{bmatrix} \begin{bmatrix} \varphi_{t-1} \\ \varphi_{t-1}^* \end{bmatrix} + \begin{bmatrix} v_t \\ v_t^* \end{bmatrix}$$

Here  $\rho$  is the damping factor, where  $0 \leq \rho \leq 1$  and the disturbances  $v_t$  and  $v_t^*$  are mutually independent white noise disturbances with zero mean and common variance  $\sigma_v^2$ . This results in a damped stochastic cycle that has time-varying amplitude and phase, and a fixed period equal to  $2\pi/\lambda$ . The parameters of this UCM are the different disturbance variances,  $\sigma_\varepsilon^2$ ,  $\delta_t$ ,  $\tau_t$ , and  $\sigma_v^2$ , the damping factor  $\rho$ , the frequency  $\lambda$ . The model is stationary if  $\rho$  is strictly less than one, and if  $\lambda$  is equal to 0 or  $\pi$  it reduces to a first-order autoregressive process.

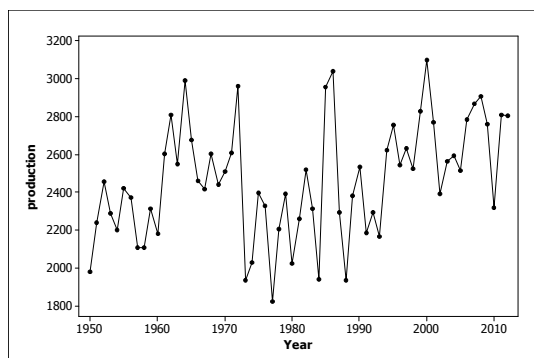
### Residual analysis and forecasting

A useful diagnostic tool for investigating the randomness of a set of observations is the correlogram. Residual diagnostic plots are useful for checking the normality and the randomness in the residuals. For the fitted models, distribution of auto correlation function residuals were examined graphically as well as tested. After verification of the assumptions of residuals of the selected model, model was used to forecast the values from 2013 to 2017. The *PROC UCM* of SAS was used for model fitting.

## RESULTS AND DISCUSSION

The time series plot of annual coconut production showed the positive trend (Fig. 1). With this positive trend in the data, all possible components such as cycle and irregular component was tested by using a UCM of the form.

$$Y_t = \mu_t + \varphi_t + \varepsilon_t$$



**Fig. 1. Time series plot of annual coconut production (1950-2012)**

At the first stage, analysis was aimed to rectify the existing component in the model by UCM technique. Error variances of the irregular, level, slope, and cyclic components were considered as free parameters of the model and their estimates are showed in the Table 1. These estimates, their corresponding t-values and the associated P values were used to test the hypothesis of the form.

$H_0$ : Corresponding component is non-stochastic

$H_a$ : Corresponding component is stochastic

According to the Table 1, disturbance variances for the level and slope components are not significant. This suggests that a deterministic trend model may be more appropriate and level and slope can be treated as constant.

**Table 1. Final Estimates of the Parameters**

Component	Parameter	Estimate	Std. Error	T value	Pr>t
Irregular	Error variance	3238.54	15623.1	0.21	0.8358
Level	Error variance	14484	8806.1	1.64	0.1000
Slope	Error variance	0.000124	0.08968	0.00	0.9989
Cycle	Damping factor	0.76777	0.1486	5.17	<.0001
Cycle	Period	3.885	0.3272	11.87	<.0001
Cycle	Error variance	17067	8258.8	2.07	0.0388

However whether model is deterministic or not, cannot be determined from estimates of parameters of stage 1 (Table 1) and it should be determined from the second stage analysis, which is the significant analysis of component. In addition, significant analysis of component helps to decide if level and slope can be dropped from the model after testing the following hypothesis,

$H_0$ : Given component is not significant

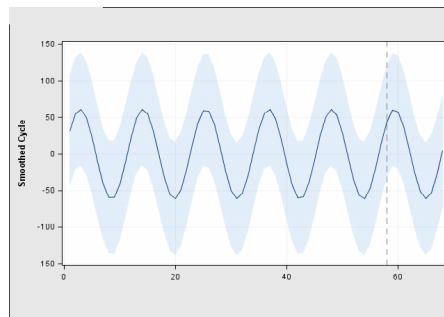
$H_a$ : Given component is significant

The goodness of fit of the analysis of components is shown in Table 2. According to Table 2, slope is not significant and can be dropped from the model. However, level is significant and cannot be dropped from the model thus the model is a stochastic model. The contribution of irregular component is also not significant, but since it is a stochastic component, it cannot be dropped from the model.

**Table 2. Significance Analysis of Component (Final State)**

Component	DF	Chi-square	P>chisq
Irregular	1	0.01	0.9320
Level	1	515.86	<0.0001
Slope	1	0.27	0.6048

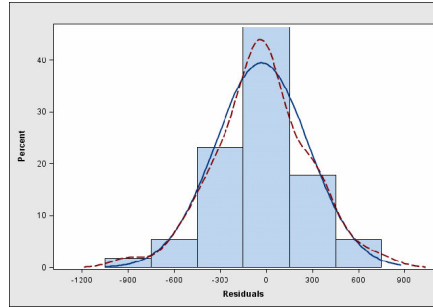
At the third stage, slope variance was fixed and free parameters were obtained. Accuracy measures, AIC and BIC with fixed slope models were recorded as 810.72 and 820.84 respectively and likelihood optimization algorithm converged at 21 iterations. After fixing the slope, MAPE was 9.66 and the estimated period of the cycle was 11.30 years. The estimate of the damping factor was 1, suggesting that the periodic pattern of production does not diminish quickly as shown in Fig. 2.



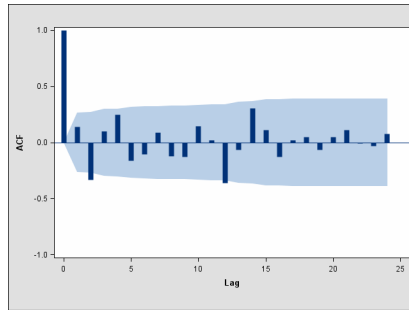
**Fig. 2. Smooth cycles with constant slope. Note that dotted line indicates the beginning of forecasted values.**

### Residual analysis

Fig. 3 clearly indicates that residuals have the normal distribution ( $P = 0.79$  from Anderson Darling test). According to Fig. 4, the indication is that there is no serious violation of auto correlation assumption.



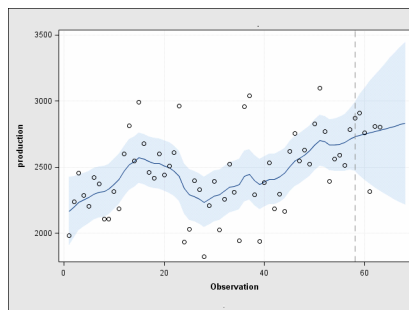
**Fig. 3.** Distribution of residual for fixed slope, — normal, - - - kernal



**Fig. 4.** ACF for fixed slope, ■ two standard errors

## Forecasting

Smooth trend for the production is shown in Fig. 5.



**Fig. 5.** Smooth trend for the production (○ actual, --begining of forecasted values, ■ 95% confidence limits)

Observed, fitted and forecasted values for fixed slope structural time series model is shown in Table 3. According to Table 3, forecasted error percentages for past five years were below 4% except in 2010. In fact, large error percentage for the year 2010 could be due to the unusual value for 2010 compared to other years. Forecasted value for 2013 is 2739.1 million nuts.

**Table 3. Observed, fitted and forecasted values for fixed slope**

Year	Observed value	Fitted / forecasted values	95% limit		Absolute % Error
			Lower	Upper	
2008	2908	2800.9	2231.7	3370.2	3.68
2009	2762	2808.5	2212.1	3404.9	1.68
2010	2317	2798.8	2176.4	3421.1	20.79
2011	2808	2777.6	2130.9	3424.3	1.08
2012	2802	2754.4	2085.2	3423.6	1.69
2013		2739.1	2048.7	3429.5	
2014		2739.3	2028.3	3450.3	
2015		2758.0	2026.2	3489.7	
2016		2792.4	2039.1	3545.8	
2017		2835.3	2059.0	3611.5	

### CONCLUSION

UCM with slope variance zero seems to fit the annual national coconut production data well. Forecasted error percentage for years of 2011 and 2012 were 1.08% and 1.69% respectively. Obtained model predicted the annual coconut production of 2739.1 million nuts in 2013 and the 95% CI is 2048.7, 3429.5. UCM models can effectively be utilized for the time series modeling of perennial crop production, especially that are of non-stationary.

### ACKNOWLEDGMENT

The authors wish to express their sincere gratitude to Coconut Research Institute, Sri Lanka for providing national coconut production data and HETC project, Ministry of Higher Education, Sri Lanka for the grant provided to pursue the study.

### REFERENCE

- Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (1994). Time series analysis: Forecasting and control, 3<sup>rd</sup> edition. Prentice Hall, New Jersey.
- Harvey, A. and Koopman, S.J. (2009). Unobserved components models in economics and finance. The role of kalman filter in Time series econometrics. pp 71 - 81.
- Harvey, A.C. and Stock, J.H. (1993). Estimation, Smoothing, Interpolation and Distribution in Structural Time Series Models in Continuous Time (with J Stock), in Models, Methods and Applications of Econometrics, P C B Phillips (ed.), pp 55 - 70.
- Harvey, A.C. and Todd, P.H.J. (1983). Forecasting economic time series with structural and Box Jenkins models: A case study. Journal of Business and Economics statistics, 1(4), 299 - 315.
- Harvey, A.C. (1993). Time series models. 2<sup>nd</sup> Edition, Harvester Wheatsheaf.



Harvey, A.C. (1996). Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge Univ. Press, U.K.

Harvey, A.C. (2001). Forecasting, Structural Time series models and the Kalman filter. Cambridge Univ. Press, U.K.

Kapombe, Crispin M. and Colyer, Dale. (1998). Modeling U.S. broiler supply response: A structural time series approach. *Agricultural and Resource Economics Review*, 27(2), 241 - 251.

Ravichandran, S. and Prajneshu. (2001). State space modeling versus ARIMA time series modeling. *Journal Ind. Soc. Ag. Statistics*, 54(1), 43 - 51.

Ravichandran, S. and Muthuraman, P. (2006). "Structural time-series modelling and forecasting India's rice production", *Agricultural situation in India*, pp.773 - 776.

West, M. and Harrison, J. (1999). Bayesian forecasting and dynamic models, 2<sup>nd</sup> edition, New York; Springer-Verlag.