Unobserved Components Model for Forecasting Sugarcane Production in Sri Lanka

B.D.S.K. Ariyawansha [1*], N.R. Abeynayake [2] and T. Sivananthawerl [3]

[1]*Postgraduate Institute of Agriculture, University of Peradeniya, Peradeniya, Sri Lanka.*

[2]*Department of Agribusiness Management, Faculty of Agriculture and Plantation Management, Wayamba University of Sri Lanka, Makandura, Gonawila, Sri Lanka.*

[3]*Department of Crop Science, Faculty of Agriculture, University of Peradeniya, Peradeniya, Sri Lanka.*

ARTICLE INFO

ABSTRACT

Unobserved Components Model (UCM) is a structural time series model and it can decompose the response series into latent components, such as trend, cycle and seasonal effects and linear and nonlinear regression effects. The UCM combines the capabilities of Autoregressive Integrated Moving Average (ARIMA) model with interpretability of smoothing models. This study was carried out to forecast sugarcane production in Sri Lanka using UCM model. The best fitting model was selected based on Akaike information criterion (AIC) and Bayesian Information Criterion (BIC), followed by residual analysis. The selected model was used to make sample period forecasts (From 1979 - 2013) and post sample period forecasts (From 2014 to 2018). Forecasting accuracy of model was evaluated using the Mean Absolute Percentage Error (MAPE). Linear trend model (adj. R^2=77 %) with zero variance slope and two cycles was selected as best among the tested UCM models for cane production data. MAPE was 10.56 % for sample period forecasts and 4.01 % for post-sample period forecasts. Predicted cane production for year 2019 was 813,888 ± 293,891 tons.

*Corresponding author : sandyariyaw@gmail.com

INTRODUCTION

National level agricultural production forecasts are important for making policy decisions. Therefore, yield-forecasting systems are available for major crops such as coconut, paddy, rubber, and tea in Sri Lanka. However, no such attempts have been reported for forecasting national level sugarcane production in Sri Lanka. Therefore, in this paper, an effort has been made to forecast sugarcane production by using time series analysis.

Time series forecasting refers to the use of statistical models to forecast future events based on known past events. Autoregressive Integrated Moving Average (ARIMA) time series methodology is widely used to model time series in agricultural production. The main limitation of the ARIMA approach is that it can be applied in the situations where either the series to be analysed is stationary, or it can be differenced into a stationary process (Koopman and Ooms, 2010). Also for small data sets, correlogram and partial auto-correlation functions produced by the ARIMA model are less informative resulting in inappropriate model specifications and predictions (Brintha *et al.,* 2014). Moreover, ARIMA methodology is empirical in nature and fails to explain the underlining mechanism (Brintha *et al*., 2014). Unobserved component models (UCM) can be used as an alternative approach to overcome these problems (Harvey, 1996; Koopman and Ooms, 2010). UCM model analyses and forecasts time series data by breaking down the response series into latent components that are useful in explaining and predicting its behavior such as trends, seasonal factors, cycles, and regression effects due to the predictor series. The UCM combines the skillfulness of the ARIMA model with interpretability of the smoothing model (Koopman and Harvey, 2003).

Some studies have been reported on effective utilization of UCM for forecasting agricultural production. Ravichandran and Muthuraman (2006) forecasted rice production of India by using the UCM approach. Brintha *et al.* (2014) employed the UCM methodology to forecast annual coconut production of Sri Lanka. Rajarathinam *et al*. (2016) found that the UCM can effectively model the wheat production of India. Singh *et al.* (2014) utilized UCM to forecast gram production of India, and discussed the relative merits of structural models compared to ARIMA.

Sugarcane production is influenced by cultivated area, environmental factors, technological advancements, changes in management, as well as yield variations due to 6-8 year ratooning cycles, and these underlining factors are not consistent over time. Therefore, structural models that have the capability of capturing latent components would be suitable to model the cane production data. The aim of this study was to model and forecast the annual national sugarcane production in Sri Lanka using UCM.

MATERIALS AND METHODS

Data used in the study

Annual sugarcane production data from 1979 to 2018 available in the Sugarcane Research Institute and Central Bank Reports of Sri Lanka were used for the study.

UCM

Basic UCM is consisted of trend, cycle, seasonal and irregular components, and specified of the form (Harvey and Stock, 1993).

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

Where $\mu_t, \varphi_t, \omega_t,$ and ε_t denotes the stochastic trend, stochastic cycle, seasonal component and irregular component respectively. Here ε_t is assumed to be a Gaussian white noise with variance σ_ε^2. Since annual data is used, seasonal effect cannot be identified. In the absence of seasonal component, equation (1) reduces to equation (2).

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

Estimating trend effect

Trend effect of the UCM can be modeled as Local liner model (LLM) i.e. Random Walk model (RWM) or Locally Linear Time (LLT) trend (Harvey, 2001).

Harvey and Koopman (2009) formulated RWM as equation (3)

$$\mu_t = \mu_{t-1} + \delta_t , \quad \delta_t \sim \text{iid } N(0, \sigma^2_\delta) \quad (3)$$

In the local linear trend model, the trend can be modelled as a stochastic component with varying level (μ_t) and slope (β_t). The LLT model can be described by the following equations.

Level: $\mu_t = \mu_{t-1} + \beta_{t-1} + \delta_t$;
$\delta_t \sim \text{iid } N(0, \sigma^2_\delta)$ (4.1)

Slope: $\beta_t = \beta_{t-1} + \tau_t$;
$\tau_t \sim \text{iid } N(0, \sigma^2_\tau)$ (4.2)

The disturbances δ_t and τ_t are assumed to be mutually independent. If σ^2_δ is set to zero then the resulting model has a smooth trend. If variance σ^2_τ set equal to zero, then the resulting model has a linear trend with fixed slope. If both variances are zero, then the resulting model has a deterministic linear time trend (equation 5).

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

Estimating cyclic effect

Cyclic effects are similar to seasonal effects but the period is not known and determined from the data. A periodic pattern can be expressed as a sum of cycles. Cyclical fluctuations with time φ_t with frequency λ are measured in radians. Period of a cycle is defined as the time taken to go through its complete sequence of values, and it is equal to $2\pi/\lambda$. Depending on two parameters (α & β), cyclical fluctuations can be expressed as a mixture of sine and cosine waves as given in equations 6 and 7 (Harvey and Stock, 1993)

$$\phi_t = \alpha \cos \lambda t + \beta \sin \lambda t \qquad (6)$$

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

Where, $\sqrt{\alpha^2 + \beta^2}$ and arc tan (β / α) represent the amplitude and phase responsively.

Following Harvey (1996), cycles can be made stochastic by allowing the parameters α and β to evolve over time.

$$\begin{bmatrix} \phi_t \\ \phi_t^* \end{bmatrix} = \rho \begin{bmatrix} \cos \lambda & \sin \lambda \\ -\sin \lambda & \cos \lambda \end{bmatrix} \begin{bmatrix} \phi_{t-1} \\ \phi_{t-1}^* \end{bmatrix} + \begin{bmatrix} \upsilon_t \\ \upsilon_t^* \end{bmatrix}, \quad t=1, 2, 3,....T \qquad (8)$$

Here the correlation coefficient ρ is the damping factor, where $0 \le \rho \le 1$. The υ_t and υ_t^* are uncorrelated white noise disturbance terms with zero mean and common variance σ^2_υ. This results in a damped stochastic cycle that has time varying amplitude and phase. The parameters of the cycle component are the damping factor ρ, the frequency λ, and the variance σ^2_υ of the disturbance term v_t. When the damping factor is less than one, ϕ_t is stationary. If λ is equal to zero or π the model reduces to the first order auto regressive process.

Assessing the model fit

The best fitting model is selected based on Akaike information criterion (AIC) and Bayesian information criterion (BIC) as shown in equations 9 and 10.

$$AIC = -2 \log L + 2n \qquad (9)$$

$$BIC = -2 \log L + n \log T \qquad (10)$$

Where, L denotes full likelihood value of the fitted model. The n is the number of free-parameters that are estimated in the chosen model. T is the total number of observations used to estimate the candidate model. These two criteria are useful for discriminating among various competing UCM models. The model that minimizes these two measures was selected as the best fitting model.

Residual analysis

Residual analysis is important to check the model adequacy with independently and identically distributed residuals. The residual diagnostic plots were used to check the normality (histogram and QQ plot) and

the whiteness (ACF and PACF) of the residuals.

Forecasting and accuracy checking

After verification of the assumptions of the residuals, the selected model was used to make sample period forecasts (From 1979-2013) and post sample period forecasts (From 2014 to 2018). Mean Absolute Percentage Error (MAPE) was used to evaluate the forecasting accuracy of the model (equation 11).

$$\text{MAPE} = \frac{1}{n} \sum \left| \frac{Actual - Forecast}{Actual} \right| X\ 100 \tag{11}$$

After evaluating the model accuracy, forecasts were updated by using the full data set (From 1979 – 2018). Final conclusions on the forecasts were made based on the full dataset.

RESULTS AND DISCUSSION

The time series plot of sugarcane production from 1979 to 2018 is depicted in Figure 1. Up to mid-1980's the cane production varied between 200,000 t to 330,000 t with the contribution of the Kantale and Hingurana sugar industries. In mid-1980's the sugar industry was expanded with the establishment of two new sugar industries, Pelwatte and Sevanagala. Therefore, cane production was increased gradually during the period 1986 to1996 with the expansion of cultivated extent due to contribution of four sugar industries. However, downfall of sugar industry was observed after its restructuring during early 1990's due to closure of sugar mills in Kantale in early 1992 and in Hingurana in 1997. Therefore, cane production gradually declined from 1998 to 2011. With restarting of Hingurana sugar industry, cane production showed an upward trend from 2012.
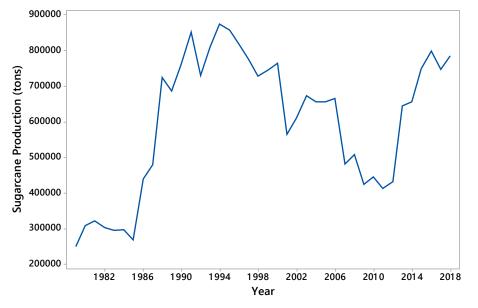

Figure 1: Time series plot of national cane production in Sri Lanka during 1979-2018.

After observing the general pattern, UCM technique was used to detect all possible time series components (level, slope, cycle, and irregular component) in the sugarcane production.

At the first stage, analysis was aimed to detect the existing time varying components in the data. The error variances of the irregular, level slope and cyclic components were called "free parameters" in the model and their estimates are given in the Table 1. These estimates and their corresponding t-values were used to test the null hypothesis that the corresponding component is non-stochastic.

Table 1: Final estimates of the free parameters

Component	Parameter	Estimate	Std. Error	t Value	P > t
Irregular	Error Variance	1085181806	809336313	1.34	0.18
Level	Error Variance	2229128429	1483599009	1.5	0.133
Slope	Error Variance	0.12479	2167	0	1
Cycle 1	Damping Factor	1	0.00019	5235.4	<.0001
Cycle 1	Period	8.60129	0.25487	33.75	<.0001
Cycle 1	Error Variance	11018	12739.2	0.86	0.3871
Cycle 2	Damping Factor	1	0.0001941	5151.9	<.0001
Cycle 2	Period	31.09	3.61842	8.59	<.0001
Cycle 2	Error Variance	148955	169803.6	0.88	0.3804

The results revealed that disturbance variance of the level and slope components were not significant. The slope has recorded the highest p value in the free parameter (Table 1). This suggests that the slope is not time varying and should be made deterministic. Therefore, slope can be treated as a constant, i.e. has a zero variance. Dampening coefficients and frequency of cycles are highly significant; therefore, cyclic term is contributing to the model and should be retained.

At the second stage, the significance analysis of components was carried out and it was used to decide the component retention in the model by testing the following hypothesis using the Chi-square statistics.

H_0 : Considered component is not significant

H_1: Considered component is significant

Table 2: Significance Analysis of Components (Based on the Final State)

Component	DF	Chi-Square	P > Chi-Square
Irregular	1	0.00	0.9459
Level	1	179.62	<.0001
Slope	1	1.58	0.2092
Cycle 1	2	13.34	0.0013
Cycle 2	2	17.08	0.0002

Results of the analysis are given in the table 2 and it revealed that, level and cycles are significant. Therefore, these components should be retained in the model. The irregular component was also retained in the model as a matter of principle since it is a stochastic component. Brintha *et al.* (2014) employed UCM to analyse coconut production in Sri Lanka and found that the level and slope components have non-
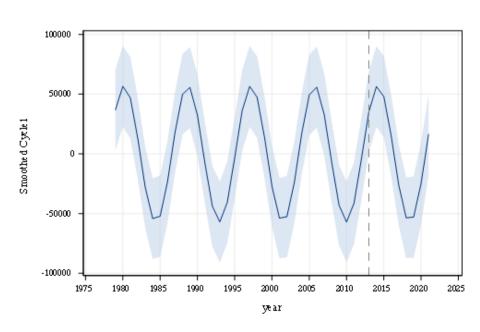
stochastic processes. Similar results were reported in their study regarding significance of level and slope components. Rajarathnam *et al*. analysed area, production and productivity of wheat crops in India using the UCM and they found that UCM with slope variance zero was suitable for forecasting wheat production.

At the next step, data set was re-analyzed by making the slope deterministic. Lower

accuracy measures (AIC=977.96 and BIC 991.06), and residual analysis suggested the linear trend model with zero variance slope and two cycles is the best among the tested UCM for cane production time series of Sri Lanka. The likelihood algorithm converged after 28 iterations. Selected model reported an adjusted R^2 value equal to 0.77, which can be considered a good fit.

Table 3: Summary statistics for cycles.

Name	Type	Period	Frequency	Damping Factor	Final Amplitude
Cycle 1	Stationary	8.6	0.73049	1.00000	57503
Cycle 2	Stationary	31.09	0.20209	1.00000	213141

Figure 2: Smoothed cycle 1 for cane production with 95 % confidence limits (dotted line indicates the beginning of forecasted values).

Summary of cycles is presented in Table 3. The estimated periods of cycles are 8.6 and 31.09 for
cycle 1 and the estimated damping factor is equal to 1, implying that the periodic pattern reflected by the cycle 1 is persistent as shown in Figure 2. The damping factor equal to one for the cycle 2 shows that the effect of

shock persists in the data. However, it was not possible to detect such persistence from the cyclic pattern depicted by the second cycle (figure 3), since, it represent only one cycle due to the 31 years long cyclic period (nearly two third fraction of the full data set).
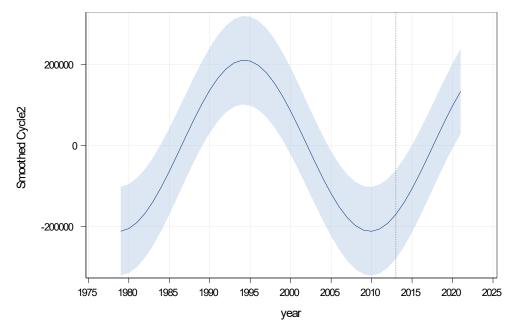
Figure 3: Smoothed cycle 2 for cane production with 95 % confidence limits (dotted line indicates the beginning of forecasted values).
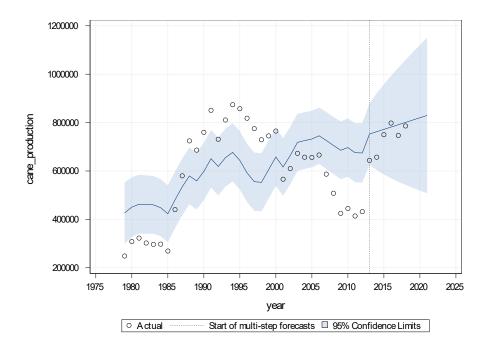

Figure 4: Smoothed trend for cane production with 95 % confidence limits (dotted line indicates the beginning of forecasted values).

Smooth trend for cane production is depicted in Figure 4. Predictions for cane production show increasing trends. Expected increase may be due to expansion of cultivated area, and higher yields due to new high yielding sugarcane varieties, and other technological advances.

Residual analysis

Panel of residual diagnostic plots for the fitted model is shown in Figure 5. Histogram and quantile plots agree with normality of residuals and it was confirmed by the non-significance with Anderson Darling

normality test (p=0.9) and Shapiro-Wilks test (p>0.1). ACF and PACF plots also do not exhibit violations of the whiteness

assumptions; the correlations at all non-zero lags seem to be non-significant.
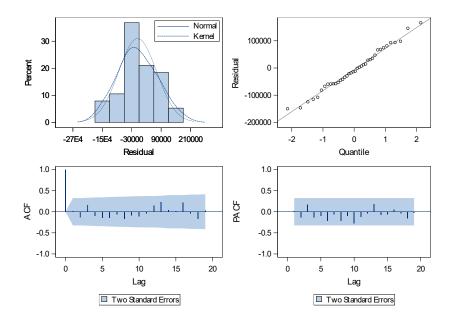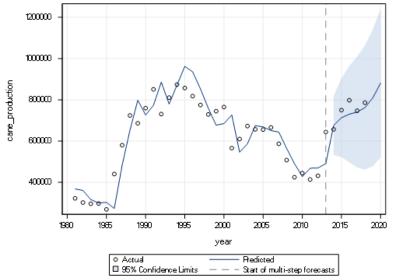

Figure 5: Residual diagnostic plots for the selected cane production model.

Forecasting and accuracy checking:

Figure 6: Forecasts for cane production (dotted line indicates the beginning of forecasted values).

Actual and predicted values of cane production showed that predicted values are closer to actuals implying the accuracy of the model (Figure 6).

Table 4 presents the forecasted values of the cane production and confidence intervals

based on the selected UCM model. Forecasted error percentages for years 2014 to 2018 were below 5 percent except in the year 2016. Reason for high deviation between expected and observed values is due to the considerable increase in cane production during 2016 compared to

previous years. Increase of harvested area was due to the expansion of sugarcane cultivation, high yielding varieties, improvements in irrigation systems and cultivation practices which were the main contributory factors for the increased production in year 2016 (Central Bank, 2016). MAPE calculated for post sample period is 4.01 %., indicating high forecasting accuracy according to the classification criteria given by Lewis based on MAPE values (Lewis, 1982).

Table 4: Forecasts for cane production (t) from 2014-2020

Year	Observed value	Forecasted value	95% Confidence Limits		Absolute % error
			Lower	Upper	
2014	657000	675723	533286	818159	2.85
2015	750000	713345	522704	903985	4.89
2016	798000	730406	498204	962608	8.47
2017	747000	740640	472140	1009141	0.85
2018	786110	762662	461906	1063417	2.98
2019		813888	519997	1107778	
2020		883760	565117	1202403	

CONCLUSIONS

UCM with Linear trend with slope variance zero and two cycles can be used to forecast sugarcane production in Sri Lanka. The model explained 77 % yield variability between years. MAPE was 10.56 % and 4.01 % for sample period forecasts and post sample period forecasts respectively. Prediction for year 2019 was 813,888 tons with lower and upper confidence interval of 519,997 and 1,107,778 tons. Future predictions showed increasing trend in cane production.

It is suggested to update the model periodically with incorporation of new data to improve the accuracy of estimates.

ACKNOWLEDEMENT

REFERENCES

Central Bank (2016). Annual Report 2016, Central Bank of Sri Lanka. 55.

Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (1994). Time series analysis: Forecasting and control, 3rd edition. Prentice Hall, New Jersey.

Brintha, N.K.K., Samita, S., Abeynayake, N., Idirisinghe, I.M.S.K. and Kumarathunga, A.M.D.P. (2014). Use of Unobserved Components Model for Forecasting Non-Stationary Time Series: A Case of Annual National Coconut Production in Sri Lanka. Tropical Agricultural Research, 25(4), 523 – 431.

Harvey, A.C. and Koopman, S.J. (2009). Unobserved components models in economics and finance. The role of Kalman Filter in time series econometrics. IEEE Control Systems Magazine, 29(6), 71-81. https://doi.org/10.1109/MCS.2009.934465

Harvey, A.C. and Stock, J.H. (1993). Estimation, Smoothing, Interpolation and Distribution in Structural Time Series Models in Continuous Time, In: P.C.B. Phillips (ed.), Models, Methods and Applications of Econometrics, Blackwell, 55-70.

Harvey, A.C. (2001). Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge Univ. Press, U.K. IEEE Control Systems Magazine, 71 – 81.

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

Koopman, S.J. and Harvey, A.C. (2003). Computing observation weights for signal extraction and filtering. Journal of Economic Dynamics and Control, 27, 1317–1333.

Koopman, S.J. and Ooms, M.V.U. (2010). University Amsterdam, Department of Econometrics FEWEB, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands, May 26, 2010. Available at. https://pdfs.semanticscholar.org/3e09/bc22ebac1ba28d6154960b6dc8f38b61c126.pdf. [Accessed on 06.07.2019.]

Lewis, C. (1981). Industrial and business forecasting methods. Butterworths Publishing, London, 40.

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

Rajarathinam, A., Vetriselvi, R. and Balamurugan, D. (2016). Unobserved components model for forecasting wheat production. International Journal of Agricultural Statistical Science. 12 (1), 161-167.

Singh, D.P., Thakur, A.K. and Ram, D.S. (2014). Application of structural time-series model for forecasting gram production in India. American International Journal of Research in Science, Technology, Engineering and Mathematics, 5(1), 60-62.