Global Journal of Modeling and Intelligent Computing (GJMIC)

Vol. 1, No. 2, July, 2021 **ISSN: 2767-1917**

UNOBSERVED COMPONENTS MODEL FOR TURMERIC PRODUCTIVITY FORECASTING

Ekta Hooda, CCS HAU, India (ektahooda@gmail.com)

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 parameter coefficients of the time series are assumed to be constant. In reality, this assumption is rarely met. The Unobserved Components Model (UCM) is a promising alternative to ARIMA in overcoming this problem as it does not assume the parameter constancy assumption. These models are time varying parameters models as they allow for known changes in the structure of the system over time and provide a flexible class of dynamic and structural time series models. Also, 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 investigation was carried out to study the trend of turmeric in India during the period from 1950-51 to 2019-20 using the Unobserved Component Model (UCM). Turmeric is an important medicinal crop and India stands at a very important place in the world turmeric production. In order to study and evaluate the trend in India's turmeric productivity over the years, time series data was used beginning from the year 1950-51. The UCM model, with slope variance zero, was found to be an appropriate model for studying the trends in turmeric. An overall increase in trend has been observed in the annual productivity of turmeric. Also, future predictions for the next five years have been made based on the selected model.

Keywords: Turmeric; Time Series Forecasting; Forecasting; Productivity; India.

1. INTRODUCTION

Turmeric "*Haridra*" is known as "Indian Saffron" and is an important commercial spice crop grown in India. Indian turmeric is well known to the world since ancient times and India being the largest producer, consumer, and exporter of turmeric in the world is also known as the "Spice Bowl of the World" due to its wide variety of premium quality spices. Turmeric has been grown throughout the country in 18 states. Among the commodities that were traded during Vedic period, spices occupied a major portion due to their superior quality and diversity which attracted foreigners to India. Turmeric – the Golden Spice – is widely cultivated in different countries such as India, China, Myanmar, Nigeria, Bangladesh, Sri Lanka, Taiwan, Burma, Indonesia, etc. Indian turmeric fetches a premium price due to its superior quality in the international market, and occupies around 60 per cent of the world trade in turmeric. Turmeric is well known for its medicinal properties and is extensively used in Ayurveda, Unani, and Siddha medicine as a home remedy for various diseases. Turmeric, derived from the rhizomes of Curcuma longa, (family-Zingiberaceae) is a perennial plant having a short stem with large oblong leaves and bears ovate, pyriform, or oblong rhizomes, which are often branched and brownish-yellow in colour. Turmeric is considered auspicious and is a part of various religious rituals. In recent times, traditional Indian medicine uses turmeric powder for the treatment of biliary disorders, anorexia, coryza, cough, diabetes, wounds, hepatic disorders, rheumatism, and sinusitis, etc.

Production of agricultural crops has always been marked by wide variations due to its dependency on many stochastic factors such as weather and insect/pest diseases responsible for determining the crop yield. In practice, it is impossible to completely control such fluctuations. Therefore, reliable estimates of forthcoming crop production have considerable significance for those who have financial interest in agriculture production. In addition, growth of agricultural trade developments in transport facilities caused farmers to become more business minded and to cease considering themselves as subsistence units. At the same time, it was realized by the governments that crop forecast are of national value as the likely supply situation for agricultural products in other countries is of great importance to all countries converting in a common world market. Dependable yield forecasting within the growing season would facilitate superior planning and more efficient management of crop production, handling and marketing.

Availability of accurate pre-harvest estimates of the benefits and risks of alternative crop management systems with knowledge of expected yield has also placed an increasing demand on crop simulation models.

India has a well-established system for collecting agricultural statistics and forecasting of crop production is one of the most important aspects of agricultural statistics system. Advance estimates of major cereal and commercial crops are issued by the central government of India through the Ministry of Agriculture, Cooperation and Farmers' Welfare. However, the final estimates are given a few months after the actual harvest of the crop. Thus, one of the limitations of the conventional methods is timeliness and quality of the estimates. Hence, there is always a considerable scope of improvement in the conventional system of estimation.

Timely and effective pre-harvest forecast of crop yield helps in advance planning, formulation and implementation of policies related to the crop procurement, price structure, distribution and import-export decisions etc. These forecasts are also useful to farmers to decide in advance their future prospects and course of action.

Indian agriculture has witnessed many policies and technological changes in the last three decades. In this context, an attempt is made in this study to examine the growth and instability in turmeric productivity in terms of area, production, and productivity in India as a country and predict its nearby future. For this purpose, time series models have advantages in many situations and can be used more easily because historical sequences on crop yield data are readily available. Dependable productivity forecasting within the growing season would facilitate superior planning and more efficient management of crop production, handling and marketing. Keeping in view the importance of the subject matter, the work has been carried out for turmeric crop in India to study its trend in India during the period from 1950-51 to 2019-20 using the Unobserved Component Model (UCM). In order to study and evaluate the trend in India's turmeric productivity over the years, time series data was used beginning from the year 1950-51. Also, forecasts for the next five years have also been made in the study using the same models after their validation testing and satisfactory results.

To achieve the targeted objectives, review of work done in India and abroad on related problems of productivity estimation and Unobserved Components Models is given in the section 2 with the title Literature Review. The third section describes the study region, data used and research methodology of the various forms of Unobserved Components Models used for the present study along with the description and discussion of the results of the investigation. The conclusions and policy implications of the study are mentioned in section 4.

2. LITERATURE REVIEW

ARIMA approach for time series forecasting dominated the statistical literature during 1970 to 2000. Recent developments in time series modelling offer further scope for extension of ARIMA models to the multivariate framework. These days, state space models are widely used in time series analysis to deal with processes that gradually change over time. State space models have their origin in engineering applications, beginning with the path-breaking paper of Kalman (1960). A good exposition of the state space approach to time series forecasting can be found in books by Kitagawa et al. (1984), Aoki (1987) and Durbin et al. (2000) etc. Unobserved Component Modeling is a promising alternative approach to model time series data as given by Harvey (1984). It is a flexible class of structural time-series models and decomposes a given time series into latent components such as trend, cyclical, seasonality, linear, and non-linear regression effects. The main feature of UCM is the latent components, which follows suitable stochastic models and provides a suitable set of patterns to capture the outstanding actions of the response series. UCM assumes that the latent components are stochastically independent of each other and allows for the inclusion of explanatory variables. All the component models in UCM can be thought of as a stochastic generalization of the corresponding deterministic time series patterns. Apart from the forecast, structural time series models give estimates of these unobserved components. In many time series, the adjacent observations are more closely correlated with each other than observations those are far apart. The UCMs are local in nature and give higher weights to the recent observations than observations in the distant past. These models tend to predict better than models that treat time-series data globally as in the deterministic time trend model.

Shumway (1982) gave a general method by combining Kalman filter and Expectation-Maximization (EM) algorithms to forecast time series using state space models by incorporating missing observations. Bayesian forecasting and state space formulation given by Harrison et al (1976) are synonymously used in time series analysis. Following simulation and empirical approaches, Fildes (1983) compared several forecasting methods from time series including the Bayesian. The modelling and estimation approach of Harrison and Stevens was further extended by West et al. (1997). Hooda et al. (1998) studied rainfall distribution for Nauni in Solan district of India in order to help in crop planning. Maravall (1994) discussed the use and misuse of unobserved components in economic time series forecasting within a general model-based approach. The author also derived analytical expressions for different types of associated errors and presented two applications on how the use of unobserved components can significantly increase forecasting precision.

Tasdemir, 2008 applied state space models to get reliable estimates and forecast the gross domestic product of Turkey during the period 1989-2007. Adubisi et al. (2019) developed multi-interval input intervention unobserved component model (MIII-UCM) to model the crude oil production series to evaluate the effects of the military crackdown, presidential amnesty program, and Niger Delta avenger insurgency, respectively in Nigeria. Hooda et al. (2019) studied the trend of sugarcane (gur) in five districts of Haryana by making use of UCM with level, trend and irregular components and concluded that the models performed very well for the forecasting purpose of sugarcane yield for most of the districts. Sariola (2019) used UCM for estimating an output model for Finland which builds a production function approach and uses several resource utilization indicators, such as capacity utilization and long-term unemployment. Hooda et al. (2020) studied state space models with weather as exogenous input in order to forecast sugarcane yield for eastern Haryana.

Karthik et al. (2014) analysed the economic status of turmeric production in Tamil Nadu state in India using the Cobb-Douglas production function and revealed that planting material, nitrogen, potash, harvesting and curing, machine hours, and irrigation contributed significant influence on the productivity of turmeric. Malik et. al. (2004) conducted the economic analysis for production and export of onion in India and also described the medicinal uses of onion for various health ailments. Angles et al. (2011) examined the production and export performance of turmeric in India using secondary data for the period from 1974-75 to 2007-08 via the exponential form of growth function. They reported the growth in production and export of turmeric to be significant, because of the high demand coupled with inflation. For the assessment of the direction of trade, they used Markov chain models and concluded that India is likely to loose its export markets in some of the countries like USA and Japan and hence, suggested appropriate steps and policies have to be evolved to maintain the market share of Indian turmeric. Angles et al. (2005) used Hazell’s decomposition model to study the instability in turmeric production were examined in terms of area and productivity in important states of South India viz. Andhra Pradesh, Tamil Nadu, Karnataka, and Kerala. Krup et al. (2013) studied the pharmacological activities of turmeric (Curcuma longa Linn), its extracts, and plausible medicinal applications of turmeric along with their safety evaluation.

Keeping in view the above points, UCMs have been developed to fit the trend in turmeric productivity of India assuming the level and trend components to be locally linear as well as when level and trend components remain constant without any persistent upward or downward drift.

3. RESEARCH METHDOLOGY

3.1 Data Description

India as a country comprises 29 states and is situated between latitudes 8° 4` N and 37° 6` N and longitudes 68° 7`E to 97° 25` E. The Tropic of Cancer (23° 30` N) divides the country into almost two equal parts. The time series data on turmeric productivity from 1950-51 to 2019-20 for India compiled from www.indiastat.com have been used for the present study. The data for the last five years i.e., 2015-16 to 2019-20 have been used to check the validity of the developed models for turmeric productivity predictions. The PROC UCM procedure available in SAS has been used for data analysis.

3.2 Unobserved Component Model

The unobserved component model can be considered as a multiple regression model with time-varying coefficients. It is based on the principles that a time series can be decomposed into trend, seasonal and cyclic components and that in many time series, the adjacent observations are more closely correlated with each other than observations that are far apart.
The UCM consists of trend, cycle, seasonal and irregular components and is expressed as

$$y_t = \mu_t + s_t + c_t + \varepsilon_t \tag{1}$$

where μ_t denotes the stochastic trend in the time series y_t at time t, s_t the stochastic seasonal effect at time t and c_t the cyclical effect at time t. Here, ε_t is the overall error or irregular component at time t, which is assumed to be Gaussian white noise with variance σ^2_ε. In case of annual time series, the seasonal and cyclical effects cannot be identified and the UCM also called the Local Linear Trend Model (LLTM) is formulated as:

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, & \varepsilon_t &\sim NID(0, \sigma^2_\varepsilon) \\ \mu_{t+1} &= \mu_t + v_t + \xi_t, & \xi_t &\sim NID(0, \sigma^2_\xi) \\ v_{t+1} &= v_t + \eta_t, & \eta_t &\sim NID(0, \sigma^2_\eta) \end{aligned} \quad (2)$$

for t = 1,2,….,n. This model contains two state equations, one each for modeling the level, and the slope. The stochastic slope v_t in equation (2) is equivalent to regression coefficient in a classical regression model and μ_t is the unobserved level at time t which is equivalent to the intercept in the classical regression model, ε_t is the observation disturbance at time *t*, ξ_t and η_t are the level and slope disturbances, respectively. For the LLTM, the slope also determines the angle of the line with the time axis. The important difference is that the regression coefficient is fixed in the classical regression model, whereas, the model in equation (2) allows the level and slope both to vary over time. In LLTM, the slope is also referred to as drift. In state space models, the unknown parameters (or hyperparameters) include the observation and the state disturbance variances, i.e. σ^2_ε, σ^2_ξ and σ^2_η.

If $\sigma^2_\eta = 0$, the model in (2) have stochastic level and deterministic slope and is known as Local Linear Model (LLM) or the Random Walk Model. This model can be written as

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, & \varepsilon_t &\sim NID(0, \sigma^2_\varepsilon) \\ \mu_{t+1} &= \mu_t + v_1 + \xi_t, & \xi_t &\sim NID(0, \sigma^2_\xi) \end{aligned} \quad (3)$$

If both of the state disturbance variances σ^2_ξ and σ^2_η are zero then model given in equation (2) reduces to the classical repression model. In this case the Linear Trend Models simplifies to

$$y_t = \mu_1 + v_1 g_t + \varepsilon_t, \qquad \varepsilon_t \sim NID(0, \sigma^2_\varepsilon) \quad (4)$$

for t = 1,2, ……, n, where, the predictor variable g_t = t-1 for t = 1,2,….,n is time effective and μ_1 and v_1 are the initial values of the level and slope (Commandeur and Koopman, 2007).

3.3 Model Selection Criteria

The following criteria have been used for comparing the performance of LLM and LLTM models developed for turmeric productivity in India:

1. Root Mean Square Error (RMSE)

   RMSE = $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2}$

2. Mean Absolute Prediction Error

   MAPE = $\frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y_i}|$

3. Relative Deviation (%)

   RD (%) = $\left(\frac{Observed - Forecast}{Observed}\right) x\ 100$

4. Akaike Information Criteria (AIC)

For state space models, the AIC takes the form

$$AIC = \frac{1}{n}[-2n\log(L_d) + 2(q + w)] \quad (5)$$

Where y_i is the actual or observed value and $\hat{y_i}$ is predicted/forecast value, n is the number of observations in the time series, log (L_d) is maximized diffuse log-likelihood function, q is the diffuse initial values in the state and w is the total number of error variances estimated in the analysis.

The time series plot of annual turmeric productivity showed a positive trend (Fig. 1) after additive outlier correction for the year 1993-94.

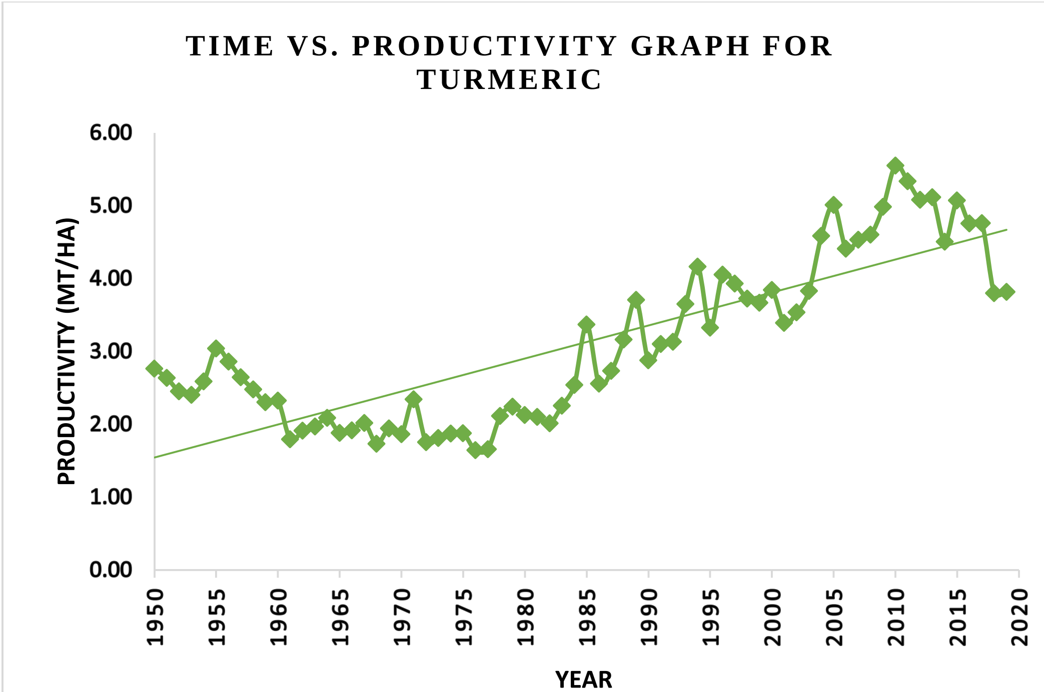

Figure 1: Time vs. Productivity Graph for Turmeric in India

Initially, all possible components viz., level, trend, and irregular were estimated and tested using the UCM or Local Linear Trend model given in Eq. (2). At the first stage, the analysis was aimed to identify the existing component in the model by UCM technique. Error variances of the irregular, level, and slope components were considered as free parameters of the model and their estimates are shown in 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_1: Corresponding component is stochastic

Table 1: Final Estimates of the Free Parameters for Turmeric Productivity

Component	Parameter	Estimate	Approx. Std. Error	t-value	Pr> |t|
Irregular	Error Variance	0.073	0.031	2.29	0.022
Level	Error Variance	0.047	0.038	1.23	0.217
Slope	Error Variance	0.00014	0.0003	0.43	0.663

According to the LLT model in 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.

However, whether the 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. Besides, 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_1: Given component is significant

Table 2: Significance Analysis of Components (based on the final state) of the Productivity of Turmeric			
Component	**DF**	**Chi-Square**	**Pr > ChiSq**
Irregular	1	2.73	0.0987
Level	1	566.09	<.0001
Slope	1	0.97	0.3237

Table 3: Trend Information (based on the final state) for the Turmeric Productivity		
Component	**India**	
	Estimate	**Std. Error**
Level	4.76	0.20
Slope	0.03	0.03

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 stochastic. The contribution of the irregular component is also not significant, but since it is a stochastic component, it cannot be dropped from the model. Thus, fixing the slope variance at zero, the free parameters were again obtained as shown in Table 3.

Table 4: Goodness of Fit Criterion values for LLTM and LLM based on Residuals and Likelihood		
Criterion	**LLT Model**	**LL Model**
MSE	0.17	0.17
RMSE	0.41	0.41
MAPE	9.46	9.30
AIC	76.12	74.61

Table 5: Post Sample Prediction Performance of UCMs for Turmeric Productivity				
Year	**Actual**	**Forecast**	**Prediction Error**	**Relative Deviation (%)**
2015-16	5.07	4.61	0.46	9.07
2016-17	4.76	4.63	0.13	2.66
2017-18	4.76	4.65	0.11	2.32
2018-19	3.80	4.68	-0.88	-23.21
2019-20	3.82	4.71	-0.89	-23.39
Avg. Abs. Percent Deviation 12.13				

Table 6: Forecast Figures for Turmeric Productivity				
Year	**Estimated Turmeric Productivity**	**Std. Error**	**95% Confidence Limits (lower)**	**95% Confidence Limits (upper)**
2020-21	4.96	0.67	3.64	6.28
2021-22	4.99	0.73	3.56	6.42
2022-23	5.02	0.78	3.49	6.56
2023-24	5.06	0.83	3.43	6.69
2024-25	5.09	0.88	3.36	6.82

Fit statistics based on residuals and likelihood are presented in Table 4. When considering the model with all components, i.e., irregular, slope and level (LLT Model), the AIC value came out to be 76.12. When the slope component was taken as constant because of its variance being approximately zero, then for the modified form of model (Random Walk Model/LLM), the AIC value was found to be 74.61. It indicated a relatively better fit model. Hence, the Random Walk Model (LLM) having only slope as constant component is found to be the best fit model for turmeric productivity forecasting in India. After fixing the slope component, the MSE, RMSE, and MAPE value obtained are also shown in Table 4. The post-sample prediction(s) using the best fit model have been given in Table 5 along with the prediction error(s) for the period 2015-16 to 2019-20. The MSE and RMSE values for LLT and LL Model came out to be the same. MAPE value for LLT model was 9.46 while for the LL Model (Random Walk Model), it came out to be 9.30 indicating that the LLM (Random Walk Model) has relatively good post-sample forecast performance. The forecast figures alongside confidence intervals for the next five years, i.e., 2020-21 to

2024-25 of annual turmeric productivity for India has been given in Table 6 along with their respective standard errors. Also, graphs in Figure 2 and 3 indicate that normality tests are also under acceptable limits.

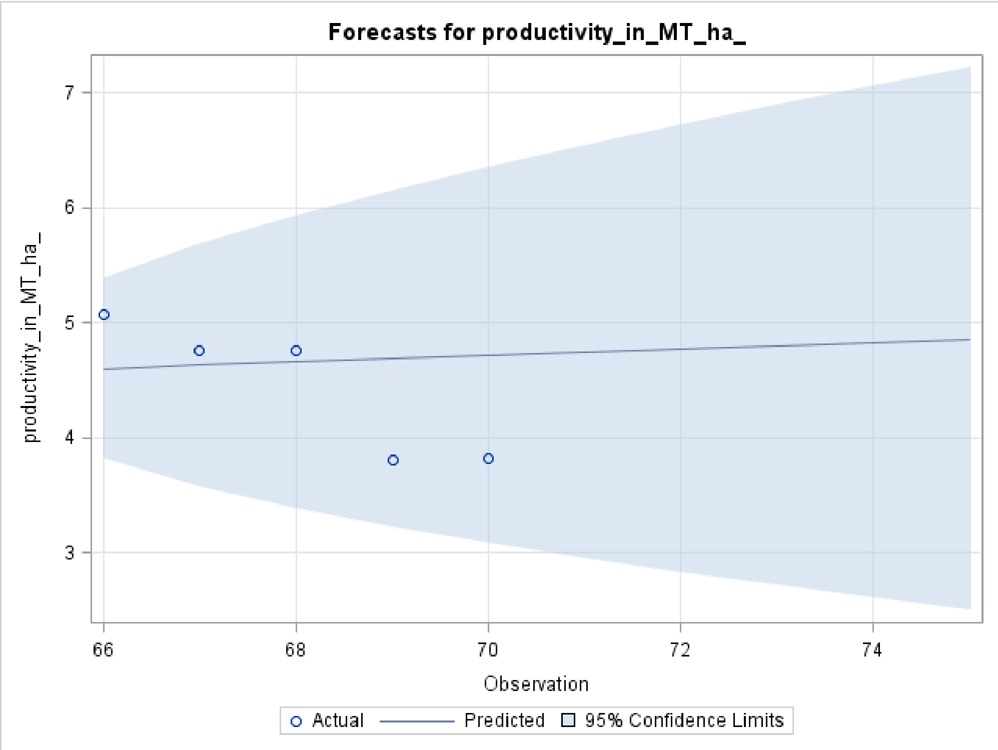

Figure 2: Normality Check for Residuals for Turmeric Productivity Forecasts

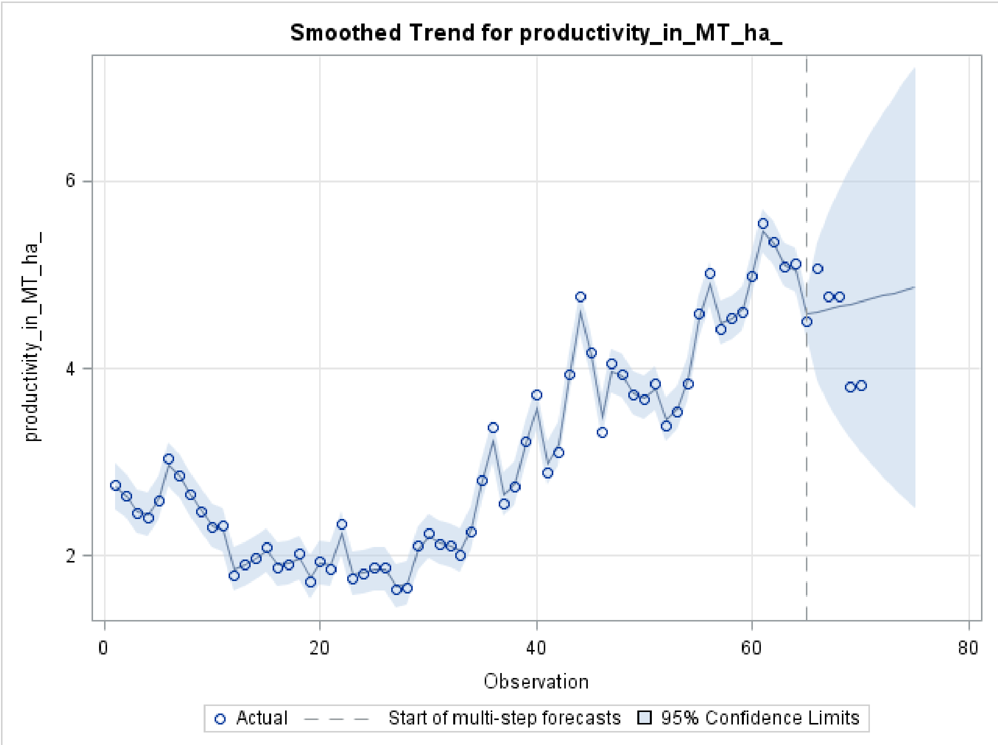

Figure 3: Graph of Actual and Forecasted Turmeric Productivity along with Confidence Interval

4. CONCLUSION

The LL Model was found out to be better than LLTM for predicting turmeric productivity of India. The UCM performed well in capturing tolerable per cent relative deviations for annual turmeric productivity forecasts in all time regimes. The developed models are capable of providing the reliable estimates of turmeric productivity well in advance of the crop harvest while on the other hand, the real-time yield estimates from Department of Agriculture and Farmers' Welfare are obtained quite late after the actual harvest of the crop.

4.1 Policy Implications

The developed models are capable of providing reliable estimates of turmeric productivity well in advance while the real-time productivity estimates issued by the central government of India are obtained quite late. Timely and effective pre-harvest forecasts of crop productivity help in advance planning, formulation, and implementation of policies related to the crop procurement, price structure, distribution and import-export decisions etc. These forecasts are also useful to farmers to decide in advance their future prospects and course of action.

REFERENCES

Adubisi, O.D., Ezenweke, C. and Adubisi, C.E. (2019). Unobserved Component Model with Multi-interval Input Interventions: An Application to Crude Oil Production, *Scientific African*, 7: e00297.

Angles, S. and Hosamani, S.B. (2005). Instability in Area, Production and Productivity of Turmeric in select South Indian States, *Madras Agricultural Journal*, 92(4-6): 271-278.

Angles, S., Sundar, A. and Chinnadurai, M. (2011). Impact of Globalization on Production and Export of Turmeric in India – An Economic Analysis, *Agricultural Economics Research Review*, 24(2), 301-308.

Aoki, M. (1987). *State Space Modelling of Time Series*, Springer, Berlin.

Commandeur, J.J.F. and Koopman, S. (2007). An Introduction to State Space Time Series Analysis, *Oxford University Press, Oxford, USA.*

Durbin, J. and Koopman, S.J. (2000). Time Series Analysis of Non-Gaussian Observations based on State Space Models from both Classical and Bayesian Perspectives (with discussion), *Journal of Royal Statistical Society. Series B*, 62, 3-56.

Fildes (1983). An Evaluation of Bayesian Forecasting: Abstract, *Journal of Forecasting*, 2(2), 137.

Harrison, P.J. and Stevens, C.F. (1976). Bayesian Forecasting, *Journal of the Royal Statistical Society. Series B*, 38(3), 205-247.

Harvey, A.C. (1984). A Unified View of Statistical Forecasting Procedures, *Journal of Forecasting*, 3, 245-83.

Hooda, B.K. and Thakur, B.C. (1998). Probability Analysis of Rainfall at Nauni, Himachal Pradesh, *Indian Journal of Soil Conservation*, 26(2), 153-155.

Hooda, E. and Verma, U. (2019). Unobserved Components Model for Forecasting Sugarcane Yield in Haryana. *Journal of Applied and Natural Science*, *11*(3), 661-665. https://doi.org/10.31018/jans.v11i3.2144

Hooda, E., Verma U. and Hooda, B.K. (2020). State space models for sugarcane yield in Eastern districts of Haryana using Weather parameters as Exogenous input. *International Journal of Agricultural and Statistical Sciences*. DocID: https://connectjournals.com/03899.2020.16.911

Kalman, R.E. (1960). A New Approach to Linear Filtering and Prediction Problems, *Journal of Basic Engineering, Transactions ASMA, Series D*, 82, 35-45.

Karthik, V and Amarnath, J.S. (2014). An Economic Analysis of Turmeric Production in Tamil Nadu, India. *Direct Research Journal of Agriculture and Food Science*; 2(6):66-76. http://directresearchpublisher.org /drjafs.

Kitagawa, G. and Gersch, W. (1984). A Smoothness Priors-state Space Modelling of Time Series with Trend and Seasonality, *Journal of American Statistical Association*, 79, 378-389.

Krup, V., Prakash, L.H. and Harini, A. (2013). Pharmacological Activities of Turmeric (Curcuma longa Linn): A Review, *Journal of Homeopathy and Ayurvedic Medicine*, 2:133. doi:10.4172/2167-1206.1000133.

Malik, D.P., Kumar, Sanjay and Hooda, B.K. (2004). An Economic Analysis of Production and Export of Onion in India, *Agricultural Marketing*, 47(2), 12-19.

Maravall, A. (1994). Use and Misuse of Unobserved Components in Economic Forecasting, *Journal of Forecasting*, 13, 157-177.

Sariola, M. (2019). An Unobserved Components Model for Finland: Estimates of Potential Output and NAWRU, *BOF Economics Review*, 2, 326-334.

Shumway, R.H. and Stoffer, D.S. (1982). An Approach to Time Series Smoothing and Forecasting using the EM Algorithm, *Journal of Time Series Analysis*, 3, 253-264.

Tasdemir, M. (2008). Estimating Monthly GDP for Turkey by State Space Approach, *International Research Journal of Finance and Economics*, 17, 144-151.

West, M. and Harrison, J. (1997) *Bayesian Forecasting and Dynamic Models* (*2nd ed.)*, Springer-Verlag, New York, USA.