International Journal of Engineering, Science and Mathematics
Vol. 6 Issue 7, November 2017,
ISSN: 2320-0294 Impact Factor: 6.238
Journal Homepage: **http://www.ijesm.co.in, Email: ijesmj@gmail.com** Double-Blind Peer
Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

Forecasting of Oilseeds production: Applications of UCM Model using SAS

Ashok Kumar Mishra
Assistant Professor, Dept. of Mathematics,
Shri.Ragvindra Singh Hajari, Govt. College, Hatta, Damho (Madhya Pradesh)

Abstract

In India, annual production of oilseeds is around 27.51 million tons and it grown in around 25.59 million hectares of area (Agricultural Statistics at a Glance 2016). Present study focused on forecasting of oilseeds production in India using Unobserved Component Models. The present paper is to discuss structural time series methodology utilized for modeling time-series data in the presence of trend, seasonal and cyclic and irregular fluctuation has been discussed. Structural time series model are formulated in such a way that their components are stochastic, i.e. they are related as being driven by random disturbances. A number of methods of computing Maximum Likelihood Estimators are considered. These include direct maximization of various times domain likelihood function. Once a model estimated, its suitability can be assessed using goodness of fit statistics and model used to predict for ten leading years. In the study the model developed for oilseeds production, from the forecasting available. Afteranalysis of production of oilseeds *by structural time series model using SAS, oilseeds production forecast for the year 2025 to be near about 35.00 million tones with upper and lower confidence limit 27.21 and 42.80 million tonnes respectively and it shows that there is an increasing trend for production of oilseeds in India.*

Keywords: *Oilseeds, Unobserved Component Model, Forecast and Kalman Filter, goodness of fit, AIC, BIC.*

International Journal of Engineering, Science and Mathematics
Vol. 6 Issue 7, November 2017,
ISSN: 2320-0294 Impact Factor: 6.238
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com Double-Blind Peer
Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

I. INTRODUCTION

This paper analyzes India's oilseeds production.In India, annual production of oilseeds is around 27.51 million tons and it grown in around 25.59 million hectares of area (Agricultural Statistics at a Glance 2016). The main purpose of thisstudy is to highlight the importance of this technique for modeling Oilseeds production data using UCM (Unobserved component Models) available in SAS.

The challenge has grown over time; the original use of time series analysis was primarily as add to forecasting. As such a methodology was developed to decompose a series into a trend, a seasonal, a cyclical and an irregular component. Uncovering the dynamic path of a series improves forecast accuracy each of the predictable components can be extrapolated into the future.

An important problem in time series model analysis concerns analyzing observation that are collected on a numeric response variable at regular time intervals, such as monthly or quarterly. In addition to the response variable, observation on some predictor variables might also be available.

The study can use a variety of statistical techniques for achieving some or all of the preceding goals. From one of them is Autoregressive Integrated moving average (ARIMA). ARIMA time series methodology is used for modeling time series data. This methodology can be applied only when either the series under consideration is stationary. Another disadvantage is that this approach is empirical in nature and does not provide any insight into the underlying mechanism.

The UCMs are also called the Structural Time Series model. The components in the model are supposed to capture the salient features of the series that are useful in explaining and predicting its behavior. Traditionally, the ARIMA models and to some limited extent, the exponential smoothing models have been the main tool in the analysis of this type of time series data. It is fair to say that the UCMs capture the versatility of ARIMA models while possessing the interpretability of the smoothing models(Harvey,1989).

The present paper is to discuss Structural time series methodology utilized for modeling time-series data in the presence of trend, seasonal, cyclic and irregular fluctuations. UCM model are formulated in such a way that their components are stochastic, i.e. they are related as being driven by random disturbances. Forecasts are made by extrapolating these components into the future.Harvey and Todd (1983) compare the forecast made by a basic form of the structural model with the forecast made by ARIMA models and conclusion out by this comparison is in favors of using structural time series models

in practice. Structural models are applicable in the same situations where Box-Jenkins ARIMA models are applicable; however, the structural models tend to be more informative about the underlying stochastic structure of the series. In another paper Harvey (1985) show structural models can be used to model cycle in macroeconomics time series. Other studied included Kitagawa and Gersch (1984). The forecast obtained from particular model depend on certain variance parameter.

II. DATA AND METHODOLOGY

This study mainly focused onbehavior of oilseeds production of India. In the present study secondary data of oilseeds production of 65 years was collected for period 1950-51 to 2014-15 from Agricultural Statistics at a glance- 2016, published by Directorate of Economics and Statistics Ministry of Agricultural, govt. of India. The data was analyzed by using the software Statistical Analysis System (SAS).For forecasting the production of oilseeds, the UCMmodel was used with the help of SAS. The UCM is generally a set upof its various components, like trend, cyclic, and seasonal and irregular variations, i.e.

$$y_t = \mu_t + \psi_t + \nu_t + \varepsilon_t,\ t=1, 2\ldots, T \qquad \ldots..(1)$$

Where $Y_t$is the observed time series at time t, μ_t, ψ_t, $\nu_t$andε_t are the trend, cyclical, seasonaland irregular components.All four components are stochastic. Since oilseeds production data are published on annual basis, so it is not possible to include seasonality component in the model. Further, a graphical representation of the data also indicates the presence of cyclical pattern. On the bases of above discussion, the UCM model is specified.

$$y_t = \mu_t + \psi_t + \varepsilon_t;\ t=1, 2\ldots, T\ldots\ldots(2)$$

The UCM procedure offers two ways to model the trend component μ_t. The first model, called the random walk (RW) model, implies that the trend remains approximately constant throughout the life of the series without any persistent upward or downward drift. The RW model can be described as

$$\mu_t = \mu_{t-1} + \eta_t, \eta_t \sim \text{i.i.d. } N(0, \sigma_\eta^2) \qquad \ldots\ldots..(3)$$

Note that if $\sigma_\eta^2 = 0$, then the RW model becomes μ_t = constant. In the second model the trend is modeled as alocally linear time trend (LLT). In the LLT model the trend is modeled using time varying level $\mu_t$and time varying slope β_t. This model is described by the following evolution equations:

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t,\ \ \eta_t \sim \text{i.i.d. } N(0, \sigma_\eta^2) \qquad \ldots\ldots..(4)$$

$$\beta_t = \beta_{t-1} + \xi_t,\ \ \xi_t \sim \text{i.i.d. } N(0, \sigma_\xi^2) \qquad \ldots\ldots..(5)$$

International Journal of Engineering, Science and Mathematics
Vol. 6 Issue 7, November 2017,
ISSN: 2320-0294 Impact Factor: 6.238
Journal Homepage: **http://www.ijesm.co.in, Email: ijesmj@gmail.com** Double-Blind Peer
Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

If you set σ_{ξ}^2 equal to zero, then the resultingmodel has linear trend with fixed slope. If you set σ_{η}^2 to zero, then the resulting model usually has a smoother trend. Ifyou set both the variances to zero, then the resulting model is the deterministic linear time trend:

$$\mu_t = \mu_0 + \beta_0 t. \quad \ldots\ldots\ldots(6)$$

Similar to the trend, the UCM procedure offers a stochastic cycle component that can capture cyclical patterns oftime-varying amplitude and phase. Note that a deterministic cycle $\psi_t$with frequency λ, $0 < \lambda < \pi$, can be written as

$$\psi_t = \alpha\cos(\lambda t) + \beta \sin(\lambda t) \quad \ldots\ldots\ldots(7)$$

Where ψ_t is a periodic function with period $2\pi/\lambda$, amplitude $(\alpha^2 + \beta^2)^{1/2}$, and phase $\tan^{-1}(\beta/\alpha)$. In time series situations it isuseful to generalize this simple cyclical pattern to a stochastic cycle that has a fixed period but time-varying amplitudeand phase. The stochastic cycle considered here is motivated by the following recursive formula for computing ψ_t:

$$\begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} = \begin{bmatrix} \cos\lambda & \sin\lambda \\ -\sin\lambda & \cos\lambda \end{bmatrix} \begin{bmatrix} \psi_{t-1} \\ \psi_{t-1}^* \end{bmatrix} \quad \ldots\ldots\ldots(8)$$

Starting with $\psi_0 = \alpha$ and $\psi_t^{*2} = \beta$. Note that ψ_0 and ψ_t^* satisfy the relation

$$\psi_T^2 + \psi_t^{*2} = \alpha^2 + \beta^2 \text{ For all t}$$

You can obtain a stochastic generalization of the cycle $\psi_t$by adding random noise to this recursion and by introducinga damping factor ρ for additional modeling flexibility. This model can be described as follows:

$$\begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} = \rho \begin{bmatrix} \cos\lambda & \sin\lambda \\ -\sin\lambda & \cos\lambda \end{bmatrix} \begin{bmatrix} \psi_{t-1} \\ \psi_{t-1}^* \end{bmatrix} + \begin{bmatrix} \upsilon_t \\ \upsilon_t^* \end{bmatrix} \quad \ldots\ldots\ldots(9)$$

Where $0 \le \rho \le 1$ and the disturbance and υ_t^* ~i.d. $N(0, \sigma_{\upsilon}^2)$ variables. The resulting stochastic cycle hasa fixed period but time-varying amplitude and phase. This cycle component depends on three parameters: the cyclefrequency (λ), the damping factor (ρ), and the variance (σ_{υ}^2)of the disturbance term υ_t. In this study the error term is taken to be Gaussian white noise.

Goodness of fit statistics is used for assessing over all models fit. Basic measure of goodness of fit in time series model is prediction error variance. Comparison of fit between different models is based on Akaike information criterion (AIC).

$$AIC = -2 \log L + 2k, \qquad \ldots\ldots\ldots(10)$$

Where Lis the likelihood function and k is the number of parameters estimated from the model. Schwartz-Bayesian informationcriterion (BIC) is also used as a measure of goodness of fit which is given as

$$BIC = -2\log L + k * \log (n), \qquad \ldots\ldots\ldots(11)$$

Where n is total number of observations. Lower the value of these statistics better is the fitted model.Basically AIC is an estimate of a constant plus the relative distance between the unknown true likelihood function of the data and the fitted likelihood function of the model, so that a lower AIC means a model is considered to be closer to the truth. BIC is an estimate of a function of the posterior probability of a model being true, under a certain Bayesian setup, so that a lower BIC means that a model is considered to be more likely to be the true model.BIC is sometimes preferred over AIC because BIC is consistent. AIC is not consistent because it has a non-vanishing chance of choosing an unnecessarily complex model as n becomes large.

III RESULT AND DISCUSSION

To judge the performance of the UCM model, we check the information criteria like Akaike information criteria (AIC), Schwartz-Bayesian information criteria (BIC) etc. In general, it is better to use all data series to develop a final best model, but one can use a silted data set to check new situation in different sets. In the present study the AIC and BIC value on the bases of analysis for oilseeds production is 29.909 and31.195respectively during the year of 2016. This inside sample verification allow model to predict forecasting. So on the basis value of AIC it can prove that our model is best fit. Therefore we have used this model foranalysis of productionof oilseedsby using software SAS. And after analysis it is observed that oilseeds production forecast for the year 2025 to be near about 35.00 million tonnes with lower and upper confidence limit 27.21 and 42.80 million tonnes respectively and it shows that this is an increasing trend for production of oilseeds in India, TheUCM shows an increasing trend for production of oilseeds in India. The reliability of these forecasts can be check when the actual data willbe available for the lead years. Forecasts for the future years from 2016 to 2025 by using the UCM model presented in the following table-1.

Table 1: Forecasts for Production of Oilseeds in India

Forecasts for Variable Production ('000' MT)				
Year	**Forecast**	**Standard Error**	**95% Confidence Limits**	
2016	29.639254	2.26383	25.202224	34.076284
2017	29.595972	2.46063	24.773232	34.418712
2018	29.699357	2.67471	24.457017	34.941697
2019	29.992149	2.89330	24.321390	35.662907
2020	30.491405	3.10613	24.403508	36.579303
2021	31.187323	3.30665	24.706407	37.668240
2022	32.045519	3.49216	25.201005	38.890032
2023	33.012337	3.66332	25.832368	40.192305
2024	34.022404	3.82334	26.528801	41.516007
2025	35.007375	3.97721	27.212196	42.802554

Source: Author's calculation

Forecasting results shows that till 2020 growth in production will at slower rate and after the year 2020 it will accelerate and production of oilseeds will reached at 35 million tonnes in 2025.

Table 2: Final Estimates of the Free Parameters

Component	**Parameter**	**Estimate**	**ApproxStd Error**	**t Value**	**ApproxPr > |t|**
Irregular	**Error Variance**	3.1950	0.71345	4.48	<.0001
Cycle	**Damping Factor**	0.9805	0.03234	30.32	<.0001
Cycle	**Period**	17.396	1.83784	9.47	<.0001
Cycle	**Error Variance**	0.052	0.05129	1.01	0.3106

Source: Author's calculation

Final estimates of all three components of time series are represented by table-2. In this table we can see that estimate of irregular component is 3.20 approximately which is statistically significant at 5% level of significance. Dumping factor of cycle component is 0.98 which is also significant at 5% level of

significance. This table shows that the cyclic period is 17.40 approximately years which is statistically significant at 5% level of significance. Error variance of cyclic component is 0.05201 which is insignificant then we can say that the cyclic model is not appropriate. Table 3 reveals the fit statistics based on residuals. Fit statistics results specially R-square that our model is best fit and it emphasized that whatever we have analyzed by this model will be similar to the actual production.

Table 3: Fit Statistics Based on Residuals

Particulars	Coefficient value
Mean Squared Error	5.16426
Root Mean Squared Error	2.27250
Mean Absolute Percentage Error	11.12184
Maximum Percent Error	31.24735
R-Square	0.92508
Adjusted R-Square	0.91851
Random Walk R-Square	0.30292
Amemiya's Adjusted R-Square	0.90931
Number of non-missing residuals used for computing the fit statistics = 63	

Source: Author's calculation

Description of model best fit present in table 3, which shows that model is best fit. About 91 per cent of variation of the model has been captures in the present model showed by the R^2 value. It is very much required that check how well the first model predicts observations from the second context.If it does fit, there is some assurance of generalisability of the first model to othercontexts. If the model does not fit, however, one cannot tell if the lack of fit isowing to the different contexts of the two data sets, or true "lack of fit" of the firstmodel.In practice, these types of validation can proceed by deriving a model and estimatingits coefficients in one data set, and then using this model to predict the Y variablefrom the second data set. One can then check the residuals, and so on. After the checking all possible options in the present study final model retain.

Table 4: Significance Analysis of Components (Based on the Final State)

Component	DF	Chi-Square	Pr > ChiSq
Irregular	1	7.96	0.0048
Level	1	642.13	<.0001
Slope	1	5.65	0.0174
Cycle	2	4.65	0.0980

Source: Author's calculation

Table-4 shows an ANOVA table in which we can see that level is statistically significant at 1 percent of significant level with one degree of freedom and cyclic component is statistically significant at 10 percentlevel of significant with two degree of freedom. In the anova table slope also significant at 5 per cent level of significant.

Table 5: Trend Information (Based on the Final State)

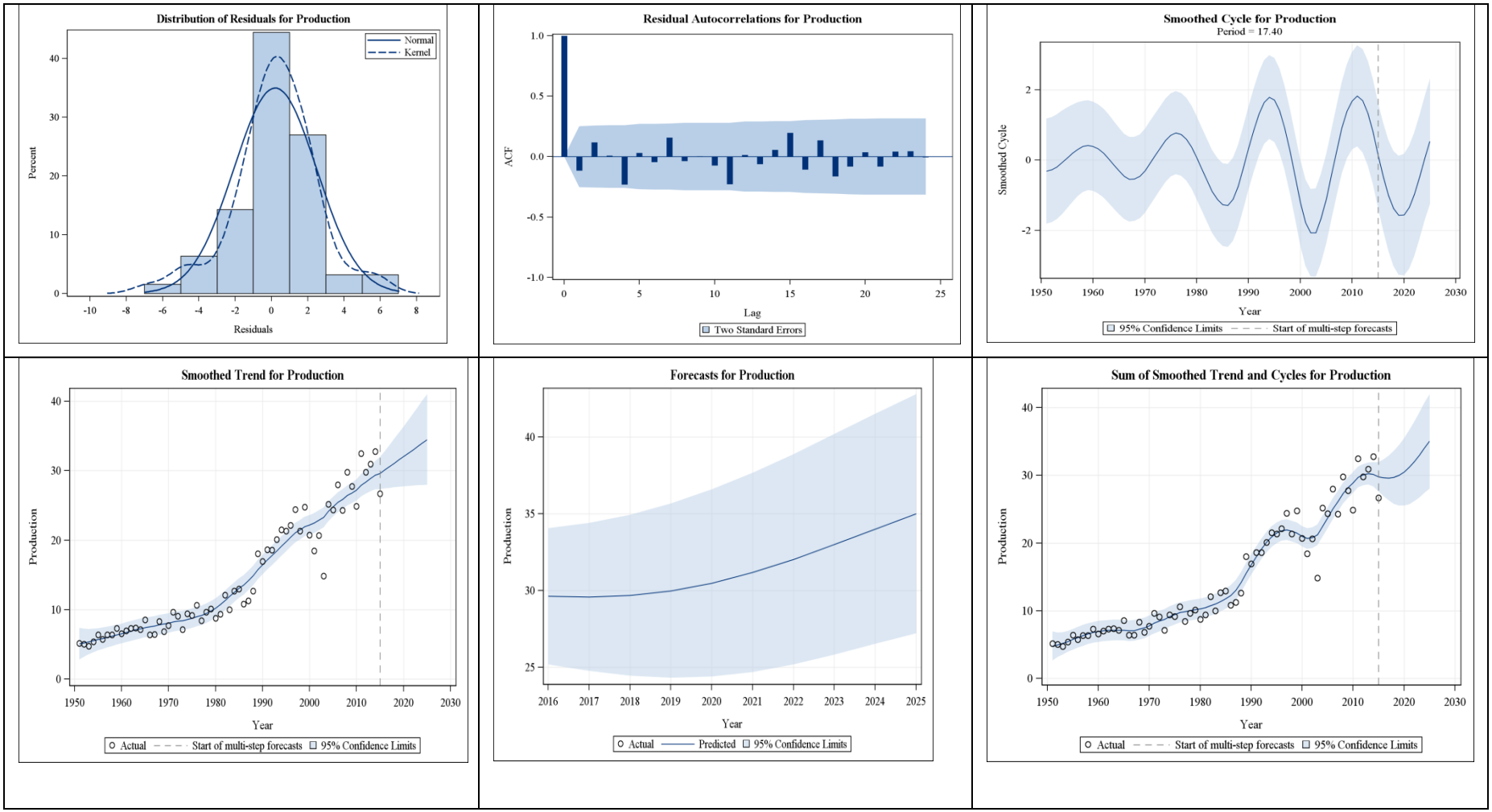
Parameter	Estimate	Standard Error
Level	29.62207186	1.168975
Slope	0.484670241	0.2038331

Source: Author's calculation

Table -5 shows the trend information based on final analysis because both are the trend component which states that the estimate of the level of the trend is 29.62 with standard error 1.16 and slope is 0.48 with standard error 0.20.

International Journal of Engineering, Science and Mathematics
Vol. 6 Issue 7, November 2017,
ISSN: 2320-0294 Impact Factor: 6.238
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com Double-Blind Peer Reviewed Refereed Open Access International Journal -
Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

Figure 1: Graph of distribution, residual, cycle, trend, trend structural break and cyclic structural break respectively



Residual graph in figure 1 is a simple auto correlations function plot (ACF). Shaded area shows the 95% confidence limit. It means that almost all units are within the 95% confidence limit. Graph 2 in the same figure shows that the residual of autocorrelation of production of oilseeds which show that one with the one lag of no autocorrelation exist in the model.In terms of the "Prediction Errors for logpass" graph, the residuals of the model(prediction errors) appear to be varying in an unsystematic way. This is verified in the third graph labeled "Prediction Error Autocorrelations for logpass"where all of the autocorrelations at the various lags are inside the blue-shaded 95%confidence intervals of zero autocorrelation. Graph-3, in figure-1, shows the smoothed cycle for production, according to it, one complete cycle if of 17.40 years approximately. The graph also indicate that the current cycle goes towards recession till 2020, one can think it means production will decreasing trend but it not because it will increasing trend but with the decreasing rate. The Graph-4 also argue the fact that the smoothed trend for production will increase in the future.Graph-5 represents the sum of smoothed trend and regression effects for production.

Therefor all these graphs are support the previous contention that modelAccording to the model oilseeds production of the country will reached at the level of 35 million tonnes till 2025.

IV CONCLUSION

Here UCM model wasused to forecast the production of oilseeds in India. From the forecast table it is clear that, the forecasts of production of oilseeds in India is expect to be slightly increasing over the next ten years with somecyclic fluctuations.According to the result, the estimated period of the cycle is about 17.40 years.The reliability of these forecasts can be check when the actual data will be available for the forecasted years. The model can be used by researchers, scientist and by others for forecastingoilseeds production in India and also it should be updated over the time with available actual data. From the above forecasts for the lead years shows that there is a small change in the forecasts of oilseeds production in India. But growth rate of population also put pressure on demand side. So it is necessary to increase the production of oilseeds in India by adopting the high yielding varieties and improved package of practices. According to the model oilseeds production of the country will reached at the level of 35 million tonnes till 2025.

V REFERENCES

Akaike, H. (1973). Information theory and an extension of the maximum ikelihood principle. In B. N. Petrov& F. Csaki (Eds.), Second international symposium on information theory (p. 267-281). Budapest, Hungary: AkademaiKiado.

Dziak John J. et al. (2012). Sensitivity and Specificity of Information Criteria. *The Methodology Center*, The Pennsylvania State University,Technical Report Series.

Government of India (2016). Agriculture *statistics at a Glance*, Directorate of economics and statistics, New Delhi.

Harvey, A. C. (1989). Forecasting, structural time series models and the Kalman Filter. Cambridge, UK: Cambridge University Press.

Kitagawa, G. and W. Gersch (1996). Smoothness Priors Analysis of Time Series. New York: Springer Verlag.

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

Koopman, S. J. and N. Shephard (1992). Exact score for time series models in state space form. Biometrika 79, 823–6.

Schwarz, G. (1978). Estimating the dimension of a model. Annals of Statistics, 6 , 461-464.