International Journal of Applied Mathematics
& Statistical Sciences (IJAMSS)
ISSN(P): 2319-3972; ISSN(E): 2319-3980
Vol. 5, Issue 4, Jun – Jul 2016; 23-28
© IASET

International Academy of Science, Engineering and Technology
Connecting Researchers; Nurturing Innovations

APPLICATION OF ARIMA MODELLING FOR MUSTARD YIELD PREDICTION IN HARYANA

RAVITA[1] & U. VERMA[2]

[1]PG Scholar, Department of Mathematics, Statistics and Physics CCS Haryana Agricultural University, Hisar, Haryana, India

[2]Professor, Department of Mathematics, Statistics and Physics CCS Haryana Agricultural University, Hisar, Haryana, India

ABSTRACT

The present study deals with application of time-series modelling in agriculture. The mustard yield data of Hisar, Bhiwani, Sirsa, Mahendergarh and Gurgaon districts of Haryana have been considered for this empirical study. Autoregressive integrated moving average (ARIMA) models have been fitted using the district-level time series yield(s) from 1966-67 to 2010-11 and the models have been validated for the post-sample period(s) 2011-12 to 2013-14. After experimenting with different lags of the moving average and autoregressive processes; ARIMA(0,1,1) for Hisar, Bhiwani and Sirsa districts and ARIMA(1,1,0) for Mahendergarh and Gurgaon districts have been fitted for mustard yield(s) estimation in Haryana. A perusal of the results indicates that the percent deviations of the forecast yield(s) from the observed yield(s) are within acceptable limits and favours the use of ARIMA modelling to get short-term forecast estimates.

KEYWORDS: Autocorrelation, Partial Autocorrelation, Differencing, Stationarity, Invertibility

INTRODUCTION

Regression analysis is the most frequently used statistical technique for investigating and modelling the relationship between variables. However, some applications of regression involve regressor and response variables that have a natural sequential order over time and then the need of time series modelling arises for the analysis of such dependence. Time series models have advantages in certain situations. They can be used more easily for forecasting purposes because the historical sequences of observations upon study variables are readily available at equally spaced intervals over discrete point of time. These successive observations are statistically dependent and time series modelling is concerned with the analysis of such dependence. The application of the Box-Jenkins (1976) univariate ARIMA models in the field of agriculture for forecasting a variety of study variables of interest for different crops / regions etc. may be of immense importance.

The importance of agriculture for the Indian society can hardly be over emphasized, as its role in economy, employment, food security, self-reliance and general well-being does not need reiteration. India has a very well established system for collection of crop statistics at village level and aggregating it at different administrative levels. However, the need for early and in-season crop production forecasting has been strongly felt. Fulfilling this requirement entails judicious planning based on information related to various aspects of agriculture. Information on crop acreage and production are important inputs for strategic planning.

Panse (1952, 59, 64) in a series of papers studied the trends in yields of rice and wheat with a view to compare the yield rates during the plan periods with that of the pre-plan periods. Just to cite a few; Balanagammal *et al*. (2000) obtained forecast values by using ARIMA models for the data related to cultivable area, production and productivity of chosen crops in Tamil Nadu, India. Verma and Grover (2006) developed ARIMA models for wheat yield forecasting in all the districts of Haryana state (India). Further, a comparison was shown with remote sensing based yield forecasts and real time yield data as well. Awal and Siddique (2011) applied the ARIMA models to efficiently forecasting Aus, Aman and Boro rice production in Bangladesh.

Rapeseed-mustard is the third important oilseed crop in the world after soybean and palm oil. India is one of the largest rapeseed-mustard growing countries in the world, occupying the first position in area and third position in production after the EU27 and China, and contributing around 11% of the world's total production. Among the seven edible oilseeds cultivated in India, rapeseed-mustard (Brassica spp.) contributes 28.6% in the total production of oilseeds. It is the second most important edible oilseed after groundnut sharing 27.8% in the India's oilseed economy. Indian mustard is predominantly cultivated in Rajasthan, UP, Haryana, Madhya Pradesh and Gujarat and collectively represent 81 per cent of the national acreage and contribute 82.9 per cent to the total rapeseed-mustard production. Rapeseed is grown on nearly 13% of the cropped land in India. It is basically a winter (i.e. the *rabi* season) crop and is grown from September-October to February-March. The crop grows well in areas receiving 25cm to 40cm of rainfall and this is provided by the monsoon rains during the sowing season of the crop in India.

In view the importance of the subject matter, an attempt has been made to develop ARIMA models for mustard yield prediction in Hisar, Bhiwani, Sirsa, Mahendergarh and Gurgaon districts of Haryana.

Data Description and Statistical Methodology

The Haryana state comprising of 21 districts (geographical area: 44212 sq. km) is situated between 74° 25' to 77° 38' E longitude and 27° 40' to 30° 55' N latitude. The time-series data of state Department of Agriculture (DOA) mustard yield spanning 1966-67 to 2013-14 (**Source:** esaharyana.gov.in/State Statistical Abstract/) were collected for ARIMA modeling. The emphasis has been given in predicting the future values on the basis of previous time-series observations. The district-level mustard yield from 1966-67(Hisar, Mahendergarh and Gurgaon)/ 1972-73(Bhiwani)/ 1975-76(Sirsa) to 2010-11 have been used for the training set and the remaining yield data i.e. 2011-12, 2012-13 and 2013-14 have been used for the post-sample validity checking of the developed ARIMA models.

Box-Jenkins Arima Modeling Procedure

Univariate Box-Jenkins(UBJ) ARIMA forecasts are based only on past values of the variable being forecast. The method applies to both discrete as well as to continuous data. However, the data should be available at equally spaced discrete time intervals. Before attempting to choose an appropriate ARIMA model for forecasting, it is necessary to make the data series stationary. One of the simplest transformations called 'differencing' is used when the mean of a series is changing over time and log transformation is used if the variance of a series is changing through time. The estimated autocorrelation function and partial autocorrelation function are very important tools at the identification stage. An estimated autocorrelation function r_k shows the correlation between ordered pairs ($\overline{Y}_t, \overline{Y}_{t+k}$) separated by various time spans ($k=1,2,3,\ldots$). An estimated partial autocorrelation function $\hat{\phi}_{kk}$ shows the correlation between ordered pairs ($\bar{Y}_t, \bar{Y}_{t+k}$) separated by various time spans (k = 1, 2, 3……) with the effect of intervening observations ($\bar{Y}_{t+1}$, $\bar{Y}_{t+2}$, ….. $\bar{Y}_{t+k-1}$) accounted for.

The general functional form of ARIMA model used is :

Autoregressive Integrated Moving Average model i.e. *ARIMA (p,d,q)* $\phi_p(B)\, \Delta^d\, Y_t = c' + \theta_q(B)\, a_t$, where c' = 0 if Y_t is adjusted for its mean

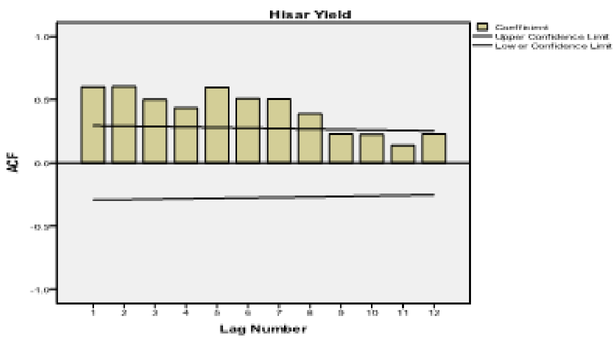
Where Y = Variable under forecasting, B - lag operator, a - error term ($Y - \hat{Y}$, where $\hat{Y}$ is the estimated value of Y), t- time subscript, $\phi_p(B)$ - non-seasonal AR component(s), $(1-B)^d$ - non-seasonal difference, $\theta_q(B)$ - non-seasonal MA component(s); ϕ's and θ's are the parameters need to be estimated

Further, an attempt is made at the estimation stage to obtain the precise estimates of a small number of parameters of the model. Linear least-squares may be used to estimate only pure AR models. All other models require a non-linear least squares (NLS) method. Thirdly, the diagnostic tests are performed to see the random shocks to be independent or not.

RESULTS AND DISCUSSION

The ARIMA methodology has been applied for estimating mustard yield(s) in Hisar, Bhiwani, Sirsa, Mahendergarh and Gurgaon disticts of Haryana. UBJ- identification involved the determination of the appropriate orders of AR and MA polynomials i.e. the values of p and q. Thus, the orders were determined from the autocorrelation functions and partial autocorrelation functions of the stationary series.

The mustard yield(s) data were found to be non stationary for all the districts. Almost all the autocorrelations upto 8/10 lags significantly different from zero shown in Figure 1 confirm non-stationarity. Differencing of order one was sufficient for getting an appropriate stationary series of all the districts however the log transformation was also applied for Hisar district.



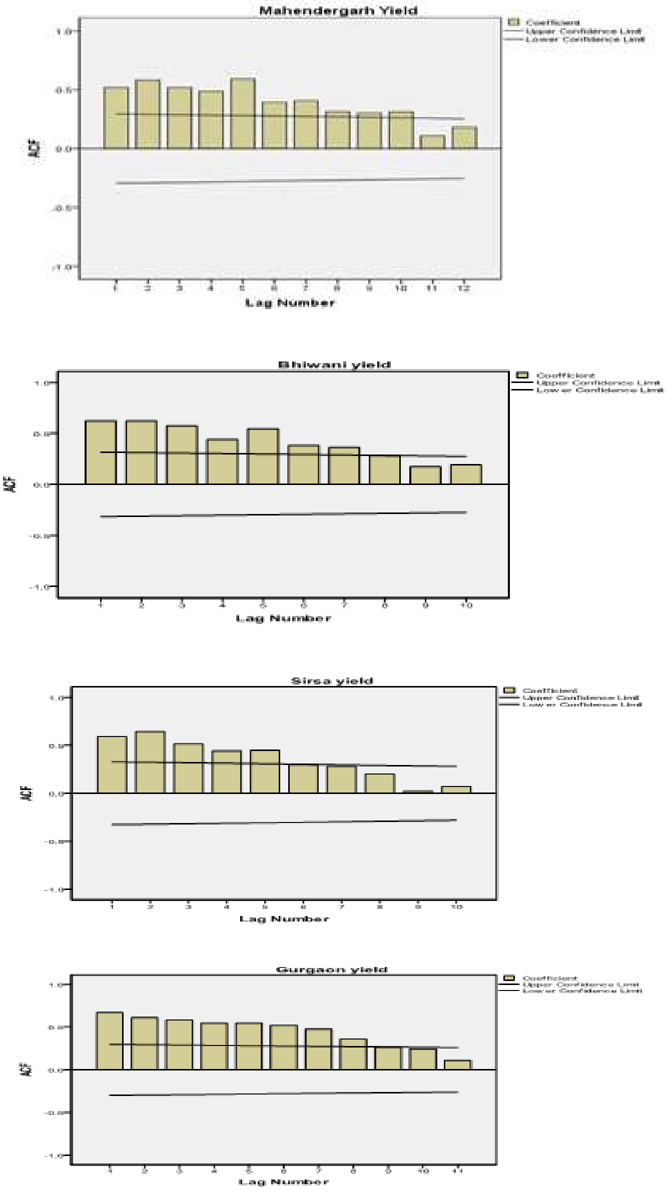


Figure 1: Autocorrelation Coefficients of Mustard Yield for Different Districts of Haryana

The models ARIMA (1,1,1), ARIMA (1,1,0) and ARIMA (0,1,1) were considered at the identification stage and parameter estimation was carried out using a non-linear least squares (NLS) approach. After experimenting with different lags of the moving average and autoregressive processes; ARIMA(0,1,1) for Hisar, Bhiwani and Sirsa districts and ARIMA(1,1,0) for Mahendergarh and Gurgaon districts were fitted for estimating district-level mustard yield(s) in Haryana. Parameter estimates of the selected models shown in Table 1 are less than one (needed for convergence) and also satisfy the stationarity and invertibility conditions under ARIMA structure.

Marquardt algorithm (1963) was used to minimize the sum of squared residuals. Log Likelihood, Akaike's Information Criterion, AIC (1969), Schwarz's Bayesian Criterion, SBC (1978) and residual variance decided to estimate AR and MA coefficients in the models. Approximate 't' values were calculated for residual autocorrelation coefficients using Bartlett's approximation for the standard error of the estimated autocorrelations. The residual acf along with the associated 't' tests and Chi-squared test suggested by Ljung and Box (1978) were used for the checking of random shocks to be white noise. All Chi-Squared statistics in this concern have been shown in Table 2.

Table 1: Results on Stationarity and Invertibility Conditions for AR and MA Coefficients of Fitted Models

Variable /Mustard Yield	Model	Stationarity	Invertibility
Hisar	ARIMA (0, 1, 1)	*	0.87
Bhiwani	ARIMA (0, 1, 1)	*	0.75
Sirsa	ARIMA (0, 1, 1)	*	0.84
Mahendergarh	ARIMA (1, 1, 0)	-0.63	**
Gurgaon	ARIMA (1, 1, 0)	-0.38	**

* Stationarity condition is not applicable since the model is MA model

** Invertibility condition is not applicable since the model is AR model

It is clear from the above table that both the stationarity and invertibility conditions are satisfied because the absolute values of AR and MA coefficients for all the districts are less than one.

Table 2: Diagnostic Checking of Residual Autocorrelations of Mustard Yield for all the Districts

District(S)	Model	Model Fit Statistic			Ljung-Box Q Statistic		
		RMSE	MAPE	SBC	Statistic	Df	Sig.
Hisar	ARIMA (0,1,1)	2.55	22.51	2.05	15.22	17	0.58
Bhiwani	ARIMA (0,1,1)	2.27	19.77	1.84	13.69	17	0.69
Sirsa	ARIMA (0,1,1)	2.26	20.32	1.8	18.71	17	0.35
Mahendergarh	ARIMA (1,1,0)	3.04	28.37	2.39	25.26	17	0.09
Gurgaon	ARIMA (1,1,0)	2.46	25.17	1.99	10.97	17	0.86

Finally, a comparison between ARIMA model based mustard yield estimates with observed yield(s) was seen in terms of percent relative deviation (RD%). The results presented in Table 3 indicate that the deviations of predicted yield(s) from the actual yield(s) are within acceptable limits and thus favouring the use of ARIMA models to get short-term forecast estimates.

Table 3: Comparison of ARIMA Model Based Mustard Yield Estimates with State Department of Agriculture Mustard Yield(s) in Different Districts of Haryana

District(s)	Forecast Years	DOA Yield (q/ha)	Estimated Yield (q/ha)	RD(%)
Hisar	**2011-12**	17.07	16.66	2.40
	2012-13	16.78	17.16	-2.30
	2013-14	16.26	17.66	-8.60
Bhiwani	**2011-12**	14.00	14.21	-1.50
	2012-13	16.40	14.47	11.77
	2013-14	15.16	14.72	2.90
Sirsa	**2011-12**	16.78	16.00	4.65
	2012-13	16.47	16.81	-2.06
	2013-14	17.37	16.77	3.45
Gurgaon	**2011-12**	20.25	18.67	7.80
	2012-13	20.26	20.02	1.18
	2013-14	15.94	18.24	-14.43
Mahendergarh	**2011-12**	18.27	15.86	13.19
	2012-13	16.99	17.36	-2.18
	2013-14	16.99	16.93	0.35

RD(%) = {(Observed yield -Estimated yield)/ Observed yield}*100

DOA = State Department of Agriculture yield estimates

REFERENCES

1. Akaike, H. (1969). Fitting autoregressive models for prediction. *Ann Inst Statist Math*, 21, 243-47.
2. Awal, M. A. and Siddique, M. A. B. (2011). Rice production in Bangladesh employing by ARIMA model. *Bangladesh J. Agril. Res.*, 36 (1), 51-62.
3. Balanagammal, D., Ranganathan, C. R, Sundaresan, R. (2000). Forecasting of agriculture scenario in Tamil Nadu - A time series analysis. *J. Ind. Soc. of Agric. Stat*, 53, 273-286.
4. Box, G.E.P. and Jenkins, G.M. (1976). Time series analysis: Forecasting and control. *Holden Day, San Franscisco.*
5. Ljung, G.M. and Box, G.E.P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65, 297-303.
6. Marquardt, D.W. (1963). An algorithm for least-squares estimation of non-linear parameters. *J Society for Industrial and Applied Mathematics*, 2, 431-41.
7. Panse, V. G. (1952). Trends in areas and yields of principal crops in India. *Agriculture Situation in India*, 7, 144-48.
8. Panse, V. G. (1959). Recent trends in the yield of rice and wheat in India. *Indian Journal of Agricultural Economics*, 14, 11-38.
9. Panse, V. G. (1964). Yield trends of rice and wheat in first two five year plans in India. *Journal of Indian Society of Agricultural Statistics*, 16(1), 1-50.
10. Schwarz, G. (1978). Estimating the dimension of a model. *The annals of Statistics*, 62, 461-64.
11. Verma, U. and Grover, D. (2006). ARIMA wheat yield modelling in Haryana. *Research Bulletin*, *Department of Mathematics and Statistics, CCS HAU, Hisar, Haryana*, 1, 1-51.