International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2023; SP-8(5): 524-527

https://www.mathsjournal.com
Received: 05-06-2023
Accepted: 06-07-2023

Delvadiya JB
Ph.D. Scholar, Department of Agricultural Statistics, N. M. College of Agriculture, Navsari Agricultural University, Navsari, Gujarat, India

Patel UB
Ph.D. Scholar, Department of Agricultural Economics, N. M. College of Agriculture, Navsari Agricultural University, Navsari, Gujarat, India

Meera Padaliya
Ph.D. Scholar, Department of Agricultural Economics, N. M. College of Agriculture, Navsari Agricultural University, Navsari, Gujarat, India

Gohil VM
Ph.D. Scholar, Department of Agricultural Statistics, N. M. College of Agriculture, Navsari Agricultural University, Navsari, Gujarat, India

Corresponding Author:
Delvadiya JB
Ph.D. Scholar, Department of Agricultural Statistics, N. M. College of Agriculture, Navsari Agricultural University, Navsari, Gujarat, India

An application of ARIMA for forecasting rapeseed and mustard area in Gujarat

Delvadiya JB, Patel UB, Meera Padaliya and Gohil VM

DOI: https://doi.org/10.22271/maths.2023.v8.i5Sh.1244

Abstract
Rapeseed and mustard are the second most important oilseed crops in India. Soybean, groundnut and rapeseed and mustard are the major oilseed crops in India contributing around 84% to its total acreage. Forecasting is used to support effective and efficient decision-making and long-term planning. The study was carried out to develop forecasting model of area of rapeseed and mustard crop in Gujarat by using the time series data of 1991-92 to 2019-20 years. The polynomial models were fitted to the original data as well as three-year, four year and five year moving average data while, Autoregressive Integrated Moving Average (ARIMA) models were fitted to the original data on area of rapeseed and mustard crop in Gujarat state. Criteria of evaluation of model was highest R^2, lowest value of RMSE and MAE, significant coefficient of model, lower value of Akaike's Information Criteria (AIC) and Schwartz-Bayesian Criteria (SBC) values, normality test and randomness test of residuals. Quadratic model on original data and ARIMA (0, 1, 3) model were found to be most suitable to explain the pattern of area of rapeseed and mustard crop in Gujarat.

Keywords: Forecasting, time series, polynomial model, ARIMA

1. Introduction

Soybean, groundnut, rapeseed and mustard are the major oilseed crops in India contributing around 84% and 88% to its total acreage and production, respectively (Average of 2014-15 to 2018-19) (Anonymous, 2023a) [3]. India is the 3rd largest producer of rapeseed and mustard after Canada, China. It contributing to around 11% of total production in world. Rapeseed and mustard are the important oilseed crops and one of the 2nd largest oilseed crops in India (Anonymous, 2023b) [4].

Gujarat ranked second in terms of productivity of rapeseed and mustard crop (1976 Kg/ha) in the country (Anonymous 2019-20) [2]. In Gujarat, rapeseed and mustard occupied an area of 1, 72, 618 hectare with production of 3, 33, 525 MT and productivity of 19.32 quintal per hectare during 2019-20. Banaskantha secured highest area of 1, 27, 220 hectare with 73.70% share in total rapeseed and mustard area in Gujarat followed by Patan (Anonymous, 2019-20) [2].

The significance of timely and precise estimates of area, production, and productivity of main crops cannot be overstated for a nation like India, where agricultural production makes up the majority of the economy. A good model most accurately predicts future values based on knowledge of past values. Forecasting is used to support effective and efficient decision-making and long term planning. This is a crucial part of economic development so that proper planning can be made for long-term, sustainable growth. Using historical data to estimate the future with the help of patterns and trends within the data, statistical forecasting models are utilized to create a suitable prediction methodology. A significant variety of univariate time series models, such as linear, quadratic and cubic are accessible in the literature. The Autoregressive Integrated Moving Average (ARIMA) model is the most relevant and popular time series model. ARIMA model is highly efficient in short term forecasting.

2. Data and methodology

To carry out this study, The time series data on area of rapeseed and mustard crop for the period of1991-92 to 2019-20 were obtained from Directorate of Agriculture, Gandhinagar, Gujarat state (Anonymous, 1991-20) [2].

The available past data has been split into testing and validation of model. The dataset has been split into testing set (1991-92 to 2014-15) and validation of model set (2015-16 to 2019-20).

The moving average concept was utilized to average out the fluctuation along with original data. In moving average technique, the trend of the observed data is expected to be much clear. The moving average technique was adopted for investigation of trend. The polynomial models were fitted to the original data as well as three year, four year and five year moving average data while, Autoregressive Integrated Moving Average (ARIMA) models were fitted to the original data on area of rapeseed and mustard crop in Gujarat state. In fitting of Univariate Box-Jenkins (UBJ) Autoregressive Integrated Moving Average (ARIMA) models, the autocorrelation up to 7 lags were worked out for under study.

2.1 Regression method

Over the last several decades, regression and time-series models play an important role in statistical modeling and data analysis. Polynomial models *viz.*, regression (linear and non-linear) provides information on relation between a response (dependent) variable and one or more predictor (independent) variables (Bishal Deya *et al.* 2022) [6]

Linear regression approach (Rangaswamy, 2006) [10]

$$\hat{Y} = a + bt$$

Quadratic regression approach (Montgomery *et al.*, 2003) [9]

$$\hat{Y} = a + bt + ct^2$$

Third degree polynomial approach (Montgomery *et al.*, 2003) [9]

$$\hat{Y} = a + bt + ct^2 + dt^3$$

2.2 Evaluation of regression model

To test the goodness of fit of the fitted polynomial model, the highest Coefficient of Determination (R^2) value with significant coefficient of model and lowest value of Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) were computed to measure the adequacy of the fitted model. Basic assumptions regarding the error term like randomness and normality was tested by Run test (Sidney and Castellan, 1988) [12] and Shapiro - Wilk test (Shapiro and Wilk, 1965) [11], respectively.

2.3 Autoregressive integrated moving average (ARIMA) method (Box and Jenkins, 1976) [5]

Autoregressive (AR)

$$Z_t = C + \varphi_1 Y_{t-1} + a_t$$

Moving average (MA)

$$Z_t = C - \theta_1 a_{t-1} + a_t$$

An ARIMA model is commonly denoted as (p, d, q), where p is the number of the autoregressive terms, q denotes the number of the moving average terms and d indicates the number of differences required for stationarity.

2.4 Evaluation of ARIMA model

The first step to apply ARIMA model is identification of the time series. An Augmented Dicky-Fuller test shows if the dataset is stationary or not. If the time series is found to be stationary then model can be estimated, diagnosed and forecast can be made. But if it is not stationary then in order to apply ARIMA it has to be converted into stationary by differencing. After identification, ARIMA models are estimated for the specific stationary time series. Parsimonious ARIMA models are estimated based on the number of significant coefficients, low Schwartz-Bayes criterion (SBC) and Akaike Information Criterion (AIC), high adjusted R^2 and lowest value of Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). After estimation, diagnostics of the preferred ARIMA model is necessary to check if the residuals are independent. Residual Q-test (Ljung and Box, 1978) and normality test (Shapiro and Wilk, 1965) [11] can be performed for diagnostics.

3. Result and Discussion

3.1 Trend on rapeseed and mustard area in Gujarat state

The average area of rapeseed and mustard in Gujarat for the period of 1991-92 to 2019-20 was 2,77,117 hectare. The maximum area was 4,14,900 hectare in the year 1992-93 and minimum area was 1,60,800 hectare in the year 2002-03.

3.2 Fitting of polynomial models

The result of fitted polynomial models are given in Table 1. The results indicated that regression constant value of all three linear, quadratic and cubic models in all approaches were found to be significant. Linear regression coefficient was found to be significant in all approaches of linear, quadratic and cubic models. Quadratic regression coefficients were found to be significant in three, four and five year moving average data approach of cubic model. While, cubic regression coefficient was found to be significant in four and five year moving average data approach of cubic model. In original data approach the value of adjusted R^2 was improved by 1.00% in quadratic regression as compared to linear regression while, in case of cubic regression, it was improved by 1.80% than quadratic model. By taking moving average of five year the improvement in adjusted R^2 was observed to be 6.90, 10.40, 21.90% in case of first, second and third degree polynomial models, respectively over original data. Thus, higher improvement in adjusted R^2 was observed due to moving average data approach.

The third degree polynomial models showed comparatively lower values of RMSE and MAE. Among these the least RMSE and MAE were observed in case of model based on five year moving averages. The criteria for testing normality (Shapiro-Wilk test) of residuals indicated that all models in all approaches except linear model in original data approach had normally distributed. The test of randomness of residuals (Run test) indicated that only linear and quadratic model of original data approach were randomly distributed.

3.2 Fitting of ARIMA models

As the series was found non-stationary, the new variable Xt was constructed by taking difference of one (d = 1) to make the data stationary.

The autocorrelation (ACF) and partial autocorrelation (PACF) coefficient of various order of Xt were computed to identify the value of p and q. On the basis of goodness of fit criteria some models were selected and are given in Table 2. From the observed model ARIMA (0, 1, 3) had lowest AIC (Akike

Information Criterion) and SBC (Schwartz-Bayesian Criterion), RMSE and MAE value with significant constant and MA term and highest adjusted R^2 (64.53%). The assumption of residuals tested by the Shapiro-Wilk test and Box-Ljung test indicated that the selected model satisfied the assumption of residuals.

Thus, among the polynomial models, quadratic model on original data approach was found satisfactory to predict the trend of rapeseed and mustard area in Gujarat state. Also, ARIMA (0, 1, 3) was found statistically suitable for prediction of trend on rapeseed and mustard area in Gujarat state.

In order to check the validity of these forecasted values, they were compared with the actual values of rapeseed and mustard area during the post sampled forecast period for quadratic model on original data approach and post sampled forecast period for ARIMA (0, 1, 3) models which are presented in Table 3.

It is observed that the mean% deviation between forecasted and actual area based on quadratic model of original data approach and ARIMA (0, 1, 3) models were 20.97 and 13.51%, respectively. This proved that the both the models were the best fit models for forecasting the area of rapeseed and mustard in Gujarat state.

4. Conclusion

The quadratic model on original data approach was satisfied all the criteria for selection of model. The significant negative linear and positive quadratic term was observed in quadratic trend for area with adjusted R^2 (49.00%) compare to linear. It was also, fulfilled the assumptions of randomness and normality of the residuals, quadratic model on original data approach was considered as suitable. The selected model is

$$\hat{Y} = 4203.17^{**} - 150.52^{*}t + 3.03t^2 \text{ (Adj. } R^2 = 49.00\%)$$

ARIMA (0, 1, 3) model was selected on the basis of significant MA coefficient, lower value of AIC, SBC, RMSE, MAE and assumptions of residuals which is as under

$$Y_t = -78.10^{**} + 0.62^{**}a_{t-1} - 0.62^{**}a_{t-2} + 1.00^{**}a_{t-3} + a_t \text{ (Adj. } R^2 = 64.53\%)$$

Table 1: Fitted linear and non-linear models for area of rapeseed and mustard crop in Gujarat state

Model	Moving average	Regression constant	Regression coefficient			Adj. R^2 (%)	RMSE	MAE	S-W test	Run test |Z|
		A	b	c	d					
Linear	Original	3874.90**	-74.76**	-	-	48.00	514.94	369.40	0.90*	1.87
	3 years	3759.26**	-72.31**	-	-	52.20	419.71	329.75	0.93	3.27**
	4 years	3702.98**	-71.25**	-	-	53.60	383.20	311.56	0.94	3.13**
	5 years	3631.62**	-68.49**	-	-	54.90	341.32	285.25	0.93	2.98**
Quadratic	Original	4203.17**	-150.52*	3.03	-	49.00	498.38	374.95	0.94	0.62
	3 years	4109.52**	-159.88*	3.80	-	55.00	396.85	312.28	0.97	2.84**
	4 years	4046.07**	-160.75*	4.06	-	56.90	359.40	293.29	0.96	3.13**
	5 years	3976.14**	-162.45**	4.47	-	59.40	314.52	263.53	0.96	2.98**
Cubic	Original	4689.33**	-362.54*	23.80	-0.55	50.80	477.43	388.82	0.94	2.29*
	3 years	4683.18**	-429.66*	32.49*	-0.83	60.70	360.97	305.16	0.97	2.40*
	4 years	4697.04**	-479.37**	39.44*	-1.07*	66.40	308.33	258.00	0.96	2.23*
	5 years	4657.38**	-510.06**	44.86**	-1.28**	72.70	250.30	211.11	0.95	2.06*

*,** indicates significant at 5% and 1% level of significance, respectively

Table 2: Fitted ARIMA models for area of rapeseed and mustard crop in Gujarat state

Model	AIC	SBC	Constant	AR (φ)				MA (θ)				Adj. R^2 (%)	RMSE	MAE	S-W test	B-L test (Q)
				AR (1)	AR (2)	AR (3)	AR (4)	MA (1)	MA (2)	MA (3)	MA (4)					
ARIMA (0, 1, 3)	349.85	354.39	-78.10**	-	-	-	-	0.62**	-0.62**	1.00**	-	64.53	395.83	264.45	0.83	3.91
ARIMA (0, 1, 4)	351.47	357.14	-79.41**	-	-	-	-	0.51	-0.56	0.89*	0.15	63.03	393.37	272.82	0.84	2.62
ARIMA (1, 1, 3)	351.48	357.16	-79.41**	0.14	-	-	-	0.66**	-0.66*	1.00	-	63.04	393.29	271.44	0.84	2.66
ARIMA (2, 1, 3)	353.48	360.29	-79.37**	0.14	-0.01	-	-	0.66**	-0.66*	1.00**	-	60.86	393.34	271.83	0.84	2.65
ARIMA (4, 1, 1)	354.54	361.35	-71.17**	0.29	0.48*	-0.26	-0.32	0.99**	-	-	-	58.71	404.00	278.56	0.87	1.12

*,** indicates significant at 5% and 1% level of significance, respectively

Table 3: Testing of forecast values for five year by using best fitted quadratic model on original approach and ARIMA (0, 1, 3) model for rapeseed and mustard area ('00 ha) of Gujarat state

Year	Observed	Predicted					
		Quadratic	Deviation	% deviation	ARIMA (0, 1, 3)	Deviation	% deviation
2015-16	1902	2333.92	431.92	22.70	2400.76	498.75	26.22
2016-17	2001	2337.93	336.93	16.83	1757.14	243.85	12.18
2017-18	2213	2348.00	135.00	6.10	1839.69	373.30	16.86
2018-19	1954	2364.13	410.13	20.98	1761.58	192.41	9.84
2019-20	1726	2386.32	660.32	38.25	1683.48	42.52	2.46
Mean	1959.2	2354.06	394.86	20.97	1888.53	270.17	13.51

5. References

1. Anonymous. Annual Report, Directorate of Agriculture, Gandhinagar, Gujarat; c1991-20c.
2. Anonymous. District-wise Area, Production and Yield of Important Food & Non-food Crops in Gujarat State, Directorate of Agriculture, Gandhinagar, Gujarat; c2019-20.
3. Anonymous. Brief about rapeseed-mustard crop; c2023a. Retrieved from https://www.drmr.res.in/about_rmcrop.php [Accessed 10 July, 2023].

4. Anonymous. From the director's desk; c2023b. Retrieved from https://www.drmr.res.in/director_desk.php [Accessed 10 July, 2023].
5. Box GEP, Jenkins GM. Time series analysis: forecasting and control, second edition, holden day; c1976.
6. Deya B, Roya B, Dattab S, Ustunc TS. Forecasting ethanol demand in India to meet future blending targets: A comparison of ARIMA and various regression models, Energy Reports. 2022;9:411-418.
7. Liew VKS, Shitan M, Hussain H. Time series modeling and forecasting of sarawak black peeper price; c2000. (https://www.researchgate.net/publication/24115696_Time_series_modelling_and_forecasting_of_Sarawak_black_pepper_price)
8. Ljung GM, Box GEP. On a measure of lack of fit in time series models. Biometrica. 1978;65:297-303.
9. Montgomery DC, Peck EA, Vining GG. Introduction to linear regression analysis. John wiley & sons; c2003. p. 221-258.
10. Rangaswamy R. A text book of agricultural statistics. New age international limited publishers; c2006. p. 158-210.
11. Shapiro SS, Wilk MB. An analysis of variance test for normality. Biometrica; c1965.
12. Sidney S, Castellan J. Nonparametric statistics for the behavioral sciences (2nd ed.). New York: Mcgraw-Hill; c1988.