Forecasting of wheat production and productivity of Ahmedabad region of Gujarat state by using ARIMA models

L. Netajit Singh*[1], V. B. Darji[2], D. J. Parmar[3]

[1]*B. A. College of Agriculture,* [2] *Information Technology,* [3]*Department of Agricultural Statistics, Anand Agricultural University, Anand, 388110, Gujarat, India.*

netajitaau@gmail.com*[1], vbdarji@aau.in[2], djparmar.a@gmail.com[3]

Abstract

Background: ARIMA models were carried out to explain the fluctuations in production and productivity for wheat crop in Ahmedabad. Data from the year 1960-61 to 2010-11 were used for model fitting and forecasting ten years ahead from the year 2010-11.

Method: The ARIMA models with different p,d and q were judged based on autocorrelation function and partial auto correlation function at various lags and different ARIMA models were fitted.

Result: Among different fitted ARIMA models, ARIMA (0,1,1) family model was found suitable to forecast the pattern of wheat production and productivity trend of Ahmedabad region of Gujarat State.

Application: Forecasted values showed an increasing pattern in production and productivity of wheat in Ahmedabad region and predicted values for production and productivity of wheat in the year 2020-21 are3113.14 thousand tons and 1757.41 kg/ha respectively.

Keywords: Production, Productivity, wheat, Ahmedabad, ARIMA.

1. Introduction

Wheat (*Triticum aestivum*) is the most important food grain crop of the world. In Gujarat wheat is grown over an area of 1.05 Mha. with the production of 3.13 Mt. and productivity of 2986 kg ha-1 [1]. Gujarat accounted for 1.75% of the total area and 1.32% of the total production of wheat in the country. Mehsana, Banaskantha, Rajkot and Kheda districts in the valleys of the Sabarmati and Mahi rivers are the main producers which together contribute about 55% of the state's production of wheat. Others include Ahmedabad, Sabarkantha, Bharuch and Bhavnagar districts where 6 to 10 per cent of the cropped area is devoted to wheat cultivation [2].

Wheat is the second most important cereal crop in India after rice and it is severely affected with abiotic factors e.g. Rainfall, humidity and other environmental factors and biotic stresses such as diseases and pest infestation which also indirectly depends upon environment. There are several statistical tools available to predict/forecast the wheat production with the help of assessing the environmental influence on yield.

ARIMA model is an extrapolation method for forecasting and like any other such method, it requires only the historical time series data on the variables under forecasting. Among the extrapolation methods, this is one of the most sophisticated method, as it incorporates the future of all such methods, does not require the investigator to choose initial values of any variables and values of the various parameters a priori. It is robust to handle any data pattern. As one would expect this is quite a difficult model to develop and apply as it involves transformation of the variable, identification of the model, estimation through nonlinear method, verification of the model and derivation of the forecasts [3].

2. Materials and Methods

In regression model, the parameters β's are assumed to be constant over the time. In the forecasting models the errors ε_t's within time period (t = l, 2, 3, . . . , n) are assumed to be uncorrelated i.e. the observations Y_t's are uncorrelated. However, this assumption is rarely met in practice. Usually serial correlations in the observations often exist in time-series data. The statistical concept of autocorrelation was used to measure the relationships between the value of Z at time t (i.e., Y_t) and Y at earlier time periods (i.e., Y_{t-1}, Y_{t-2} ...). The algebraic forms of Autoregressive (AR) and Moving average (MA) processes are:

2.1.1. Autoregressive (AR) process

$$Z_t = C + \varphi_1 Y_{t-1} + a_t \quad .....(1)$$

Where Z_t= time sequenced random variable

C = constant term related to mean (μ) such that C = μ(1-ϕ_1)

ϕ_1 = relationship of Y_t with Y_{t-1}

a_t = a random shock element at time t

Similarly, the MA (q) model is again the generalizations of moving average model may be specified as.

2.1.2. Moving average (MA) process

$$Z_t = c - \theta_1 a_{t-1} + a_t \quad ....(2)$$

Where C = constant term related to mean μ and

θ = relation of at with a_{t-1}

Combining both the model is called ARIMA model, which has general form

$$Z_t = c + \phi_1 Y_{t-1} + \theta_1 a_{t-1} + a_t$$

2.2. Fitting of Box-Jenkins ARIMA Models

Box-Jenkins time-series models i.e. ARIMA (p, d, q) is known as "Univariate Box-Jenkins technique" [4] ARIMA model is an algebraic statement telling how observations on a variable are statistically related to past observation.

This model amalgamates three types of process, viz., Autoregressive of order p; differencing to make a series stationary of degree d and moving average of order q. This method applied only to a stationary time series data. When the data is non-stationary then it has to be brought into stationary by the method of differencing.

2.3. Test for Stationarity

The stationarity requirement ensures that one can obtain useful estimates of the mean, variance and ACF from a sample. The stationarity condition of a series was tested by examining the

1. The change of mean and variance over time.
2. The coefficients of AR and MA process i.e. in case of AR (1) and MA (1) process it should be $|\phi_1|<1$ and $|\theta|<1$.
3. The estimated ACF values which should be tails-off towards zero rapidly.

The significance of autocorrelation was tested by t-test. The standard error of autocorrelation [5] was calculated as under

$$s(r_k) = \left(1 + 2\sum_{j=1}^{k-1} r_j^2\right)^{1/2} n^{-1/2} \qquad \text{.....................(3)}$$

$$t_{r_k} = \frac{r_k - \rho_k}{S(r_k)} \qquad . \qquad \text{....................(4)}$$

k=1,2,3,…

The significant value of "t" indicates the presence of autocorrelation. The process of time series modelling involves transformation of data in order to achieve stationary, followed by identification of appropriate models, estimation of parameters, validation of models and finally for prediction. The complete description of these process and steps of time series modelling is clearly explained below.

2.4. ARMA modelling consists of three operational steps

Identification, Estimation and Diagnostics checking

2.4.1. Identification

Identification involves the techniques to determine the values of p, q and d. The values are determined by using Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF). For any ARIMA (p, d, q) process, the theoretical PACF has non-zero partial autocorrelations at lags 1, 2, ..., p and has zero partial autocorrelations at all lags, while the theoretical ACF has non zero autocorrelation at lags 1,2, …, q and zero autocorrelations at all lags. The nonzero lags of the sample PACF and ACF are tentatively accepted as the p and q parameters. For a non-stationary series the data is differenced to make the series stationary. The number of times the series is differenced determines the order of d. Thus, for a stationary data d = 0 and ARIMA (p, d, q) can be written as ARMA (p, q).

2.4.2. Estimation

The main approaches for fitting Box-Jenkins models are non-linear least squares and maximum likelihood estimation which was estimated by using SPSS (version 17) software

2.4.3. Diagnostic Checking

The best model was selected on the basis of minimum values of Schwartz-Bayesian Information Criterion (SBC), Akaike Information Criterion (AIC) and Root Mean Square Error (RMSE). Residuals were tested by run test and for randomness by Shapiro – Wilk test for normality and The Ljung and Box for independent were used.

2.5. Test for normality of the residual [6]

The Shapiro - Wilk statistic was used to test the normality of residuals,
The required test statistics W was defined as

$$W = \frac{S^2}{b} \text{ Where } S^2 = \sum a(k)\left[e(n+1-k) - e(k)\right]. \qquad \text{............(5)}$$

The parameter k takes the values

$$K = \begin{cases} 1,2,3,4,\text{..........}n/2 & \text{when n is even} \\ 1,2,3,4,\text{.........}(n-1)/2 & \text{when n is odd} \end{cases}$$

$$\text{and } b = \sum_{i=1}^{n} (e_i - \bar{e})^2$$

2.6. Test for independence of errors (Chi-square test)

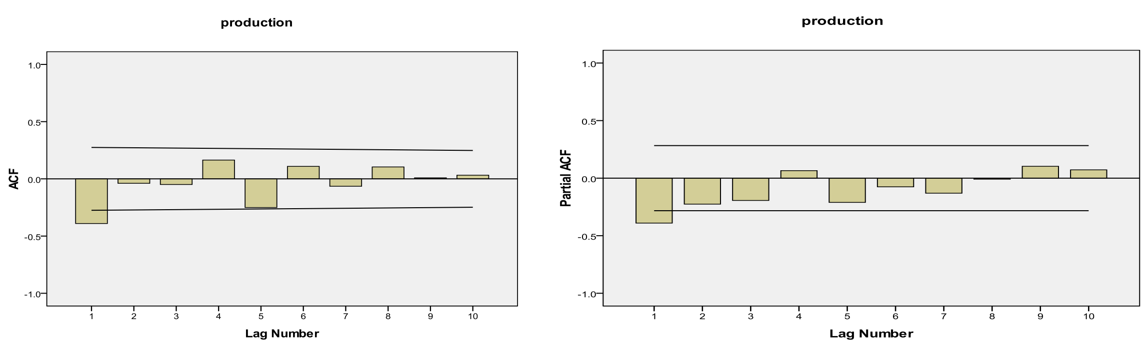
$$Q = n(n+2)\sum_{i=1}^{K}(n-k)^{-1} r_i^2(\hat{a}) \qquad \text{.........................(6)}$$

Where, n is the number of observations. The statistic Q approximately follows a Chi – square (χ^2) distribution with (K-m) degrees of freedom, where K is the number of residual autocorrelation and m is the number of parameters estimated in the ARIMA model.

3. Results and Discussion

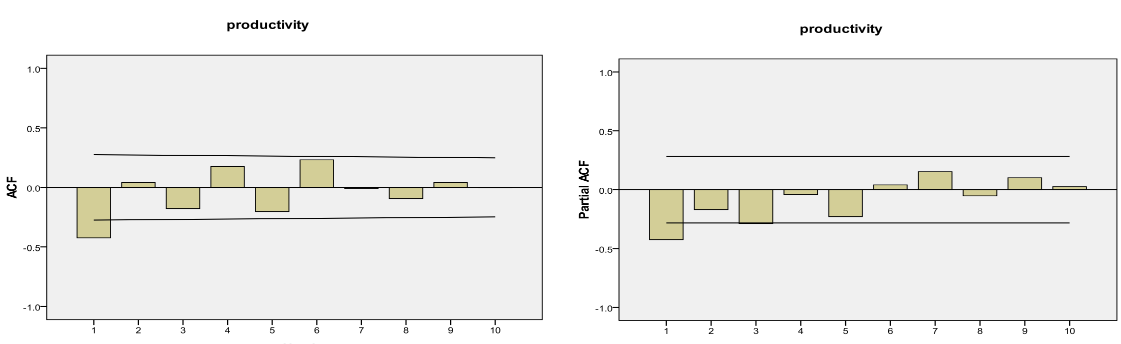
ARIMA models were fitted for wheat production and productivity of Ahmedabad region of Gujarat state. The data for the year 1960-61 to 2010-11 from the published reports by the Directorate of Agriculture, Gujarat State, Gandhinagar were used. In fitting of Univariate Box-Jerikins (UBJ) ARIMA models, the autocorrelation up to 10 lags were worked out. If the spikes did not sharply tails-off towards zero and if the visual inspection of the realization indicates that the mean, variance and autocorrelation were not constant over time then the series was considered as non-stationary. Therefore, the new variable X_t was constructed by taking difference of one (i.e. d = 1) to make the series stationary.

Figure 1. ACF and PACF of the different series for wheat production in Ahmedabad region of Gujarat



(a) ACF for production of wheat in Ahmedabad (b) PACF for production of wheat in Ahmedabad

Figure 2. ACF and PACF of the different series for wheat productivity in Ahmedabad region of Gujarat



(a) ACF for productivity of wheat in Ahmedabad (b) PACF for productivity of wheat in Ahmedabad

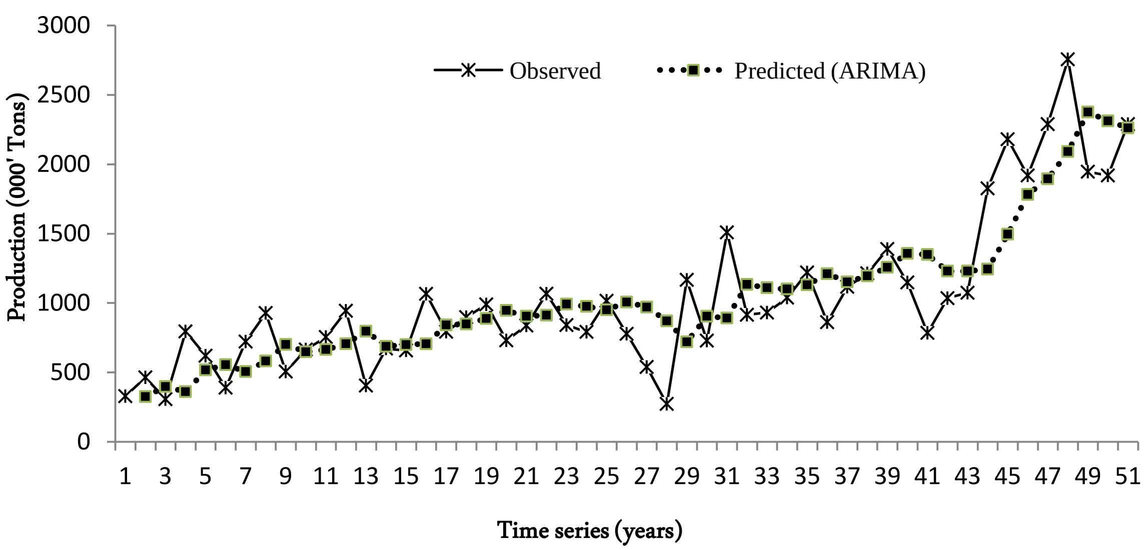
3.1. Fitting trend on wheat Production in Ahmedabad region in Gujarat state by using ARIMA models

The series was made stationary by taking differences of one (*i.e.* d=1). The ACF (Υ_k) of the transformed variables were tails off toward zero with cut- off first and fifth spikes and PACF (ϕ_{kk}) of the transformed variables tails off toward zero with cut- off first spike (fig. 1). This suggested that the algebraic family of ARIMA on p =0,1 d=1 and q=0,1,2,3,4,5 can be used. The different models among the different value of p and q were fitted. Among the models, those models having lower value of AIC and SBC are given in Table 1. Among these models, only ARIMA (0,1,1) had significant coefficients *i.e.* MA (θ) coefficient. The assumptions of residuals *i.e.* normality and independence of residuals were tested by Shapiro- Wilk test and Box-Ljung (Q) test indicated that ARIMA (0,1,1) models satisfied the assumptions of normality and independence residuals and also ARIMA (0,1,1) model had comparatively lower value of AIC, SBC and RMSE . So, ARIMA (0,1,1) model was found suitable to explain the trend of wheat production in Ahmedabad region of Gujarat state. The trend of wheat production by using selected model *i.e*. ARIMA (0,1,1) is given in fig. 3.

Table 1. Fitted ARIMA models for Wheat production in Ahmedabad region of Gujarat

ARIMA	AIC	SBC	AR(ϕ)	MA(θ)	CONS	RMSE	SW-TEST	BLQ- TEST
(0,1,1)	720.114	722.026	-	0.680**	33.039	312.391	0.971	18.103
(1,1,0)	722.609	724.521	-0.393**	-	5.043	330.834	0.986	20.579
(1,1,1)	722.131	725.955	-0.357*	1.00	-12.764	306.046	0.987	15.147
(2,1,1)	724.073	702.889	0.384*,0.216	0.997	27.261	242.772	0.986	15.805
(1,1,2)	695.422	703.070	-0.969*	-0.420, 0.576	40.505	243.857	0.990	15.626
* Significant at 5% level, ** Significant at 1% level								

Figure 3. Trend in wheat production based on ARIMA (0,1,1) model in Ahmedabad region



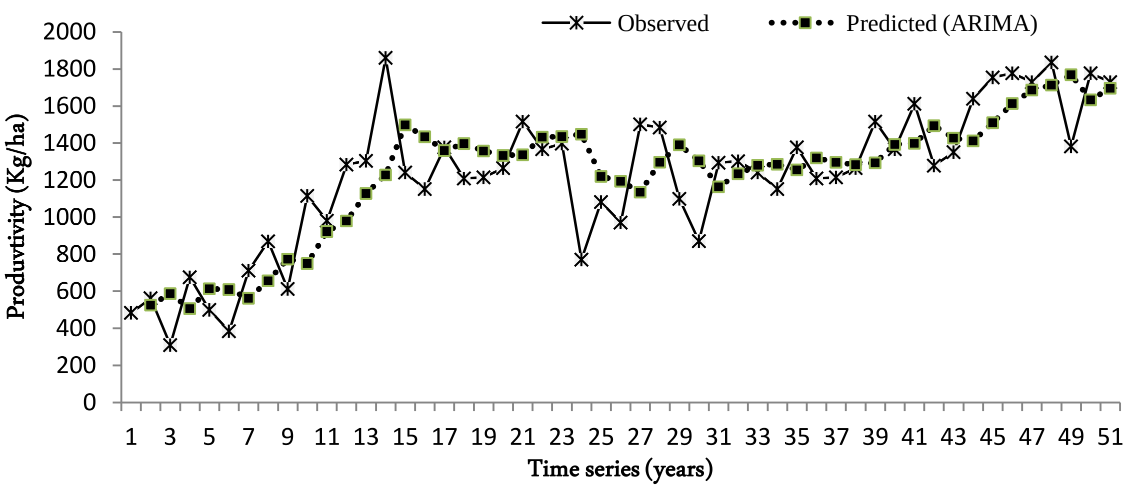
3.2. Fitting trend on wheat productivity in Ahmedabad region of Gujarat state by using ARIMA models

The series was made stationary by taking differences of one (*i.e.* d=1). The value of p and q were identified using ACF and PACF coefficients of various order of X_t. The ACF (Υ_k) of transformed variables were dumping-off towards zero with cut-off initial spike and the PACF (ϕ_{kk}) also cut-off at first and third lag (Fig.2). This suggested that the algebraic family of ARIMA on p =0,1,2,3, d=1 and q=0,1 can be used. The different models with these different values of p and q were fitted. Among the models, those models having lower value of AIC and SBC are given in Table 2. From the fitted models, ARIMA (0,1,1) and ARIMA (1,1,0) model had significant MA (θ) and AR (ϕ) coefficient term of which ARIMA (0,1,1) model had lower values of AIC, SBC and RMSE. The assumptions of residuals (*i.e.* normality and independence of residuals) were tested by Shapiro- Wilk test and Box-Ljung (Q) test indicated that ARIMA (0,1,1) satisfied all the assumptions. Therefore, ARIMA (0,1,1) model were found suitable to explain the trend of wheat productivity in Ahmedabad region of Gujarat state. The trend of wheat productivity by using selected model *i.e*. ARIMA(0,1,1) is given in fig. 4.

Table 2. Fitted ARIMA models for Wheat productivity in Ahmedabad region in Gujarat

ARIMA	AIC	SBC	AR(ϕ)	MA(θ)	CONS	RMSE	SW-TEST	BLQ- TEST
(0,1,1)	691.290	695.114	-	0.625**	43.183	239.033	0.985	15.208
(1,1,1)	693.206	698.942	0.90	0.625**	43.222	241.272	0.983	14.484
(1,1,0)	695.839	699.717	-0.417	-	37.154	251.644	0.976	22.38
(2,1,1)	695.241	702.889	0.384*,0.216	0.997	27.261	242.772	0.986	15.805
(1,1,2)	695.422	703.070	0.969*	-0.420, 0.576	40.505	243.851	0.990	15.626
* Significant at 5% level, ** Significant at 1% level								

Figure 4. Trend in wheat productivity based on ARIMA (1,1,0) model in Ahmedabad region



Finally ten year ahead forecast was made for production and productivity by using ARIMA (0,1,1) models. Table 3, show the forecast values for production and productivity at the 95% confidence limit. From table 3 Forecast for the 2011-12 for production and productivity were 2347.73 thousand tons with a 95% confidence limit of (1720.57, 2974.89) thousand tons and 1716.05 kg/ha with a 95% confidences limit of (1236.04, 2196.06) kg/ha respectively. For the year 2020-21, the forecast for production and productivity were 3113.14 thousand tons and 1757.41 kg/ha with a 95% confidences limit of (2244.04, 3982.24) thousand tons and (1034.69, 2480.12) kg/ha respectively.

Table 3. Forecast for production and productivity of wheat in Ahmedabad region of Gujarat by using selected models

Years	Production (000'Mt)	Productivity (Kg/ha)
2011-12	2347.73±627.16	1716.05±480.01
2012-13	2426.36±658.45	1723.35±512.68
2013-14	2506.59±688.32	1729.98±543.40
2014-15	2588.43±716.94	1735.93±572.46
2015-15	2671.87±744.46	1741.2±600.12
2016-17	2756.92±771	1745.8±626.57
2017-18	2843.57±796.66	1749.71±651.93
2018-19	2931.82±821.52	1752.96±676.35
2019-20	3021.68±845.64	1755.52±699.91
2020-21	3113.14±869.1	1757.41±722.72

4. Conclusion

The Box- Jenkins approach was used to model and forecast production and productivity of wheat crop in Ahmedabad region of Gujarat. Forecasted values showed an increasing pattern in production and productivity of wheat crop in Ahmedabad region. The main result behind the increase in production is due to the increase of productivity by using new technology in cultivation of wheat crops in Ahmedabad district in Gujarat. For further improvement of production and productivity if wheat in this area selection of high yielding variety, adequate input supply at right time and farmer education is required.

5. Acknowledgement

The author thanks Dept. of Agril. Statistics, BACA, AAU, Anand, Directorate of Agriculture, Gujarat State, Gandhinagar and Dept. of Agril. Economics, BACA, AAU, Anand for their kind help and active support.

6. References

1. Anonymous. Wheat Scenario – A Snippet, Directorate of Wheat Research Karnal, Haryana (India). 2013.
2. Anonymous. Complete information on Area and Production of wheat in India, Preserve Articles. 2012.
3. G. S. Gupta. ARIMA model for forecasting on tea production in India. *Indian Economic Journal*. 1993; 41(2), 88-110.
4. G. E. P. Box, G. M. Jenkins. Time Series Analysis: *Forecasting and Control*, Second Edition, Holden Day. 1976.
5. M. S. Bartlett. On the theoretical specification of sampling properties of autocorrelated time series. *Journal of the Royal Statistical Society*. 1946; B (8),27.
6. S. S. Shapiro, M.B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*. 1965; 52, 591-611

The Publication fee is defrayed by Indian Society for Education and Environment (www.iseeadyar.org)

Cite this article as:

L. Netajit Singh, V. B. Darji, D. J. Parmar.Forecasting of Wheat Production and Productivity of Ahmedabad Region of Gujarat State by Using ARIMA Models*. Indian Journal of Economics and Development.* Vol. 3 (6), 2015.