Journal of Crop and Weed, 9(2):96-101(2013)

Instability and forecasting using ARIMA model in area, production and productivity of onion in India

P. MISHRA, [1]C SARKAR, VISHWAJITH K.P, B. S. DHEKALE AND P.K. SAHU

Department of Agricultural Statistics
Bidhan Chandra Krishi Viswavidyalaya
Nadia, Mohanpur-741252, West Bengal
[1]Office of the ADA, Soil Survey, Silguri, Govt. of West Bengal

Received: 18-09-2013, Revised: 29-10-2013, Accepted: 15-11-2013

ABSTRACT

Onion is a bulbous spice crop which is produced and consumed largely in India as well as in the world. It has numerous medicinal uses for treating diseases. Amongst the onion producing countries in the world, India ranks second in area and production, the first being China. India being a second major onion producing country in the world has a productivity of 10.16 t ha^{-1} only (FAO 2008). Time Series data covering the period of 1978-2008 was used for the study. Among the parametric trend models cubic trend model is found to adequately delineate trends in area, production and productivity of onion. Onion showed high instability in period II in area, production and productivity. The study also focuses on forecasting the cultivated area and production of onion in India using Autoregressive Integrated Moving Average (ARIMA) model. The analyses forecast onion production for the year 2020 to be about 23.02 million tonnes. These projections will help information of good policies with respect to relative production, price structure as well as medicine frontier at least of onion in the country.

***Keywords*:** ARIMA, forecasting, instability, modelling and production

Onion is one of the most popular vegetable that form of daily diet. In India, onion an important commercial crop.Onion is an important vegetable crop grown all over the world. Its demand is worldwide .Onion is carminative, melt the phlegm and oil extracted from these is volatile. Besides, onion has several medicinal uses described by various authors as remedy or cure for different ailments. Onion is the largest vegetable produced and consumed not only in India but also in the world. India being a second major onion producing country in the world has a productivity of 10.16 t ha^{-1} only (FAO, 2008). Onion is exported from India to 38 countries in varying quantities. Thus, from food, nutritional, medicinal and economic security point of view, study of production behaviour and its future prediction is of utmost importance. Auto Regressive Integrated Moving Average (ARIMA) is the most general class of models for forecasting a time series. Appearance of lags of the forecast errors in the model is called “moving average”. (ARIMA) model was introduced by Box and Jenkins. (1976) for forecasting variables. Sahu and Mishra (2013) studied forecasting the production, import–export (both in quantity and value) and trade balance of total spices in India and China alongwith world using Autoregressive Integrated Moving Average (ARIMA) model for time series data and forecasted for year 2020. Agricultural growth and instability has remained subject of intense debate not only in India but throughout the world. While the need for increasing agricultural production or growth is obvious, the increase in instability in agricultural production is considered sometimes adverse for several reasons(Chand and Raju, 2009). Hasan *et. al.,* (2008) measured the change and instability in area, production, and yield of two major cereal crops wheat and maize in Bangladesh.

Email: pradeepjnkvv@gmail.com

MATERIALS AND METHODS

Data related to area, production and yield of onion in India since 1978 to 2008 were collected from Agriculture at Glance, 2010. The whole period was divided into two period’s *viz*., period-I from 1978 to 1992 and period-II from 1993 to 2008 to compare in area, production, and yield between the two periods. In order to examine the nature of change, instability, and degree of relationship in area, production, and yields of onion in India, various statistical measures, such as mean, correlation co-efficient and co-efficient of variation were worked out. ; Box-Jenkins ARIMA modelling has been used to forecast series under consideration.

Descriptive statistics

To examine the nature of each series these have been subjected to get various statistics. Descriptive statistics are typically distinguished from inferential statistics. With descriptive statistics one simply describes what is or what the data shows. With inferential statistics, one tries reach conclusions that extend beyond the immediate data alone. Statistical tools used to describe the above series are minimum, maximum, average, standard error, skewness, kurtosis.

Parametric trends models

To get an overall movement of the time series data, trend equations are fitted. In this exercise different idea about the models like, polynomial, exponential, linear, compound etc are used for the purpose.

Models	Form
Polynomial	$Y_t = b_0 + b_1t + b_2t^2 + b_2t^3 + \ldots\ldots\ldots + b_kt^k$
Linear	$Y_t = b_0+b_1t$
Quadratic	$Y_t = b_0+b_1t+b_2t^2$
Cubic	$Y_t= b_0+ b_1 t + b_2t^2 + b_2t^3$
Compound	$Y_t= b_0\, b_1^t$
Exponential	$Y_t = b_0e^{(b_1t)}$
Logarithmic	$Y_t = b_0 + b_1\ln(t)$
Growth	$Y_t = e^{b_0 + b_1\ln(t)}$

To identify the significant change in area, production, and yield between two periods, the following formula was used: $t = \dfrac{\bar{x}_1 - \bar{x}_2}{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$

where n_1 and n_2 are the sample sizes, $\bar{x}_1$ and $\bar{x}_2$ are the sample means, and S_1^2 and S_2^2 are the sample variances. The production of onion is likely to be influenced by the area used for onion in order to estimate the parameter; simple linear regression models were fitted to examine the change of production by the change of area. The model can be expressed as:

$Y = a + bX – e$ Where, Y = Production in ton, a = intercept, b = regression coefficient, X = area in ha, e = error term. The growth rates of area, production, and yield of onion were worked out by fitting a semi-log function of the following type:

$y=e^{a+bt}$ *or* $I_n y=a+bt$

Where y = Area (ha), production (t), yield ($t.ha^{-1}$), t = time period (year). An index of instability was computed for examining the nature and degree of instability in area, production, and yield in India. For measuring the instability in area, production and yield the index given by Cuddy and Della (1978) and used by Larson *et al.* (2004): $CV_t = (CV) \times \sqrt{1-R^2}$

where, $C.V. = \dfrac{\sigma}{\bar{X}} \times 100$

Where σ = Standard Deviation,

CV_t = CV around trend,

$\bar{X} = Mean$

More general option is to use ordinary CV value but in presence of trend, ordinary CV fails to explain the inherent trend component in a time series properly (Hasan *et al*., 2008). As such, this study obtained for CV around trend *i.e.* CV_t,

After the evaluation of trend and instability of each and every series, our next task is to forecast the series for the year to come. For the purpose the study adopted the Box –Jenkins (1976) methodology. Data for the period 1970-2006 has been used for the model building, while data for years 2007-08 are taken for model validation. On the basis of best fitted model forecasting has been made up to 2020.

Autoregressive model

ARIMA models, which stands for Autoregressive Integrated Moving Average models. Integrated means the trends has been removed; if the series has no significant trend, the models are known as ARMA models. The notation AR (*p*) refers to the autoregressive model of order *p*. The AR (*p*) model is written $X_t = c + \sum_{i=1}^{P} \alpha_i X_t + \mu_t$

where $\alpha_1, \alpha_2 \ldots \alpha_p$ are the parameters of the model, *c* is a constant and μ_t is white noise. Sometimes the constant term is omitted for simplicity.

Moving average model

The notation MA (*q*) refers to the moving average model of order *q*: $X_t = \mu + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t$

Where the $\theta_1, \ldots, \theta_q$ are the parameters of the model, μ is the expectation of X_t (often assumed to equal 0), and the ε_t is the error term.

Given a set of time series data, one can calculate the mean, variance, autocorrelation function (ACF), and partial autocorrelation function (PACF) of the time series. Based on the nature of the above appropriate ARIMA models are worked out, but the final decision is made once the model is estimated and diagnosed. In this step one can see the whether the chosen model fits the data reasonably well. One simple test of the chosen model is to see if the residuals estimated from this model white noise; if they are, one can accept the particular fit; if not, one start the process afresh, thus the BJ Methodology is an iterative process. Models are compared according to the minimum values of Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Square Error (MSE) and Mean Absolute Percentage Error (MAPPE) and maximum value of Coefficient of determination (R^2).

$$MAE = \frac{\sum_{i=1}^{n} \left| X_i - \hat{X}_i \right|}{n},$$

$$R^2 = \frac{\sum_{i=1}^{n} \left(\hat{X}_i - \overline{X} \right)^2}{\sum_{i=1}^{n} \left(X_i - \overline{X} \right)^2},$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \left(X_i - \hat{X}_i \right)^2}{n}},$$

$$MAPE = \frac{\sum_{i=1}^{n} \left| \frac{X_i - \hat{X}_i}{X_i} \right|}{n} \mathrm{x}100$$

RESULTS AND DISCUSSION

Since 1978, the area under onion has been increased from 0.21 million ha to 0.83 million ha till 2008, registering a growth of almost 295% (Table 1).The average area under onion being 0.40 million hectare. In fact the effect of green revolution is being reflected. The effect of expansion of area is clearly visible in the production scenario of onion. With a mere 2.20 million tonnes of production it has reached to 13.97 million tonnes during the year 2008 and registering growth of almost 516%. B_2 (kurtosis) value (4.74) of production indicates there leptokurtic in nature. Increased production of onion would not been possible without a substantial increasing per hectare yield of the crop. Starting with only 9091kg of onion per hectare, it has reached to 16260 kg ha^{-1} during the year 2008. Positive values of skewness reveal that starting from the initial years of the period under investigation a continuous effort was there to increase the yield of onion. Thus the joint effect of expansion area and yield has resulted in a brighter picture of onion production scenario in India. There has been substantial growth in area, production and yield of onion during the period under investigation.

Table 1: *Per se* behaviour of onion production in India during 1978-2008

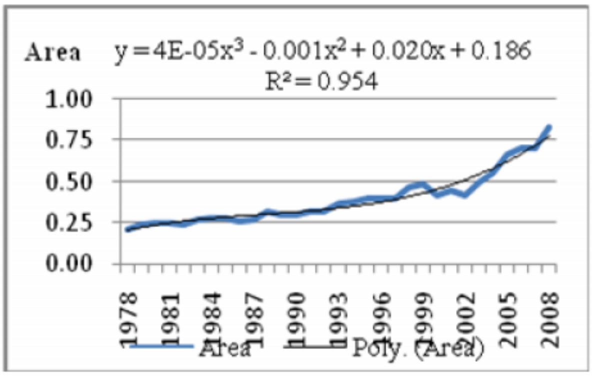
	Area (million ha)	Production (million t)	Yield (kg ha^{-1})
Mean	0.400	4.49	10910.38
Kurtosis	0.970	4.74	7.31
Skewness	1.220	2.07	2.34
Minimum	0.210	2.20	9091.00
Maximum	0.830	13.56	16260.00
SGR (%)	295.240	516.36	56.30

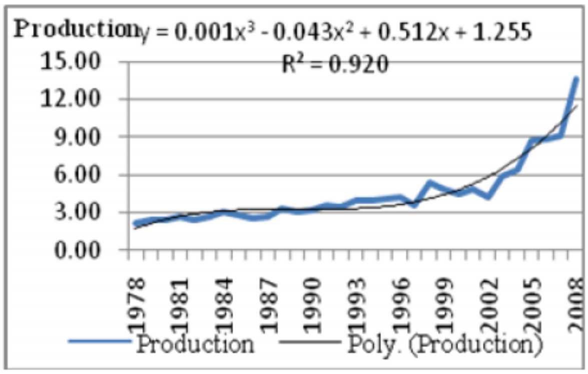
Trends in production behaviour of onion

To work out the trends in area, production and yield of Onion in India different parametric models like linear, logarithmic, quadratic, cubic, compound, growth, and exponential models were attempted to among the competitive models. The best model was selected on the basis of the maximum R^2 value, significance of the model and its coefficient (Table 2). In most of the cases, the non-linear patterns are revealed (Fig.1). In all cases cubic model is found to be best, the coefficients of cubic time factor are negative in nature and thereby indicating the tendencies of the series to decline in recent past.

Table 2: Model summary of trend model fitted

	Model	Constant	b1	b2	b3	R^2	F	Sig.
Area	Cubic	0.187	0.021	-0.001	0.000	0.955	189.526	0.0
Production	Cubic	1.256	0.513	-0.043	0.001	0.921	104.783	0.0
Yield	Cubic	9455.242	367.829	-35.003	0.924	0.761	28.644	0.0





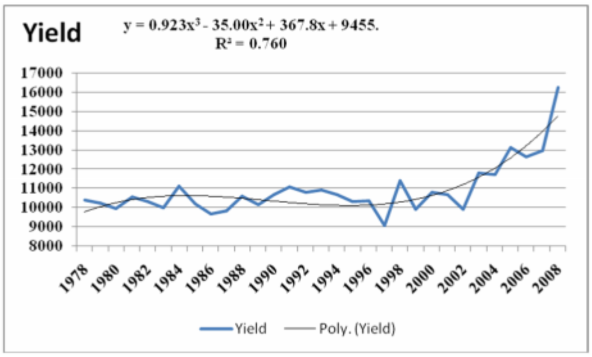


Fig.1: Trends in area, production and yield of onion of India

Change in area, production and yield of onion over the periods

Onion area and production increased from period-I to period-Il significantly (at 1% level of significance). The increase in yield from period-I to period-II was also significant at 10% level of significance (Table 3). So the yield of onion is also increased from period-I to period-II satisfactorily as with the enhancement of area.

Table 3: Change in area, production and yield of onion in India

	Period –I (1978-1992)	Period –II (1993-2008)	t- value	P (T<t) two tail
Area (million ha)	0.274	0.509	-6.520	0.000
Production (million tonnes)	2.859	6.023	-4.580	0.000
Yield (kg ha^{-1})	10379.200	11408.350	-2.310	0.017

Regression analysis

From table- 4, the simple linear regression functions were fitted for estimating the response of production of onion due to the change of their respective area. Productions of onion were significantly increased by 15.69 and 19.09 times during the whole study period and period II, respectively, by unit change in area. In the case of period-I, the production of onion was significantly increased as the co-efficient of production on area was found to be 12.20.

Table 4: Testing dependency of production of area in onion

Periods	Intercept	Regression co-efficients	t- value	P (T<t) two tail)
Whole period	-1.71	15.69	22.491	0.000
Period I	-0.48	12.2	13.644	0.000
Period II	-3.692	19.097	18.168	0.000

Instability in area, production and yield of onion

Fluctuations in area and production onion are interrelated as wider area gives greater production. But variation in yield may be due to weather condition, technological changes, etc. The instability of onion in area, production and yield are shown in table- 5. Area of onion during the whole period shows highest degree of instability. Excluding the whole period area, production, and productivity of onion showing higher instability in period II (1993-2008). Production during whole period and period-II reflect higher degree of instability, at the same time yield

also shows remarkable degree of instability during these two periods.

Box-Jenkins modelling and forecasting

After the evaluation of trend of each and every series, our next task is to forecast the series for the year to come. For the purpose of forecasting Box –Jenkins methodology as discussed in the material and method section was adopted. Data for the period

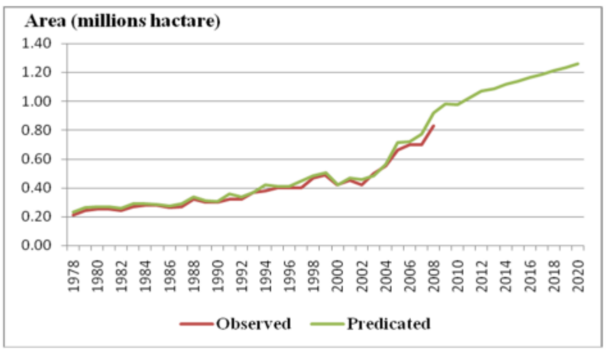
1978-2006 was used for the model building, while data for years 2007 and 2008 are taken for model validation. None of the series was stationary in nature and first order differencing was required for all the series. Best fitted models are used to forecast the series for the years to come. Though different series has been fitted with different ARIMA models but one thing is clear that table- 6 reveals that ARIMA (1, 1, 5) model is best suited for modelling of onion area data.

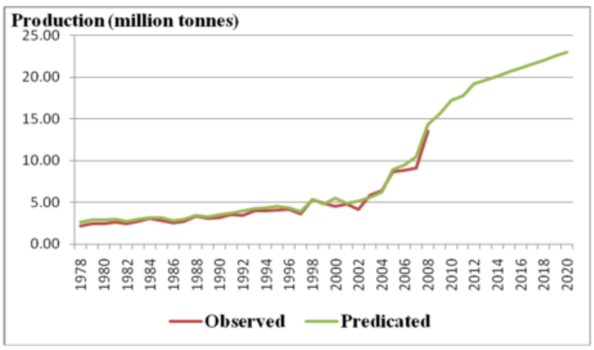
Table 5: Instability in area, production and yield of onion in India

	Measurement statistics	Whole period	Period I	Period II
Area (million ha)	CV	39.630	12.167	27.490
	R^2	0.832	0.846	0.787
	CV_t	15.850	4.750	12.650
Production (million tonnes)	CV	56.240	14.713	45.272
	R^2	0.671	0.764	0.691
	CV_t	32.058	7.062	24.899
Yield (kg ha^{-1})	CV	12.450	4.200	15.100
	R^2	0.363	0.111	0.543
	CV_t	9.839	3.945	10.119

Table 6: Model selection criteria for area, production and yield of onion in India

	Best fitted ARIMA model (p,d,q)	R^2	RMSE	MAPE	MAE	Max. APE	Max. AE	Normalized BIC
Area (Million ha)	(1,1,5)	0.95	0.04	6.92	0.03	20.35	0.10	-5.61
Production (million tonnes)	(0,1,5)	0.91	0.85	11.24	0.53	27.87	2.42	0.35
Yield (kg.ha^{-1}**)**	(1,1,4)	0.56	946.41	5.76	645.03	19.57	3181.42	14.05





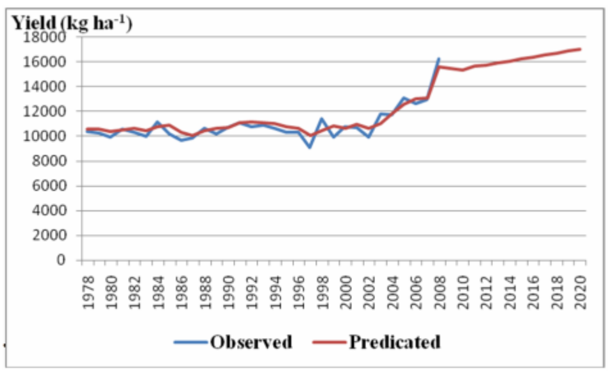


Fig. 2A: Observed and forecasted of area, production and yield of onion of India

It is clear from the observed and predicated values, and also from the figures (Fig. 2) that by and large the models have adequately been identified. From the forecast values obtained it can be said that forecasted area will increases to some extent in future and it would be 1.26 million hectare during 2020. In case of production of onion the ARIMA (0, 1, 5) model is best fitted, it can be said that forecasted production will increases to 23.02 million tonnes during 2020 (Table 7).

Table 7: Model validation and forecasting of area, production and yield of onion of India

Year	Area (million ha)		Production (million tonnes)		Yield (kg ha^{-1})	
	Observed	Predicated	Observed	Predicated	Observed	Predicated
2007	0.70	0.77	9.14	10.48	12974	13079
2008	0.83	0.92	13.56	14.40	16260	15613
2010	--	0.98	--	17.28	--	15365
2020	--	1.26	--	23.02	--	17030

In case of yield of onion the ARIMA (1, 1, 4) model is best fitted and forecasted yield would be 17030 kg ha^{-1} in 2020 increases to some extent in future i.e. in 2008-09 production of onion was 16260 kg ha^{-1}. However, demand for onion is expected to exceed production in future as well, creating severe problem of shortage in the province. Thus, it calls for an increase in the supply of onion to fulfil the high demand of onion in the province. The need is to exploit the existing potential to boost onion production. To meet the onion demand in country it is needed to induce vegetable growers for bringing additional area under onion.

REFERENCES

Box, G.E.P. and Jenkins, G.M. 1976. *Time Series Analysis: Forecasting and Control*. Holden-Day, San Fransisco.

Chand, R. and Raju, S.S. 2009. Instability in Indian Agriculture during Different Phases of Technology and Policy. *Indian J. Agril Econ.,* **64**: 283-88.

Cuddy, J.D.A. and Della, V.P A. 1978. Measuring the Instability of Time Series Data. *Oxford Bull. Econ. Stat.,* **40**:79-85

FAO.2008. http:// faostat.fao.org./www.fao.org.,Last access 11 april, 2012.

Hasan, M.N., Mohayem, M.A. Miah, Islam, M.S., Islam, Q.M. and Hossain, M.I. 2008. Change and Instability in area and production of Wheat and Maize in Bangladesh. *Bangladesh J. Agril. Res.* **33**:409-17.

Larson, D.W., Jones, E., Pannu, R.S. and Sheokand, R.S.2004. Instability in Indian Agriculture – A Challenge to the Green Revolution Technology. *Food Policy,* **29**: 257-73.

Sahu,P.K. and Mishra, P. 2013. Modelling and forecasting production behaviour and import-export of total spices In two most populous countries of the world. *J. Agric. Res.,* **51**: 81-97.