INNSPUB

International Journal of Biosciences | IJB |
ISSN: 2220-6655 (Print) 2222-5234 (Online)
http://www.innspub.net
Vol. 10, No. 4, p. 249-254, 2017

RESEARCH PAPER OPEN ACCESS

Forecasting dates production in Khyber Pakhtunkhwa, Pakistan

Syed Asghar Ali Shah*[1], Nagina Zeb[1], Alamgir[2]

[1]*Agricultural Research Institute Tarnab, Peshawar, Pakistan*

[2]*Department of Statistics, University of Peshawar, Pakistan*

Key words: Forecast evaluation criteria, Diagnostics measures, ARIMA modeling, Parameter estimates, ADF test

http://dx.doi.org/10.12692/ijb/10.4.249-254 Article published on April 30, 2017

Abstract

The present study was undertaken to forecast Date's production for onward ten years in Khyber Pakhtunkhwa, Pakistan. The study was based on secondary data covering a period of about 34 years i.e. starting from 1980-81 to 2013-14. Whereas auto regressive integrated moving average (ARIMA) modeling has been employed to fit the best time series model for Dates production. It revealed through the results that for Dates production, the time series models ARIMA (3, 1, 1) was found adequate having least values of forecast evaluation criteria. Hence it can be recommended that time series models was suitable for forecasting Date's production in Khyber Pakhtunkhwa.

* **Corresponding Author:** Syed Asghar Ali Shah ✉ shah_asghar@yahoo.com

Introduction

Pakistan is blessed with vast agricultural resources on account of its fertile land, well-irrigated plains, extremes of weather, and centuries old tradition of farming. It is because of its central importance in the economy that the Government has identified agriculture as one of the four major drivers of growth. According to an estimate, the total value of agriculture crops at current factor cost is estimated at Rs.550.268 billion, divided into major crops Rs.407.623 billion and minor crops including horticulture Rs. 142.645 billion. The horticulture crops (fruits, vegetables & condiments) alone contribute Rs.116.645 billion, equivalent to US$ 2 billion, which is 26% of the total value of all crops and 81.8% of the total value of minor crops.

Date Palm (*Phoenix dactylifera* L.) belongs to the family of Palmaceae and mostly well adapted to the dry and semi dry regions of the world. Date palms like a warm climate where summers are considerably longer than winters. It is generally said that its feet should be in water and its head is in the fire.

In several regions of the world date palm or "Khajoor" in Urdu phonologic has been reflected as very old crop (Kwaasi, 2003). Date have delicious taste, contain rich amount of nutrients such as vitamin A, B, C, E, K, iron, calcium, potassium, magnesium, Phosphorus, sodium and zinc which are very important for human health (Al-Rawahi *et al*., 2005). Also, in preparation of several delicious food stuffs dates are used as significant component like sweet, finger food, sweltering products, formal feeding and health foods (Yahaya *et al*., 2015).

In the production of dates are approximately 5.4mt which are mostly produced in the sub-tropical regions, desert, semi-arid and dry areas with conducive environmental settings. Pakistan is the major producer of dates and approximately 90% of total production used up in the vicinity and also contributes approximately 10% in the export sector of the country. Every year Pakistan produces approximately 650,000 tons dates and stands 4th largest producer of dates in the world. During the year 2007-08, Pakistan sold 88,451 tons dehydrated and 4,687 tons fresh dates to other countries and got US Dollar36.033 million (EPB, 2008-09). In Pakistan, major dates producing areas are D.I. Khan, Khairpur Turbat and Gwadar,. Moreover, in Pakistan there are 300 varieties of dates in which the prominent varieties are Aseel, Karbalai, Fasliand Dhakki of Dera Ismail Khan (PHDEB, 2008). Ata *et al.* (2014) examined that date palm was the third most important source of revenue for farmers and broadly consumed in profitable and domestic products made by date palm which is the source of income generation as well as to develop the living standards of the growers.

Dates are generally associated with health foods. Peak season for date consumption is during the month of Ramadan. Entire Muslim community around the world, currently numbering 1.6 billion people is loyal consumer of dates. Consumption is also quite high during Christmas. Similarly, the fruit enjoys enormous significance on the occasion of Divali and such festivals in other religions.

Dates have found their way into sweets, confectionery, chocolates, baking products, preservatives, salads, sauces, and breakfast cereals. Dates also have bulk industrial uses. In literature several studies have been made for forecasting different trend behavior of fruits. According to Abid *et. al.* (2014) and El-Juhany (2010) who examined the status of date palm trees as well as date production in Arab countries and concluded increasing production patterns. Moreover, Farah (2012) used ARIMA modeling approach to forecast the date fruit exports in Pakistan and concluded the increasing pattern for future dates exports.

The main objective of the study was to fit an adequate model for forecasting dates production in Khyber Pakhtunkhwa, Pakistan on the basis of past trends by using ARIMA modeling technique. This study will help the date fruit producers, planners and policy makers for improving their future policies and resources as well.

Materials and methods

The present study is conducted to forecast production for onward ten years regarding Dates Production in Khyber Pakhtunkhwa, Pakistan by using time series data with effect from 1980-81 to 2013-14 i.e. time series data of 34 years. The time series data were collected from secondary sources of various issues of Fruits, Vegetables and Condiments Statistics, Crop Reporting Service of Khyber Pakhtunkhwa and were analyzed in Statistical Package Gretl1. 9.4.

Analytical Techniques

Generally, ARIMA model technique has extensively been employed in literature to forecast the specific area as well as production related to different major crops (Munir, 2008; Gujrati, 2003).

Autoregressive Integrated Moving Average (ARIMA)

This model is a generalized form of ARMA model introduced by Box and Jenkins (1976) which includes both autoregressive as well as moving average parameters, and also includes the differencing in the formulation of this model. ARIMA model is summarized as ARIMA (p, d, q). In ARIMA (p, d, q) model where p, d and q are the non-negative integers referred to as the order of the autoregressive integrated moving average process. It is an important part of Box Jenkins approach to time series modeling. It can be written as;

$$\Delta^d Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \cdots + \alpha_p Y_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \cdots + \beta_q \varepsilon_{t-q} \text{----------(1)}$$

Where, Δ^d represents differencing of order d i.e. $\Delta Y_t = Y_t - Y_{t-1}, \Delta^2 Y_t = \Delta Y_t - \Delta Y_{t-1}$ and so forth, $Y_{t-1} \ldots Y_{t-p}$ shows lags of the variables.

Stationarity test

The first step in Box- Jenkins methodology is to find whether data is stationary or not. There are a number of tests which can be used to decide about the stationary of the variables. Augmented- Dickey Fuller (1981), abbreviated as ADF, is the more popular test in literature due to its simplicity and powerfulness.

Mathematically,

$$\Delta y_t = a_0 + \lambda y_t + a_1 t + \sum_{i=1}^{p} \beta_i y_{i-1} + e_t \text{--(2)}$$

There are three options in this equation:

- a_0 is the intercept or drift parameter of the time series.
- T is the time trend in time series. There may be downward or upward linear trend in the data.
- It is also possible that both drift and time trend exist in the data.

Diagnostic Measures for Selection of Best Forecasting Model

There are few diagnostic checks which each estimated model has to fulfill and are as follows;

- Residuals are normally distributed
- The corresponding projected model is stable
- Residuals of the projected model are not serially correlated

a) The Q- Statistic:

The Q-Statistic is used to test whether the set of autocorrelation is significant i.e. diverse from zero. Box and Pierce (1970) make use of sample autocorrelation to form the statistics.

$$Q = T \sum_{k=1}^{s} r_k^2 \text{----------(3)}$$

In the hypothesis testing procedure, the null hypothesis is that every values of r_k=0, and Q has asymptotically χ^2 distributed with s degrees of freedom. Moreover, it is better to use Ljung-Box (1978) in case of small samples in support of modified Q-statistic designed as;

$$Q = T(T+2) \sum_{K=1}^{S} r_k^2 / T - K \text{----------(4)}$$

It has χ^2distribution with s degree of freedom.

b) *Jarque-bera test:*

To check the normality of residuals Jarque-Bera (1978) test is used. It is based on the fact that skewness and kurtosis of normal distribution are equals to zero. The corresponding test therefore an absolute value of these parameters and a measure of deviation from normal distribution. The Jarque-Bera statistic is calculated as follows;

$$Jarque - Bera = \frac{N-P}{6}\left[S^2 + \frac{(K-3)^2}{4}\right] \text{----------(5)}$$

Where S and K represents skew ness and kurtosis respectively, of the distribution while p denotes estimated coefficients involved in the Jarque-Bera statistic, having asymptotic χ^2distributionwith "2" degree of freedom.

Model Selection Criteria

Generally, the model selection criteria statistics are used to compare the fits of different forecasting and smoothing method and also contributes a great deal of information by comparing the fits obtained through different methods. These measures including Mean absolute percentage error (MAPE), Mean absolute deviation (MAD), Mean squared deviation (MSD). Akaike information criteria (AIC) and Schwartz information criteria (SIC). Smaller values of these accuracy measures indicate a good fitted model with minimum forecasting error (Karim *et al*. 2010).

Results and discussion

ARIMA Modeling for Dates Production

At first, it is very essential to find out the stationarity of the data for dates production. It is important to consider that for which order difference series of date production satisfies the stationarity condition.

Production satisfies the stationarity conditions. The plot of time series for date production is shown in Fig. 1.

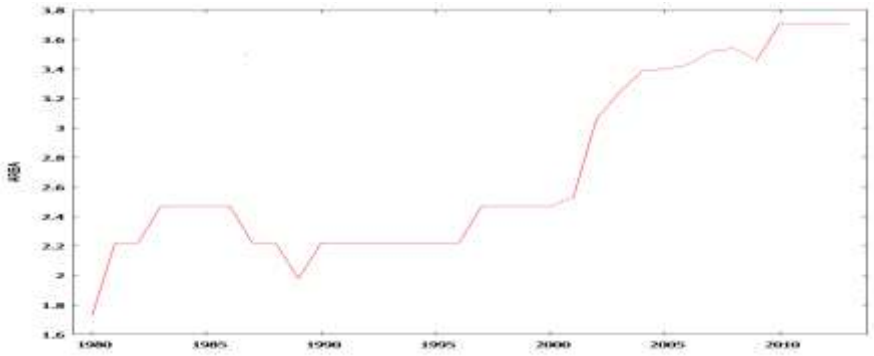

Fig. 1. Graph of original series for dates production.

From the results of Dickey fuller test show that the original series is non-stationary as there is a unit root in the data presented in Table 1.

Table 1. Augmented dickey fuller test for dates production.

Production Series	Null hypothesis	P-Value	Remarks
Original series	a=1	0.8214	Non-Stationary
Ist order difference	a=1	0.000008709	Stationary

By taking first order difference, it is found that stationarity condition is satisfied with the p-value =0.000008709, which strongly suggests that there is no unit root.

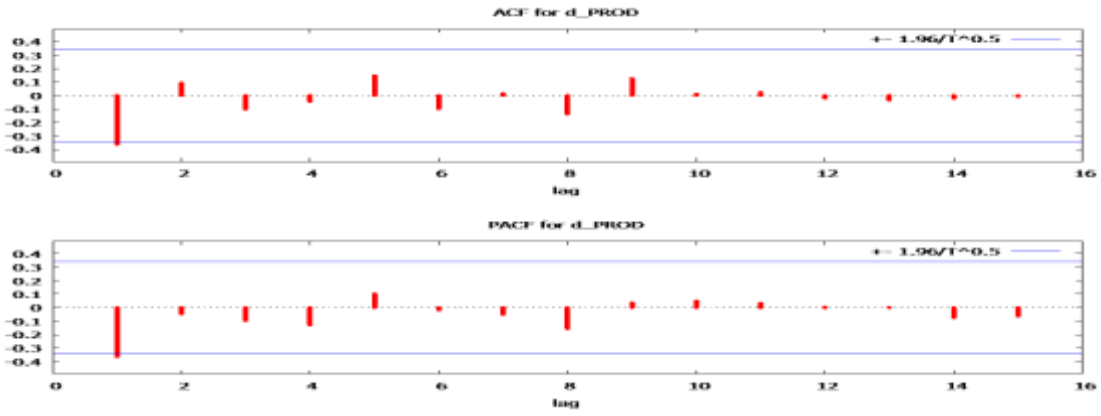

Fig. 2. ACF & PACF Plot of 1st order differenced series for dates production

From Fig. 2, it is evident that ACF & PACF of the differenced series, the adequate tentative selected ARIMA model to forecast the date's production in Khyber Pakhtunkhwa is ARIMA (3, 1, 1) at 5% level of significance. Table 2. presents model parameter estimates along with their significance.

Table 2. Model, ARIMA (3, 1, 1), using observations from 1981-2013 (t= 33). Dependent variable: 1st order difference of Dates Production.

	Coefficient	Std. Error	Z	p-value	
Const	0.172411	0.10163	1.6965	0.08980	*
phi_1	-0.298721	0.597104	-0.5003	0.61688	
phi_2	-0.117866	0.314889	-0.3743	0.70817	
phi_3	-0.245474	0.235485	-1.0424	0.29722	
theta_1	-0.146047	0.573515	-0.2547	0.79899	
Mean dependent var	0.224242		S.D. dependent var	1.212361	
Mean of innovations	0.030588		S.D. of innovations	1.079606	
Log-likelihood	-49.54379		Akaike criterion	111.0876	
Schwarz criterion	120.0666		Hannan-Quinn	114.1087	

The best selected model has smallest MAE, MSE, MAPE, AIC and SIC presented in Table 3.

Model Diagnostics

To check the auto correlation assumption, the "Ljung-Box" test is used. Test statistic gives Q' = 17.34, with p-value = 0.500, which suggests that we may accept the assumption that there is no autocorrelation among the residuals of the fitted ARIMA model at 5% level of significance.

To check the normality assumption, the Jarque Bera test is used resulting test statistic value = 1.1414, with p-value = 0.565129, which suggests that normality assumption of the residuals is valid.

Thus, it can be concluded from the formal tests that the selected model ARIMA (3, 1, 1) is an adequate model to forecast dates production in Khyber Pakhtunkhwa.

Forecast for Dates Production

The selected model is used for forecasting the date's production. In Table-1.4 the predicted values, standard errors, lower and upper confidence limits for ten years onward values are given, for Khyber Pakhtunkhwa based on the sample data.

Table 3. Forecasting criteria for best selected model in dates production.

Forecasting Criterion						
Fruit	Model	MAE	MSE	MAPE	AIC	SIC
Dates	ARIMA (3,1,1)	145.97	28873	1789.6	111.0876	120.0666

Table 4. Forecast for Dates Production.

Year	Predicted Production	Std. Error	95% interval
2014	11.15	1.080	9.03 - 13.26
2015	11.64	1.235	9.22 - 14.06
2016	11.59	1.380	8.88 - 14.29
2017	12.14	1.437	9.32 - 14.96
2018	12.15	1.551	9.11 - 15.18
2019	12.38	1.645	9.15 - 15.60
2020	12.46	1.748	9.03 - 15.88
2021	12.69	1.829	9.11 - 16.28
2022	12.84	1.912	9.10 - 16.59
2023	13.04	1.988	9.14 - 16.93

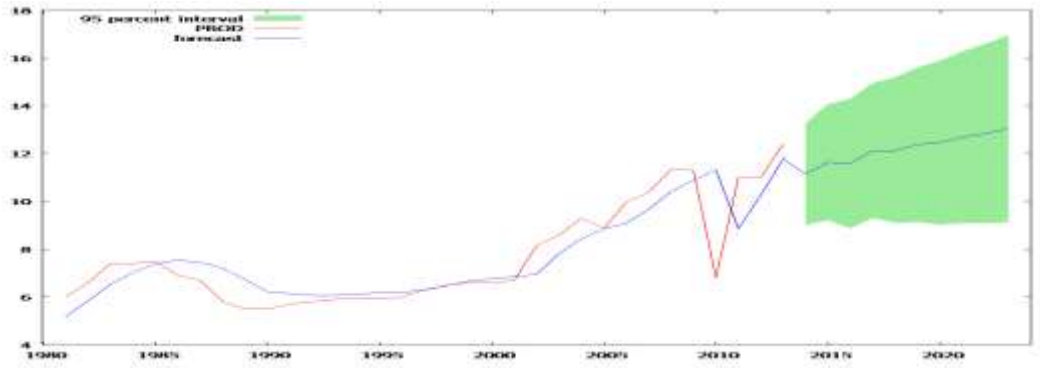

Fig. 3. Forecast plot for dates production.

By comparing the original and forecasted series it is obvious that the original series of Date's production shows increasing tendency with the passage of time and then decreases production and finally showing increasing pattern. Similarly the forecasted series shows the same pattern. In the forecasted plot in sample and out sample forecasting part is shown in Fig.-3.

Conclusions and recommendations

The instant results suggest that the time series modeling for Date's production was appropriate and gave best forecast for onward ten years. From the results of analyzed data it can be concluded that for date's production the forecasting model ARIMA (3, 1, 1) was found adequate for forecasting purpose, based on forecast evaluation criteria. Hence, it can be recommended that these selected models could be used by researchers, business men, policy makers and date fruit producers for information, planning their resources as well as decision making regarding fruit production in Khyber Pakhtunkhwa. Also, at the same time Box-Jenkins ARIMA model give good representation of short time forecasting.

References

Abid SI, Raza M, Masood A. 2014. Trend analysis and forecasting of maize area and production in Khyber Pakhtunkhwa. Pakistan. Eur. Acad. Res **2(4),** 4653-4664.

Al-Rawahi AS, Kasapis S. Al-Bulushi IM. 2005. Development of a date confectionery: Part 1. Relating formulation to instrumental texture. Intern. J. Food Properties **8(3),** 457-468.

Ata S, Shahbaz B, Khan IA, Iftikhar M. 2014. Role of date palm in livelihoods of farmers of marginal areas: A case study of South Punjab, Pakistan. J. Agric. Res **52(3),** 453-462.

El-Juhany LI. 2010. Degradation of date palm trees and date production in Arab countries: Causes and potential rehabilitation. Aust. J. Basic Appl. Sci **4(8),** 3998-4010.

Farah N. 2012. A univariate time modelling of dates exports in Pakistan J. Contemporary Issues in Business Res **1(2),** 37-48.

Karim R. Awala A, Akhter M. 2010. Forecasting of Wheat production in Bangladesh. Bangladesh J. Agric. Res **35(1),** 17-28.

Kwaasi, A. A. A. 2003. Date palm and sandstorm borne allergens. Clinical and Experimental Allergy **33(4),** 419-426.

PHDEB. 2008. Dates marketing strategy. Pakistan Horticulture Development and Export Board. p. 11-12.

Yahaya SA, Omokhudu CA, Koloce I.M, Hamza AM, Abdullahi MA. 2015. Proximate composition and fruit weight of fresh date fruits (*Phoenix dactylifera* L.) varieties in wet season of Nigeria. Pakistan J. Nutr **14(11),** 834-836.