Modeling Barley Production in Punjab

Muhammad Bilal
Department of Statistics & Computer Science
University of Veterinary & Animal Sciences, Lahore

Muhammad Qaiser Shahbaz
Department of Mathematics
COMSATS Institute of Information Technology, Lahore, Pakistan

Abstract

Barley has been an important commodity of the country. In this paper Forecast model for production of Barley in Punjab has been fitted. Methodologies for fitting of the model has been used these include ARIMA model. Diagnostic test has been carried out to see the adequacy of fitted models. Forecasted production has been obtained for coming five years.

1. Introduction

Pakistan has a rich and vast natural resource base covering various ecological and climatic zones; hence the country has great potential for producing all types of commodities. Agriculture is the hub of economic activity in Pakistan. It lays down foundation for economic development and growth of the economy. It directly contributes 25% to Gross Domestic Product (GDP) and provides employment of 44% of the total labour force of the country.

During the last five years (2000-01 to 2004-05), against the normal surface water availability at canal heads of 103.5 million acre feet (MAF), the overall (both for Kharif and Rabi) water availability has been less in the range of 5.9 percent (2003-04) to 29.4 percent (2001-02). (Source: Ministry of Food, Agriculture and Livestock; Bureau of Statistics).

Barley (Hordeum vulgare) is a major food and animal feed crop, a member of the grass family. Barley is the fifth largest cultivated cereal crop in the world (530,000 km^2 or 132 million acres). Cultivated barley is descended from wild barley, which still can be found in the Middle East. Both forms are diploid (2n=14 chromosomes). All variants of barley have fertile bastards and are thus considered to belong to one and the same species today. The major difference between wild and domesticated barley is the brittle rachis of the former, which is conductive to self-propagation. The earliest finds of barley come from Epi-Paleolithic sites the Levant beginning in the Natufien. The first domesticated barley has been found in the aceramic neolithic layers (PPN B) of Tell Abu Hureyra in Syria. The domestication seems to be contemporaneous to that of wheat.

The most proper seed season for spring barley is any time in March or April, though good crops produced have seen, the seed of which was sown at a much later period. Barley is widely adaptable and is currently a major crop of the temperate and tropical areas.

Table 1.1: Area Production of other Major Kharif and Rabi Crops

	2003-04		2004-05(P)		
Crops	Area (000 hectares)	Production (000 tons)	Area (000 hectares	Production (000 tons	% Change In production
KHARIF					
Maize	947	1897	896	2775	46.3
Bajra	539	274	343	193	-29.6
Jawar	392	238	308	186	-21.8
RABI					
Gram	982	611	1038	761	24.5
Barley	102	98	97	96	-2.0
Rapeseed & Mustard	259	221	244	227	2.7
Tobacco	46	86	45	84	-2.3

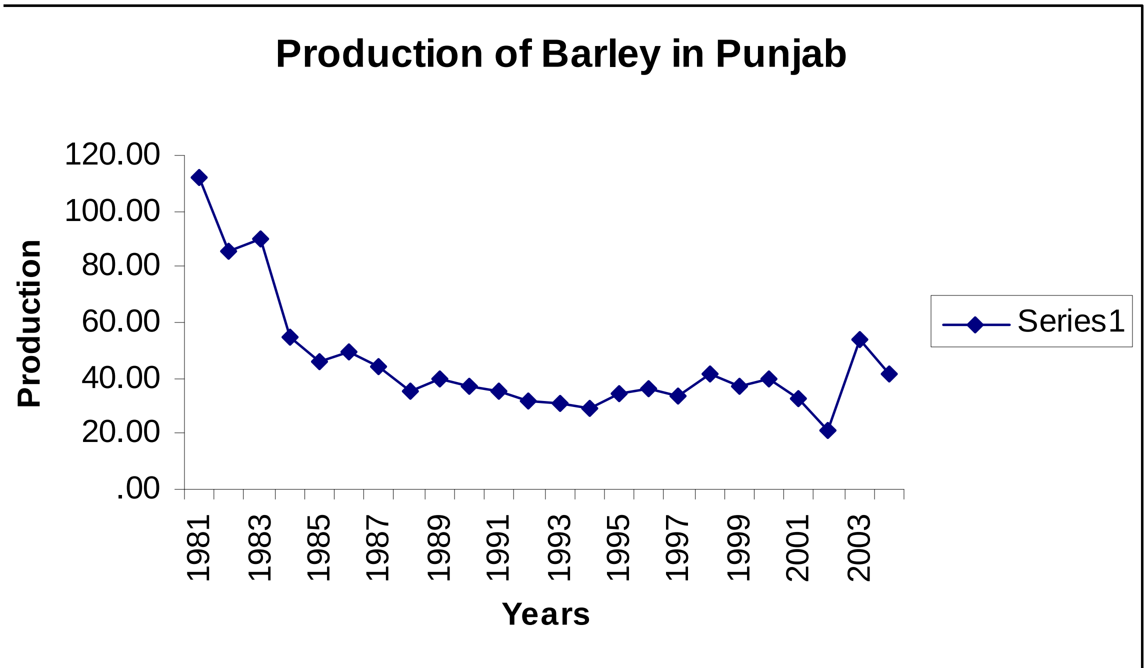
P=Provisional (July-March), Source: Ministry of Food, Agriculture and Livestock; Federal Bureau of Statistics.

Table 1.2: Agriculture growth in Pakistan (1990-2003)

YEAR	*AGRICULTURE*	*MINOR CROPS*
1990-91	*4.96*	*3.51*
1991-92	*9.50*	*2.37*
1992-93	*-5.29*	*3.95*
1993-94	*5.23*	*12.62*
1994-95	*6.57*	*6.91*
1995-96	*11.72*	*4.89*
1996-97	*0.12*	*0.94*
1997-98	*4.52*	*8.13*
1998-99	*1.95*	*4.23*
1999-00	*6.09*	*-9.10*
2000-01	*-2.64*	*0.11*
2001-02	*-0.07*	*-1.82*
2002-03	*4.15*	*0.41*

P=Provisional (July-March), Source: Ministry of Food, Agriculture and Livestock; Federal Bureau of Statistics.

Figure 1: Trend of Production of Barley in Punjab with different Time periods



2. Methodology

2.1 Stationarity Test for Barley

The stationarity of the data has been checked by using the Unit Root test the results are given below:

Table 2.1: Stationarity Summary

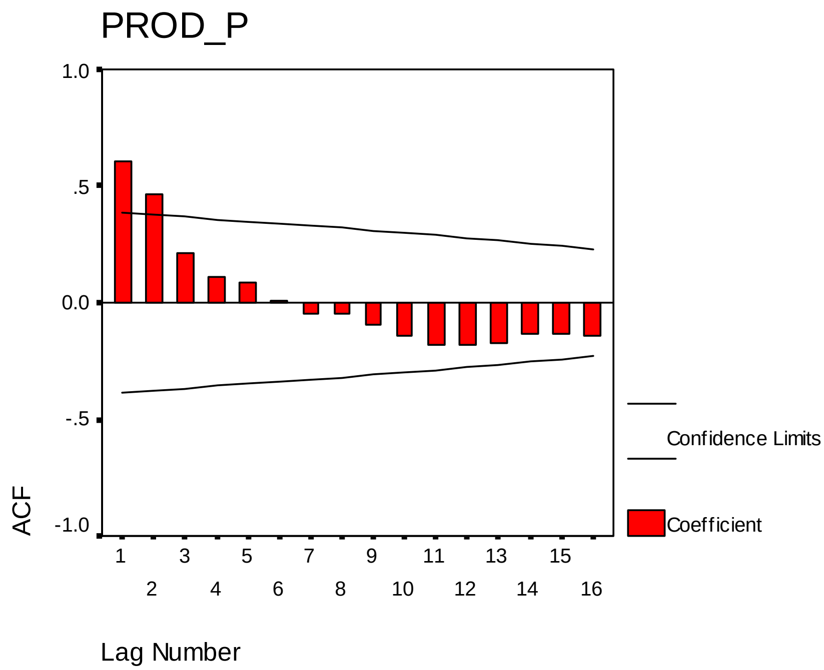
		Punjab
Level	Co-efficient	-0.45461
	P-value	0.0012
First difference	Co-efficient	
	P-value	

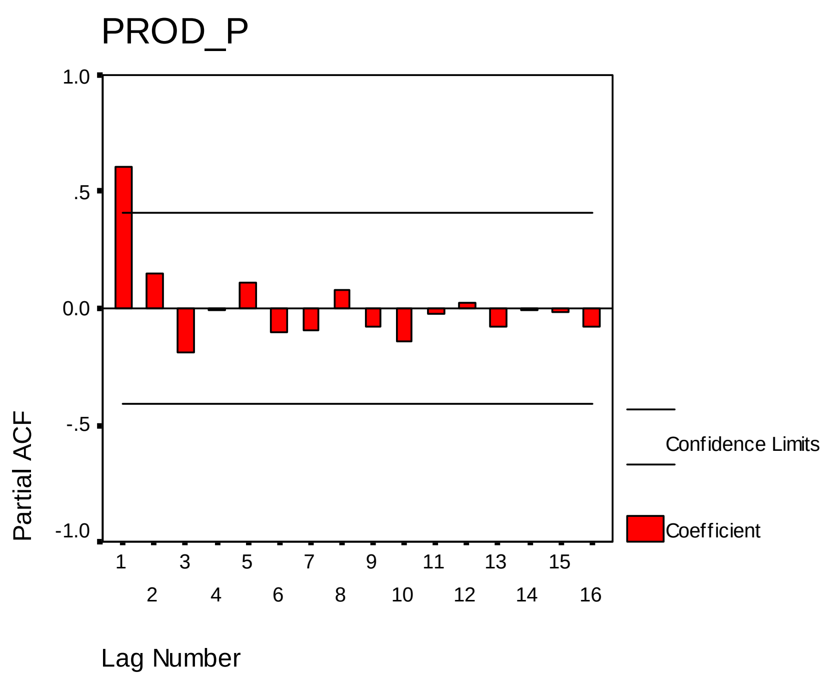
From the above table it is concluded that production of Barley in Punjab become stationary at the level, so ARIMA model with d at "0" will be used.

2.2 ACF and PACF for Production of Barley in Punjab

For the data of production of Barley for Punjab the data is stationary at level so the "ACF" and "PACF" of original data are plotted and from the plot it is observed that for this the value of "p" and "q" are respectively "1", "1".

Figure 2: ACF and PACF For Production of Barley in Punjab





2.3 Auto-Regressive Integrated Moving Average (ARIMA) Model

The ARMA (p, q) model applied on the ‘d’ differences of Y_t is called Auto-Regressive Integrated Moving Average (ARIMA) Model. It is denoted by ARIMA (p, d, and q). Where “p” is the order of AR process, “q” is the order of MA process and “d” is the order of differencing.

The ARMA models are generalization of the simple AR model that uses three tools for modeling series correlation in the disturbance.

The model can also be checked for adequacy by doing a chi-square test, known as the Box-Pierce Q statistic, on the autocorrelations of the residuals. The test statistic is:

$$Q = \left(N - d\right)\sum_{k=1}^{m} r_k^2 \qquad (1)$$

Which is approximately distributed as a chi-square variate with “k-p-q” degree of freedom. In this equation

N = length of the time series.
K = First k autocorrelation being checked.
M = Maximum no. of lags checked.
r_k = Sample autocorrelation function of the kth residual term.
d = Degree of differencing to obtain a stationary series.

If the calculated value of Q is larger than the chi-square for k-p-q degree of freedom, the model should have been considered inadequate. It is possible that two or more models have been judge to be approximate, yet none of the models may be an exact fit for the data. In this case, the principle of parsimony should prevail, and simpler model should have chosen.

2.4 ARMA Models

These are the mixture of AR and MA process or models. The time series x_t for t = $0, \pm 1, \pm 2, \pm$.............. is said to be ARMA (p, q) if x_t is stationary and

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + ... + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + ... + \theta_q w_{t-q} \qquad (2)$$

With

$$\phi_p \neq 0 \quad and \quad \theta_q \neq 0 \quad and \qquad \sigma_w^2 \succ 0$$

The parameter p and q are called the auto regressive and moving average orders. If x_t has non-zero mean μ then ARMA (p, q) can be written as given bellow

$$x_t = \alpha + \phi_1 x_{t-1} + \phi_2 x_{t-2} + ... + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + ... + \theta_q w_{t-q} \qquad (3)$$

The ARMA models become AR if q=0 and if p=0 these become MA models. The ARMA models can be written as bellow:

$$\phi(B)x_t = \theta(B)w_t \quad (4)$$

2.5 ARIMA Models

ARIMA stands for Auto-Regressive Integrated Moving Average. These models deal with non-stationary time series, while ARMA (p, q), AR (p) and MA (q) models are used to deals with second order stationary time series. By using different operation on non-stationary population the population becomes stationary. The ARIMA (p, d, q) models assume that the d^{th} difference

$$\nabla^d y_t = (1-B)^d y_t \quad (5)$$

is a stationary ARMA (p, q) process.

3. Analysis

The analysis of the data has been carried out in order to obtain suitable model for forecasting production of Barley in Punjab. The data is used for this purpose is from 1981 to 2004. The data is taken from "fifty years of Pakistan in statistics" and "statistical year book". The production is taken to be dependent variable, and Area, Temperature and Rain are taken to be independent variables.

Table 3.1: ARIMA Model Parameters Table for Barley

AR(1)	0.8243
MA(1)	-0.2163
Area	0.56191
Temperature	4.0725
Rain	0.01425
Constant	44.173

The P- values and standard error values of the co-efficients of ARIMA Model for Production of Barley in Punjab are given below:

Table 3.2: P-values and Standard Error ARIMA Model co-efficients Table for Barley

	AR1	MA1	Area	Temperature	Rain	Constant
P – value	0.00100	0.49262	0.0000	0.095	0.234	0.116
Standard Error	0.2103	0.3088	0.1030	2.4376	0.0119	61.6295

ARIMA Model for Barley is

$$\phi(B)Y_{Be,t} = \theta(B)z_t + \sum_{j=1}^{k} B_j X_{Be,jt} + \varepsilon_t \qquad (6)$$

The estimated ARIMA Model for Production of Punjab is

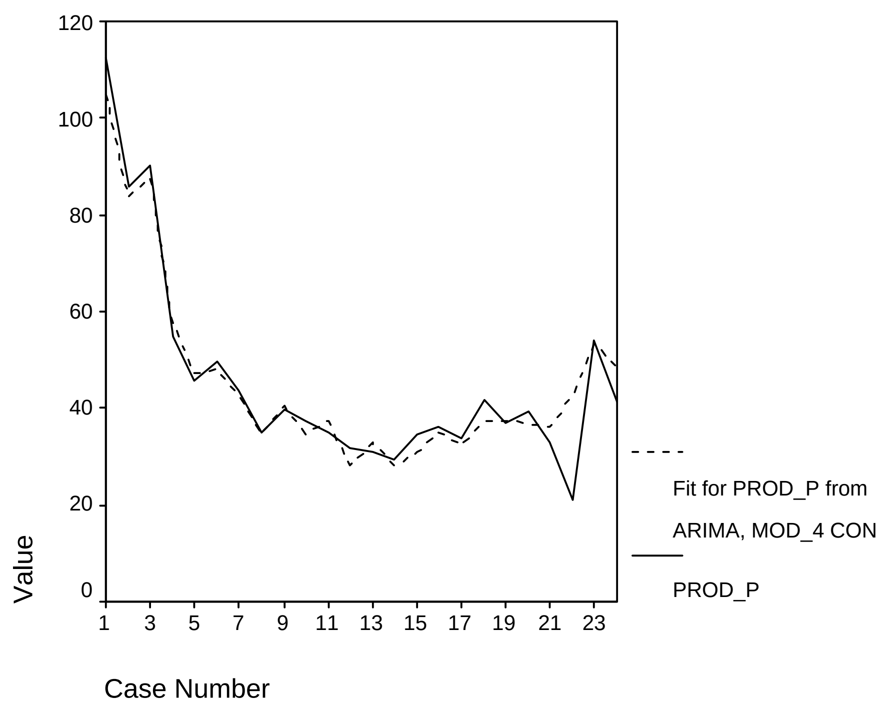
$$Y_{be,t} = 44.173 + 0.8243y_{t-1} + 0.5619X_{Abe,pt} + 4.0725X_{Tbe,pt} + 0.01425X_{Rbe,pt} - 0.2163z_{t-1} \qquad (7)$$

The Wald chi–square test is 31.40 with a p–value of 0.000, indicating that the overall model for Punjab is significant and therefore it can be used to forecast the production of Barley in the province.

From the (Table-3.2) Model it can be easily seen that the coefficients of Area is significant and Temperature and Rain are insignificant.

The graphs of fitted values from the models along with the actual values are given below:

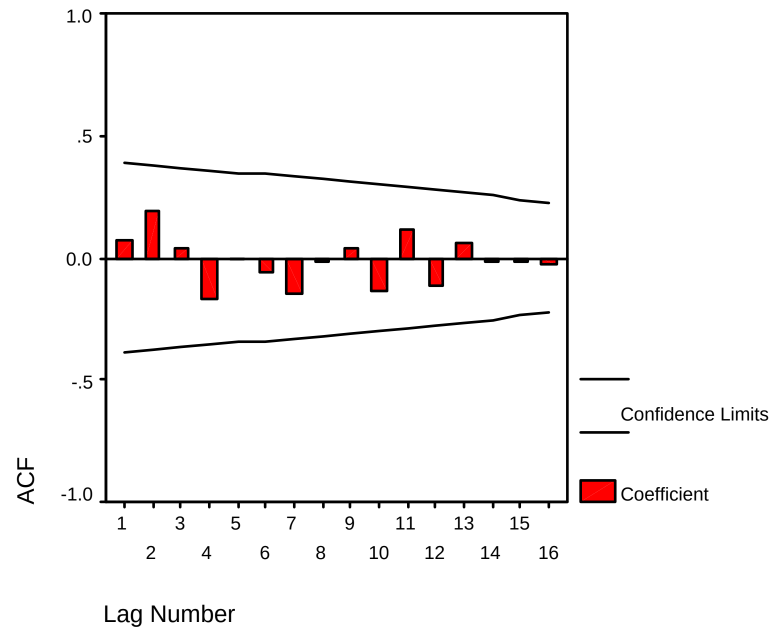
Figure 3.1: Fitted and Actual values for Production of Barley in Punjab

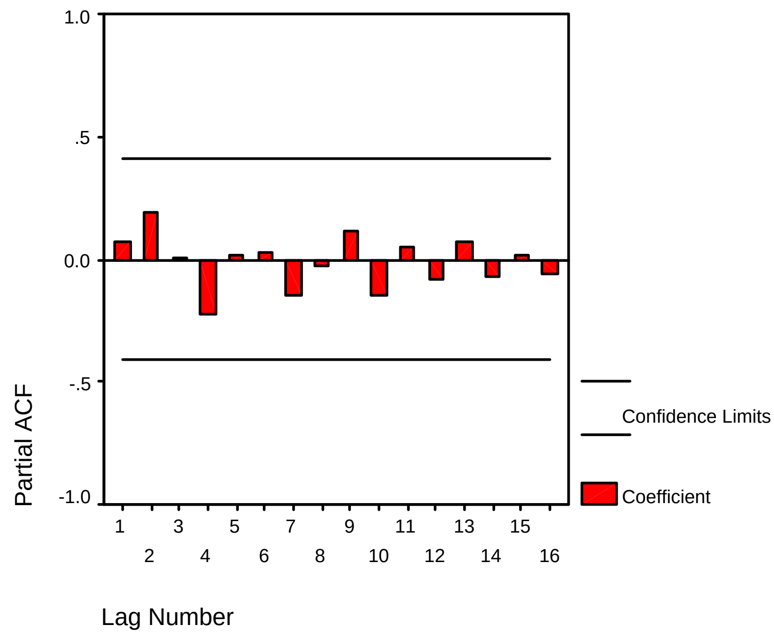


From the above graph it can be easily seen that the fitted and actual values are close to each other.

Further, obtaining the ACF and PACF of the residuals has plotted for diagnostics of the residuals after the fitted model.

Figure 3.2: Figures of ACF and PACF of Error of Barley in Punjab





The above graph shows that the residuals after fitted model are not auto correlated.

Also the Kolmogrov Smirnov test has been carried out at the errors to test the normality of the errors. The p-value of errors of Punjab is 0.21 and also which is shown (Figure-3.2), that the errors for fitted model are normal. From these diagnostic tests it can be concluded that the fitted model is adequate for forecasting of Production of Barley in Punjab.

References

1. 50 Years of Pakistan in statistics, Federal Bureau of Statistics-Statistics Division Government of Pakistan.
2. Akaike, H. (1969), Fitting auto regressions for predictions, Annals of the Institute of Statistical Mathematics, Tokyo.
3. Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, 2nd International Symposium on Information Theory, B.N. Petrov and F. Csaki(eds.), Akademiai Kiado, Budapest.
4. Akaike, H. (1978). Time series analysis and control through parametric models, Applied Time Series Analysis, D.F. Findley (ed.), Academic Press, New York.
5. Anderson, T. W. (1971). The Statistical Analysis of Time Series, John Wiley, New York.
6. Anderson, T. W. (1980). Maximum likelihood estimation for vector autoregressive moving average models, Direction in time series, D.R. Brillinger and G.C. Tiao (eds.), Institute of Mathematics Statistical
7. Box, G. E. P. and Jenkins, G. M. (1970). Time Series Analysis: Forecasting and Control, Holden-Day, San Francisco.
8. Box, G. E. P. and Pierce, D. A. (1970). Distribution of residual autocorrelations in auto-regressive integrated moving average time series models. J. Amer. Statist. Assoc.
9. Brillinger, D. R. (1981). Time Series Analysis: Data Analysis and Theory, Holt, Rinehart & Winston, New York.
10. Brockwell, P. J. and Davis, R. A. (1985). Applications of a recursive prediction algorithm in time series analysis, Statistics Dept., Colorado State University.
11. Chatfield, C. (1984). The Analysis of Time Series: An introduction, 3rd ed., Chapman and Hall, London.
12. Economic Surveys of Pakistan for the years 1962 to 2001.
13. Ehsan Ahmad and Rosser J. B. (1995). Non-linear speculation Bubbles in the Pakistani stock Market, The Pakistan Development Review 34:1, 25-41.
14. Fuller, W. A. (1976). Introduction to Statistical Time Series, John Wiley, New York.
15. Kinal, T. and Lahiri, K. (1988). A model for Ex Ante Real Interest Rates and Derived Inflation Forecasts, Journal of the American Statistical Association, Vol.83, No. 403.
16. Ljung, G. M. and Box, G. E. P. (1978). On a measure of lack of a fit in time series models, Biomertika.
17. Walker, A. M. (1964). Asymptotic properties of least squares estimates of the parameters of the spectrum of a stationary non-deterministic time series, J. Aust. Math. Soc.