Journal of Biology, Agriculture and Healthcare
ISSN 2224-3208 (Paper) ISSN 2225-093X (Online) DOI: 10.7176/JBAH
Vol.9, No.16, 2019 

# Regression Analysis for the Forecasting of Production and Yield of Wheat Crop in Sindh and Punjab.

A. FAHEEM. MALIK, A. A. SHAIKH, A. SALEEM. MEMON

Department of Basic Science and Related Studies, Mehran University of Engineering and Technology, Jamshoro

**ABSTRACT**
Wheat is an important and major agriculture crop of Pakistan. The purpose of this study is to fit the Simple regression model and to forecast the value of each indicator for planning purpose. The data set of 30 years (i.e. 1988 – 2018) for Punjab Wheat and Sindh Wheat are collected from Pakistan Bureau of Statistics, Agricultural Statistics of Pakistan and internet also investigated. Econometrics techniques (Trend Curves, Lagged Models, Simple Regression, Correlation and Moving Averages) are applied using Minitab software and MS Excel and observed that: The changes in all indicators with respect to time are positive. The changes in Production for Sindh are less than for Punjab which shows better consistency towards Production of Sindh than that of Punjab. The changes in Yield for Punjab Wheat are larger than for Sindh Wheat. Yield of Wheat Sindh is more consistent than Wheat Punjab. For Production, Poly-2(PP), Poly-6(PS) Exponential and lagged-1 models are preferred due to better results.
**Key words:** Simple Regression, Exponential Model, Lagged Model, Data, Production, Forecasting, etc.
**DOI**: 10.7176/JBAH/9-16-06
**Publication date**: August 31st 2019

## 1. INTRODUCTION

Econometricians & Statisticians have tried to develop models for Production, Price and Yield of different crops in their countries. Some of them used Regression Analysis to develop forecasting models of different crops with the help of past data depending upon the constraints like availability of land, type of land, water, climate, temperature, etc. I have considered only Production and Yield of Wheat crop in two main provinces Punjab and Sindh which are playing a vital role for providing wheat in Pakistan. I have forecasted the value for production and yield of wheat crop in both provinces using simple regression, exponential, lagged – 1 and Moving Averages and observed above key points. Some of the econometricians had considered only Production. Production indicators help to ensure an adequate supply of food and fiber, whereas and Yield indicators help to determine the numerical value of the effects of technological developments. Econometrics is the study of economics theory with economics statistics and attempts using mathematical and statistical methods to explore the actual support of the common schematic law settled by economic theory. According to Gerhard Tinter Econometrics is the result of a definite outlook on the economics role, contains of the applications of mathematical statistics to economical data to advance actual support to the models formed by mathematical economics for getting the results which are numerical.

## 2. METHODOLOGY

Index Number, Ratios, Growth Rates, Moving Averages, Trend Curves (Polynomial of different order and Exponential Curves), Lagged regression, i.e. regression of variable with its previous values, Correlation, Linear Regression will be used. Regression Analysis Curve Fitting (Trend Curves) As the scatter diagram shows that the time series data for Production and Yield but not always genius of parabola, cubic equation and polynomial is up to 5 degree and exponential precipitate using time as the trend variable. The curve with largest adjusted $R^2$ square and maximum number of statistically significant Coefficient (using t-test at 5% level) is taken as the best fit. Value of t-test, F-test, size and sign of the estimated parameters is given due weight age. These are the formulas for regression analysis from linear to 5 degree polynomials:
Linear Polynomial $y = a + bt$, Quadratic Polynomial $y = a + bt + ct^2$, Cubic Polynomial $y = a + bt + ct^2 + dt^3$, Bi-quadratic Polynomial $y = a + bt + ct^2 + dt^3 + et^4$, 5th Degree Polynomial $y = a + bt + ct^2 + dt^3 + et^4 + ft^5$, 6th Degree Polynomial $y = a + bt + ct^2 + dt^3 + et^4 + ft^5 + gt^6$, Exponential Model $y = ae^{bt}$.

**Lag Linear Model** One period lag linear model will be fitted for each indicator. For lagged - 1 model data for an indicator of crop for 30 years (i.e. from 1988 - 2018) in taken as regressand and its data for 30 years from 1988 - 2018, i.e. its values is taken as regressor. It works as a simple linear model. Lagged -1 Model $y(t) = a + by_{t-1}$.

**Moving Averages**

(3 years), (5 years) and (7 years) moving averages for Production and Yield are calculated. The period of moving averages is chosen in a way that the period over which fluctuations occur is covered.

**Forecasting**

(a) Substitute the given value of the explanatory variable in the estimated model and obtained corresponding $\hat{y}$.

(b) Calculate the absolute percentage error in y as $\left(\frac{\text{actual value} - \text{forecasting value}}{\text{actual value}} * 100\right)$

(c) Find 95% confidence interval for prediction $y_F \pm t_{0.025}\ S_F$ where t has n-2 d.f.

## 3. RESULTS AND DISCUSSION

### Fitting Trend Curves

### Production

The results of fitting Trend Curves, i.e. Poly-1, Poly-2, …, Poly-6 and Exponential Models, on the data for Production of two different commodities are presented in corresponding Trend Tables. The Production in (000) tones is taken as regressand and time is taken as regressor. Regression Equation, Values of t-test, Adj $R^2$ , F-test , p , d-test, r or $\sqrt{R^2}$ , number of significant and insignificant coefficients, number of estimated parameters with wrong signs, etc, for each model, are presented in relevant tables.

**Trend Table.1 (Production)**
**Econometric Models For Wheat Punjab (1988-2018)**

| Curve | Regression Equation | Adj. $R^2$ | F/P Values | r/ $\sqrt{R^2}$ | Sig | Insig | Neg | Remarks |
|---|---|---|---|---|---|---|---|---|
| 1. Poly -1 1.701 | y = 10019.266 + 355.659 t<br>(33.57) (21.16) | 0.939 | 447.57 | 0.970 | 2 | 0 | 0 | Better |
| 2. Ploy -2 1.703 | y = 9411.917 + 469.537 t – 3.673 $t^2$<br>(20.87) (7.00) (-1.75) | 0.943 | 241.80 | 0.973 | 3 | 0 | 1 | Best |
| 3. Ploy -3 1.706 | y = 9825.587 + 321.41 t + 8.1 $t^2$ – 0.253 $t^3$<br>(15.30) (1.82) (0.62)* (-0.91)* | 0.943 | 160.43 | 0.947 | 2 | 2 | 1 | W.M |
| 4. Ploy -4 1.708 | y = 10290.362 + 59.924 t + 44.604 $t^2$<br>(11.71) (0.16)* (0.92)*<br>– 2.065 $t^3$ + 0.029 $t^4$<br>(-0.88)* (0.78)* | 0.942 | 118.67 | 0.975 | 1 | 4 | 1 | I.V |
| 5. Expon-ential | y = 10504.612 $e^{0.0238\ t}$<br>(2.62) (1.28)* | 0.926 | 363.41 | 0.964 | 1 | 1 | 0 | W.M |
| 6. Lagged-1 | $y(t) = 1658.506 + 0.913\ y_{(t-1)}$<br>(1.49)* (12.85) | 0.854 | 165.20 | 0.927 | 1 | 1/ 2 | 0 | Good |

**Trend Table.2 (Production)**

**Econometric Models For Wheat Sindh (1988-2018)**

| Curve | Regression Equation | Adj. $R^2$ | F/P Values | r/ $\sqrt{R^2}$ | Sig | Insig | Neg | Remarks |
|---|---|---|---|---|---|---|---|---|
| 1. Poly -1 1.701 | $y = 1864.174 + 68.313\ t$ (12.85) (8.36) | 0.704 | 69.94 | 0.854 | 2 | 0 | 0 | Good |
| 2. Ploy -3 1.706 | $y = 2641.9 - 155.2\ t + 14.89\ t^2 - 0.278\ t^3$ (9.20) (-1.97) (2.54) (-2.24) | 0.765 | 32.48 | 0.888 | 4 | 0 | 2 | Better |
| 3. Ploy -4 1.708 | $y = 1940.040 + 239.653\ t - 40.266\ t^2$ (5.72) (1.64)* (-2.15) $+2.458\ t^3 - 0.044\ t^4$ (2.72) (-3.04) | 0.822 | 34.49 | 0.920 | 4 | 1 | 2 | W.M |
| 4. Ploy -6 1.714 | $y = 3046.765 - 871.819\ t + 279.599\ t^2$ (5.66) (-1.98) (2.41) $- 36.568\ t^3 + 2.247\ t^4 - 0.064\ t^5 + 0.0007\ t^6$ (-2.69) (2.86) (-2.92) (2.91) | 0.859 | 30.45 | 0.942 | 7 | 0 | 3 | Best |
| 5. Exponential | $y = 1993.38\ e^{0.0229\ t}$ (2.18) (0.91)* | 0.694 | 66.69 | 0.839 | 1 | 1 | 0 | W.M |
| 6. Lagged-1 | $y(t) = 300.031 + 0.911\ y_{(t-1)}$ (1.24)* (11.20) | 0.816 | 125.54 | 0.907 | 1 | 1/2 | 0 | W.M |

* **Indicates:-**

- **Insignificant coefficient at 5% level, in regression equation.**
- **Failure of F-test or that the model is invalid in F/P values**
- **Very low value of $\sqrt{R^2}$**

**YIELD**

The results of fitting Trend Curves are presented in corresponding Trend Tables. Here Yield (kgs per hectare) is taken as regressand and time (in years) is taken as the regressor.

**Trend Table.3 (Yield)**

**Econometric Models For Wheat Punjab (1988-2018)**

| Curve | Regression Equation | Adj. $R^2$ | F/P Values | r/ $\sqrt{R^2}$ | Sig | Insig | Neg | Remarks |
|---|---|---|---|---|---|---|---|---|
| 1. Poly -1 1.701 | $y = 1841.504 + 39.772\ t$ (41.31) (15.84) | 0.896 | 250.86 | 0.948 | 2 | 0 | 0 | Better |
| 2. Ploy -2 1.703 | $y = 1738.259 + 59.131\ t - 0.624\ t^2$ (26.25) (6.00) (-2.03) | 0.906 | 141.39 | 0.955 | 3 | 0 | 1 | Best |
| 3. Ploy -3 1.706 | $y = 1709.9 + 69.3\ t - 1.43\ t^2 + 0.017\ t^3$ (17.91) (2.64) (-0.73)* (0.42)* | 0.903 | 91.44 | 0.956 | 2 | 2 | 1 | I.V |
| 4. Ploy -4 1.708 | $y = 1876.184 - 24.277\ t + 11.640\ t^2$ (15.28) (-0.46)* (1.71) $-0.631\ t^3 + 0.010\ t^4$ (-1.93) (2.00) | 0.913 | 77.50 | 0.962 | 4 | 1 | 2 | Good |
| 5. Exponential | $y = 1877.154\ e^{0.0166\ t}$ (2.58) (1.18)* | 0.886 | 227.10 | 0.943 | 1 | 1 | 0 | W.M |
| 6. Lagged-1 | $y(t) = 324.853 + 0.883\ y_{(t-1)}$ (1.49)* (9.97) | 0.778 | 99.39 | 0.887 | 1 | 1/2 | 0 | W.M |

Journal of Biology, Agriculture and Healthcare
ISSN 2224-3208 (Paper) ISSN 2225-093X (Online) DOI: 10.7176/JBAH
Vol.9, No.16, 2019 

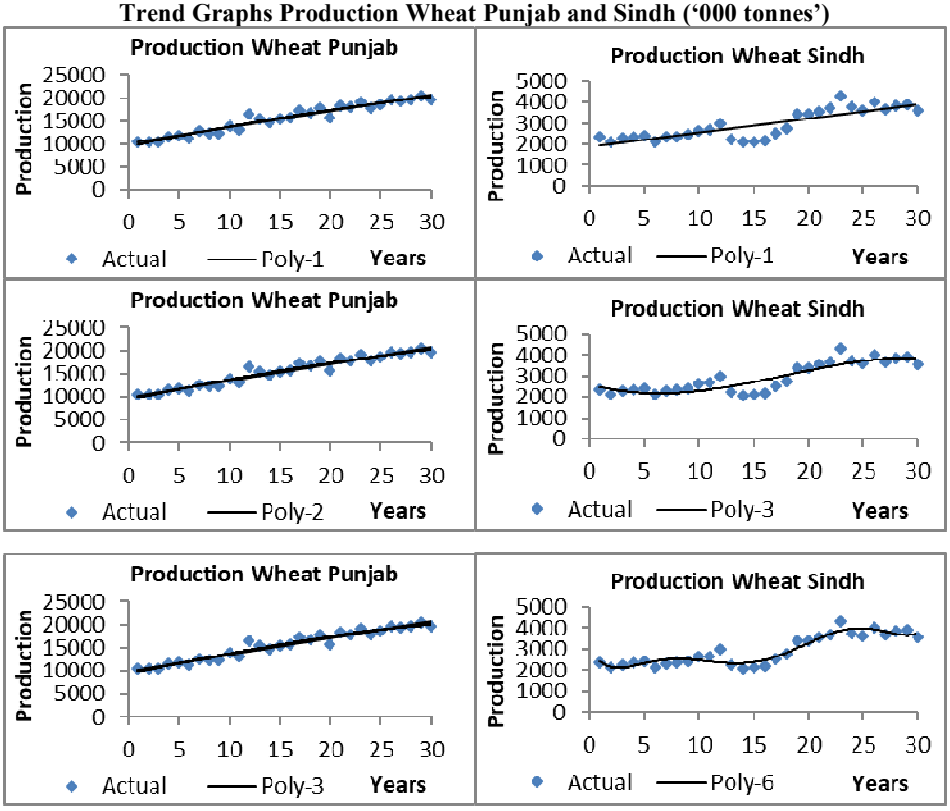
**Trend Table.4 (Yield)**
**Econometric Models For Wheat Sindh (1988-2018)**

| Curve | Regression Equation | Adj. $R^2$ | F/P Values | r/ $\sqrt{R^2}$ | Sig | Insig | Neg | Remarks |
|---|---|---|---|---|---|---|---|---|
| 1. Poly -1 1.701 | $y = 1865.787 + 59.907\ t$ (19.40) (11.06) | 0.807 | 122.32 | 0.902 | 2 | 0 | 0 | Better |
| 2. Ploy -2 1.703 | $y = 1831.313 + 66.371\ t - 0.209\ t^2$ (11.96) (2.92) (-0.29)* | 0.801 | 59.21 | 0.902 | 2 | 1 | 1 | Good |
| 3. Ploy -3 1.706 | $y = 2429.8 - 147.9\ t + 16.8\ t^2 - 0.366\ t^3$ (16.52) (-3.66) (5.59) (-5.74) | 0.909 | 97.08 | 0.958 | 4 | 0 | 2 | Best |
| 4. Poly -6 1.714 | $y = 2571.134 - 407.390\ t + 117.428\ t^2$ (7.85) (-1.52)* (1.66)* $-14.859\ t^3 + 0.941\ t^4 - 0.028\ t^5 + 0.0003\ t^6$ (-1.79) (1.97) (-2.11) (2.20) | 0.923 | 58.59 | 0.969 | 5 | 2 | 3 | W.M |
| 5. Expon-ential | $y = 1950.743\ e^{0.0218\ t}$ (2.35) (1.06)* | 0.820 | 133.22 | 0.909 | 1 | 1 | 0 | W.M |
| 6. Lagged-1 | $y(t) = 202.357 + 0.940\ y_{(t-1)}$ (1.11)* (14.60) | 0.883 | 213.42 | 0.942 | 1 | 1/2 | 0 | W.M |

* **Indicates:-**

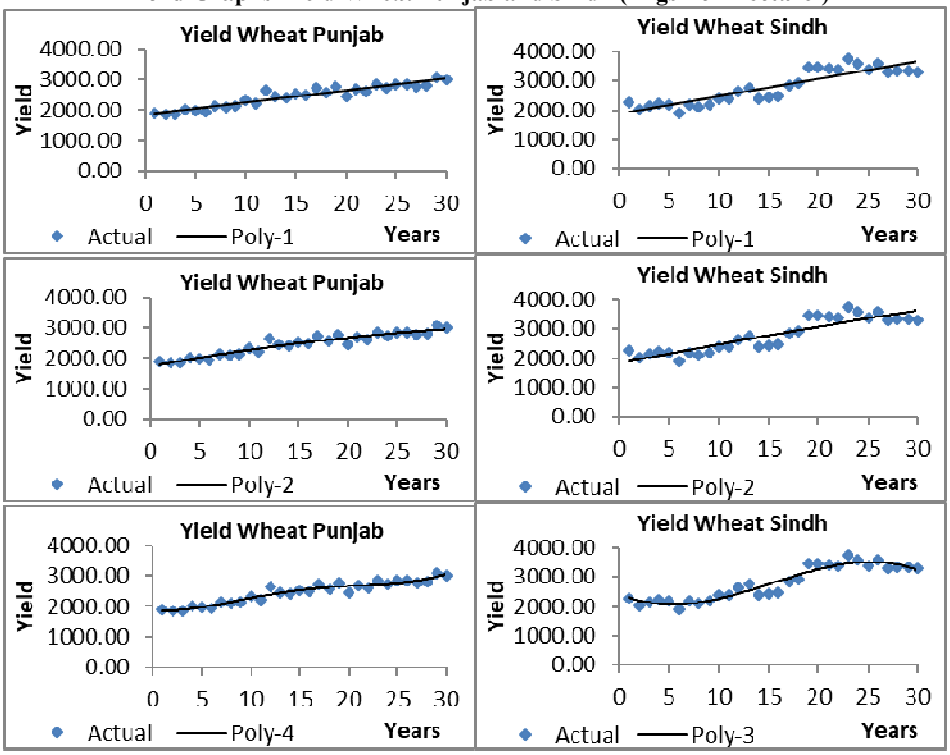
- **Insignificant coefficient at 5% level, in regression equation.**
- **Failure of F-test or that the model is invalid in F/P values**
- **Very low value of $\sqrt{R^2}$**

**Trend Graphs Production Wheat Punjab and Sindh ('000 tonnes')**



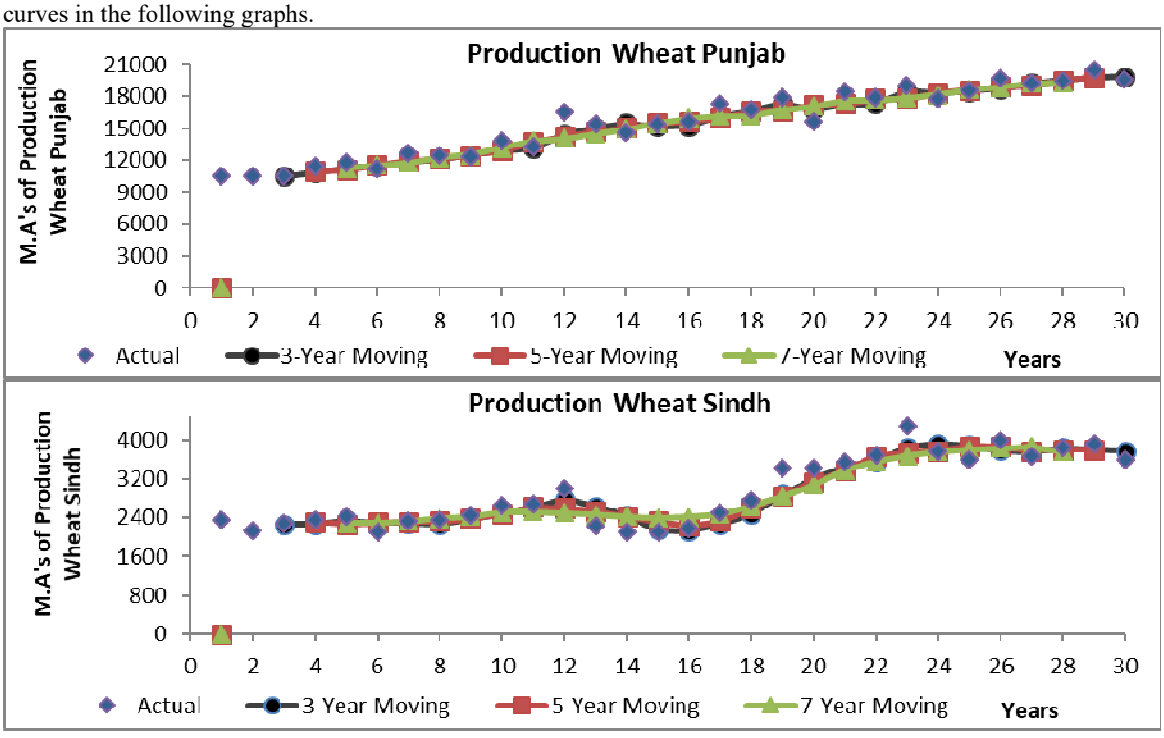
 

**Trend Graphs Yield Wheat Punjab and Sindh ('Kgs Per Hectare')**

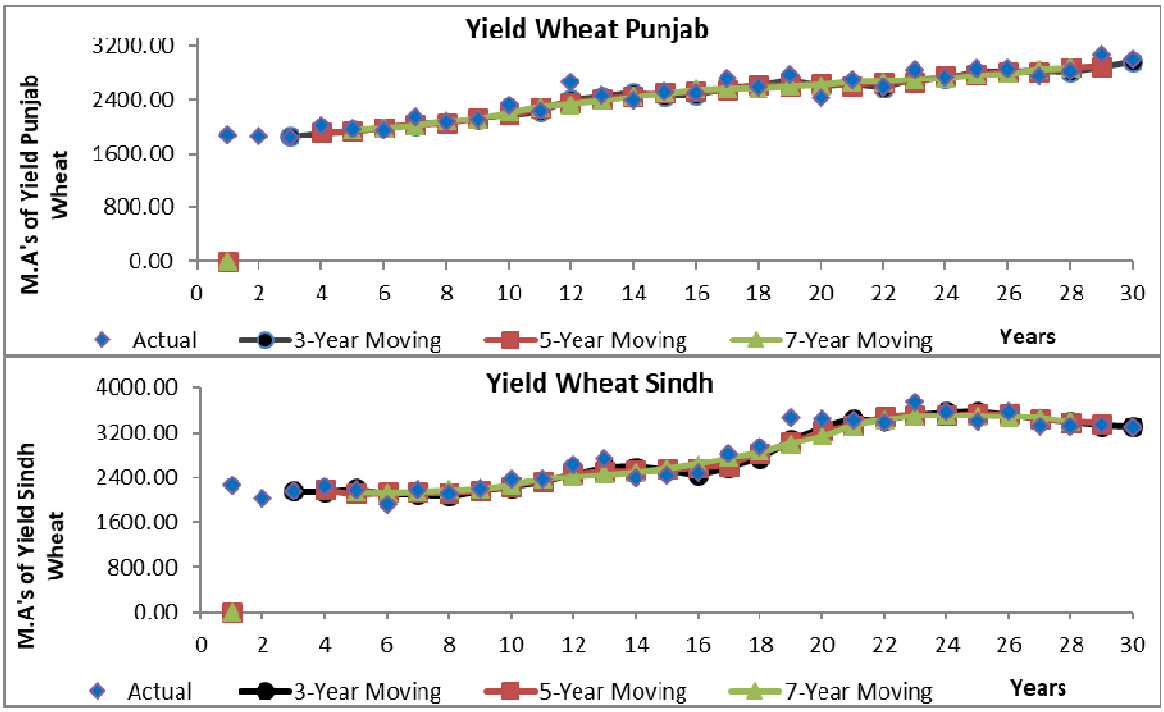



**MOVING AVERAGES (Production)**

The results of 3-years, 5-years and 7-years Moving Averages Trends on the data for Production are shown as curves in the following graphs.






**Forecasting Using Trend Curves And Moving Averages**

The results for forecasting by the regression models including forecasting intervals (95% Class Intervals), percentage errors from true values for year 2019 extrapolation by Moving Averages and ranking (including Adj $R^2$ , etc) are presented in the corresponding tables and discussed here:

**Forecasting Table Production Wheat Punjab (2018, 19607)**

| Model/ Methods | AdjR²/ C.V | No of Insig Coef: | Forecast-ing Year 2019 | Forecasting Intervals Year 2019 (95%) | % error | Comments | |
|---|---|---|---|---|---|---|---|
| | | | | | | Rank By Forecast | Rank By Regression |
| **Poly-1** | 0.939 | 0 | 21044.7 | 19412.1 – 22677.2 | 7.33 | 5th | 2nd |
| **Poly-2** | 0.943 | 0 | 20437.3 | 18859.1 – 22015.6 | 4.23 | 4th | 1st |
| **Poly-3** | 0.943 | 2 | 20023.7 | 18445.4 – 21601.9 | 2.13 | 3rd | W.M |
| **Expon:** | 0.926 | 1 | 21728.9 | 20521.9 – 22935.9 | 10.82 | 6th | W.M |
| **Lagged-1** | 0.854 | 1/2 | 19552.1 | 17093.0 – 22011.1 | 0.28 | 1st | 3rd |
| **M.A 7-Years** | 0.17 | _ | 19866.7 19526.7 | 18774.2 – 20959.1 | 1.32 | 2nd | - |

**Forecasting Table Production Wheat Sindh (2018, 3582.1)**

| Model/ Methods | AdjR²/ C.V | No of Insig Coef: | Forecast-ing Year 2019 | Forecasting Intervals Year 2019 (95%) | % error | Comments | |
|---|---|---|---|---|---|---|---|
| | | | | | | Rank By Forecast | Rank By Regression |
| **Poly-1** | 0.704 | 0 | 3981.9 | 3188.6 – 4775.2 | 11.16 | 4th | 3rd |
| **Poly-3** | 0.765 | 0 | 3848.5 | 3143.0 – 4554.0 | 7.44 | 3rd | 2nd |
| **Poly-6** | 0.859 | 0 | 4115.4 | 3562.6 – 4668.2 | 14.89 | 6th | 1st |
| **Expon:** | 0.694 | 1 | 4097.5 | 3831.3 – 4363.6 | 14.39 | 5th | W.M |
| **Lagged-1** | 0.816 | 1/2 | 3563.6 | 2934.0 – 4193.1 | 0.52 | 1st | W.M |
| **M.A 7-Years** | 0.21 | - | 3775.7 3834.6 | 3518.2 – 4033.2 | 5.40 | 2nd | - |

**Forecasting Table Yield Wheat Punjab (2018, 3005.64)**

| Model/ Methods | AdjR²/ C.V | No of Insig Coef: | Forecast-ing Year 2019 | Forecasting Intervals Year 2019 (95%) | % error | Comments | |
|---|---|---|---|---|---|---|---|
| | | | | | | Rank By Forecast | Rank By Regression |
| **Poly-1** | 0.896 | 0 | 3074.4 | 2830.6 – 3318.3 | 2.29 | 4th | 2nd |
| **Poly-2** | 0.906 | 0 | 2971.2 | 2739.5 – 3202.9 | 1.15 | 2nd | 1st |
| **Poly-4** | 0.913 | 1 | 3165.9 | 2942.2 – 3389.6 | 5.33 | 5th | 3rd |
| **Expon:** | 0.886 | 1 | 3173.3 | 3035.2 – 3311.3 | 5.58 | 6th | W.M |
| **Lagged-1** | 0.778 | 1/2 | 2977.9 | 2631.3 – 3324.6 | 0.92 | 1st | W.M |
| **M.A 7-Years** | 0.11 | - | 2967.6 2824.3 | 2848.0 – 3087.2 | 1.27 | 3rd | - |

**Forecasting Table Yield Wheat Sindh (2018, 3287.54)**

| Model/ Methods | AdjR²/ C.V | No of Insig Coef: | Forecast-ing Year 2019 | Forecasting Intervals Year 2019 (95%) | % error | Comments | |
|---|---|---|---|---|---|---|---|
| | | | | | | Rank By Forecast | Rank By Regression |
| **Poly-1** | 0.807 | 0 | 3722.9 | 3196.9 – 4248.9 | 13.24 | 6th | 2nd |
| **Poly-2** | 0.801 | 1 | 3688.4 | 3152.7 – 4224.1 | 12.19 | 4th | W.M |
| **Poly-3** | 0.909 | 0 | 3089.9 | 2728.4 – 3451.5 | 6.01 | 3rd | 1st |
| **Expon:** | 0.820 | 1 | 3709.6 | 3490.9 – 3928.2 | 12.84 | 5th | W.M |
| **Lagged-1** | 0.883 | 1/2 | 3345.1 | 2934.6 – 3755.6 | 1.75 | 2nd | 3rd |
| **M.A 7-Years** | 0.19 | - | 3317.4 3321.1 | 3092.0 – 3542.8 | 0.91 | 1st | - |

## 4. CONCLUSION

**i.** Production and Yield of Wheat show fluctuation in both Provinces.
**ii.** No single trend curve is preferred for regression, as the results vary over crops, Provinces and indicators.
**iii.** Polynomials of higher order are rejected in majority of cases.
**iv.** All the indicators are strongly correlated for Wheat with time and with previous values.
**v.** The models better for regression are not always better for forecasting.
**vi.** 7-years Moving Averages work better than 3-years and 5-years Moving averages.
**vii.** Moving Averages are better than Polynomials of higher order for forecasting.
**viii.** Yield depends on rain, temperature, seed, climate, irrigation water, crop diseases etc.

## REFERENCES:-

Saleem Abid, Nasir Jamal, Muhammad Zubair Anwar and Saleem Zahid (2018) "Exponential Growth Model for Forecasting of Area and Production of Potato Crop in Pakistan".

N. T. Khan[1], J. Park[2] and Y. B. Kim[1*] The J. Anim. Plant Sci. 28(1):2018. An integrated model for forecasting wheat consumption in Pakistan: emphasizing social and economic factors.

Sajid Ali1, Nouman Badar1, Hina Fatima2 (2015) Forecasting Production and Yield of Sugarcane and Cotton Crops of Pakistan for 2013-2030.

M. Amin, M. Amanullah and A. Akbar (2014) "Time series modeling for forecasting wheat production of pakistan".

Falak Sher and Eatzaz Ahmad* (summer 2008) "Forecasting Wheat Production in Pakistan".

Siraj Muhammad Pandhiani (2001) "Some Regression Models and Trend Curves for the forecasting of production, price and yield of Rice and Sugarcane in Punjab and Sindh".

Brunstad, R. J.; Gaasland, I.; Vardal, E. (1998). "Deregulation on the Norwegian Market for Milk Production" Norwegian School of Economics and Business Administration, Hellevien 30, 5035 Bergen Sandviken Norway. 16p.

Jh, R.; Murthy, K.V.B.; Nagarajan, H. k.; Seth, A.K. (1997). "Market Integration in Indian Agriculture, Rice and Wheat"; Indira Gandhi Institute of Development Research, Gen. Vaida Marg. Goregaon (E) Bombay-400 065 India. 17p.

M. N Sajjad, Lal Chand and N.M Shaikh (1996). "Econometric models for the prices of Wheat Flour, Rice, Mongs and Pulses in Baluchistan"; (1973-1992), Research Journal, Islamia University, Bahawalpur.