# Fitting and Forecasting of Trend Models for the HYV Boro Yields of Dinajpur District

Provash Kumar Karmokar*, Rangan Kumar and Ranjan Kumar Kundu

Department of Statistics
University of Rajshahi, Rajshahi-6205, Bangladesh
*Corresponding Authors Email: sprovash@yahoo.com
†Research Student

**Abstract**

Bangladesh is a densely populated country and the main food of the country is rice. Although the High Yielding Variety (HYV) Bro rice is being cultivated in almost all areas of Bangladesh it is enormously cultivated in Dinajpur district of Bangladesh. Trend is very important to know the HYV Boro rice yields of the country. Hence the objective of this research is to fit and forecast of trend models of HYV Bro rice yields of Dinajpur district for the year from 1971 to 2007. The regression diagnostics revealed in this study indicate that the data is autocorrelated. Hence the Cochrane-Orcutt method was employed to fit such data set. Finally, we forecasted the yields of the three trend models for HYV Boro rice yields of Dinajpur up to 2021. On basis of the regression diagnostics, the quadratic trend model is an appropriate model for this data set and in this model would be useful for the decision makers for their agriculture and food policy formulation.

***Keywords:*** *HYV Boro rice, Trend Models, Ordinary Least Square, Cochrane and Orcutt method, Forecasting.*

## 1. Introduction

Bangladesh is an agro-based developing country and the economic development of the country is mainly based on agriculture which contributes about 20.24% to the Gross Domestic Product. About 43.53% of the labor force is employed in agriculture (BER, 2010). Therefore, agriculture is a most important sector of Bangladesh and rice is the staple food of the country. Rice is the principal sources of food, calorie, and protein intake for most of the people. Rice contributes positively in the economy of the country and about 84% people of Bangladesh are directly or indirectly engaged in a wide range of agricultural activities (Rahman 2004). In the agricultural sector rice is a major source of livelihood in terms of providing food, income and employment in Bangladesh. Rice occupies 10.58 million hectares of land which is about 77 percent of the cultivated area (BBS, 2008). In Bangladesh, a country currently experiencing rapid population growth and serious food shortages, an effort to increase crop yields through the introduction of High Yielding Varieties (HYV) of rice was initiated in 1966. Among the all other rice varieties, Boro is the major rice crop of Bangladesh which provides about 55% of the total rice production. Boro is a variety of rice which is cultivated in almost every area of Bangladesh. This type of rice is cultivated in nearly 35% of the 10.80 million ha of rice harvested area, and contributed 50% of the 38.7 million tons of rice produced in 2001/2002 (Singh et al 2003b). Normally rice is grown in Bangladesh in three distinct seasons like, Aus (from April to August), and Amon (from August to December) and Boro (from January to June). Irrigated rice or Boro rice is a potential area for increasing rice yield, which currently accounts for about 57% of total rice production (BBS estimate 2008).

In the literature, a number of trend analyses have been carried out to measure the secular trend of some agricultural productions (e.g., Sahu (2003); Gupta *et al* (1999); Karmokar and Imon (2008)). These studies mainly based on the traditional OLS method of estimation. However, in this study we found that the rice production data of the selected area are affected with autocorrelation problem. Hence the objective of the study is to analyse and validate results of the High Yielding Variety (HYV) Boro rice production data of Dinajpur district of Bangladesh in presence of autocorrelation

This paper is organised as follows. The description of the data and methods are presented in Section 2. The results and discussions are reported in Section 3 and finally Section 4 concludes the paper.

## 2. Data and Methods

The rice production data for Dinajpur district of Bangladesh comprising of 37 years HYV Boro rice yields (y) from 1971 to 2008 have been used in this study. The data have been collected from Bangladesh Bureau of Statistics (BBS), government of Bangladesh.
A varieties of nonparametric tests like Kruskal–Wallis's test, Wald–Wolfowitz's test, Whitney-Pettitt's test (Pettitt, 1979), Mann–Kendall's test, Spearman's test are seen in the literature to detect the trend of a data. To investigate the trend in a data set, Kendall and Stuart (1961) suggested Mann–Kendall's test and Spearman's test as a powerful test when the most likely alternative to randomness is linear or non linear trend. In this study we measure the existence of trends using the non-parametric tests Mann-Kendall rank tests and Spearman's rank-correlation. The Mann-Kendall rank test and Spearman's rank-correlation (see Kendall and Stuart, 1961; Mann, 1945) are briefly discussed bellow.

The Mann-Kendall rank statistic $t_m$ is computed, by first replacing the observations $x_i$' s by their ranks $k_i$ s such that each term is assigned a number ranging from 1 to $n$ which reflects its magnitude relative to the magnitudes of all other terms. For each element $k_i$ the number $N_i$ *is* calculated as the number of $k_j$ terms preceding it such that $kj > k_i$ . Then $t_m$ *is* given by,

$$t_m = \frac{4\sum_{i=1}^{n-1} N_i}{n(n-1)} \tag{1}$$

tm is distributed very nearly as a normal distribution for large n and can be used as the basis of a significance test,

$$r_m = \pm r_g \sqrt{\frac{4n+10}{9n(n-1)}} \tag{2}$$

where rg is the desired probability point of the normal distribution appropriate to a two-tailed test. If tm lies inside the range ± rm then the time series does not contain a trend.

Spearman's rank-correlation method is simple and distribution-free which is also used to test the trends among the data sets. The test statistics of Spearman rank-correlation test is defined as:

$$\rho = 1 - \frac{6\sum_{1=1}^{n} d_i^2}{n(n^2-1)} \tag{3}$$

where, $d_i = kx_i - ky_i$ when two or more observations, $d_i$, have the same value, the average rank $ky_i$ is calculated.

A test-statistic $t_t$ is used to test the null hypothesis $H_0 : \rho = 0$ (there is no trend) against the alternative hypothesis $H_1 : \rho \neq 0$ (there is a trend). The test statistic is defined as

$$t_t = \sqrt{\frac{n-2}{1-\rho^2}} \tag{4}$$

For large $n$, the value of $r_s$ can be tested for significance by calculating the quantity $t_s$ given by the equation

$$t_s = r_s \sqrt{\frac{n-2}{1-r_s^2}} \tag{5}$$

The computed $t_s$ indicate the boundary value for the test.

Method of Ordinary Least Square (OLS) is an estimation technique in statistics (Harper, 1974) that was first described by German mathematician Carl Friedrich Gauss around 1794 (Bretsche, 1995). It is more convenient and widely used method among various method of estimation of unknown parameters on regression analysis technique. In the OLS the estimation of parameters of a linear regression model is based on the assumption that the errors are normally independently distributed with zero means and constant variance. Sometime data of a regression model involve regressors and response variables that have a natural sequential order over time. Further, the errors of the regression model may have a special form of relationship which indicator of autocorrelation. In such cases the traditional OLS estimators may have some faulty solution of the parameters and consequently the result may mislead to explain the whole scenario. Because the OLS estimators are still unbiased and consistent this is because both unbiasedness and consistency do not depend on assumption σ2 which is in this case violated. The OLS estimators will be inefficient and the best linear unbiased estimator (BLUE) property will be not maintained. The estimated variances of the regression coefficients will be biased and inconsistent, and therefore hypothesis testing may not be valid. Since $\sum u_i^2$ is affected in an autocorrelated situation the $R^2$ will also be affected. Thus the formal autocorrelation can be tested by DW statistics which may be falls in the inconclusive region or may falls in the positive or negative autocorrelation region. The use of the Cochrane and Orcutt method (see Cochrane and Orcutt, 1949) may be a solution of such problem may gives a meaningful estimate and interpretation. Consequently we can get the best fit of the model. In these viewpoints in this study we employ the Cochrane and Orcutt method to forecast the HYV Boro yields of Dinajpur district of Bangladesh.

To investigate the trend of yield of HYV Boro rice for Dinajpur district of Bangladesh we consider Linear, Compound and Quadratic trend models. The mathematical model of Linear, Compound and Quadratic models are as follows.

i. linear trend model

$$y_{ij} = a + bt + \epsilon_i \ , \quad (6)$$

where $y_i$ is the yield of HYV Boro, $t$ is the time and $\epsilon_i$ is the random error.

ii. quadratic model

$$y_i = a + bt + ct^2 + \epsilon_i \quad (7)$$

iii. compound trend model

$$y_i = ab^t \epsilon_i \quad (8)$$

Taking logarithmic on both sides of Eq.(8) we get,

$$y_i^* = a^* + b^* t + \epsilon_i^* \quad (9)$$

where $y_i^* = \ln y_i$ , $a^* = \ln a$ , $b^* = \ln b$ , $\epsilon_i^* = \ln \epsilon_i$ .

## 3. Results and Discussions

A time series consists of a trend component can be ensured using any time series (TS) plot. The TS plot of the yield of HYV Boro rice of the selected district for the year from 1971 to 2007 is shown in Fig.1.

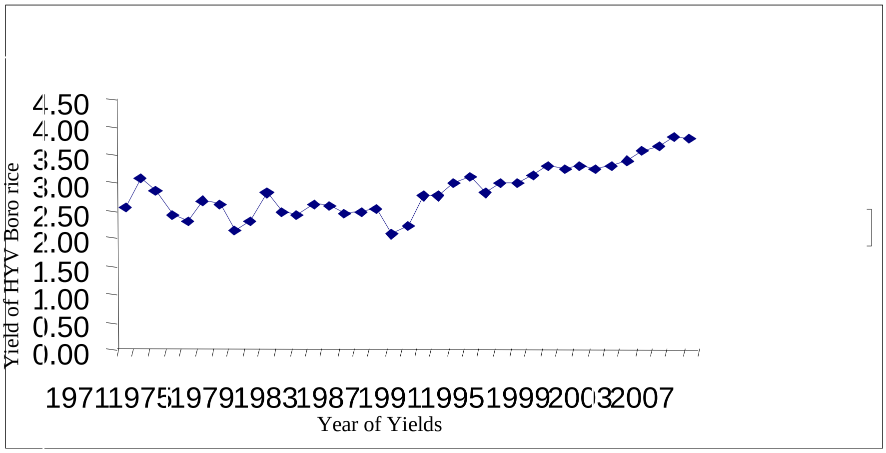


Fig 1. TS Plot of HYV Boro production for the district Dinajpur

The descriptive statistics of yield of HYV Boro rice of Dinajpur District is shown in Table 1.

Table 1. Descriptive Statistics for HYV Boro yields of Dinajpur

| Min | Max | Mean | SE | CV (%) | Skewness | Kurtosis |
|---|---|---|---|---|---|---|
| 2.03 | 3.81 | 2.715 | 0.420 | 15.45 | 0.917 | 3.859 |

From the Table 1, the skewness and Kurtosis are 0.917 and 3.859 respectively which indicate that the data are positively skewed and the kurtosis value is greater than 3 imply that the distribution of the frequency curve is leptokurtic. Since we are dealing with the yield data it is important to determine either the values of the yields increase or decrease over time or not. Basically it is a matter of investigation that whether the probability distribution of such variable is changes over time or not. It is interesting to note whether or not there is a trend in the time series data and for this purpose, we used the Mann-Kendal and Spearman rank correlation tests. Specially we wish to test the hypothesis that,

$H_0$ : There is no trend in the data set

$H_1$ : There exits trend in the data set.

The results of these tests appear in Table 2.

Table 2. Mann-Kendal and Spearman test of trend

| Test | Range (Years) | Test statistic | Value at 95% level $\pm r_m$ | Trend Exist |
|---|---|---|---|---|
| Mann-Kendal $t_m$ | 37 | −0.935 | 0.338 | Yes |
| Spearman rank $t_S$ | 37 | −2.146 | 0.834 | Yes |

The 95% level of $\pm r_m$ values will indicate that either there exist a trend or not. If the computed test statistics falls outside the band $(-r_m, +r_m)$ would indicate that the trend exist in the data set. From Table 2 the value of Mann-Kendal statistic, tm is −0.935 and its rm value at 95% is 0.338. Therefore the test statistic value falls outside of 95% interval of rm indicating that the H0 is rejected. Therefore, there exists a trend in the data set of HYV Boro rice yields of Dinajpur. The Spearman rank statistic, tS and rm values are −2.146 and 0.834 respectively. Since the tS value also falls outside of the 95% level value indicating the existence of trend by Spearman rank test.

The OLS results of the coefficients of the three fitted models together with their corresponding p values are presented in Table 3 to Table 5.

Table 3. OLS estimates of parameters for Linear Trend model

| Variable | Co-efficient | Standard Error | T Ratio | P-Value |
|---|---|---|---|---|
| t | 0.019 | 0.004 | 4.458 | 0.000 |
| Constant | 2.292 | 0.867 | 26.440 | 0.000 |

Table 4. OLS estimates of parameters for Compound Trend model

| Variable | Co-efficient | Standard Error | T Ratio | P-Value |
|---|---|---|---|---|
| t | 0.007 | 0.002 | 4.457 | 0.000 |
| Constant | 0.831 | 0.033 | 25.140 | 0.000 |

Table 5. OLS estimates of parameters for Quadratic Trend model

| Variable | Co-efficient | Standard Error | T Ratio | P-Value |
|---|---|---|---|---|
| t | −0.020 | 0.017 | −1.221 | 0.231 |
| $t^2$ | 0.001 | 0.000 | 2.460 | 0.020 |

| | | | | |
|---|---|---|---|---|
| Constant | 2.528 | 0.126 | 20.150 | 0.000 |

The regression diagnostics $R^2$, adjusted $R^2$, Durbin-Watson statistic and Jarque-Bera (JB) statistic ( Jarque and Bera in 1981) with *p* values are reported in the Table 6 for the selected data set.

Table 6. $R^2$, adjusted $R^2$, JB and DW statistic for the trend models by OLS

| Model | $R^2$ | Adjusted $R^2$ | JB | | DW |
|---|---|---|---|---|---|
| | | | Statistic | P value | |
| Linear | 0.383 | 0.364 | 3.882 | 0.144 | 1.373 |
| Quadratic | 0.484 | 0.451 | 1.049 | 0.592 | 1.624 |
| Compound | 0.383 | 0.364 | 2.074 | 0.355 | 1.403 |

We observe from the results presented in Table 6 that the quadratic trend model possesses the highest $R^2$ and adjusted $R^2$. The Jarque-Bera clearly shows that the three trend models obey the normality assumption of errors.

The presence of autocorrelation in a data set is measured by the DW statistics of the residuals is compared to respective critical values (see, the tables in Draper and Smith (1981)). If the DW value is less than a lower limit then the null hypothesis, H0 of no serial correlation in the residuals can be rejected, but if DW exceeds an upper limit, the null hypothesis, H0 can't be rejected. But if the DW statistic is found to be in between these boundary levels it will be indeterminate. If the DW statistic is found to be near to 2 mean that the data are not affected by autocorrelation problem.

With regards to the measurement of the OLS parameters of selected models the DW test revealed that in the three trend model, the H0 of non-existence of serial correlation in the respective residuals were accepted. So the data set has autocorrelation problems and autocorrelation may complicate the application of statistical tests by reducing the number of independent observations. It can also affect the predictions because future values depend on current and past values. Durbin-Watson (Durbin and Watson, 1950) test to detect the presence of autocorrelation which affects the efficiency of the estimators can be used in autocorrelated data set. To overcome the autocorrelation we employed the Cochrane and Orcutt method for three trend models according to our objectives of the study.

The results of Cochrane and Orcutt method for Linear, Quadratic and Compound trend models are displayed in Table 7- Table 9. Finally the regression diagnostics and forecasted yield of HYV Boro rice for Dinajpur is presented in Table 10 and Table 11 respectively.

Table 7. Cochrane and Orcutt estimates by Linear Trend model

| Variable | Co-efficient | Standard Error | T Ratio | P-Value |
|---|---|---|---|---|
| t | 0.020 | 0.006 | 3.526 | 0.001 |
| Constant | 2.286 | 0.114 | 20.000 | 0.000 |

Table 8. Cochrane and Orcutt estimates by Compound Trend model

| Variable | Co-efficient | Standard Error | T Ratio | P-Value |
|---|---|---|---|---|
| t | 0.008 | 0.021 | 3.531 | 0.001 |
| Constant | 0.830 | 0.043 | 19.270 | 0.000 |

Table 9. Cochrane and Orcutt estimates by Quadratic Trend model

| Variable | Co-efficient | Standard Error | T Ratio | P-Value |
|---|---|---|---|---|
| t | -0.021 | 0.019 | -1.080 | 0.289 |
| $t^2$ | 0.001 | 0.001 | 2.171 | 0.038 |
| Constant | 2.528 | 0.145 | 17.390 | 0.000 |

In Table 7, the estimated constant is 2.286 with standard error 0.114. The estimated coefficient for t is 0.020 with standard error 0.006. The t-ratio of time t and constant (estimates are 3.526 and 20.000 respectively) are significant for the Linear trend model. From Table 8, the estimated constant is 0.830 with standard error 0.043. The estimated coefficient for time is 0.008 with standard error 0.021. The t-ratio of time t and constant (estimates are 3.531 and 19.270 respectively) are significant for the Compound trend model. In the Table 9, the estimated constant is 2.528 with standard error 0.145. The estimated coefficient for t and t2 are –0.021 and 0.001 respectively with standard errors 0.019 and 0.001. The t-ratio of t, t2 and constant are 17.390, 2.171 and -1.080. These coefficients are significant.

Table 10. Regression diagnostics by Cochrane-Orcutt method

| Model | $R^2$ | Adjusted $R^2$ | JB | | DW |
|---|---|---|---|---|---|
| | | | Statistic | P value | |
| Linear | 0.431 | 0.413 | 2.073 | 0.355 | 1.867 |
| Quadratic | 0.499 | 0.467 | 1.776 | 0.411 | 1.935 |
| Compound | 0.429 | 0.411 | 1.401 | 0.496 | 1.877 |

Table 11 Method wise forecasted HYV Boro rice production of Dinajpur

| Year of Production | Actual Production (yhdin) | Forecasted by Linear Trend Model | Forecasted by Compound Trend Model | Forecasted by Quadratic Trend Model |
|---|---|---|---|---|
| 2004 | 3.39 | 2.965 | 2.963 | 3.156 |
| 2005 | 3.52 | 2.985 | 2.985 | 3.215 |
| 2006 | 3.75 | 3.005 | 3.008 | 3.276 |
| 2007 | 3.81 | 3.025 | 3.031 | 3.339 |
| 2008 | - | 3.044 | 3.054 | 3.405 |
| 2009 | - | 3.064 | 3.077 | 3.473 |
| 2010 | - | 3.084 | 3.100 | 3.543 |

| | | | | |
|---|---|---|---|---|
| 2011 | - | 3.104 | 3.123 | 3.616 |
| 2012 | - | 3.124 | 3.147 | 3.691 |
| 2013 | - | 3.144 | 3.171 | 3.768 |
| 2014 | - | 3.164 | 3.195 | 3.848 |
| 2015 | - | 3.184 | 3.219 | 3.930 |
| 2016 | - | 3.204 | 3.243 | 4.014 |
| 2017 | - | 3.224 | 3.268 | 4.100 |
| 2018 | - | 3.244 | 3.293 | 4.189 |
| 2019 | - | 3.264 | 3.318 | 4.280 |
| 2020 | - | 3.284 | 3.343 | 4.373 |
| 2021 | - | 3.304 | 3.368 | 4.469 |
| MSE | | 0.056 | 0.053 | 0.051 |
| AME | | 0.184 | 0.183 | 0.184 |

It is observed from Table 10 that the $R^2$ and adjusted $R^2$ values for Linear trend model are 0.431 and 0.413 respectively. The $R^2$ and adjusted $R^2$ values for Quadratic trend model are 0.499 and 0.467. For Compound trend model the $R^2$ and adjusted $R^2$ values are 0.429 and 0.411. Therefore the quadratic trend model possesses the highest $R^2$ and adjusted $R^2$ values. The Jarque-Bera clearly shows that the three trend models obey the normality assumption of errors. The DW statistics of the three models are close to 2 indicating that the data set is not affected by autocorrelation problem.

## 4. Conclusion

In this research, the objective was to fit and forecast of HYV Bro rice yields for Dinajpur. It is evident from the statistical tests that the data is autocorrelated. Hence the Cochrane-Orcutt method has been employed to fit such data set. Finally, we forecasted the yields of the three trend models for HYV Boro rice yields of Dinajpur up to 2021. On basis of the regression diagnostics, the quadratic trend model is an appropriate model for this data set and in this model would be useful for the decision makers for their agriculture and food policy formulation.

## References

Bangladesh Bureau of Statistics, BBS (1971-2008). Yearbook of Agricultural Statistics of Bangladesh, Ministry of Planning, Government of the People's Republic of Bangladesh, Dhaka.

BER. (2010). Bangladesh Economic Review, Ministry of Finance, Dhaka, Bangladesh.

BRRI. (2006). Improvement of standard Boro rice. BRRI (Bangladesh Rice Research Institute) Annual Report for July 2005-June 2006. Plant Breeding Division, BRRI, Gazipur, Bangladesh.

Cochrane, D. and G. H. Orcutt (1949). Application of least squares regression to relationships containing autocorrelated error terms, Journal of the American Statistical Association. 44, 32-61.

Draper, N. and H. Smith (1981). Applied Regression Analysis, Second Edition, New York: John Wiley & Sons.

Durbin, J. and G. Watson (1950). Testing for serial correlation in least squares regressions. Biometrika. 37, 409-428.

Fischer, K. S. and V. G. Cordova (1998). Impact of IRRI on rice science and production, in Impact of Rice Research, P.L. Pingali and M. Hossain (Eds.), Thailand Development Research Institute, Bangkok, and International Rice Research Institute (IRRI), Philippines. 27-50.

Gupta, D. S., P. K. Sahu, M. K. Sanyal and S. R. Pal (1999). Growth and trend analysis of potato production in West Bengal, J. Interacad. 3 (3&4), 345-350.

Harper, H. L. (1974). The method of least squares and some alternatives. Part I. International Statistical Review. 42, 147–174.

Hossain, M., M. A. Quasem, M. A. Jabbar and M. Mokaddem (1994). Production environments, MV adoption, and income distribution in Bangladesh. In Modern rice technology and income distribution in Asia, eds. C. C. David and K. Otsuka. Boulder, Colo., U.S.A.: Lynne Rienner Publishers, Inc.

Jarque, C. M. and A. K. Bera (1981). An Efficient Large-Sample Test for Normality of Observations and Regression Residuals. Australian National University Working Papers in Econometrics No. 40, Canbera.

Karmokar, P. K. and A.H.M.R. Imon (2008). Trend Analysis of the Production of Rice in Bangladesh. Int. J. Stat. Sci., 8: 103-110.

Kendall, M. (1938). A New Measure of Rank Correlation. Biometrika. 30 (1–2), 81–89.

Kendall, M. G. and A. Stuart (1961). The Advanced Theory of Statistics (vol. 2), New York: Charles Griffi

Sahu, P. K. (2003). Production model: In search of alternative food, Presented in Prof. K M Hossain International Seminar-Cum-Workshop, Department of Statistics, University of Rajshahi, Bangladesh, October 2003.

Singh, R.K., R. Thakur and S.D. Chatarjee (2003b). Harnessing boro rice potential for increasing rice production in deepwater areas of Eastern India, An overview. In: Boro rice (RK Singh, M Hossain and R Thakur, eds). IRRI-India Office. 51-72.