Int.J.Curr.Microbiol.App.Sci (2017) *6*(8): 107-115

International Journal of Current Microbiology and Applied Sciences
ISSN: 2319-7706 **Volume 6 Number 8 (2017) pp. 107-115**
Journal homepage: http://www.ijcmas.com

Original Research Article https://doi.org/10.20546/ijcmas.2017.608.014

Analysis of Growth and Instability in Pulse Production of Odisha during Rabi Season: A Statistical Modelling Approach

A. Dash*, D.S. Dhakre and D. Bhattacharya

College of Agriculture, Chiplima (OUAT) District – Sambalpur, Odisha 768025, Institute of Agriculture, PSB, Visva-bharati, Santiniketan, 731236, West Bengal, India
**Corresponding author*

Keywords

Spline regression, Kink, Economic reforms, Growth rate, Instability, Coefficient of variation.

Article Info

Accepted:
04 June 2017
Available Online:
10 August 2017

A B S T R A C T

Agriculture in Odisha during rabi season is based mainly on pulse crops. The economic reforms in the 1991-92, affected the agriculture scenario of Odisha following which our study period (1970-71 to 2013-14) has been divided into pre-reform (1970-71 to 1991-92) and post reform (1992-93 to 2013-14) period. Attempts have been taken to make a comparative study of the growth rate and instability of area, production and yield of rabi pulses in the two periods. For studying the growth rate and instability, appropriate model which could best possibly describe the behavior of the phenomenon are fitted by applying spline regression technique with kink or knot placed at the year of transition from pre–reform period to post- reform period which is considered to be 1991-92. The possible spline regression models which could fit very well to the area, production and yield of pulses are identified from the scatter plot of the data. Then by help of residual diagnostics and model fit statistics, the best fit models are obtained. Using the best fit model, average growth rates of area, production and yield are found. Coefficient of variation is used as a measure of instability. The difference in growth rate of area, production and yield from the pre-reform to post –reform period is found to be highly significant but negative in case of area and production, whereas, in case of yield it is negative non-signifiacnt. The difference in coefficient of variation between the two periods is non-significant for area, production and yield of rabi pulses.

Introduction

Agriculture is the backbone of rural economy and livelihood of Odisha. Though rice is the important crop of Odisha in *kharif* season, pulses are the important rabi season crops.

The most important pulses grown in Odisha are green gram, black gram, cowpea and arhar. Pulses share 12 per cent and 51per cent of the total cropped area in the state, in *kharif* and rabi season respectively. They share about 14 per cent and 81 percent of total area under food grains in *kharif* and rabi season respectively (Odisha Agricultural Statistics). The economic reformation in 1991-92, is expected to have substantial effects in Indian agriculture which in turn affect the state agriculture.

As economic reform is said to have brought about a clear shift in the focus on growth strategy, it may be useful to analyze the scenario of agriculture in the state by comparing the growth rate and instability in area, yield and production of cereals in pre-reform and post-reform period. To describe the behavior of data and to take care of any significant jumps in the data over a long period of time, there is need for fitting of spline regression models. Keeping these perspectives in mind, the study has been made with objective of exploring appropriate model that best fits the area, production and yield of cereals, in Odisha, and making a comparative study of growth rate and instability in pre-reform and post reform period.

Materials and Methods

Period of study

The study pertains to area, production and yield of cereals in the state of Odisha from the year 1970-71 to 2013-14. The entire study period is divided into two periods – Pre-reform period (1970-71 to 1991-92) and Post-reform period (1992-93 to 2013-14). Pre-reform period is referred to as Period I and Post-reform period is referred to as Period II.

Sources of data

The analysis is based on secondary source data relating to the area, production and yield of cereals in Odisha for the period from 1970-71 to 2013-14. The data are obtained from various volumes of Odisha Agricultural Statistics published by the Directorate of Economics and Statistics, Government of Odisha. The area, production and yield are expressed in '000 ha, '000 MT and kg ha^{-1} respectively. (1 ha = 10000 $m^{2,}$ 1 MT = 1000 kg)

Analytical techniques

Fitting of appropriate model to the data:

A model is an equation or a set of equations which represents the behavior of the system. Models that are used to describe the behavior of the variables that vary with respect to time are called the growth models. Fitting a single regression line for a long period to estimate the growth, may be misleading due to significant change (*i.e.* jumps or breaks) in direction of the growth path, which would make the estimate biased. To avoid this situation, the models are fitted by spline regression technique. A spline regression model avoids the inappropriate "jump" (*i.e.*, break) in the data for a long period of time by placing a kink in the line at the point of jump without allowing a break in the line.

Spline regression models are fitted with the help of dummy variables. Since the whole period of study (1970-71 to 2013-14) is divided into two periods – Pre-reform period (190-71 to 1991-92) and post-reform period (1992-93 to 2013-14) basing on the economic liberalization in 1991-92, the spline or kinked models have been fitted with one break point (i.e. kink) which is considered at the year 1991-92.

The time variable for the year 1970-71 to 2013-14 is referred to as t = 1,2,3,…..22,23,24……43,44 *i.e.,* t =1 for the year 1970-71, t = 2 for the year 1971-72, ……., t = 22 for the year 1991-92,……, t = 44 for the year 2013-14.

The time variable for the year in which kink is placed (i.e. 1991-92) is referred as k. Thus k = 22.

The kinked models are obtained with the help of dummy variables (Paltasingh, 2013).

The kinked linear model, power model, compound model and quadratic model are obtained as:

Linear model

$$Y_t = \beta_0 + \beta_1 . t. I_{(1 \leq t \leq 22)} + \{\beta_1 . t + A_1 (t - k)\}. I_{(23 \leq t \leq 44)} + \varepsilon_t$$

Power model

$Y_t = \beta_0 . t^{\beta 1} . I_{(1 \leq t \leq 22)} \{t^{\beta 1} . (t-k)^{A1}\}. I_{(23 \leq t \leq 44)} . Exp(\varepsilon_t)$

The power model obtained is transformed to linear model by natural log transformation as,

$\ln Y_t = \ln \beta_0 + \beta_1 . \ln t. I_{(1 \leq t \leq 22)} + \{\beta_1 .\ln t + A_1 \ln(t-k)\} I_{(23 \leq t \leq 44)} + \varepsilon_t$

Compound model

$Y_t = \beta_0 . \beta_1^{t.} I_{(1 \leq t \leq 22)} . \{\beta_1^{t} . A_1^{(t-k)}\}. I_{(23 \leq t \leq 44)} . Exp(\varepsilon_t)$

The compound model obtained is transformed to linear model by natural log transformation as,

$\ln Y_t = \ln \beta_0 + t. \ln \beta_1 . I_{(1 \leq t \leq 22)} + \{t. \ln \beta_1 + (t-k) \ln A_1\} I_{(23 \leq t \leq 44)} + \varepsilon_t,$

Quadratic model

$Y_t = \beta_0 + \{\beta_1 .t + \beta_2 .t^2\} I_{(1 \leq t \leq 22)} + \{\beta_1 .t + A_1 (t-k) + \beta_2 .t^2 + A_2 (t-k)^2\} I_{(23 \leq t \leq 44)} + \varepsilon_t$

Where $I_{(A)}$ is the indicator function which is 1 if A holds and 0 else.

Using Ordinary Least Square technique, the estimated values of the coefficients β_0, β_1 and A_1 are found out. The estimated values of β_0, β_1 β_2, A_1 and A_2 are written as b_0, b_1, b_2 and a_1, a_2 respectively.

The significance of the estimated coefficient is tested by applying t test statistic

$$t = \frac{\beta_j}{S.E.(\beta_j)}$$

Which follows 't' distribution with n – p degrees of freedom, n is the number of observations.

The overall significance of the model is tested by applying F statistic

$$F = \frac{MSM}{MSE}$$

Which follows F distribution with (p-1, n-p) degrees of freedom

MSM is the mean square of the model, MSE is the error mean square;

$$MSM = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{p-1}, \quad MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-p}$$

Where, n is the number of observations and p is the number of parameters involved in the model.

Assumptions in the model are: (i) Errors should be independent (ii) Error should have zero mean. (iii) Errors should have constant variance i.e. errors should be homoscedastic, (iv) Errors must follow normal distribution

The assumptions regarding the errors are tested by using: (i) Durbin-Watson test for testing independence of residuals (ii) Park's test for testing homoscedasticity of residuals. (iii) Shapiro-Wilk's test for testing normality of residuals,

Durbin-Watson test

This test considers the first order autocorrelation among the residuals. (Montgomery, *et al.*, 2001) Durbin-Watson test statistic *i.e.*, D-W statistic,

$$d = \frac{\sum_{t=2}^{n}(\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^{n}\varepsilon_t^2},$$

Where, $\varepsilon_t = y_t - \hat{y}_t$, y_t and $\hat{y}_t$ are respectively the actual and estimated values of the response variable in time t.

The value of 'd' ranges from 0 to 4. Upper and lower critical values, dU and dL have been tabulated for different values of k (no. of explanatory variables) and n (no. of observations) for corresponding level of significance (α) in the Durbin – Watson statistical table.

If $d < d_L$, it is significant. If $d > d_U$, then it is insignificant and the residuals are independent.

If $d_L < d < d_U$, test is inconclusive. For testing negative autocorrelation, the statistic 4 - d is used to compare with d_U and d_L.

Park's test

In this test, natural logarithm of the residual (ε_t) is regressed with natural logarithm of the independent variable (which is time, t) by fitting linear regression, *i.e.*, $\ln(\varepsilon_t) = a + b \ln(t)$. If the slope of the regression coefficient, b is found to be non-significant, then it is concluded that residuals are homoscedastic (i.e. constant error variance), otherwise, residuals are heteroschedastic (error variance not remaining constant). (Gujarati, D.N., 2003)

Shapiro-Wilk's test

Shapiro-Wilk's test statistic *i.e.*, S-W test staitistic, $w = s^2 / b$

Where, $s^2 = \Sigma\, a(k) \{ x_{(n+1-k)} - x_{(k)} \}$; $b = \sum_{t=1}^{n}(y_t - \bar{y})^2$ (Thode, 2012)

The parameter k takes the values 1, 2,….,n/2, when n is even and 1,2,.....,(n-1)/2, when n is odd, n is the number of observations. $X_{(k)}$ is the k^{th} order statistic of the set of residuals.

The values of coefficients a(k) for different values of n and k are obtained from the table of Shapiro-Wilk. If w is non-significant, then the residuals are normally distributed.

The model fit statistics, viz., R^2, adjusted R^2 and RMSE (Root mean Square Error) are also computed. Among the models fitted for the dependent variable, which satisfy the error assumptions and show overall significance and significant parameter estimates, the one having highest adjusted R^2 and lowest RMSE is considered to be the best fit model for that variable.

$R^2 = \frac{SSM}{SSE}$, where, SSM is the sum of square due to model; SSE is the sum of square due to error.

$SSM = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$; $SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$

Adjusted $R^2 = 1 - (1 - R^2) \times \frac{n-1}{n-p}$; RMSE (Root Mean Square Error) $= \left\{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-p}\right\}^{1/2}$

Where 'n' is the number of observations; p is the no. of parameters involved in the model.

Estimation of growth rates

Using this best fit model, the estimated/predicted values ($\hat{Y}_t$) of the dependent variable (Area/Production/Yield) are found for the period I, period II and the whole period. By using the predicted values, the annual growth rates are found.

Annual Growth Rate for the year t, $AGR_t = \left(\frac{\hat{Y}_t - \hat{Y}_{t-1}}{\hat{Y}_{t-1}}\right) \times 100$

Average Growth rate for the period I (1970-71 to 1991-92), period II (1992-93 to 2013-14) and the whole period is obtained by taking arithmetic mean of the annual growth rates of the respective periods (Prajneshu, 2005).

Also the difference in average growth rates of period I and period II is obtained as,

$\Delta GR = GR_2 - GR_1$

Study of instability in area/production/yield

Coefficient of variation is used as a measure of instability. The simple C.V. often contains the trend component and thus overestimates the level of instability in time series data characterized by long term trends. So to eliminate the effect of trend, C.V. is estimated from the detrended values (*i.e.,* trend eliminated values).

For linear and quadratic model, where the effects are assumed to be additive in nature, the detrended values are obtained by subtracting the predicted values (obtained from the model) from the actual values.

Thus, detrended value, $y_D = y_t - \hat{y}_t$ (assuming additive model)

Where,

y_t is the actual value of the variable in time t.
$\hat{y}_t$ Is the predicted value obtained from the best fit model?

For power, compound and logarithmic model, where the effects are assumed to be multiplicative in nature, the detrended values are obtained by dividing the actual values with predicted values (obtained from the model).

Thus, detrended value, $y_D = \frac{y_t}{\hat{y}_t}$ (assuming multiplicative model)

Where,

y_t is the actual value of the variable in time t.

$\hat{y}_t$ is the predicted value obtained from the best fit model?

The detrended values are then centered by adding the mean of the actual values ($\bar{y}_t$)

The C.V. is found from these detrended and centered values.

$$CV = \frac{\sigma_{yD}}{\bar{y}_D} \times 100$$

C.V. is found for period I, period II and for the whole period. Also the difference in CV (ΔCV) between period I and period II is found.

Results and Discussion

For selecting the appropriate models to be fiited for the variables, such as, area, production and yield of pulses in Odisha, the scatter plot is made for the variables against the time.

Figure 1 shows the different scatter plots for area, production and yield of foodgrains. From the observation of the pattern of dots in the scatter plot, it is found that some linear and non-linear models are found suitable for being fitted to the data. The models like linear, power, compound and quadratic are fitted to each of the nine variables, *i.e.,* area, production and yield of pulses of odisha.

Table 1 shows the parameter estimates, residual diagnostics and model fit statistics of the fitted models for area under rabi pulses. From the table it is found that among all the models fiited for area under rabi pulses,quadratic model is the only model in which all the parameter estimates are significant, all the assumpptions regarding the error are satisfied and has the highest adjusted R^2 and lowest RMSE than other fitted models. So, quadratic model is found to be the best fit model for area under pulses during rabi season in Odisha.

Table 2 shows the parameter estimates, residual diagnostics and model fit statistics of the fitted models for production of rabi pulses. From the table it is found that among all the models fiited for production of rabi rpulses, power model is the only model in which all the parameter estimates are significant, all the assumpptions regarding the error are satisfied and has the highest adjusted R^2 and lowest RMSE than other fitted models. So, power model is found to be the best fit model for area under pulses during rabi season in Odisha.

Table 3 shows the parameter estimates, residual diagnostics and model fit statistics of the fitted models for yield of rabi pulses. From the table it is found that among all the models fiited for yield of rabi pulses, quadratic model is the only model in which all the parameter estimates are significant, all the assumpptions regarding the error are satisfied and has the highest adjusted R^2 and lowest RMSE than other fitted models. So, quadratic model is found to be the best fit model for area under pulses during rabi season in Odisha.

Table 4 shows the average growth rates and coefficient of variation of area, producftion and yield of rabi pulses in period I, period II and whole period. From deep perusal of table 4 regarding the average growth rates, it is found that the average growth rates of area uder rabi pulses is positive and highly signifiacnt in period I and in the whole period but positive and non-significant in period II. The difference in average growth rates of area from period I to period II is negative and significant. Same is the case with average growth rate of production which is positive and highly signifiacnt in period I and in the whole period but positive and non-significant in period II. The difference in average growth rates of production from period I to period II is negative and significant.

In case of yield of rabi pulses, the average growth rate is positive and highly whereas, in period II it is negatiuve but non-significant. In whole period it is positive but non-significant. Te difference in growth rates from period I ti period II is negative and non-significant.

Fig.1 Scatter diagrams of area, production and yield of pulses bi season during rin Odisha

Area **Production** **Yield**

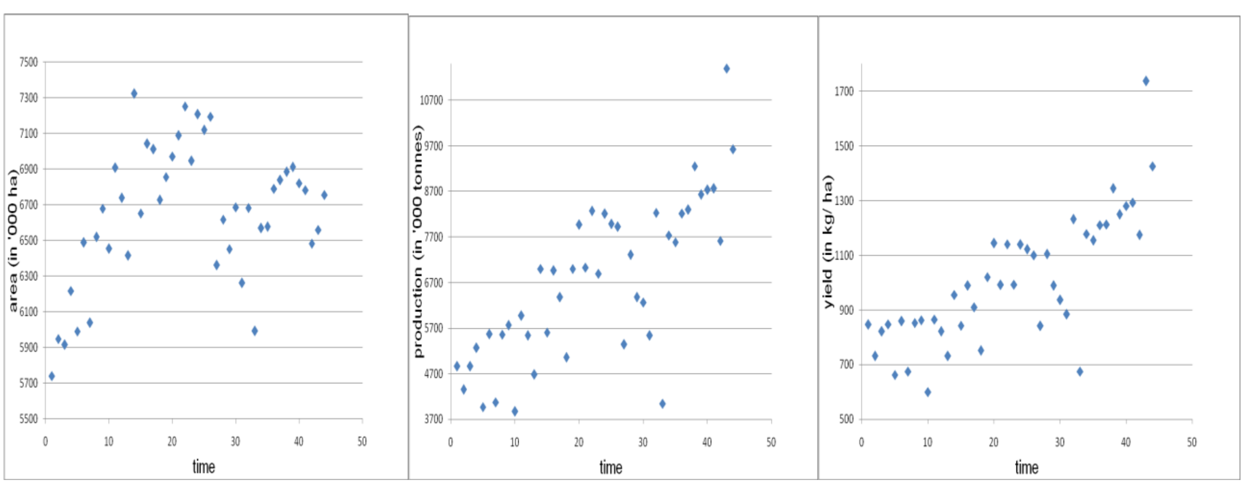

Table.1 Parameter estimates, residual diagnostics and model fit statistics of the fitted models for area under pulses during rabi season in Odisha

Fitted Models →		Linear	Power	Compound	**Quadratic**
Parameter estimates	b_0	729.69** (73.63)	573.51** (91.1)	742.65** (83.12)	**486.06** (96.42)**
	b_1	39.24** (5.61)	0.311** (0.034)	1.038** (0.005)	**100.15** (19.31)**
	a_1	-73.08** (11.79)	-0.417** (0.059)	0.935** (0.011)	**-40.77* (18.08)**
	b_2	-	-	-	**-2.65** (0.82)**
	a_2	-	-	-	**5.28** (1.17)**
Residual diagnostics	D-W Statistic	0.61	1.21	0.56	**1.98**
	Coefficient of ln(t) in Park's test	0.528	0.447	0.906*	**0.268**
	S-W Statistic	0.973	0.97	0.971	**0.945**
Model fit statistics	R^2	0.06	0.521	0.024	**0.721**
	Adjusted R^2	0.038	0.509	0.016	**0.707**
	RMSE	245.23	175.13	255.93	**133.62**
	F Value	2.62	45.62**	1.11	**24.54****

(Figures in the parentheses indicates the standard errors)
* Significant at 0.05 level of significance ** Significant at 0.05 level of significance

Table.2 Parameter estimates, residual diagnostics and model fit statistics of the fitted models for production of pulses during rabi season in Odisha

Fitted Models →		Linear	**Power**	Compound	Quadratic
Parameter estimates	b_0	332.83** (42.27)	**268.54** (57.52)**	346.27** (47.51)	224.08** (61.21)
	b_1	22.59** (3.22)	**0.338** (0.047)**	1.043** (0.006)	49.78** (12.24)
	a_1	-49.03** (8.39)	**-0.524** (0.095)**	0.913** (0.016)	-53.09* (24.33)
	b_2	-	-	-	-1.18* (0.52)
	a_2	-	-	-	3.59* (0.8)
Residual diagnostics	D-W Statistic	**0.53**	**1.07**	**0.57**	1.82
	Coefficient of ln(t) in Park's test	1.328*	0.524	1.611*	0.077
	S-W Statistic	0.986	0.979	0.988	0.985
Model fit statistics	R^2	0.131	0.451	0.015	0.695
	Adjusted R^2	**0.104**	**0.438**	**0.009**	**0.681**
	RMSE	174.36	122.94	167.13	91.56
	F Value	3.88	34.498*	1.61	21.69**

(Figures in the parentheses indicates the standard errors)
* Significant at 0.05 level of significance ** Significant at 0.05 level of significance

Table.3 Parameter estimates, residual diagnostics and model fit statistics of the fitted models for yield of pulses during rabi season in Odisha

Fitted Models →		Linear	Power	Compound	Quadratic
Parameter estimates	b_0	469.61** (21.82)	468.24** (31.85)	466.37** (21.85)	**480.73** (35.49)**
	b_1	2.55 (1.66)	0. 026 (0.029)	1.005** (0.004)	**-0.233* (.098)**
	a_1	-10.97** (3.18)	-0.107* (0.048)	0.977** (0.007)	**-30.34* (13.35)**
	b_2	-	-	-	**0.12** (0.03)**
	a_2	-	-	-	**0.95* (0.38)**
Residual diagnostics	D-W Statistic	0.53**	1.07**	0.57**	1.82
	Coefficient of ln(t) in Park's test	1.328*	0.524	1.611*	0.077
	S-W Statistic	0.991	0.963	0.990	0.961
Model fit statistics	R^2	0.131	0.451	0.040	0.695
	Adjusted R^2	0.104	0.438	0.015	0.681
	RMSE	174.36	122.94	167.13	91.5
	F Value	**3.88**	**34.49****	**0.61**	**21.69****

(Figures in the parentheses indicates the standard errors); * Significant at 0.05 level of significance ** Significant at 0.05 level of significance

Table.4 Average growth rates of coefficient of variation of area, production and yield of pulses of Odisha during rabi season in period I, period II and the whole period

(In per cent)

Variables	Average Growth Rate				Coefficient of Variation			
	Period I (GR_1)	Period II (GR_2)	Whole Period (GR)	Δ GR = GR_1 - GR_2	Period I (CV_1)	Period II (CV_2)	Whole Period (CV)	Δ CV = CV_1 - CV_2
Area	4.38** (1.01)	0.27 (0.55)	2.18** (0.66)	-4.11* (1.55)	10.8** (1.65)	11.04** (1.68)	10.82 ** (1.14)	0.24 (1.71)
Production	5.01** (1.03)	0.13 (1.11)	2.35** (0.86)	-4.89* (2.11)	13.63** (2.09)	16.09** (2.49)	15.81** (1.64)	2.44 (2.34)
Yield	0.5** (0.06)	-0.07 (0.63)	0.11 (0.33)	-0.58 (0.58)	9.41** (1.43)	7.52** (1.14)	9.16** (0.97)	-1.89 (1.31)

(Figures in the parentheses represent the standard errors)
***** Significant at 5 % level; ****** Significant at 1 % level

From the scrutiny of table 4 regarding the coefficient of variation, it is found that the coefficient of variation is significant at 0.01 level of significance for all the variables under study *i.e.*, area, production and yield in period I, period II and the whole period. The difference in coefficient of variation from period I to period II is positive but non-significant in case of area and production of rabi pulses, whereas, for yield it is negative and non-significant.

The above discussion hightlighted that though some models perform better in terms of model fit statistics, such as, high value of adjusted R^2 and low value of RMSE, they are not worthy of being selected as the best model if they do not satisfy the error assumptions. This is the case with the statistical modelling in case of production of pulses. Though the linear model, compound model and quadratic model fitted to the production of pulses data, have higher values of adjusted R^2 than the power model fiited to the same data, but these models show lack of normality of errors as indicated by their significant S-W statistic. The spline regression technique also takes care of any abrupt jumps in the data.

From the study of average growth rate found with the help of best fit model, it is found that area under rabi pulses shows an increasing trend in the whole period which occurs at a higher rate in pre-reform period, whereas, in post reform period the trend is almost stable. This stable trend in area during post – reform period may be assigned to the rapid growth of industrialisation and shift of agriculture from pulses to commercial crops. The production of rabi pulses also shows the similar pattern as shown by area under rabi pulses. The yield shows increasing trend only in the pre-reform period. In post reform period it started showing negative trend though it is non-significant. Due to this the trend in yield of rabi pulses in whole period remains non-significantn and can be considered to be almost stable.

From instability point of view, there is increase in instability of area, production and yield of pulses in both the periods. But the increase in period II over period I is non-significant. In case of area, production and yield of rabi pulses.. This shows that in case of rabi pulses, both high and low average growth rates are prone to high instability which may be due to the fact that the policy decisions of growing rabi pulses vary year to year.

References

Gujarati, D.N., 2003. Basic Econometrics, Fourth Edition, Published by McGraw-HiII/lrwin.

Montgomery, D. C., Peck, E. A., Vining, G. G. 2001. Introduction to Linear Regression Analysis. 3rd Edition. New York. John Wiley & Sons.

Paltasingh, R.K., Goyari, P. (2013) Agricultural Economics Research Review. Vol. 26 (Conference number) 67-78.

Prajneshu, Chandan, K.P., 2005. Computation of compound growth rates in agriculture, Revisited. *Agricultural Economics Research Review*. 18: July-December, Pp. 317-324.

Thode Jr., H.C., 2012. Testing for Normality. Marcel Dekker, Inc., New York.

How to cite this article:

Dash, A., D.S. Dhakre and Bhattacharya, D. 2017. Analysis of Growth and Instability in Pulse Production of Odisha during Rabi Season: A Statistical Modelling Approach. *Int.J.Curr.Microbiol.App.Sci.* 6(8): 107-115. doi: https://doi.org/10.20546/ijcmas.2017.608.014