Madras Agric. J. 76 (4) : 212-216 April, 1989

https://doi.org/10.29321/MAJ.10.A02051

FORECASTING IN AGRICULTURE : AN APPROACH THROUGH INTERVAL ESTIMATION

G. NAGESWARA RAO and B. S. KULKARNI

GNT Road, Near Cotton Mill, Sullurpet-524 121 Nellur Dist.

ABSTRACT

In this paper two new methods are attempted for forecasting in agriculture through interval estimation. The first method relaxes the assumption of constant variance for residual errors, which is used in the usual least squares method in curve fitting. The second method involves fitting of three growth curves for normal, above normal and below normal values of the time series data. The above methods were illustrated along with the usual least-squares for the data of Production and Productivity of Paddy for Andhra region in Andhra Pradesh from 1961-62 to 1980-81. The first method gave lower and upper boundaries with minimum bias as compared to other two methods.

Key words : Least squares method, Growth curves, Interval estimation

Several attempts have been made for forecasting area, production and productivity of crops in a region by various authors by fitting different growth curves. The estimates obtained through this approach are the point estimates and are in general far away from the actual values. Rao, *et al.* (1980) attempted measurement of growth and fluctuations in the crop ontput by using the concept of Non-Systematic component, which is a different approach hitherto unattempted by earlier workers. In this approach they have identified 'Normal' years separately, 'Peak aud Trough' years separately in the time series data and fitted linear models using the least squares technique. They have defined the peak year as the one during which the crop output is greater than the output in both preceding and succeeding years and the excess in each situation is more than 5 percent. Similarly, a trough year is that for which the crop output is less than that corrasponding to both the adjacent years with the deficit in each case being more than 5 percent.

All these attempts did not make much head-way due to regular and irregular fluctuations in nature and technology on crops. Therefore, in general, the point estimates predicted by these fitted curves are not much efficient for policy making decisions.

The purpose of this paper is to attempt a new approach of forecasting by using interval estimation instead of point estimation which generally is usful and necessary for policy making bodies.

MATERIALS AND METHODS

In the classical approach of time-series analysis, moving average method is followed for smoothening seasonal, cyclic and random fluctuations by taking constant period. However the assumption of constant period is not realistic for application, as it (period) varies from data to data. Hence two alternative approaches are attempted here.

METHOD - I : This method involves identifying the periods in which 'peaks' and' troughs' have occurred. The length of these periods may differ depending upon the nature of curve and of the time series data. The arithmetic mean $(\bar{Y}_i)$ and standard deviation $(\hat{s}_i)$ for each i-th period (i= 1, ..., k) will be obtained and a linear regression model of the following type is fitted :

$$s_i = a + b\,\bar{Y}_i,\ i = 1, \ldots, k \text{ periods.}$$

This equation can be used for predicting the standard deviation of the predicted year or period by using the predicted mean (Hills and Morgan, 1981). The predicted mean is the point estimate obtained through the trend equation which is fitted to the given data. This point estimate $(\hat{Y}_t)$ and its corresponding estimate of s.d. (s_t) together give the prediction interval as :

$$\hat{Y}_t - s_t < \hat{Y}_t < \hat{Y}_t + \hat{s}_t$$

where, $\hat{Y}_t - \hat{s}_t$ is the lower bound and $\hat{Y}_t + \hat{s}_t$ is the upper bound Thus the difference between these two bounds is the interval estimate.

Here, one need not fit the usual growth curve but even use any crop model by including weather variables besides time when the variable under study is production or productivity.

It is thus clear that this method is based on the assumption of a variable standard deviation for the residual error in the time series model which is more realistic than the classical approach of trend fitting which assumes a constant variance for the residual errors. Further, the assumption of a variable s.d. for the periods, in a way, also smooths out the cyclic, seasonal and random flucuations.

METHOD - II : In this method, the time series data are classified into 'Above Normal', 'Normal' and 'Below Normal' catagories on the basis of the periods of fluctuations in which the peak and troughs have oocurred. Thus within each period 'i' (i = 1, ... , k), Y_{ij} (j= 1, ... , m ; m = number of observations in i-th period) is classified according to the three categories as follows :

$Y_{ij} < \bar{Y}_i + s_i$: Above Normal (AN)

$\bar{Y}_i - s_i < Y_{ij} < \bar{Y}_i + s_i$: Normal (N)

$Y_{ij} > \bar{Y}_i - s_i$: Below Normal (BN)

Separate growth curves are then fitted for each of these three categories which provide three point estimates for a given year (t) as : $\hat{Y}_t$ (AN), $\hat{Y}_t$ (N), $\hat{Y}_t$ (BN), for Above Normal, Normal and Below Normal categories. The prediction interval is then :

$$\hat{Y}_t\,(BN) < \hat{Y}_t < \hat{Y}_t\,(AN)$$

where, $\hat{Y}$ (BN) and $\hat{Y}$ (AN) are the lower and upper bounds respectively for the variable under study.

RESULTS AND DISCUSSION

The two methods outlined earlier along with the classical least squares approach (LS) are illustrated by taking

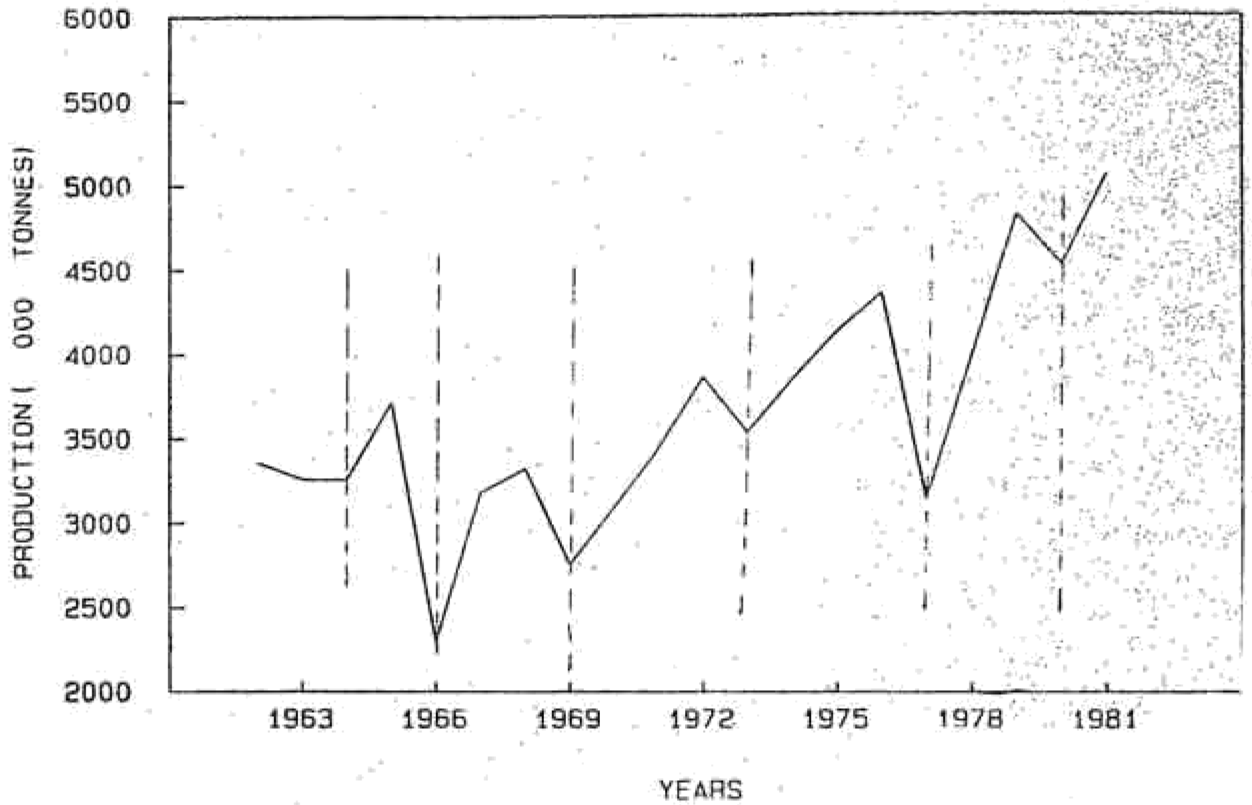

FIG 1 ANDHRA REGION (A P) - PADDY PRODUCTION

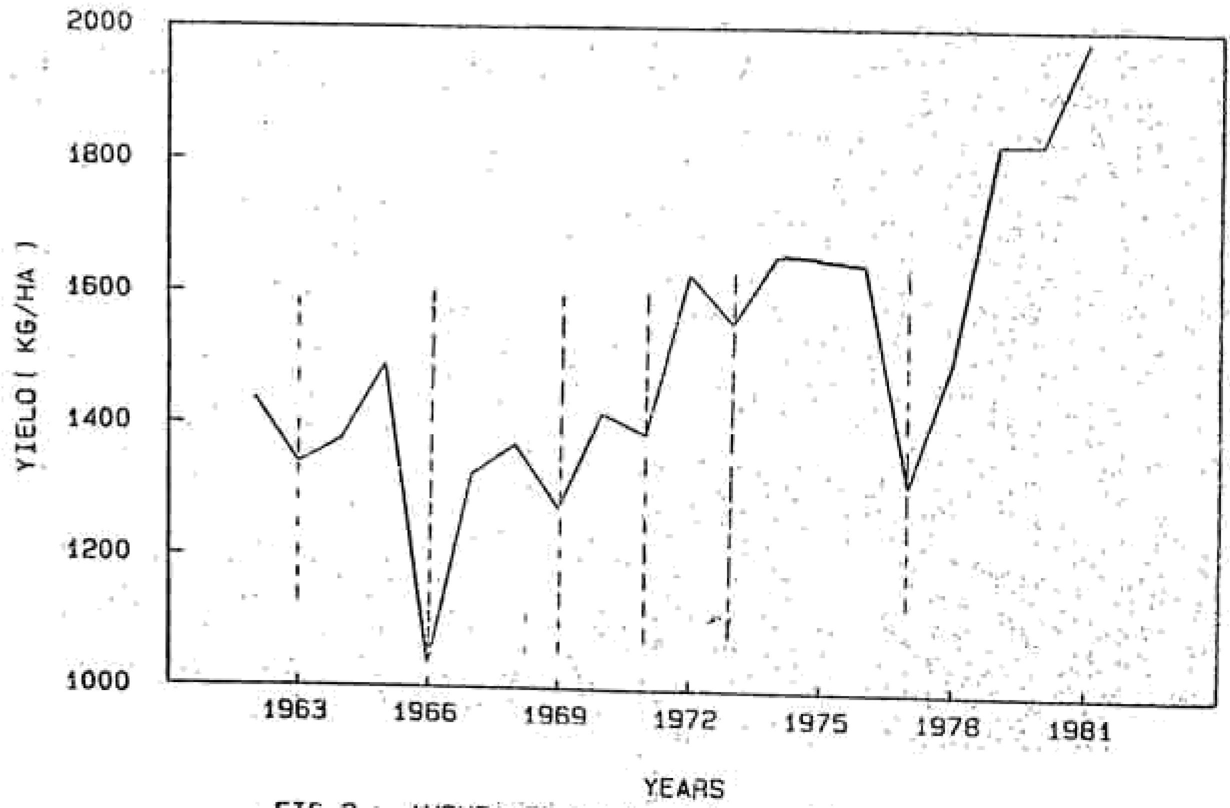

FIG 2 : ANDHRA REGION (A.P.) - PADDY YIELD (KG/HA) time series data of Production and Productivity (Yield/ha) of paddy of Andhra region of Andhra Pradesh from 1961-62 to 1980-81. The relevant data were collected from the Season and Crop reports. The following growth equations were fitted :

1. $Y = a' + b' t$ ... Linear
2. $Y = b_0 + b_1 t + b_2 t^2$ ... Quadratic

In order to study the prediction capacity of the interval estimation through these methods, three observations were deleted from the data i. e. the data from 1961-62 to 1977-78 were used for fitting of growth curves, while, forecasts $(\hat{Y})$ were obtained for the remaining period from 1978-79 to 1980-81. The relative bias in estimation was computed by using the formula :

$$\text{Bias (percent)} = [\,(\hat{Y} - Y).\,100]/y.$$

The fitted growth curves as well as regression equations are presented in Table 1. The results of interval estimation corresponding to Production and Productivity of paddy in Andhra region are presented in Tables 2 and 3, while a graphical representation of the data can be seen in Figs. I and II.

It can be seen that there are a total of 6 and 7 periods of peak and trough years in respect of Production and Productivity respectively of paddy of Andhra region (Figs. I and II). These identified periods give the corresponding equation(s) for estimating 's_i' and also the relevant growth equations (Table 1).

From Table 2, it can be seen that the prediction interval for production of paddy in Andhra region can be more efficiently obtained with Method I than with Method II and the classical least squares approach. The percentage bias of the bounds as well as the range of the estimates obtained with method I is considerably less than that with the other two methods. The observed values for all the three years (predicted) are lying in between the lower and upper bounds for Method I. Though the observed values are lying within the bounds for LS method also, its range is much larger than Method I. The percentage bias in Method I is ranging from —14.50 to 38.70. In Method II, the percentage bias is ranging from 0.28 to 31.60. However, the observed values are within the bounds only in one of the three predicted years in this method, i. e., Method II.

A similar trend can be observed about the prediction efficiency of Method I in prediction interval estimation of Productivity of paddy in Andhra region (Table 3). The percentage bias of the two bounds of Method I is much less than those of the other two methods and it ranges from 2.53 to 27.90. For other two methods, the range of the bias is from —5.85 to 37 10 and from 9.56 to 25.70 for LS Method and Method II respectively.

From the results of Table 2 and Table 3, it can be inferred that the Method I is superior to other two methods both in respect of percentage bias and range. Hence this method can be successfully used for forecasting purpose with respect to interval estimation.

Table 1 : Regression models and fitted growth curves for interval estimation

S. No.	Production of paddy (Andhra Region)	Productivity of paddy (Andhra Region)
1.	*Regression Equation :* $s = -134.31 + 0.18\ \bar{Y}$	$s = -180.03 + 0.23\ Y$
	Growth Curves : Method I / LS Method	
2.	$Y = 2961.23 + 53.71\ t$ $R^2 = 0.2802^*$	$Y = -1291.57 + 16.21\ t$ $R^2 = 0.2458^*$
3.	$Y = 3273.03 - 44.74\ t + 5.47\ t^2$ $R^2 + 0.3533^*$	$Y = 1315.20 + 8.75\ t + 0.41\ t$ $R^2 = 0.2489$
	Growth Equations : Method II	
4.	$Y_{(AN)} = 3252.08 + 67.77\ t$	$Y =$
5.	$Y_{(N)} = 3084.73 + 50.02\ t$	$Y_{(N)} = 1341.08 + 17.46\ t$
6.	$Y_{(BN)} = 2046.12 + 70.71\ t$	$Y_{(BN)} = 1043.01 + 21.56\ t$

* Significant at 5% level of probability.

Table 2 : Prediction Interval for Production of paddy, Andhra Region (A. P.) ('000' Tonnes)

Year	Observed	LS Method*		Method-I		Method-II		
		Lower bound	Upper bound	Lower bound	Upper bound	Lower bound	Normal	Upper bound
1978-79	4825	2957 (38.7)	5523 (−14.5)	3611 (25.1)	4869 (−0.9)	3319 (31.2)	3985 (17.4)	4472 (7.32)
1979-80	4527	3066 (32.3)	5728 (−26.5)	3740 (17.4)	5054 (−11.61)	3390 (25.1)	4035 (10.9)	4540 (0.28)
1980-81	5059	3149 (37.8)	5983 (−18.3)	3879 (23.3)	5253 (−3.83)	3461 (31.6)	4085 (19.3)	4607 (8.93)

Table 3 : Prediction Interval for Productivity of paddy, Andhra Region (A. P.) ('00' kg/ha)

Year	Observed	LS Method*		Method-I		Method-II		
		Lower bound	Upper bound	Lower bound	Upper bound	Lower bound	Normal	Upper bound
1978-79	1830	1229 (32.8)	1937 (−5.8)	1403 (22.3)	1763 (3.36)	1431 (21.8)	1655 (9.56)	—
1979-80	1820	1240 (32.2)	1960 (−7.1)	1416 (22.[illegible])	1784 (2.53)	1453 (20.6)	1673 (8.58)	—
1980-81	1984	1249 (37.1)	1983 (0.1)	1429 (27.9)	1803 (9.12)	1474 (25.7)	1690 (14.8)	—

(Figures in parenthesis are the percentage bias of the corresponding estimates)
* 95 percent Confidence Interval

REFERENCES

Hills, R. C. and Morgan, J. H. T. 1981. Rainfall Statistics : An Interactive approach to analysing rainfall records for Agricultural purposes. Experimental Agriculture, 17 : 1-16.

Rao, V. M., Nadkarni, M. V. and Deshpande R. S. 1980. Measurement of Growth and Fluctuations in Crop Output-An Approach based on the concept of Non-Systematic Component. Indian Journal of Agricultural Economics, 35 : 21-30