Effect of Urbanization on Agriculture in India

Pramanik, Chanchal
National Academy of Agricultural Research Management, Information and Communication Management Division
Rajendranagar
Hyderabad (500407), India
E-mail: cpramanik@gmail.com

Sarkar, Ananta
National Academy of Agricultural Research Management, Information and Communication Management Division
Rajendranagar
Hyderabad (500407), India
E-mail: anantasarkar@naarm.ernet.in

Introduction

Agriculture is the backbone of Indian economy, which is providing livelihood for about 65 to 70 per cent of total population and employs about 52 per cent of country's workforce and presently contributing nearly 17.5 per cent to GDP (2009). India is the second most populated country in the world after China. Presently the study on India is concerned with effect of urbanization, population growth and total cropped area on total food grain production based on 60 years of data (1950-51 to 2009-10). The data of the above mentioned variables is collected from www.indiastat.com. Here Multiple Linear Regression, Auto Regressive Integrated Moving Average (ARIMA) and Markov Chain modeling techniques are used. Similar studies were carried out on Andhra Pradesh state of India which showed that the urbanization and population growth have adverse effect on the agriculture of the state. It is the aim of this research investigation to see if the condition of overall India is similar to that of Andhra Pradesh.

Methodology

The data have been collected online from www.indiastat.com from 1950-51 to 2009-10 on Total Food Grain production (TFG), Total Cropped Area (TCA), Urbanized Area (UA) and Population. Then on the basis of 60 years of data a Multiple Linear Regression (MLR) has been fitted taking TFG as dependent variable and TCA, UA and Population as independent variables. The statistical significance of the model is determined based on Coefficient of Determination (R^2). The significance of the estimated independent variables is tested based on the respective p-values, taking 5 percent level of significance as standard.

Next after fitting the MLR equation ARIMA models are fitted to forecast all the variables involved in the study. Based on the respective ARIMA models TFG, TCA, UA and Population are forecasted for next 11 years (2010-11 to 2020-21).

Now based on the TCA, UA and Population forecasted values TFG values are obtained based on the previous fitted MLR equation for 2010-11 to 2020-21. Here two sets of TFG values have been obtained, one from the time series analysis of ARIMA model and other from fitted MLR equation where independent variables are estimated based on ARIMA model. Differences of two sets of TFGs are calculated and fitted with a suitable trend equation.

The forecasted TFGs based on MLR up to 2020-21 consider the trend of TCA, UA and Population which signifies the possible TFGs. The forecasted TFGs based on ARIMA technique up to 2020-21 signify the expected TFGs considering the present trend of TFG. The differences between two sets of TFGs signify the shortage of Total Food Grain production due to the present trend of UA, TCA and Population.

Statistical Methodology

Multiple Linear Regression analysis (MLR) has been explored to establish relation between the TFG and TCA, UA and Population. Ordinary Least Square (OLS) technique has been used to estimate the coefficients of independent variables. Coefficient of Determination (R^2) has been calculated to test the significance of the MLR equation. To test the significance of the estimates of independent variables respective p-values has been compared with 5 percent level of significance. The general form of the MLR equation is as follows:

$$Y = b_0 + b_1 X_1 + \cdots + b_k X_k \quad (1)$$

where,

Y = Dependent variable

X_i = Independent variable (i =1, 2,…, k)

b_0 = Constant

b_i = Regression coefficient of X_i

Fitting of Auto Regressive Integrated Moving Average (ARIMA) Model

Time series models are developed under the assumption of stationarity of the data.

Stationarity: A stochastic process is said to be stationary if its mean and variance are constant over time and the value of the covariance between the two time periods depends only on the disturbance or lag at which the covariance is computed (Damodar Gujarati, 1995).

ARIMA model is a combination of AR and MA models with suitable order of differencing. Hence before describing ARIMA model, it is essential to know AR and MA models.

Auto Regressive (AR) Model:

$$Y = a + b_1 Y_{t-1} + \cdots + b_p Y_{t-p} + U_t \quad (2)$$

where,

Y_t = The values of the variable for forecasting at time 't'

a = Constant

b_i = i^{th} regression Coefficient, (i = 1,2,…,p)

U_t = Random error

This above model is called AR (p) model or AR (p) process.

Moving Average (MA) Model:

Sometimes residuals with different lags may exhibit relationships with the dependent variable, as follows:

$$Y_t = \mu + \varphi U_t + \varphi_1 U_{t-1} + \varphi_2 U_{t-2} + \cdots + \varphi_q U_{t-q} \quad (3)$$

where,

Y_t = The values of the variable for forecasting at time 't'

μ = Constant

ϕ_l = Partial regression coefficient, (l = 0, 1, 2,…,q)

This model is known as MA (q) model.

Auto Regressive Moving Average (ARMA) Model

In this model Y_t depends on AR as well as MA variables and can be specified as:

$$Z_t - b_1 Z_{t-1} - \cdots - b_p Z_{t-p} = U_t - \varphi_1 U_{t-1} - \cdots - \varphi_q U_{t-q} \quad (4)$$

where,

$Z_t = Y_t - \bar{Y}$ (deviation of Y_t from mean $\bar{Y}$)

The above model is ARMA(p,q) model.

Auto Regressive Integrated Moving Average (ARIMA) Model

In the above models it was assumed that the error U_t is random (white noise) *i.e.*, the data is stationary. However in general the data are not stationary [A stochastic process is said to be stationary if its mean and variance are constant over time and the value of the covariance between the two time periods depend only on the disturbance or lag between the two line periods and not on the actual time at which the covariance is computed (Gujarati, 1995)]. Data under study may often exhibit trend, seasonal and cyclic fluctuations. In the presence of these components the time series models described above cannot be applied. These components can be eliminated. For instance if the data exhibits trend it can be estimated by fitting a suitable trend equation and finally it can be eliminated from the data. Similarly, when seasonality is present it can be removed by successive differencing. These two situations are respectively referred as trend stationary and difference stationary models.

The difference stationary model leads to ARIMA model. Here, the non-stationary time series is reduced to stationary by selecting suitable order of differencing. An ARIMA model is therefore an ARMA model with suitable differencing (d) (Box and Jenkins, 1978). ARIMA model essentially requires identification of the three constants p, d, q *i.e.*, the order of AR terms (p), order of differencing (d) and the order of MA terms (q).The ARIMA(p, d, q) model is then formulated as:

$$Z_t - b_1 Z_{t-1} - \cdots - b_p Z_{t-p} = U_t - \varphi_1 U_{t-1} - \cdots - \varphi_q U_{t-q} \quad (5)$$

where,

$Z_t = Y_t - \bar{Y}$ (deviation of Y_t from mean $\bar{Y}$)

Results and Discussions

A Multiple Linear Regression (MLR) equation has been fitted by taking Total Food Grain (TFG) production as dependent variable and Total Cropped Area (TCA), Urbanized Area (UA) and Population as independent variables based on 60 years of data (1950-51 to 2009-10). The MLR equation is as follows:

TFG = -159119909.30 + 1.18*TCA -0.63*UA + 0.20*Population (6)

Table 1: Model Summary of the MLR

Parameter	*Estimate*	*Standard Error*	*T-Statistic*	*P-Value*
CONSTANT	159119909.30	37445949.11	-4.25	<0.0001
UA	-0.63	1.77	-0.35	0.7245
TCA	1.18	0.35	3.33	0.0015
Population	0.20	0.03	8.03	<0.0001

Table 2: ANOVA of the MLR

Source	*Df*	Sum of Squares	*Mean Square*	*F-Ratio*	*P-Value*
Model	3	$1.70*10^{17}$	$5.68*10^{16}$	696.929	<0.0001
Residual	56	$4.56*10^{15}$	$8.15*10^{13}$		
Total	59	$1.75*10^{17}$			

$R^2 = 0.9709$ which justifies that 97.09% variability of TFG is explained by Population, UA, TCA. The source of rest variation of 2.91% of TFG may be due to some errors or other factors involved that are not included in the model which influences TFG.

Explanation of the estimates:

- For 1 person increase in population, Total Food Grain (TFG) production must be increased by 0.20 tonne to meet the demand in a year.
- For increase in 1 ha urbanized area, in present trend, 0.63 tonne of Total Food Grain production will be decreased.
- For 1 ha increase in Total Cropped Area (TCA), in present trend, 1.18 tonnes of Total Food Grain (TFG) production can be increased.

Table 3: Summary of ARIMA Models for UA, AA, Population & TFG

Variables	**ARIMA(p,d,q)**	**ARIMA Model Summary**				
		Parameter	*Estimate*	*Stnd. Error*	*t-statistic*	*P-value*
UA	ARIMA(1,1,2)	AR(1)	0.96037	0.0393	24.4114	<0.0001
		MA(1)	1.10111	0.1657	6.6467	<0.0001
		MA(2)	-0.32670	0.1618	-2.0194	0.0664
AA	ARIMA(1,0,1)	AR(1)	0.98660	0.0241	40.9075	<0.0001
		MA(1)	0.30218	0.1522	1.9853	0.0705
		Intercept	132456083	11900195	11.1306	<0.0001
Population	ARIMA(1,2,1)	AR(1)	-0.05113	0.1875	-0.2727	0.7893
		MA(1)	0.80934	0.1175	6.8883	<0.0001
TFG	ARIMA(0,1,1)	MA(1)	0.68246	0.1216	5.6146	<0.0001
		Intercept	3054288	440476	6.9341	<0.0001

Table 4: Forecasted values of different variables based on ARIMA Models

Year	**TFG**	**TCA**	**UA**	**Population**
2010-11	229675474.8	140868248.9	25791525.01	1239186469
2011-12	232729762.5	140755490.7	25931822.87	1261932520
2012-13	235784050.2	140644243.9	26066560.91	1284671012
2013-14	238838337.9	140534488.3	26195959.45	1307409892
2014-15	241892625.7	140426203.9	26320230.09	1330148751
2015-16	244946913.4	140319370.9	26439576.04	1352887611
2016-17	248001201.1	140213970	26554192.47	1375626471
2017-18	251055488.8	140109981.9	26664266.79	1398365332
2018-19	254109776.5	140007387.6	26769979	1421104192
2019-20	257164064.2	139906168.6	26871501.98	1443843052
2020-21	260218352.0	139806306.3	26969001.73	1466581913

Now based on the forecasted values of TCA, UA and Population TFGs are estimated on the basis of the basis of the equation (6) for 12 years (2009-10 to 2020-21).

Table 5: Comparative study of two sets of forecasted TFGs based on MLR & ARIMA

Year	**TFG** *(in tonnes)* (MLR method)- (I)	**TFG** *(in tonnes)* (ARIMA method) -(II)	**D** (I)-(II) *(in tonnes)*
2010-11	241924814.8	229675474.8	12249340.06
2011-12	246317526.6	232729762.5	13587764.12
2012-13	250713966.5	235784050.2	14929916.32
2013-14	255115584.0	238838337.9	16277246.07
2014-15	259522140.6	241892625.7	17629514.97
2015-16	263933490.1	244946913.4	18986576.76
2016-17	268349487.2	248001201.1	20348286.11
2017-18	272769991.9	251055488.8	21714503.09
2018-19	277194869.2	254109776.5	23085092.67
2019-20	281623988.9	257164064.2	24459924.63
2020-21	286057225.3	260218352.0	25838873.30

The difference (D) values exhibits a linear trend which is as follows:

$$D=10855451.32+1359046.57*T \quad (7)$$

where, T = time

Table 6: D- trend model summary

Parameter	*Estimate*	*Standard Error*	*t-statistic*	*P-value*
Constant	10855451.32	14558.07	745.67	<0.0001
Slope	1359046.57	2146.47	633.15	<0.0001

Thus it is quite evident that the differences of the possible and expected TFGs are increasing in a steep linear trend. Hence if the same trend of TCA, UA and Population exists there will be no serious problem for regarding food security in next 10 years.

Conclusion

The detailed study on India shows that the urbanization has no effect on Indian agriculture in near future. Probably the Green Revolution and innovative agricultural technologies have dampened the effect of urbanization and population growth on agriculture. But in the earlier studies on Andhra Pradesh state of India (Pramanik *et al.*, 2010) it was found that the urbanization and population growth has serious effect on agriculture. Andhra Pradesh besides agriculture also contributes significantly on industry sector in particular to software industries. So Andhra Pradesh is a second home state for many people. This may be the reason why urbanization and population have adverse effect on agriculture. Since Andhra Pradesh ranks third in total agricultural production it may have some ripple effect on India in distant future. The similar project can also be run for other countries. This may be bookmarked for a future assignment.

References

Bhattacharya A. (2006). A goal Programming Approach for Developing Pre-Harvest Approach in Crop Yield. The Journal of Operational Research Society. 57(8), 1014-1017.

Dey S. K., C Pramanik and C Dey. Mathematical Modelling of Effects of Urbanization and Population Growth on

Agricultural Economics. Journal of Intellectual Economics. 2(8), 7-20.

Dobre I and Alexandru A A M (2008). Modeling unemployment rate using Box-Jenkins Procedure. Journal of Applied Quantitative Methods, 3(2), 156-166.

Ferguson C. E. (1958). A Statistical Study on Urbanization. Social Forces, 37(1), 19-26.

Gopinath C. K. (2000). An indirect method for forecasting the annual food production in India. Current Science. 79(1), 23-24.

Iqbal N, Bakhsh K, Maqbool A and Ahmad A S (2005). Use of ARIMA model for forecasting wheat area and production in Pakistan. Journal of Agriculture and Social Sciences. http://www.ijabjass.org. 1813–2235/2005/01–2, 120-126.

Kumari G. K. S. (1997). Forecasting and Estimation of Rainfall. M. Sc. (Ag.) Statistics thesis. Acharya N. G. Ranga Agricultural University.

Pramanik Chanchal (2010). Study of Fertilizer Consumption Pattern in Districts of Northern Telangana Region of Andhra Pradesh. M.Sc. (Ag.) Statistics thesis. Acharya N. G. Ranga Agricultural University.

Pramanik Chanchal, S K Dey and Sarkar A. (2010). Effect of Urbanization on Agriculture: A Special Scenario on Andhra Pradesh, India. International Journal of Applied Science & Computations, 17(2), 121-128.

Abstract

Increase in urbanization in the state of Andhra Pradesh, India has a negative impact in total food grain production in near future (Pramanik et al., 2010). Therefore, the present study has been formulated to account for the effect of urbanization, population growth and total cropped area on total food grain production in India based on 60 years of data (1950-61 to 2008-09). Multiple Linear Regression has been explored to explain the relationship between total food grain production and urbanized area, population growth and total cropped area. Autoregressive Integrated Moving Average (ARIMA) models have been used to develop forecasting models and to forecast the variables involved in the study up to 2020. Later the multiple linear regression equation has been used to forecast the possible total food grain production based on the forecasted figures of the urbanized area, population growth and total cropped area. Separate forecast obtained for the total food grain production from the ARIMA model based on the trend is the expected total food grain production. Differences between possible and expected total food grain productions are obtained for the years 20010-11 to 2020-21. It has been observed that differences have a steep upward linear trend. This is a positive aspect in case of Indian agriculture. So within the next 10 years urbanization may not be a serious issue but the situation in Andhra Pradesh may have some ripple effect in distant future, as the state has significant contribution in Indian agriculture.