IOSR Journal of Mathematics (IOSR-JM)
ISSN: 2278-5728. Volume 4, Issue 1 (Nov. - Dec. 2012), PP 01-06
www.iosrjournals.org

Modeling of Sugar Production in India based on Principal Component

Shivagaje Ashok.J.[1] , Kasture Madhukar.S.[2]
[1]*Department of Statistics, Mahatma Phule Krishi Vidyapeeth, Rahuri, (M.S), India*
[2]*Department of Statistics, New Arts, Commerce & Science College, Ahmednagar,(M.S), India*

ABSTRACT: *Sugar as a luxurious product now becomes an essential one. Sugarcane is the leading cash crop in India used for sugar extraction. Sugarcane was cultivated on 4 million hectares of land shows increasing trend of production over last 10 years, in India. Sugar production scenario for Pre-WTO and Post-WTO period in India, is studied. The effect of eight predictors on sugar production in Pre-WTO (1985-86 to 1994-95) and Post-WTO period (1995-96 to 2008-09) is analyzed. Matrix plot exhibits the scatterness in predictor and response variables. Principal component analysis (PCA) is used to reduce the dimensionality of the data. Multiple regression models, for sugar production on combined score of selected predictors, are obtained. The significance of auto-correlation in residuals of sugar production in India using multiple regressions is tested by Durbin-Watson test. The analysis reveals that in Pre-WTO, Post-WTO and overall period sugar production in India increases as a combined effect of four selected predictors. Durbin-Watson test indicates that autocorrelation in residuals is insignificant in these periods. The calculations of uncrushed sugarcane in Pre-WTO period is of the order of 50 percent which reduces to 40 percent in Post-WTO period.*

Keywords - *Autocorrelation, eigen values, matrix plot, principal component, residual.*

I. Introduction

"*Honey produced without bees*", means producing sugar by evaporating the cane juice, the process is known to Indian in around 500 BC. India is the second largest producer of sugar in the world after Brazil and is indulged in the production of cane sugar whereas Brazil in beet sugar. In 1960-61 the sugar production in India was 3.021 million tons (MT) which rose to 24.5 MT in 2010-11. Global sugar scenario is studied and then discussed the issue regarding India's sugar export potential and concluded that India was sugar bowl of the world having carryover stock of 11.7 million tons in 2001-02 and became rising sugar exporter and proposed four issues for the adjustment of sugar economy to trade dominated scenario [1]. The sugarcane acreage of India was increased from 2415 thousand hectares to 4944 thousand hectares during the same period. Further, the productivity of sugarcane increased from 45.5 tons per hectare to 68.6 tons per hectare. Meaning thereby with doubling acreage, production becomes eight times more but productivity is just 1.5 times. Naturally, the credit of this success goes to breeding and evolving high yielding varieties, crop production technologies, creation of water resources and establishment of sugar factories.

The impact of WTO on agricultural development of India is studied and highlighted the implications of world trade organization on agricultural sectors in India [2]. The sugarcane scenario of India on the basis of percent compound growth rate of area, production and productivity during Pre-WTO and Post-WTO is summarized in table 1

TABLE 1: PERCENT COMPOUND GROWTH RATES OF AREA, PRODUCTION AND PRODUCTIVITY OF SUGARCANE IN INDIA.

Periods	Area	Production	Productivity
Pre-WTO	2.92**	4.41**	1.57**
Post-WTO	1.34	0.81	-0.51
Overall	1.86	2.31**	0.38*

(**, * - Significant at 1% and 5% level of significance respectively)

From table 1 it is observed that in Pre-WTO period area, production and productivity was significantly increasing every year, whereas such a consistency is not observed in Post- WTO period, which may be addressed as a one of the cause of slow increasing trend of sugar production in India.

Apart from area, production and productivity there are some of the factors like number of sugar factories in operation, cane crushed, percent sugar recovery, average actual capacity of sugar factories and average duration of sugar factories that may have effect on sugar production. As on 30-06-2009, 626 sugar factories including 62 public sectors, 247 private sectors and 317 co-operative sectors were installed in India. In

Pre-WTO total cane crushed ranges between 68556 thousand tons to 147643 thousand tons, the corresponding range in Post- WTO period is 124771 thousand tons to 279249 thousand tons. The percent recovery of sugarcane in Pre-WTO period lies between 9.70 to 10.30, the corresponding range in Post- WTO period was 9.40 to 10.55.The average actual capacity (tons/24 hour) in Pre-WTO period was between 1862 to 2483, which rose to 2531 to 3725 in Post- WTO period. In the present article using principal component analysis an attempt is made to extract the important factors and model the sugar production based on these extracted factors.

II. MATERIAL AND METHOD

The secondary data relating to India is collected on area, production, productivity of sugarcane, number of sugar factories in operation, cane crushed, percent sugar recovery, average actual capacity of sugar factories and average duration in days of sugar factories, for Pre-WTO and Post-WTO period, from different publications of Indian Sugar, Indian Sugar Mills Association, Centre for Industrial and Economic Research (CIER) and the website http:\\www.indiastat.com which is used for the present analysis.

In many practical applications predictor variables are correlated and hence multi co linearity plays the role in multiple regressions. In order to justify the validity of assumptions in multiple regressions author had used principle component technique [4], which is also one of the powerful data reduction tool.

Among all variables sugar production is response variable and remaining eight are predictor variables. In order to investigate the possible scatterness in predictor variables and response variable period wise matrix plot is drawn. The usual multiple regression of sugar production on the eight predictor variables is applied to study the significance of the model as well as the significance of regression coefficients of eight predictor variables. Since all the eight predictor variables logically seems to be useful for modeling sugar production, the technique of principal component analysis (PCA) is applied to classify predictor variables into significant groups. To test the validity of assumption of normality, the significance of auto-correlation in residuals of sugar production in India using multiple regressions is tested by Durbin-Watson test. Period wise geometric mean of percentage uncrushed sugarcane is also calculated.

III. RESULT AND DISCUSSION

3.1 Multiple Regression Analysis

The matrix plot of order 9x9 showing the possible scatterness between response as well as within predictor variables for Pre-WTO and Post-WTO period is shown in Figure 1 and Figure 2 respectively. Matrix plots in Figure 1 and Figure 2 exhibits that response variable sugar production shows increasing trend with respect to increase in predictor variables namely; area, production, number of factories in operation and cane crushed. The scatterness in other variables does not show any trend of sugar production on the eight predictor variables is obtained as a part of explanatory analysis. The multiple regression models for Pre-WTO, Post-WTO and overall period of 24 years are presented in Table 2.

Figure 1 : Matrix plot for Pre-WTO period

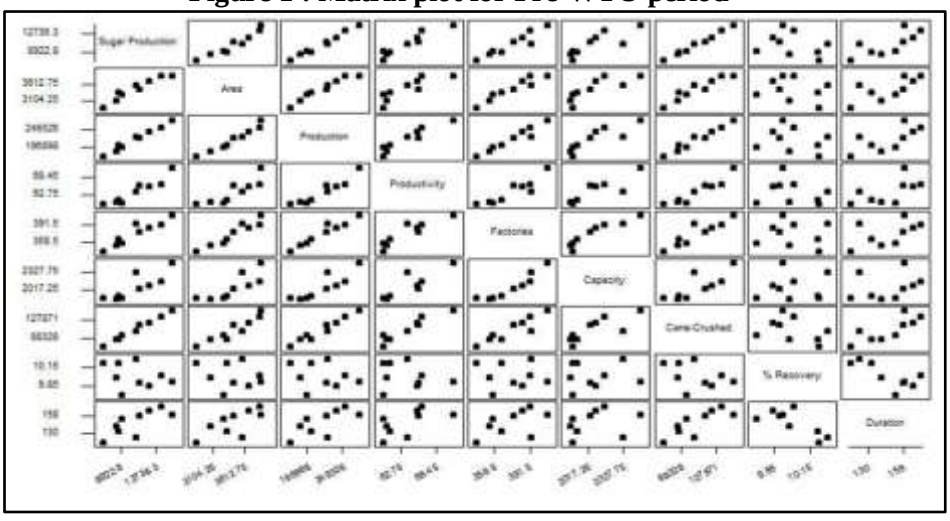

Figure 2 : Matrix plot for Post-WTO period

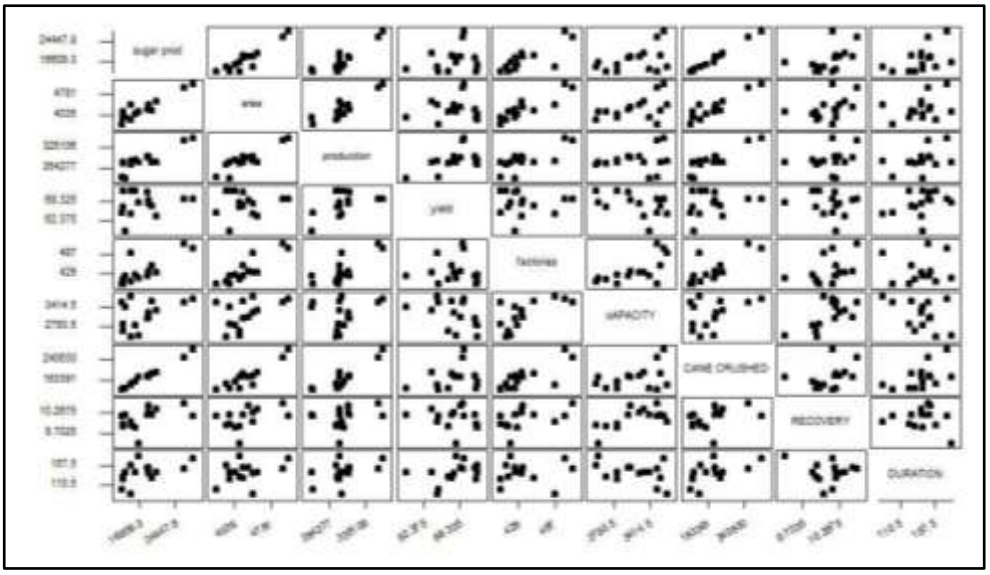

TABLE 2: PERIOD WISE MULTIPLE REGRESSION MODELS OF SUGAR PRODUCTION ON PREDICTORS

Sr. No	Predictor Variables	Pre-WTO period		Post-WTO period		Overall Period	
		Coefficient	P-value	Coefficient	P-value	Coefficient	P-value
1	Constant	31788.87	0.2305	-15783.64	0.0157	-14086.29	0.0049
2	Area	-12.46	0.2061	-0.53	0.5233	-0.52	0.5855
3	Production	0.21	0.1840	0.01	0.6772	0.01	0.4640
4	Productivity	-789.88	0.1774	-34.18	0.5518	-49.89	0.4147
5	No. of factories in operation	18.42	0.5159	2.51	0.3000	0.00	0.9987
6	Capacity	0.22	0.8650	-0.49	0.0786	-0.04	0.7907
7	Cane Crushed	0.08	0.0630	0.11	0.0000	0.10	0.0000
8	% Recovery	925.05	0.2080	1924.46	0.0000	1651.04	0.0000
9	Duration in days	11.65	0.3016	-5.30	0.1510	0.14	0.9650
R^2		0.99989		0.99910		0.99947	
P-value		0.02297		1.098E-09		3.486E-23	

From Table 2 it is observed that all the three multiple regression models are acceptable at 5 percent level of significance with predictive power of each model is more than 99.9 percent.

The predictor variable wise partial regression co-efficient indicates that in Pre-WTO period none of the predictor significantly contributes to sugar production, whereas, in Post-WTO period and Overall Period except cane crushed and percent recovery none of the predictor significantly contributes to sugar production. It implies that even though three regression models are acceptable but they are not suitable because in all the three models sugar production decreases as area and productivity of sugarcane increases. This conclusion seems to be contradictory because model considers the effect of area and productivity on sugar production individually when other variables are treated as constant. Further the regression coefficients in the model of Post-WTO period shows that increase in capacity and duration in days of sugar factories affects sugar production. In order to overcome the contradiction observed in application of multiple regression models and to suggest the adequate model, technique of principal component analysis is used to classify the predictor variables into different groups having information contained in it in decreasing order.

3.2 Principal Component Analysis based Multiple Regression Model

The data mining unsupervised learning methods includes principle component analysis as a tool of dimensionality reduction. The conceptual basis and detail mathematical treatment of principal component is used as a methodology for identifying the redundancy is available in literature [3]. As different predictor variables are measured in different scales principal component analysis is carried out on the basis of correlation matrix. The principal component analysis for Pre-WTO, Post-WTO and overall period reveals that first two eigen values are greater than unity and hence first two principal components are dominating. Table 3 presents eigen values and loads of different predictor variables in first two leading principal components.

TABLE 3: PERIOD WISE LOADS OF DIFFERENT PREDICTOR VARIABLES IN FIRST TWO PRINCIPAL COMPONENTS

Sr.No	Predictor Variables	Pre-WTO period		Post-WTO period		Overall Period	
		PC-I	PC-II	PC-I	PC-II	PC-I	PC-II
1	Eigen value	6.092	1.545	4.234	2.172	5.054	1.626
2	Proportion	0.761	0.193	0.529	0.272	0.632	0.203
3	Cumulative proportion	0.761	0.954	0.529	0.801	0.632	0.835
4	Area	0.392*	-0.031	0.470*	0.007	0.435*	0.024
5	Production	0.403*	-0.060	0.450*	0.207	0.435*	0.129
6	Productivity	0.385	-0.094	0.143	0.477	0.292	0.347
7	No. of factories in operation	0.388*	-0.206	0.433*	-0.183	0.425*	-0.105
8	Capacity	0.345	-0.374*	0.227	-0.561*	0.378	-0.314*
9	Cane Crushed	0.402*	0.083	0.466*	0.085	0.425*	0.106
10	% Recovery	-0.130	-0.727*	0.234	-0.404*	0.169	-0.539*
11	Duration in days	0.296	0.580*	0.211	0.460*	0.061	0.671*

From table 3 it is observed that in Pre-WTO, Post-WTO and overall period the loads of predictor variables area, production, number of factories in operation and cane crushed are dominating in first principal component, whereas except productivity the loads of remaining predictor variables is dominating in second principal component. In Pre-WTO period the two principal components together explains 95.4 percent variation in sugar production. Similarly in Post-WTO and overall period the two principal components together explain 80.1 and 83.5 percent variation in sugar production respectively. In the further analysis the predictor variable productivity is dropped because area and production will jointly take care of productivity.

The score equations of two principal components in three study periods based on standardized predictor variables [(variable-mean)/SD] are given in table 4.

TABLE 4: PERIOD WISE SCORES OF FIRST TWO PRINCIPAL COMPONENTS

Period	Score	Score Equation
Pre-WTO	Z_1 =	0.392 *Area +0.403*Production +0.388*No. of factories in operation +0.402* Cane Crushed
	Z_2=	-0.374*Capacity -0.727*% Recovery +0.580*Duration in days
Post-WTO	Z_1 =	0.470 *Area +0.450*Production +0.433*No. of factories in operation +0.466* Cane Crushed
	Z_2=	-0.561*Capacity -0.404*% Recovery +0.460*Duration in days
Overall	Z_1 =	0.435 *Area +0.435*Production +0.425*No. of factories in operation +0.425* Cane Crushed
	Z_2=	-0.314*Capacity -0.539*% Recovery +0.671*Duration in days

Using the score equations of two principal components as shown in table 4, scores are computed for three study periods and multiple regression of sugar production on scores is revisited. The results of which are shown in table 5.

TABLE 5 : PERIOD WISE MULTIPLE REGRESSION MODELS OF SUGAR PRODUCTION ON SCORES

Sr.No.	Predictor Variables	Pre-WTO period		Post-WTO period		Overall Period	
		Coefficient	P-value	Coefficient	P-value	Coefficient	P-value
1	Constant	10490.70	8.22E-11	17707	2.33E-13	14700.2	1.4E-23
2	Z_1	1452.395	6.01E-06	2696.63	6.5E-07	3080.5	5.5E-14
3	Z_2	211.2728	0.2378	350.016	0.5980	84.3207	0.7542
R^2		0.9586		0.9045		0.9374	
P-value		1.44E-05		2.45E-06		2.3E-13	

From Table 5 it is observed that the score Z_2 is non significant, implies that predictors capacity, percent recovery and duration in days does not have any effect on sugar production. This is logically acceptable because the information on capacity does not throw light on its full or partial utilization, which is practically depends upon availability of sugarcane. Similarly percent recovery is the feature of sugarcane indicating the quality of sugarcane, having no contributory information on sugar production. Since Z_2 is non significant simple linear regression of sugar production on Z_1 is revisited with results as shown in table 6.

TABLE 6: PERIOD WISE MULTIPLE REGRESSION MODELS OF SUGAR PRODUCTION ON SCORE Z_1

Sr.No.	Predictor Variables	Pre-WTO period		Post-WTO period		Overall Period	
		Coefficient	P-value	Coefficient	P-value	Coefficient	P-value
1	Constant	10490.7	7.8E-12	17707	2.1E-14	14700.2	1.3E-24
2	Z_1	1485.47	1.9E-06	2671.27	2.1E-07	3089.01	1.1E-14
R^2		0.9488		0.9020		0.9371	
P-value		1.9E-06		2.1E-07		1.1E-14	
Durbin-Watson Statistic		1.83		0.91		1.06	

From Table 6 it is observed that in Post-WTO period the unit rise in combined effect (Z_1) of area, production, number of factories in operation and cane crushed increases sugar production by 2671.27 thousand tons, which is 1485. thousand tons in Pre-WTO period and 3089.01 thousand tons in overall period of 22 years. From table 5 it is also concluded that the multiple regression models of sugar production on Z_1 explains 94.88 percent variation in sugar production in Pre-WTO period. The predictive power of models in Post-WTO and overall period is 90.20 and 93.71 percent respectively.

The above analysis shows that out of eight predictor variables only half of the variables are significantly contributing towards sugar production. The regression models of sugar production on these predictor variables shown in table 5 also have predictive power of 95 percent or more, showing the suitability of model. Therefore an attempt has been made to test significance of autocorrelation in residuals. The Durbin-Watson Statistic as shown in table 5 supports the non significance of autocorrelation in residuals.

3.3 Uncrushed Sugarcane

The number of sugar factories in operation in Pre-WTO period was 342 in 1985-86 which rose to 408 in 1994-95, whereas in Post-WTO period 416 number of sugar factories in 1995-96 rose to 516 in 2007-08. The increase in number of sugar factories has considerably reduced the percentage of uncrushed sugarcane. The geometric mean of percentage uncrushed sugarcane in Pre-WTO period was 53.31 percent which drop down to 39.55 percent in Post-WTO period. This decrease definitely increases the sugar production in order to meet the growing demand of it due to increasing population. The uncrushed sugarcane even it is not used for sugar production but it is used for other byproducts such as *Gur*, Juice or it can also be used as a fodder for domestic animals like cow, buffalo, and bullocks.

IV. CONCLUSION

On the basis of above discussed principal component analysis based multiple regressions, it can be concluded that, combined effect of area, production, number of factories in operation and cane crushed seems to have double effect on sugar production in Post-WTO period as compare to Pre-WTO period.

The agriculture in India have great prospects for sugarcane production and hence sugar production. The percentage uncrushed sugarcane in Pre-WTO period was 53.31 percent which drop down to 39.55 percent in Post-WTO period, shows that India can reduce the gap between demand and production of sugar.

Lastly, like Brazil the feasibility of sugar production through beet may be investigated to keep another option open for sugar production.

References

[1] J.Puri. Global sugar Scenario-is India – a rising sugar Export Star? *Co-operative Sugar*, 33(5), 2002, 389-391.
[2] R.N. Tiwary, Impact of World Trade Organization on Agricultural Development of India. *Co-operative Sugar*. 35(7), 2004, 537-540.
[3] A.M. Kshirsagar, *Multivariate Analysis*, (Marcel and Dekker New York,1972)
[4] A.J.M. Sufian, Analyzing Collinear Data by Principal Component Regression Approach – An Example from Developing Countries, *Journal of Data Science* 3, 2005, 221-232.