ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jan851

An Advanced Econometric Analysis of Sugarcane Cultivation in the South Gujarat Region of Gujarat

Dr. Gunjan B. Shah[1]

[1]Assistant Professor, Department of Statistics Government Arts, Commerce & Science College Kachhal, Ta- Mahuva, Dist- Surat

Publication Date: 2026/01/24

Abstract: Sugarcane cultivation plays a pivotal role in the agrarian economy of South Gujarat, particularly in the South Gujarat region of Surat district. The present study develops an advanced econometric framework to examine the structural relationship between sugarcane output, cultivated area, and production expenditure. Using cross - sectional farm-level data collected from 90 sugarcane growers during the agricultural year 2018–19, log-linear regression models are estimated to capture elasticity - based responses of output to key inputs. Beyond conventional regression estimation, the study rigorously investigates classical linear regression assumptions through diagnostic tests for multicollinearity, heteroscedasticity, and autocorrelation. The findings reveal strong scale effects in sugarcane production, the presence of cost-output responsiveness, and structural econometric issues that limit the reliability of naïve Ordinary Least Squares estimates. The study emphasizes the need for model refinement and policy-oriented interpretation for sustainable sugarcane development in the region.

Keywords: *Sugarcane Economics, Log-Linear Regression, Multicollinearity, Heteroscedasticity, ANOVA.*

How to Cite: Dr. Gunjan B. Shah (2026) An Advanced Econometric Analysis of Sugarcane Cultivation in the South Gujarat Region of Gujarat. *International Journal of Innovative Science and Research Technology*, 11(1), 1826-1829. https://doi.org/10.38124/ijisrt/26jan851

I. INTRODUCTION

Sugarcane is one of the most important commercial and industrial crops in India, playing a decisive role in rural employment generation, agro-industrial development, and income stabilization of farming households. The South Gujarat region, comprising districts such as Surat, Navsari, Tapi, Bharuch, and Valsad, represents one of the most productive sugarcane belts of the country due to assured irrigation facilities, fertile alluvial soils, humid climatic conditions, and a strong cooperative sugar industry network.

The sugar production in India is in cycles. For every two to three years of high sugar production, there are two to three years of low sugar production. The sugarcane production in the state of Gujarat is expected to grow by 13% in the year 2025. Sugar industries in the state of Gujarat expect an average production level of sugar this year because the sugar recovery from sugarcane is lower this year. The average annual production of sugarcane in the state of Gujarat in India was approximately 82 thousand kilograms per hectare in the year 2025. The production of sugarcane in India was about 83 thousand kilograms per hectare in 2025.

In applied econometrics, the Cobb–Douglas type log-linear production function is extensively used to estimate output elasticities and to evaluate the nature of returns to scale in agricultural production. The logarithmic transformation not only facilitates elasticity interpretation but also helps stabilize variance and linearize inherently non-linear economic relationships. However, farm-level cross-sectional data often exhibit heterogeneity across holdings, leading to violations of classical linear regression assumptions such as homoscedasticity and absence of multicollinearity.

Therefore, modern econometric analysis necessitates a systematic diagnostic framework involving tests for heteroscedasticity, multicollinearity, and model adequacy before drawing economic or policy conclusions. The present study applies an advanced econometric approach to examine sugarcane production dynamics in South Gujarat, with special emphasis on statistical theory, elasticity-based interpretation, and robustness of regression results.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jan851

II. DATA BASE AND SAMPLING DESIGN

The empirical analysis is based on primary data collected from the South Gujarat taluka of Surat district, which comprises approximately 105 villages with substantial sugarcane cultivation. A stratified random sampling technique was adopted to ensure representation across different farm-size categories. Farmers were classified into three groups based on operational landholding under sugarcane cultivation: (i) 1–10 acres, (ii) 10–20 acres, and (iii) 20–30 acres.

A total sample of 90 sugarcane-growing farmers was selected from various villages. Data pertain to the agricultural year 2018–19 and were collected through structured personal interviews. The dataset captures both physical and monetary aspects of sugarcane production, enabling a comprehensive econometric examination.

III. MODEL SPECIFICATION AND METHODOLOGY

➢ *Definition of Variables*

The econometric analysis is based on the following variables:

- Y: Total sugarcane production (in tonnes)
- X1: Area under sugarcane cultivation (in acres)
- X2: Total cost of cultivation (in rupees), including seed cost, human and machine labour, irrigation, fertilizers, plant protection chemicals, and other operational expenses

To capture proportional responsiveness and stabilize variance, all variables are transformed into their natural logarithmic form.

➢ *Econometric Framework*

The general log-linear regression model employed in the study is specified as:

$$\log Y = \alpha + \beta 1 \log X1 + \beta 2 \log X2 + u$$

Where (u) denotes the stochastic disturbance term satisfying classical regression assumptions. The estimated coefficients (β1) and (β2) are interpreted as elasticities of output with respect to cultivated area and production cost, respectively.

Single-variable and multivariate regression models are estimated to isolate individual and joint effects. Model adequacy is assessed using coefficient of determination (R^2), adjusted R^2, F-statistics, and t-tests for parameter significance.

IV. CORRELATION STRUCTURE AND PRELIMINARY ANALYSIS

Pair-wise correlation analysis reveals a strong positive association between sugarcane output and cultivated area, as well as between output and production expenditure. However, a high correlation between cultivated area and total cost indicates potential multicollinearity, which may inflate standard errors and distort coefficient estimates in multivariate models. These findings justify the need for careful diagnostic testing prior to policy interpretation.

To determine the nature of linear relationship among the dependent and independent variables, two methods are applied. (i) Graphical method and (ii) Pair-wise correlation method.

➢ *Graphical Representations:*

To examine the nature of linear relationships among the variables, scatter diagrams were plotted for the pairs of variables: (log Y1, log X1) and (log Y1, log X2) based on the collected data. These scatter diagrams are presented in Figure 1.

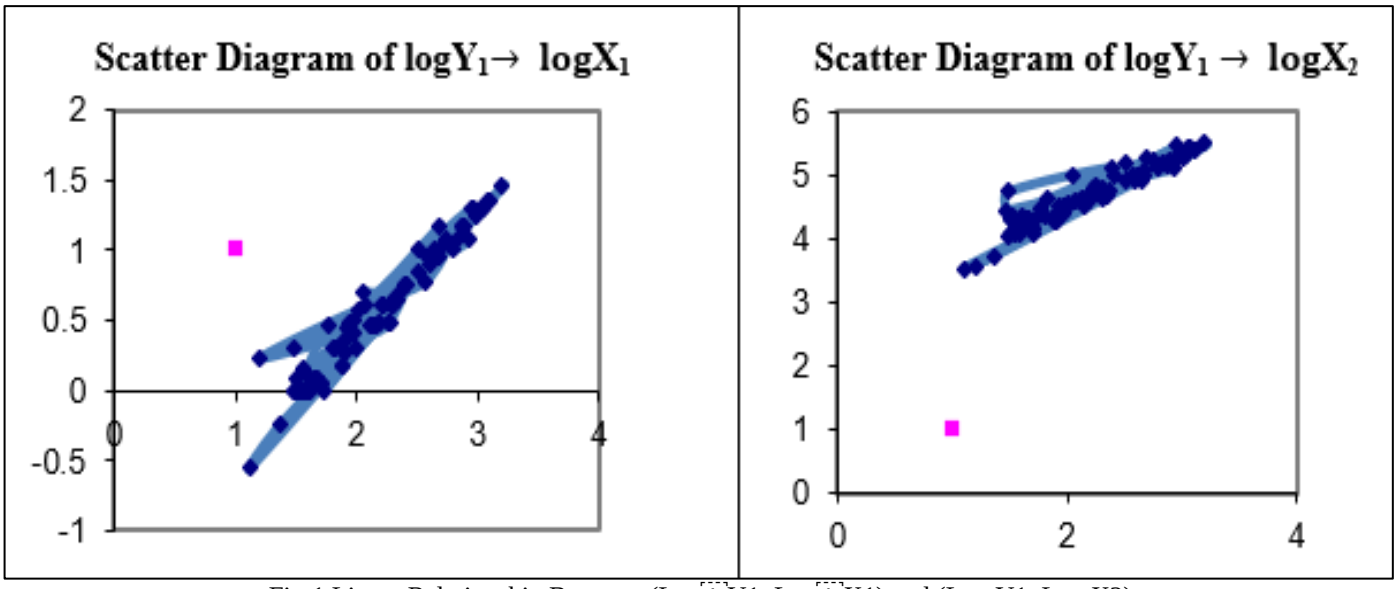

Fig 1 Linear Relationship Between (Log Y1, Log X1) and (Log Y1, Log X2)

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jan851

From the figure 1, researcher can interpret that there is a linear relationship among independent and dependent variables under examination.

➢ *Pair-Wise Correlations Study:*

Correlation co-efficient of logX and logY are calculated from the independent and dependent variables and are presented in table 1.

Table 1 Correlation Co-efficient of logX and logY

	$logX_1$	**$logX_2$**
$logY_1$	0.9736	0.9194
$logX_1$	-	0.9231

Here the variables ($logY_1$, $logX_1$) and ($logY_1$, $logX_2$) are highly perfect correlated. Therefore, there is a positive relationship among them. The variables $logX_1$ and $logX_2$ have very perfect correlation. So it is not advisable to include both these variables in studying the variations in $logY_1$.

- The variables: (log Y1, log X1) and (log Y1, log X1) are highly correlated, Conforming a strong linear relationship with the dependent variables.
- The Variables log X1 and log X2 are also highly correlated (r= 0.9231) indicating multicollinearity.

V. FITTING OF LINEAR REGRESSION MODELS (ADVANCED APPROACH)

This section investigates the underlying linear relationship between the dependent variable log Y1 and the independent variable log X1 using regression analysis. Regression modeling provides insights into the strength, direction, and statistical significance of these relationships.

➢ *Regression Model of log Y1 on log X1*

- *Model Specification*

We consider the simple linear regression model:

$$\log Y1 = \beta 0 + \beta 1 \log X1 + \varepsilon$$

Where:

✓ β0 = intercept,
✓ β1= slope coefficient of log X1\log X_1logX1,
✓ ε = random error term with E[ε]=0 and Var(ε)=σ2.

- *Regression Statistics*

Table 2 Summary of Regression Fit

Metric	**Value**
Multiple R	0.9737
R^2	0.9480
Adjusted R^2	0.9474
Number of Observations	90

✓ The Multiple R indicates a very strong linear correlation between log Y1 and log X1.
✓ Adjusted R^2 = 0.9474 suggests that approximately 95% of the variation in log Y1 is explained by log X1, confirming the model's explanatory power.

- *ANOVA for Regression Model*

Table 3 ANOVA Summary

Source	**d.f.**	**Sum of Squares (SS)**	**Mean Square (MS)**	**F-value**
Regression	1	23.7598	23.7598	1605.402
Residual	88	1.3024	0.0148	-
Total	89	25.0622	-	-

✓ The high F-value (1605.402) with p<0.01p < 0.01p<0.01 indicates that the regression model is statistically significant.
✓ Thus, log X1\log X_1logX1 is a strong predictor of log Y1.

- *Regression Coefficients*

Table 4 Estimated Coefficients

Coefficient	**Estimate**	**Std. Error**	**t-statistic**	**Significance**
Intercept	1.55153	0.01927	80.53	p < 0.01
log X1	1.11428	0.02781	40.07	p < 0.01

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jan851

- ✓ The slope coefficient (β1=1.11428) is highly significant (t=40.07), confirming that logf_0X1 has a substantial effect on logf_0Y1.

- *Fitted Regression Equation:*

$$\log f_0 Y1 = 1.55153 + 1.11428 \log f_0 X1$$

➢ *Diagnostics and Model Assumptions5*

- *Autocorrelation Check*

Since this dataset is cross-sectional (no time-series component), autocorrelation is unlikely to be present.

- *Heteroscedasticity Test (Goldfeld-Quandt Test)*

- *Hypotheses:*

- ✓ H0: $\sigma 1^2 = \sigma 2^2$(homoscedasticity)
- ✓ H1: $\sigma 1^2 \neq \sigma 2^2$(heteroscedasticity)

- *Procedure:*

- ✓ Sort observations in ascending order of logf_0X1.
- ✓ Divide into two equal groups (33 observations each, middle 24 excluded).
- ✓ Fit regression for each group and calculate residual sum of squares (RSS1,RSS2).

- *Test Statistic:*

$$Fc = \frac{RSS1}{RSS2} = 2.2500$$

- ✓ Degrees of freedom: n1=n2=31
- ✓ Critical F-values: $F_{0.05}$=1.74, $F_{0.01}$=2.20

- ✓ Decision: Since Fc>Ft , the null hypothesis H0H_0H0 is rejected.
- ✓ Conclusion: Residuals exhibit heteroscedasticity. Appropriate remedial measures (e.g., weighted least squares or robust standard errors) may be considered.

➢ *Interpretation*

- The fitted model demonstrates a strong linear relationship between logf_0Y1 and logX1.
- The explanatory variable logf_0X1 accounts for over 94% of the variation in logf_0Y1.
- Regression diagnostics indicate heteroscedasticity, but no autocorrelation.
- The model is reliable for predictive and inferential purposes, provided heteroscedasticity is addressed.

VI. DISCUSSION AND POLICY IMPLICATIONS

The empirical findings underscore the dominance of scale effects in sugarcane production in the South Gujarat region. Cultivated area emerges as the most influential determinant of output, while production cost reflects intensity of input use. However, the presence of heteroscedasticity and multicollinearity highlights structural complexities inherent in farm-level agricultural data.

From a policy perspective, the results suggest that productivity enhancement strategies should focus on cost-efficient input management rather than mere expansion of cultivated area. Extension services, mechanization support, and rational input pricing can improve output responsiveness while containing cost escalation.

VII. SUMMARY AND CONCLUSION

Analysis of the above statement gives the fact that the relation of the variable is linear with two independent variables and and the line of regression is a good fit in both the situations. But, in the two situations, hetroscedasticity is found to exist and hence it is not recommendable to use the two models in determining the relation of log Y1 and log X1 and .

In case of regression of log Y1 on log X1 and log X2 together, it has been observed that log linear regression model is a good fit. In this case, as logX1 and logX2 are largely correlated to each other, a multicolinearity situation has arisen. Therefore, it is not a good model in which the dependence relation of log Y1 on log X1 and log X2 together can be examined.

This study provides an advanced econometric examination of sugarcane cultivation in the South Gujarat region using farm-level primary data. While log-linear regression models reveal strong and statistically significant relationships, diagnostic tests caution against uncritical reliance on conventional OLS estimates. Future research may employ remedial measures such as heteroscedasticity-consistent estimators or ridge regression techniques to address identified econometric issues. Overall, the study contributes to evidence-based agricultural planning and sustainable sugarcane development.

REFERENCES

[1]. Damodar Gujarati (1988): Basic Econometrics, McGraw-Hill Book Company
[2]. J. Johnston (1995): Econometric Methods, John Wiley and Co.
[3]. Ramu Ramanathan (2001): Introductory Econometrics with Applications, 5th Edition, Course Technology Inc
[4]. Potluri Rao and R.I. Miller (1972): Applied Econometrics, Prentice Hall Co.
[5]. Tintner Gerhard (1968): Methodology of Mathematical Economics and Econometrics, the University of Chicago Press