

BSc (Economics, Mathematics and Statistics)

Batch: 2021 – 2024

Introduction to Econometrics - ECO631

CIA – 3

Understanding the Dynamics of Rural Economic Development in India: A Comprehensive Analysis (2000-2019)

Submitted by

Aleena Patrick (2140823)

Vasudha Rajaram (2140842)

Madhumitha Ramakrishnan (2140838)

Submitted to

Dr Rajeshwari U R

CHRIST (Deemed to be University)

BENGALURU

Introduction:

Rural development refers to the process of improving the economic, social, and environmental conditions in rural areas, with the aim of enhancing the quality of life for rural residents and promoting sustainable growth. It involves the implementation of policies, programs, and projects focused on various aspects such as agriculture, infrastructure, education, healthcare, and community empowerment. The goal of rural development is to address disparities between urban and rural regions, alleviate poverty, create employment opportunities, and ensure access to basic services and amenities. By fostering inclusive and equitable development, rural development contributes to overall national progress and the well-being of rural communities.

India's rural economy plays a pivotal role in the nation's overall economic development, particularly in sectors such as agriculture, forestry, and fisheries. Over the past two decades, India has witnessed significant transformations in its rural landscape, driven by various factors such as urbanization, government policies, and environmental changes. Understanding the dynamics of India's rural economy development during this period is crucial for policymakers, researchers, and practitioners seeking to promote sustainable growth and enhance livelihoods in rural regions.

This study explores the complex dynamics of India's rural economy Development over the period of 20 years, from 2000 to 2019, using a wide range of datasets that include employment in agriculture, exports of agricultural products, urbanization trends, government infrastructure spending, and the extent of forests, our goal is to interpret the complex interactions between these variables.

Our study adopts a lin-log linear regression approach to model and analyse the relationships between these variables comprehensively. Lin-log linear regression offers a robust framework for capturing the nuanced interactions and dependencies among various socio-economic indicators, thereby providing valuable insights into the drivers of rural economic development in India.

Objectives of the study:

1. To examine the relationships and interactions between different variables, including urbanization, government spending on rural infrastructure, employment patterns, and environmental factors, in shaping rural economic outcomes.
2. To provide evidence-based insights and policy recommendations that can inform decision-making processes aimed at promoting sustainable growth and improving livelihoods in rural India.

Review of Literature

Statement of Theory:

Our economic theory, is based on the idea that the health of rural sectors—that is, agriculture, forestry, and fishing—contributes substantially to the overall structure of the rural economy. A number of critical factors influence the economic trajectory of rural areas, including urbanization, government investment in rural infrastructure, the makeup of agricultural employment, and the ecological balance represented by the amount of forest cover. A larger urban population, for example, can affect the market for agricultural goods. Concurrently, it is thought that effective government spending on rural infrastructure increases employment and production in the agriculture industry. The selection of each variable is predicated on its theoretical significance with respect to the dynamics of rural economic growth.

Dependent Variable (Y): Agriculture, Forestry, and Fishing, Value Added per Worker (constant 2015 US$): Due to its crucial significance in rural economies, agriculture, forestry, and fishing—value added per worker—were chosen as the dependent variable. Investments in agricultural research and rural infrastructure, for example, have had significant effects on growth and poverty reduction.

Urban Population (% of Total Population): An independent variable that captures the urbanization component of rural development is the percentage of the population living in urban areas. Urbanization becomes crucial to comprehending the dynamics of rural economic growth because it affects the patterns of demand for agricultural products.

Agricultural Raw Materials Exports (% of Merchandise Exports): Raw Materials for Agriculture The dimension of international trade is captured by exports as a percentage of merchandise exports. This variable emphasizes how crucial institutional change, well-structured public and private investment, and higher investment efficiency are to boosting agricultural growth and lowering poverty.

Govt Expenditure on Rural Infrastructure (In Rs. Crores): The indicator used to assess the effect of public investment on rural development is government expenditure on rural infrastructure. It emphasizes how more funding for agricultural infrastructure is required in order to attain sustainable growth. The presence of regressors expressed in different units, such as percentages and currency (Rs Crores), requires careful consideration in the Multiple Linear Regression model. One approach to address this issue is to transform the variables into comparable units. Hence, a logarithmic transformation was applied to the variable. By log-transforming X3, its values are scaled proportionally, making it more comparable to the other variables in the regression model. The decision to use a linear-log model, where the dependent variable remains linear while X3 is log-transformed, is justified by the nature of the data. This approach accommodates the varying scales of the regressors and facilitates the interpretation of coefficients in the regression analysis. Overall, the transformation strategy ensures consistency in the measurement scale across regressors and enhances the robustness of the regression model.

Employment in Agriculture (% of Total Employment): One measure of the workforce composition in rural areas is the percentage of employment in agriculture. Emphasizing the significance of employment-related variables in rural development, research indicates that increased investments in rural roads, agricultural research, and education have the biggest impacts on poverty reduction per rupee spent.

Forest Area (% of Land Area): The environmental aspect of rural development is reflected in the Forest Area as a percentage of Land Area. This variable recognizes the value of forest conservation in promoting rural economic growth, which is consistent with the increasing emphasis on sustainable development.

The variables have been chosen based on the body of current literature that emphasises their significance in comprehending and advancing the development of India's rural economy.

Mathematical Model:

The model is consistent with the economic theory, which suggests that the dependent variable and the selected independent variables have a linear connection.

Multiple Linear Regression model, expressed as:

Econometric Model:

Y is the dependent variable, representing Agriculture, Forestry, and Fishing, Value Added per Worker (constant 2015 US$).

X1, X2, X3, X4, X5​ and ​ are the independent variables: Urban Population (% of Total Population), Agricultural Raw Materials Exports (% of Merchandise Exports), Govt Expenditure on Rural Infrastructure (In Rs. Crores), Employment in Agriculture (% of Total Employment), and Forest Area (% of Land Area) respectively.

are the coefficients to be estimated.

is the error term, capturing unobserved factors affecting the dependent variable.

Methodology:

We have opted for a multiple linear regression model after analyzing scatter plots of the dependent variable (Y) against the independent variables (X1, X2, X3, X4, X5). These plots exhibit a consistent linear trend, with data points aligning closely along straight lines, indicating a linear association between the variables.



The scatter plots reveal positive linear trends for X1 (Urban Population % of Total Population), indicating that as urban population percentage, increase, there is a corresponding increase in the value added per worker in agriculture-related sectors. As urban population increases, there is typically greater demand for agricultural products to meet the needs of urban consumers. This increased demand can stimulate agricultural production and investment, leading to higher productivity and value added per worker in agriculture.

X4 (Employment in Agriculture % of Total Employment) exhibits a negative linear association, suggesting that higher percentages of employment in agriculture relative to total employment are linked to lower value added per worker. When a larger proportion of the workforce is engaged in agriculture relative to other sectors, it often indicates a lower level of labour productivity in agriculture. This could be due to reasons such as outdated farming techniques, limited access to modern technology and machinery, or insufficient investment in agricultural infrastructure and research. High dependence on agriculture for employment may indicate a lack of diversification in the economy, with fewer opportunities for employment in higher value-added industries. This can result in a lower overall productivity level within the agricultural sector, as resources are spread thinly across a large labour force.

X5 (Forest Area % of Land Area) displays a positive linear trend, indicating that higher percentages of forest area relative to land area are associated with increased value added per worker in agriculture-related sectors. Forests play a crucial role in providing ecosystem services such as soil fertility, water regulation, and biodiversity conservation, which are essential for supporting agricultural productivity. A higher percentage of forest area relative to land area indicates better environmental conditions and ecosystem health, which can positively influence agricultural yields and contribute to higher value added per worker.

X3 (Government Expenditure on Rural Infrastructure) and Y (Value Added per Worker in Agriculture-related Sectors) exhibit an upward sloping trend without strictly adhering to linearity. The effectiveness of government expenditure on rural infrastructure in enhancing agricultural productivity can be influenced by various contextual factors such as the quality of infrastructure projects, governance issues, and the efficiency of resource allocation. These factors may introduce variability in the relationship between infrastructure spending and agricultural value added per worker, leading to deviations from strict linearity.

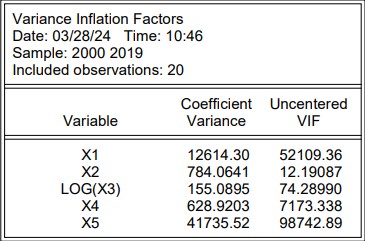
X2 (Agricultural Raw Materials Exports % of Merchandise Exports) and Y (Value Added per Worker in Agriculture-related Sectors) exhibit an upward sloping trend however, their relationship is not linear. The relationship between agricultural raw materials exports and agricultural productivity may be subject to fluctuations and seasonality, leading to variability in the observed trend. Changes in export volumes, prices, and market conditions can impact the profitability of agricultural production and, consequently, its value added per worker. The composition of agricultural raw materials exports may vary in terms of value-added content and market competitiveness. Higher-value agricultural products may contribute more significantly to agricultural productivity than lower-value commodities, resulting in variations in the observed relationship between X2 and Y.

Hence, Linear regression can accommodate such relationships and provide estimates of the relationships between variables, making it an appropriate choice for this analysis. Overall, the linear regression model is suitable for capturing the relationships between the variables and providing valuable insights into factors influencing value added per worker in agriculture-related sectors. The decision to utilize a lin-log model, where the dependent variable remains linear while X3 is log-transformed, is made in recognition of the inherent characteristics of the dataset. This approach acknowledges the need to maintain consistency in the measurement scale across regressors and enhances the reliability of the regression model's findings.

Test for Multicollinearity:

Multicollinearity occurs when independent variables in a regression model are highly correlated, which can lead to unstable coefficient estimates and inflated standard errors. One common approach to assess multicollinearity is to calculate the variance inflation factor (VIF) for each independent variable. The VIF measures how much the variance of an estimated regression coefficient is increased due to multicollinearity. Generally, a VIF greater than 10 indicates a problematic level of multicollinearity, although some researchers may use a threshold of 5. While severe multicollinearity can distort coefficient estimates and inflate standard errors, it's important to note that in the context of economic variables, some degree of multicollinearity may be expected and tolerated. Economic phenomena are inherently complex and interconnected, and variables often co-vary due to common underlying factors. In such cases, the focus shifts from eliminating multicollinearity entirely to understanding and interpreting the relationships between variables more cautiously.

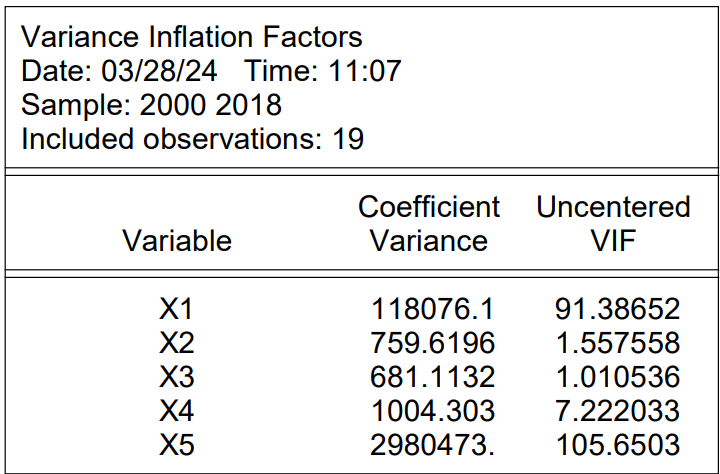
In this study, tests for multicollinearity were conducted using EViews software. The results indicate high multicollinearity for several independent variables, as evidenced by their elevated VIF values. Specifically, variables X1, X4, and X5 exhibited VIF values well above the threshold of 10, suggesting severe multicollinearity.



The high multicollinearity observed in these variables could be attributed to the inherent nature of economic variables, which often exhibit interdependencies and shared influences. Economic indicators are interconnected through complex networks of causality and feedback loops. Changes in one variable can ripple through the economy, affecting other related variables. For example, investments in rural infrastructure (X3) may stimulate agricultural productivity, which in turn can impact employment levels and environmental outcomes. These interconnected relationships can reinforce correlations among economic variables, contributing to multicollinearity in regression models.

To address the issue of multicollinearity, we employed the method of first differencing. Differencing involves computing the differences between consecutive observations of a time series variable. In this case, first-order differencing subtracts each observation from its preceding value. By doing so, we aim to remove the linear trends and correlations present in the original data, which can help mitigate multicollinearity among the independent variables in the regression model.

VIF after first differencing:

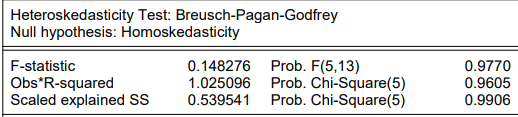


Variables X2, X3, and X4 show VIF values below 10 suggesting little to no multicollinearity. Hence through the method of first differencing, we have reduced the severity of multicollinearity among the variables. From this point, all the analysis is done using the differenced data.

Test for Homoscedasticity:

Heteroscedasticity refers to the unequal spread of residuals (errors) across the range of predictor variables in a regression analysis. To test for heteroscedasticity, one common method is the Breusch-Pagan test. These tests examine whether the variance of the residuals is dependent on the values of the independent variables. The null hypothesis of the Breusch-Pagan test is homoscedasticity, meaning that the variance of the residuals is constant across all values of the independent variables. A rejection of the null hypothesis suggests the presence of heteroscedasticity, indicating that the variance of the residuals varies systematically with changes in the independent variables.

The following is the output of BPG test carried out on the differenced data done on E Views software.



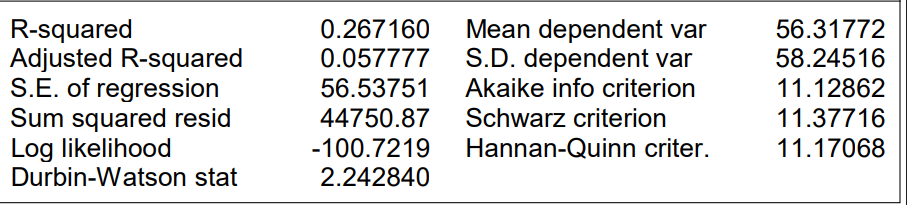
The results of the heteroscedasticity test, show that there is no significant evidence to suggest that the variance of the residuals varies systematically with changes in the independent variables. The F-statistic, which measures the overall significance of the test, has a value of 0.148276 with a corresponding probability of 0.9770. Since this p-value significance level of 0.05, we fail to reject the null hypothesis. Similarly, the observed R-squared has high p-value, further supporting the conclusion of homoscedasticity.

Overall, based on these test results, there is no evidence to suggest the presence of heteroscedasticity in the regression model. Therefore, the assumption of constant variance of residuals across the range of predictor variables is reasonable and does not violate the assumptions of classical linear regression analysis.

Test for Autocorrelation:

Autocorrelation, also known as serial correlation, refers to the correlation between observations within a regression model at different time points. In simpler terms, it's a measure of the degree to which the values of a variable are related to themselves over time. Autocorrelation is important in regression modelling because it can affect the accuracy of statistical estimates and predictions. Positive autocorrelation indicates that values tend to be similar to preceding values, while negative autocorrelation suggests an inverse relationship. On the other hand, no autocorrelation implies that there is no systematic relationship between observations at different time points.

The Durbin-Watson test is a statistical test used to detect the presence of autocorrelation in the residuals of a regression analysis. It specifically assesses whether there is a correlation between adjacent residuals in a time series or regression model. The test statistic for Durbin-Watson ranges between 0 and 4. A value near 2 indicates no significant autocorrelation (i.e., the residuals are independent), while values significantly below 2 suggest positive autocorrelation and values significantly above 2 suggest negative autocorrelation.



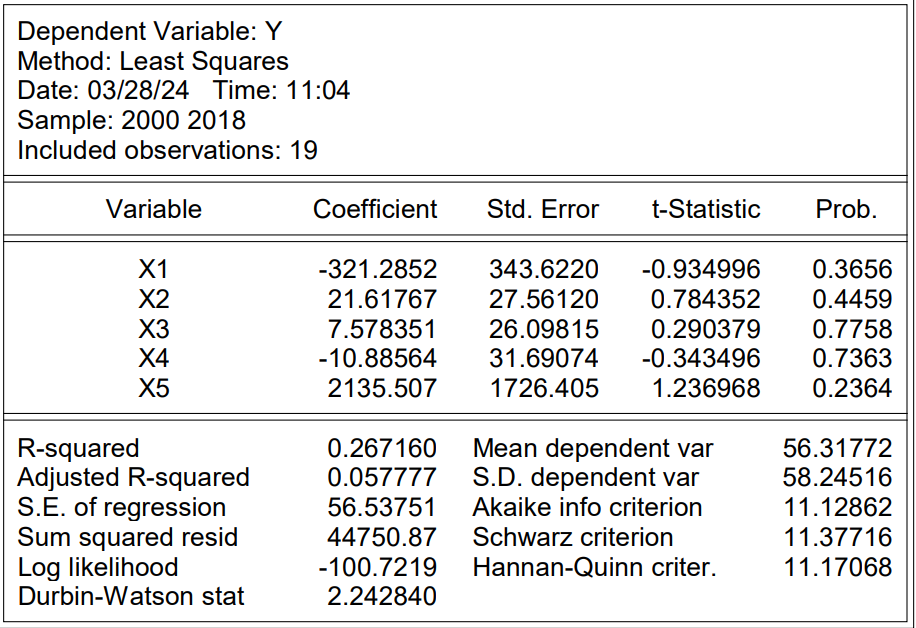
With Durbin-Watson statistic of 2.24284, which is very close to the expected value of 2, there is no significant evidence of autocorrelation. This means that the residuals, which represent the difference between the observed and predicted values of the dependent variable, do not exhibit any systematic patterns over time or across observations. Hence, the assumptions of the regression model regarding the independence of errors are met, and the model's estimates can be considered reliable.

Fitting the model:

Fitting the model involves estimating the parameters of the regression equation using the available data. In multiple linear regression, this typically involves using techniques such as ordinary least squares (OLS) to find the coefficients that minimize the sum of the squared differences between the observed and predicted values of the dependent variable.

Once the model is fitted, it can be used to make predictions about the dependent variable based on new values of the independent variables. Additionally, the fitted model allows for the interpretation of the relationship between the dependent variable and each independent variable, as captured by the estimated coefficients. These coefficients indicate the magnitude and direction of the effect of each independent variable on the dependent variable, holding other variables constant.

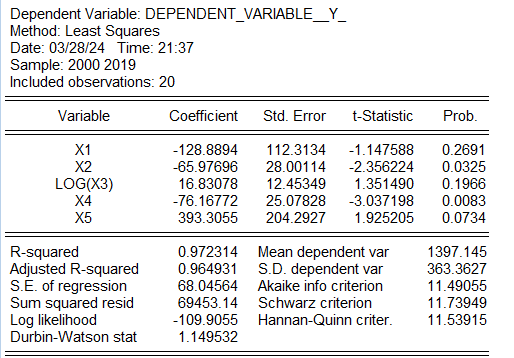
Regression results after first differencing:



Regression Equation after first differencing:

Y = -321.28516171\*X1 + 21.6176681603\*X2 + 7.57835147357\*X3 - 10.8856445736\*X4 + 2135.50688084\*X5.

Regression equation before first differencing



Regression equation before first differencing:

Y= -128.889441853\*X1 - 65.9769579089\*X2 + 16.8307763205\*LOG(X3) - 76.1677184661\*X4 + 393.305505231\*X5

Where,

