A Statistical Analysis of Economic Recession in India

Arya Vijayan, Richu Rajesh *

[a]Department of Economics and Statistics, Kerala University of Fisheries and Ocean Studies, Panangad, Kochi, 682506, Kerala, India
[b]Department of Statistics, Government Victoria College, Palakkad, 678001, Kerala, India

Abstract

This study conducts a comprehensive statistical analysis of economic recessions in India, emphasizing GDP forecasting, sectoral influences, and regional economic variations. It employs a combination of time series modeling, multiple regression analysis, and cluster analysis to capture the multifaceted nature of the Indian economy. Historical GDP data from 1965 to 2023 is modeled using ARIMA techniques to generate forecasts. Multiple regression analysis is conducted for the period 2011 to 2022 to examine the impact of various economic sectors on GDP fluctuations. Additionally, clustering methods are applied to state-wise GDP data from 2022 to 2023 to identify regional economic patterns. The study aims to provide a holistic understanding of macroeconomic behavior, offering evidence-based insights to support policymakers in developing strategies for balanced and sustainable economic growth.

Keywords: Economic recession, ARIMA, Multiple regression, K-Means clustering

1. Introduction

Since independence, India has undergone profound economic transformation, evolving from a predominantly agrarian economy to one of the fastest-growing major economies in the world (Daga et al., 2004). This rapid growth has been accompanied by structural changes in various sectors, technological advancements, and increased integration with the global economy. However, despite these significant strides, India remains susceptible to economic recessions—periods characterized by a broad and sustained decline in economic activity extending over several months or longer. Such downturns are typically evidenced by deteriorating macroeconomic indicators including gross domestic product (GDP), employment rates, industrial output, and retail sales. While India has demonstrated considerable resilience in navigating these downturns, it has experienced multiple phases of economic slowdown, each distinguished by unique causes and far-reaching consequences. A thorough understanding of recessionary dynamics is thus critical, as recessions often precipitate adverse outcomes such as widespread job losses, diminished household incomes, declining investment levels, and an escalation in poverty and inequality. If these effects are not effectively mitigated, they can cascade across various sectors, ultimately undermining the country's overall economic growth and social stability (Bhatt, 2011).

India's economic fluctuations are influenced by a complex interplay of domestic and international factors. On the domestic front, inflationary pressures, shifts in government policies, alterations in consumer demand, and disruptions in production processes serve as critical triggers for economic slowdowns. Concurrently, the Indian economy is increasingly exposed to global economic conditions, including international financial crises, volatility in oil prices, imposition of trade restrictions, and fluctuations in foreign direct investment. This dual exposure renders the economy highly sensitive to both internal disturbances and external shocks, heightening the importance of conducting a detailed and systematic analysis of past recession episodes. Such analyses are instrumental in uncovering recurrent patterns, assessing sector-specific impacts, and identifying reliable early warning indicators that can inform timely policy interventions. To this end, key macroeconomic and sectoral indicators—such as GDP growth rate, consumer price index

*Richu Rajesh (Corresponding author)
Email address: richurajesh@gvc.ac.in (Richu Rajesh *)

(CPI), fiscal and revenue deficits, capital expenditure, employment trends, and output across sectors including services, agriculture, manufacturing, construction, banking, trade, and public administration—provide valuable insights into the evolving economic landscape and potential vulnerabilities during downturns (Daga et al., 2004; Maiti et al., 2024).

The nature and impact of recessions in India have varied considerably in terms of both intensity and duration. The industrial sector typically experiences immediate and pronounced effects, including sharp declines in investment and production activities. The agricultural sector, while less directly exposed, often suffers indirectly through decreased demand and tighter credit availability. Meanwhile, the services sector—although relatively more resilient due to its growing share in the economy—remains vulnerable, particularly in sub-sectors sensitive to consumer sentiment and global economic conditions, such as tourism, hospitality, and financial services. Over time, policy responses to economic recessions have evolved significantly. The balance of payments crisis in 1991 marked a watershed moment that ushered in comprehensive liberalization reforms aimed at stabilizing the economy and attracting foreign investment. During the 2008 global financial crisis, policymakers deployed fiscal stimulus measures alongside monetary easing to cushion the economic impact. More recently, the COVID-19 pandemic-induced recession necessitated unprecedented fiscal interventions, credit guarantee schemes, and structural reforms designed to bolster economic resilience. Despite these concerted efforts, structural challenges persist, including high fiscal deficits, inflation volatility, vulnerabilities within the banking sector, and the precarious conditions faced by the large informal workforce, which lacks adequate social protection (Bhatt, 2011; Maiti et al., 2024).

Although there exists a substantial body of literature focusing on India's overall economic growth, there remains a notable gap in focused statistical investigations addressing the patterns, determinants, and sector-specific consequences of recessions. This study seeks to fill this gap by employing advanced quantitative techniques—specifically time series modeling, regression analysis, and cluster analysis—to comprehensively examine the drivers and historical trends of recessions in India, as well as their differentiated effects across various sectors and states. The significance of this research lies in its potential to generate data-driven, actionable insights that can enhance the capacity of policymakers to detect early signs of economic distress and implement targeted interventions. Such evidence-based approaches are essential for reinforcing economic resilience, safeguarding employment, and promoting sustained, inclusive growth amidst both domestic challenges and an increasingly volatile global economic environment (Daga et al., 2004; Maiti et al., 2024).

In alignment with these objectives, the study focuses on three key aims: first, to develop a robust time series model of India's GDP to facilitate accurate forecasting; second, to investigate the interrelationships between GDP and its constituent economic sectors to deepen understanding of their mutual influences; and third, to classify Indian states according to GDP and its contributing factors, thereby illuminating regional disparities and dynamics. Together, these objectives underpin a comprehensive framework for analyzing recession dynamics at both the national and subnational levels.

The remainder of the article is organized as follows. Section 1 introduces the study's background, objectives, and significance. Section 2 provides a detailed review of relevant literature, critically evaluating prior research to identify gaps and justify the present study. Section 3 outlines the data sources and the methodological framework adopted for empirical analysis. Section 4 presents the study's results alongside an in-depth discussion of their implications. Section 5 summarizes the principal conclusions and offers policy recommendations based on the findings. Finally, Section 6 provides a complete list of references cited throughout the article.

2. Literature Review

Daga et al. (2004) analyzed India's GDP behavior and forecasting, advocating the use of GDP at factor cost to avoid distortions from indirect taxes. They found GDP to be a stationary process, challenging the belief that reforms always raise growth. Their regression model (Adj. R^2 = 99.7%) showed strong alignment between estimated and actual GDP, with the Trade, Transport, Storage, and Communication sector contributing most and the Financing and

Business Services sector the least. Gupta (2009) assessed the global financial crisis's impact on India, finding significant effects through trade and financial channels, which worsened an ongoing slowdown. Limited fiscal space and weak monetary transmission reduced the effectiveness of stimulus measures, and recovery depended on boosting exports, private investment, and maintaining competitive exchange rates. Singh (2009) explored the services sector's relationship with GDP, finding a stable long-run equilibrium with non-services sectors and evidence that growth in services preceded and drove overall growth, while also acting as a buffer against shocks.

Bhatt (2011) found that the 2008 crisis caused capital outflows, rupee depreciation, and losses in banking and IT sectors, with fiscal measures proving insufficient. The study recommended more spending on agriculture and infrastructure. Singhania and Gupta (2011) used ARIMA models to analyze FDI inflows, finding GDP, inflation, and patents as significant predictors, with policy reforms in 1995–1997 boosting inflows. Micheal and Abiodun (2014) examined Nigeria's GDP composition, addressing multicollinearity and finding positive contributions from agriculture, industry, and services, with industry contributing the most. Agrawal (2018) applied ARIMA models to quarterly GDP data, preferring AR(1) and MA(2) for forecasting, while noting short-term deviations during economic shocks. Bhattacharya et al. (2019) used a time-varying parameter regression model, which outperformed constant-parameter and dynamic factor models. Gupta and Minai (2019) evaluated GDP forecast accuracy and found RBI forecasts generally stronger than others.

Basantwani et al. (2021) studied GNI, public expenditure, and employment, finding long-run equilibrium and that private employment was influenced by its lagged values and private spending. Hassan and Mirza (2021) used Prophet and ARIMA for GDP forecasting, highlighting the importance of data-driven policymaking. Jethwani et al. (2021) examined agricultural GDP and NPAs, showing that loan waivers harmed credit culture and long-term growth. Malik and Mirza (2021) analyzed GDP trends from 1979–2020, noting a decline before COVID-19 and using ARIMA/Prophet models for policy guidance. Satpathi and Hasan (2021) linked manufacturing productivity growth to export expansion, suggesting productivity improvement as a recession-mitigation strategy. Raj et al. (2022) applied shape-based clustering to COVID-19 data, identifying distinct patterns and offering policy insights. Majumder et al. (2023) used Wroclaw Taxonomy and K-means clustering to classify states into development tiers, revealing regional disparities.

Taneja et al. (2023) found natural resource rents significantly impact GDP, recommending a unified resource policy. Maiti et al. (2024) developed a Factor-Augmented Error Correction Model to forecast post-COVID GDP growth between 4–8%, identifying oil prices and climate change as risks, and reforms and digitalization as growth drivers. Manimala and Kurian (2024) studied HR strategies in downturns, finding internal beliefs more influential than external pressures, and classifying managers into five types. Pandey et al. (2024) applied Bayesian regression, achieving higher predictive accuracy than OLS. Pervez and Ali (2024) used robust regression to evaluate public sector banks, improving reliability by reducing the influence of outliers. Sinha (2024) identified consumption, investment, exports, and employment as main GDP drivers, recommending inclusive and sustainable growth policies.

3. Data and Methodology

This study employs a robust and comprehensive dataset primarily sourced from the Reserve Bank of India (RBI), recognized as an authoritative provider of macroeconomic and regional economic statistics. The data were retrieved from the RBI Handbook of Statistics on the Indian Economy, which offers extensive annual and quarterly economic indicators essential for the analysis. Specifically, the dataset includes national-level annual GDP figures spanning from 1965 to 2023, sector-wise GDP contributions from 2011 to 2022, and state-wise economic indicators for the fiscal year 2022–2023. These data encompass a wide array of variables such as GDP growth rates, inflation (measured via the Consumer Price Index), fiscal deficits, employment statistics, and sectoral outputs covering agriculture, manufacturing, trade, services, and other relevant economic activities.

The dataset has been systematically organized and processed in formats compatible with statistical software such as R and Excel to facilitate efficient analysis. Three primary methodological approaches are employed to examine different dimensions of economic recessions in India: time series modeling, multiple regression analysis, and cluster analysis.

The time series analysis utilizes the univariate annual GDP data from 1965 to 2023. This long-term dataset enables the identification of economic trends, cyclical fluctuations, and structural shifts through the application of the AutoRegressive Integrated Moving Average (ARIMA) model. This model facilitates forecasting of future GDP trends and helps in understanding the temporal dynamics of India's economic growth and recessions.

The regression analysis investigates the relationship between GDP (as the dependent variable) and various sectoral outputs (independent variables) using quarterly data from 2011 to 2022. Key sectors examined include Agriculture, Forestry and Fishing; Manufacturing; Construction; Trade, Hotels, Transport, Communication; Financial, Real Estate and Professional Services; and Public Administration. Given the potential for multicollinearity among explanatory variables, appropriate diagnostic tests and variable selection techniques are applied to ensure the validity and reliability of the model. This analysis aims to identify the significant contributors to GDP fluctuations and to elucidate sector-specific impacts during recessionary phases.

To capture spatial economic heterogeneity, cluster analysis is conducted on state-wise economic indicators for the fiscal year 2022–2023. Using K-means clustering, states and union territories are grouped based on similarities across multiple variables including GDP growth rate, inflation, employment levels, fiscal and revenue deficits, capital expenditure, and sectoral contributions. This unsupervised learning approach reveals regional patterns and disparities, informing targeted policy interventions to address development gaps and enhance inclusive growth.

All datasets were carefully verified for completeness, consistency, and continuity prior to analysis to ensure robustness of the findings. The integration of national, sectoral, and regional data through these complementary methods provides a multifaceted understanding of India's recession dynamics, offering valuable insights for policymakers and researchers.

4. Results and Discussion

4.1. Time Series Analysis

This section presents the results of the time series analysis performed on the Gross Domestic Product (GDP) data from 1965 to 2023. The primary focus is on identifying trends, stationarity, and forecasting GDP for the upcoming years using an ARIMA model. The findings are interpreted in the context of economic trends and recessions.

4.1.1. Time Series Plot

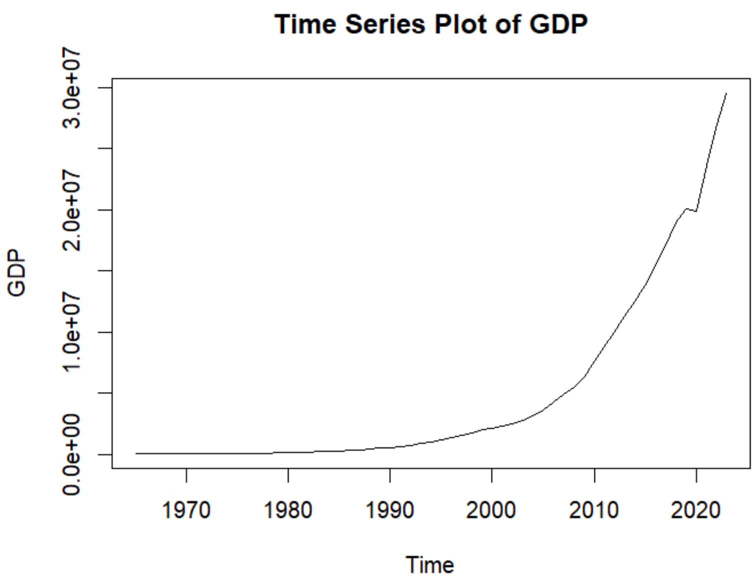

Figure 1: Plot of Time Series Analysis of GDP

From Figure 1, the original time series plot of GDP showed that India's GDP grew steadily from 1965 to 2019, with a noticeable decline in 2020 due to the COVID-19 pandemic, followed by a gradual recovery in subsequent years.

4.1.2. Augmented Dickey-Fuller Test

Table 1: Augmented Dickey-Fuller Test

Dickey-Fuller	4.9148
Lag order	3
p-value	0.99

From Table 1, the ADF test shows a p-value of 0.99, which is greater than 0.05. Hence, it indicates that the time series is not stationary.

4.1.3. Autocorrelation Function Plot of GDP

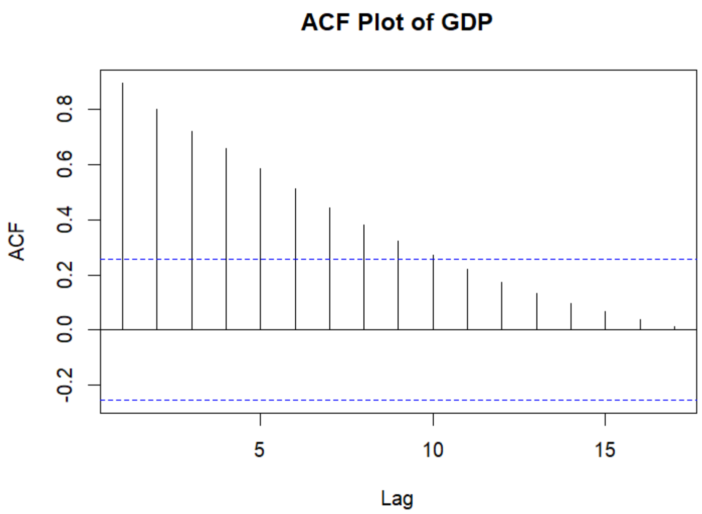

Figure 2: Plot of Autocorrelation Function of GDP

In Figure 2, the exponential decay structure of the ACF shows that the data are not stationary. This is confirmed by the Augmented Dickey-Fuller test.

4.1.4. Partial Autocorrelation Function Plot of GDP

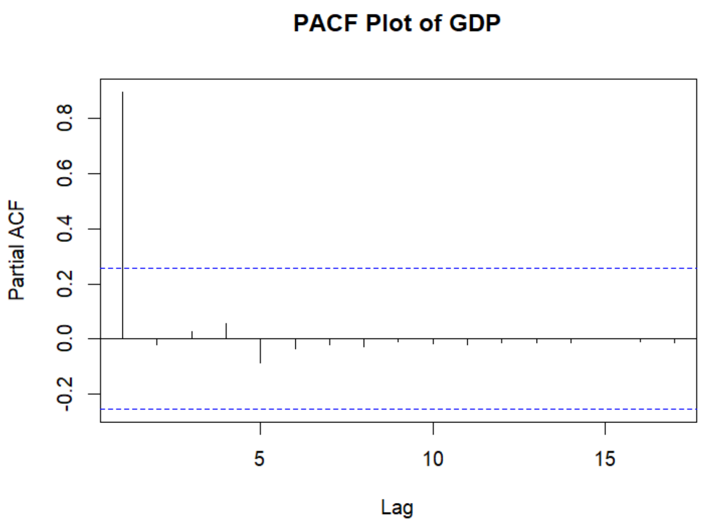

Figure 3: Plot of Partial Autocorrelation Function of GDP

4.1.5. Augmented Dickey-Fuller Test on the Differenced GDP

Dickey-Fuller	-5.1087
Lag order	3
p-value	0.01

Table 2: Augmented Dickey-Fuller Test on the Differenced GDP

From Table 2, after applying log-transformation and differencing, the series was found to be stationary. ADF test confirmed the stationarity of the differenced data.

4.1.6. Autocorrelation Function Plot of Differenced GDP

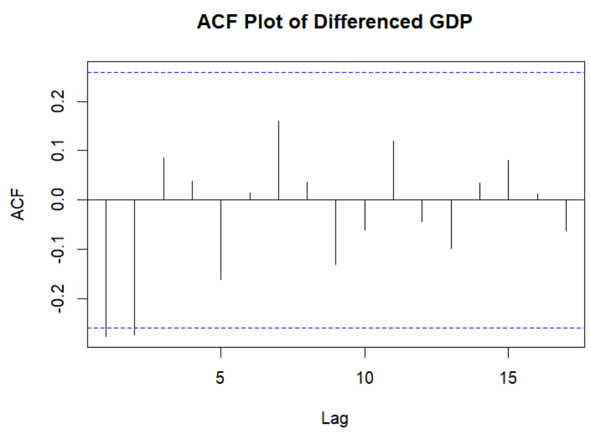

Figure 4: Plot of Autocorrelation Function of Differenced GDP

4.1.7. Partial Autocorrelation Function Plot of Differenced GDP

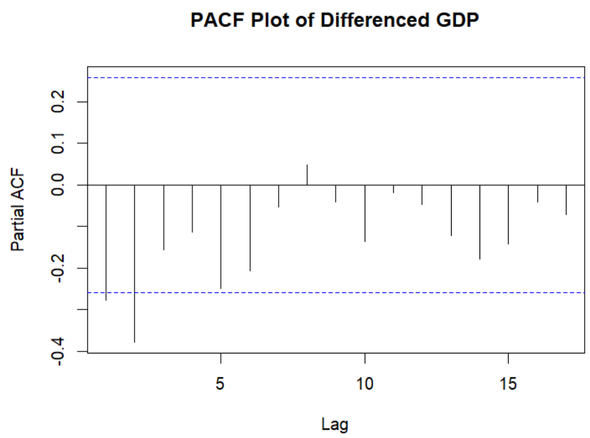

Figure 5: Plot of Partial Autocorrelation Function of Differenced GDP

From Figures 4 and 5, the ACF and PACF plots of the differenced GDP series show rapid decay within confidence limits, confirming stationarity and suitability for ARIMA modeling.

4.1.8. ARIMA Model

The ARIMA(2,2,2) model was identified as the best fit for the GDP time series data based on the minimum AIC value. The parameters of the ARIMA model were estimated using the Conditional Sum of Squares – Maximum Likelihood (CSS-ML) method, which combines the computational efficiency of the CSS approach with the accuracy of Maximum Likelihood estimation.

Table 3: Estimated Parameters of ARIMA(2,2,2) Model for GDP

Parameter	**Estimate**	**Standard Error**
AR(1)	0.8186	0.1506
AR(2)	-0.6920	0.1674
MA(1)	-1.5730	0.2142
MA(2)	0.9493	0.2112

4.1.9. Different ARIMA Models and Their AIC Values

Table 4: Different ARIMA Models and Their AIC Values

Model No.	**ARIMA Model (p,d,q)**	**AIC Value**
Model 1	ARIMA(2,2,1)	1665.19
Model 2	ARIMA(1,2,2)	1666.04
Model 3	ARIMA(1,2,0)	1673.99
Model 4	ARIMA(1,2,1)	1666.36
Model 5	ARIMA(1,2,3)	1658.97
Model 6	ARIMA(2,2,2)	**1660.09**

From Table 4, the ARIMA(2,2,2) model yielded the lowest AIC value of 1660.09, indicating the best fit for the GDP time series. Consequently, this model was selected for further analysis and forecasting.

4.1.10. Residual Diagnostics

Autocorrelation Function Plot of Residuals

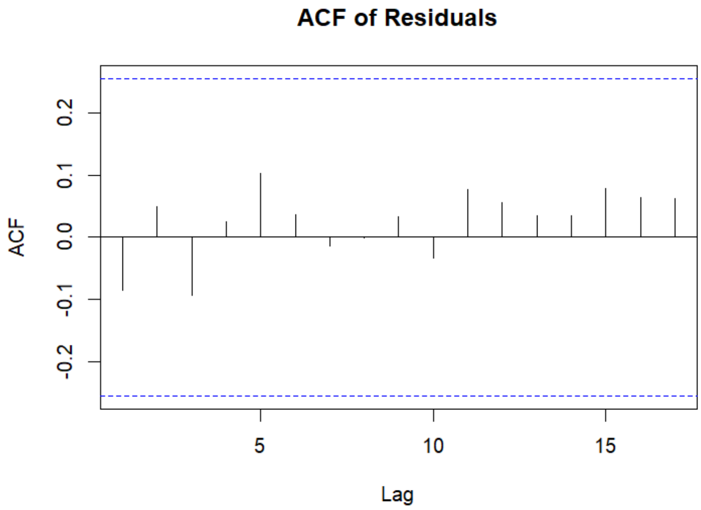

Figure 6: plot of Autocorrelation Function (ACF) of Residuals

Partial Autocorrelation Function Plot of Residuals

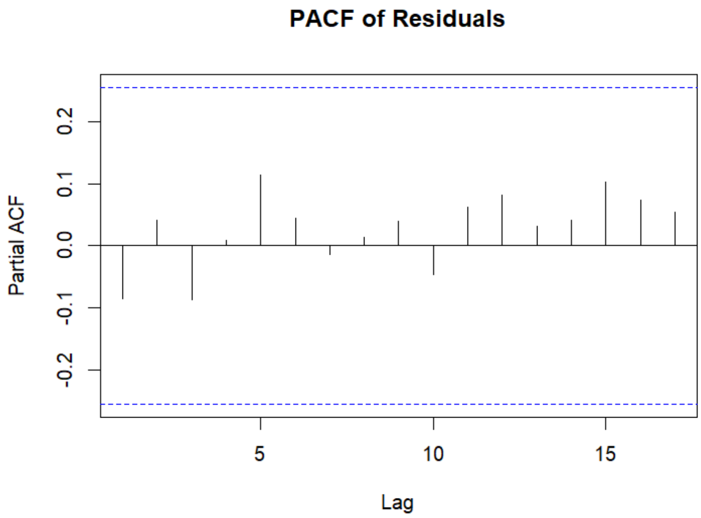

Figure 7: plot of Partial Autocorrelation Function (PACF) of Residuals

From Figure 6 and 7, the ACF and PACF values of the residuals are within the confidence bands, and the residuals are white noise, meaning that the ARIMA model is a good fit with no leftover patterns.

4.1.11. Ljung-Box Test

The Ljung–Box test produced a statistic of 0.45348 with one degree of freedom, corresponding to a p-value of 0.5007. As this p-value exceeds the 0.05 significance level, the null hypothesis of no autocorrelation cannot be rejected. This indicates that the residuals are free from significant serial correlation and resemble a white-noise process, suggesting the model adequately captures the data's temporal structure.

4.1.12. Forecasting

Using data from 1965–2023, the model forecasted GDP for 2024–2030, with 95% confidence intervals shown in the Table 5.

Year	**Forecast**	**Lo 95**	**Hi 95**
2024	30695659	29783154	31608163
2025	32260612	30800218	33721007
2026	35143881	33352687	36935076
2027	38826155	36704236	40948074
2028	42250269	39572075	44928464
2029	44910141	41371290	48448991
2030	47123012	42567862	51678162

Table 5: Forecasted GDP values

4.1.13. Forecast Accuracy Evaluation

Metric	Value
Mean Percentage Error (MPE)	2.696818
Mean Absolute Percentage Error (MAPE)	3.599664
Mean Absolute Scaled Error (MASE)	0.3386245
Autocorrelation of Residuals at Lag 1 (ACF1)	-0.08548694

Table 6: Accuracy of the forecasted values

Table 6 presents the model's performance metrics, indicating a Mean Percentage Error (MPE) of 2.70%, which suggests a slight positive bias. The Mean Absolute Percentage Error (MAPE) of 3.60% reflects low forecast error, while the Mean Absolute Scaled Error (MASE) of 0.34 demonstrates strong predictive accuracy. Furthermore, the lag-1 autocorrelation coefficient (ACF1 = -0.085) is close to zero, providing evidence of the model's validity and adequacy.

4.1.14. Plot of Forecasting

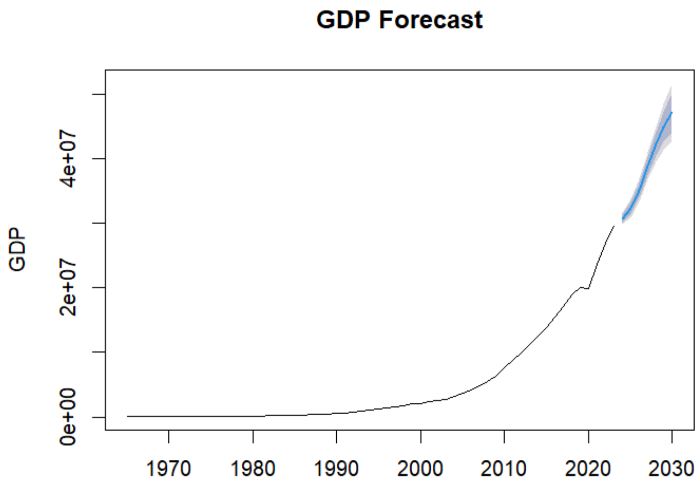

Figure 8: Plot of GDP Forecast

From Figure 8, the GDP forecast shows a steady upward trend, indicating sustained economic growth and positive momentum for India's near future.

4.2. Regression Analysis

Regression analysis was conducted to study the relationship between India's GDP and five key economic sectors from 2011 to 2022 (quarterly). The model is:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon \quad (1)$$

where
Y = Gross Domestic Product (GDP)
X_1 = Agriculture, Forestry, and Fishing
X_2 = Manufacturing
X_3 = Trade, Hotels, Transport, Communication and Services Related to Broadcasting
X_4 = Financial, Real Estate, and Professional Services
X_5 = Public Administration, Defense, and Other Services
The model is based on 48 quarterly observations spanning 12 years (2011–2022).

4.2.1. Multiple linear regression Initial Model

The multiple linear regression model included all the selected variables: Agriculture, Forestry, Fishing; Manufacturing, Trade, Hotels, Transport, Communication and Services Related to Broadcasting; Financial, Real Estate, and Professional Services; and Public Administration, Defense, and Other Services. The regression output is as follows:

Table 7: Regression Coefficients for Initial Multiple Linear Regression Model

Variable	**Estimate**	**Std. Error**	**t value**	**Pr(> $|t|$)**
(Intercept)	68480	27560	2.485	0.0174
Agriculture, Forestry, and Fishing	0.9198	0.06856	13.417	< 0.001
Manufacturing	1.789	0.1695	10.556	< 0.001
Trade, Hotels, Transport, Communication and Services Related to Broadcasting	1.599	0.1143	13.987	< 0.001
Financial, Real Estate, and Professional Services	0.6033	0.06635	9.093	< 0.001
Public Administration, Defense, and Other Services	1.317	0.1959	6.724	< 0.001

Multiple R-squared	0.9974
Adjusted R-squared	0.9971
F-statistic	3031
p-value	< 0.01

Table 8: Multiple Linear Regression Initial Model

From Table 8, the model has a high explanatory power with an Adjusted R^2 of 0.9971. All variables are significant at 1%, rejecting the null hypothesis and showing the sectors significantly impact GDP.

4.2.2. Variance Inflation Factor

To assess multicollinearity among the explanatory variables, the variance inflation factors (VIF) were calculated for the regressors. The results are presented below.

Agriculture, Forestry, Fishing	2.305130
Manufacturing	13.503520
Trade, Hotels, Transport, Communication, and Broadcasting Services	8.449513
Financial, Real Estate, Professional Services	4.768611
Public Administration, Defence, and Other Services	10.247766

Table 9: VIF Initial Values

From Table 9, the Variables with VIF > 10 indicate high multicollinearity and were removed to improve model reliability.

4.2.3. Multiple linear regression Reduced Model

After removing multicollinear variables, the reduced model retained only three sectors: Agriculture, Forestry & Fishing, Trade, Hotels, Transport, Communication & Broadcasting Services, and Financial, Real Estate & Professional Services.

Table 10: Regression Coefficients for Reduced Multiple Linear Regression Model

Variable	**Estimate**	**Std. Error**	**t value**	**Pr(> $|t|$)**
(Intercept)	67020	64480	1.039	0.305
Agriculture, Forestry, Fishing	1.320	0.1350	9.778	< 0.001
Trade, Hotels, Transport, Communication and Broadcasting Services	2.991	0.1407	21.256	< 0.001
Financial, Real Estate, Professional Services	1.358	0.09289	14.625	< 0.001

Table 11: Multiple Linear Regression Reduced Model

Multiple R-squared	0.984
Adjusted R-squared	0.9828
F-statistic	840.2
p-value	< 0.01

Table 11 shows that the reduced model explains over 98% of the variance and is statistically significant with an F-value of 840.2 and a p-value less than 0.01.

4.2.4. Variance Inflation Reduced Factor

Table 12: VIF Reduced Values

Agriculture, Forestry, Fishing	1.507227
Trade, Hotels, Transport, Communication, and Broadcasting Services	2.157796
Financial, Real Estate, Professional Services	1.575453

In Table 12, all the values are below the threshold of 5, indicating that multicollinearity is not an issue in the reduced model.

Now, the fitted regression model is

$$\hat{Y} = 67020 + 1.3202 \times X_1 + 2.9910 \times X_2 + 1.3585 \times X_3 \tag{2}$$

where

Y = GDP
X_1 = Agriculture, forestry, and fishing
X_2 = Trade, Hotels, Transport, Communication, and Broadcasting Services
X_3 = Financial, real estate, and professional services.

4.2.5. *ANOVA Table*

Table 13: ANOVA Table of Multiple Linear Regression Model

Source of Variation	df	Sum of Squares	Mean Square	F-value
Agriculture, Forestry and Fishing	1	4.8887e+12	4.8887e+12	894.82
Trade, Hotels, Transport, Communication and Broadcasting	1	7.7130e+12	7.7130e+12	1411.79
Financial, Real Estate and Professional Services	1	1.1685e+12	1.1685e+12	213.88
Residual (Error)	41	2.2399e+11	5.4633e+09	—
Total	44	1.39942e+13	—	—

The ANOVA results presented in Table 13 indicate that all variables significantly affect GDP, with high F-values demonstrating a strong model fit for the period 2011–2022.

4.2.6. *Shapiro-Wilk normality test*

The Shapiro–Wilk normality test returned a W statistic of 0.97871 with a p-value of 0.5684, which is greater than the 0.05 significance threshold, indicating that the residuals do not deviate significantly from normality.

4.2.7. *Breusch-Pagan Test*

The Breusch–Pagan test produced a statistic of 7.962 with 3 degrees of freedom and a p-value of 0.0468. Since this p-value is below 0.05, the null hypothesis of homoscedasticity is rejected, suggesting the presence of mild heteroscedasticity in the model residuals.

Robust standard errors were used to correct for heteroscedasticity. The results showed that the regression estimates remained significant, confirming the reliability of the model.

4.2.8. *Residual Analysis*

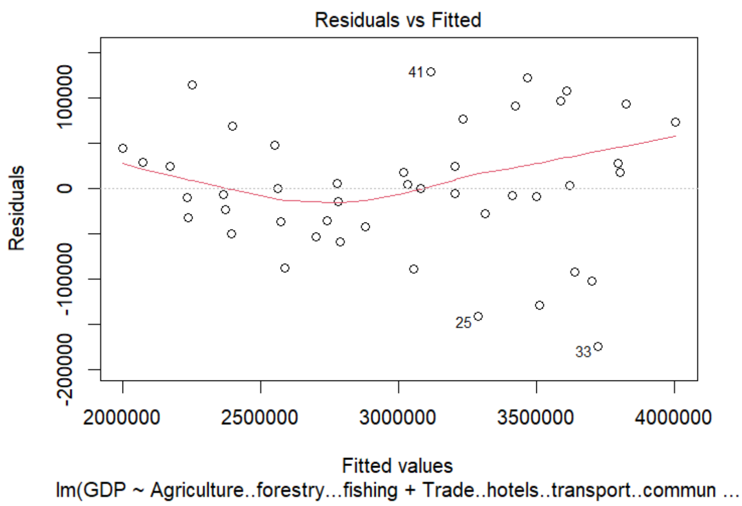

Figure 9: Plot of Residuals vs Fitted values

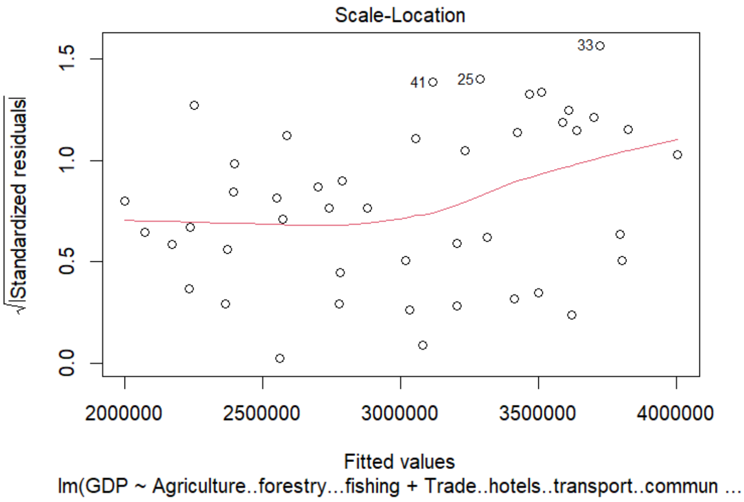

Figure 10: Plot of standardised residual vs fitted values

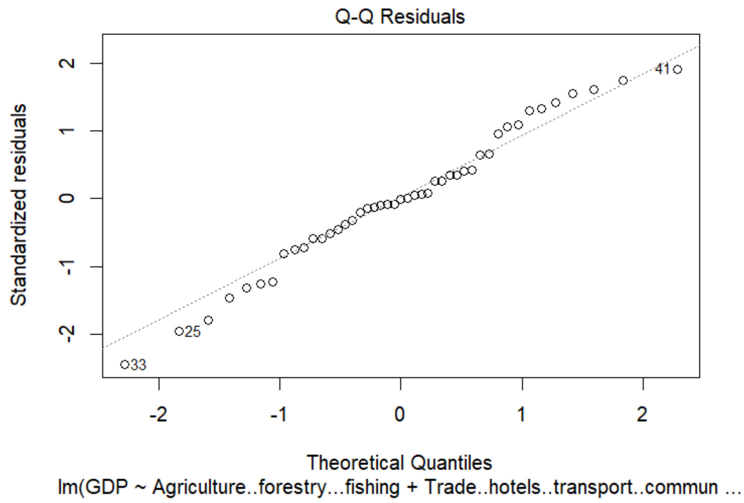

Figure 11: Plot of Q-Q Analysis

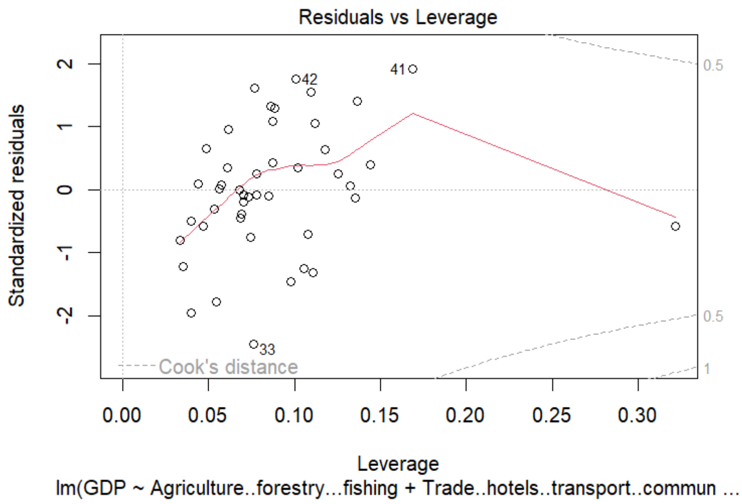

Figure 12: Plot of Residuals vs Leverage

The plot of residuals versus fitted values (Figure 9) shows that residuals are randomly scattered around zero with no clear pattern, indicating linearity and constant variance. Similarly, the plot of the standardized residuals versus the fitted values (Figure 10) confirms that the residuals are well centered around zero with no apparent outliers or heteroscedasticity. The Q-Q plot (Figure 11) demonstrates that the standardized residuals closely follow the theoretical normal distribution line, validating the assumption of normality. Lastly, the residuals versus leverage graph (Figure 12) reveals that all observations lie within Cook's distance of ±0.5, suggesting the absence of any influential data points. Overall, the model appears to be well-fitted.

4.3. Cluster Analysis

This section examines the state-wise economic data of India for 2022–2023 using the K-means clustering method. The purpose is to group states and union territories based on similarities in economic indicators, primarily focusing on GDP and related variables. This classification helps uncover regional development trends, enabling more precise policy interventions and effective allocation of resources. By identifying clusters of states with similar economic characteristics, such as GDP growth, inflation (CPI), employment, fiscal deficits, capital expenditure, and sectoral outputs in agriculture, manufacturing, construction, industry, banking, insurance, and services, this approach supports the design of balanced, region-specific development strategies.

4.3.1. Elbow method

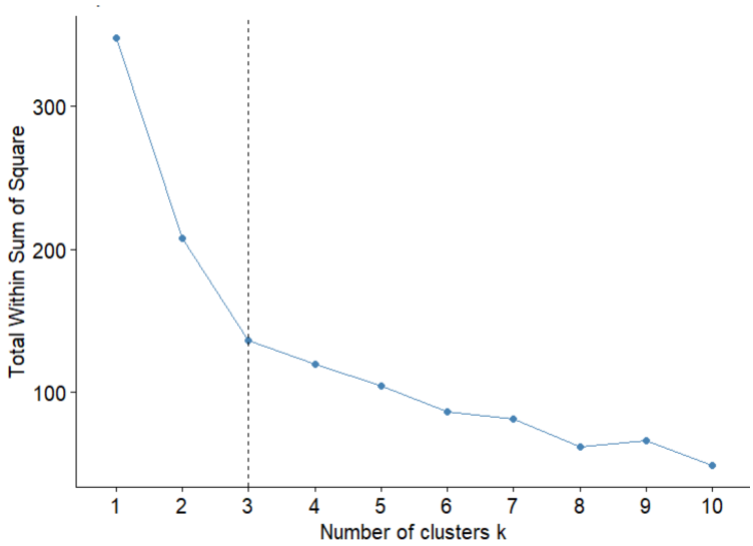

Figure 13: Plot of Scree Analysis

Figure 13 shows that as the number of clusters increases from 1 to 10, the within-cluster sum of squares (WSS) decreases, indicating tighter groupings. The optimal number of clusters is determined to be K = 3.

4.3.2. K-Means Clustering

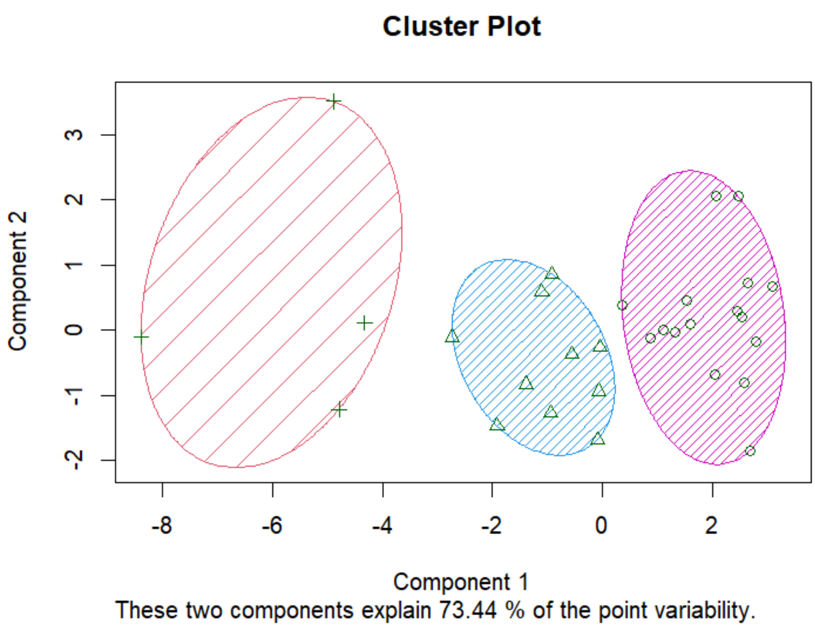

Figure 14: Plot of Cluster Analysis

In Figure 14, the algorithm grouped the 30 regions into three distinct clusters to show how the states are distributed based on their similarity.

4.3.3. State-wise Cluster Distribution

Table 14 shows the states/UTs in each cluster based on their economic similarity.

Table 14: State-wise Cluster Distribution

Cluster 1	**Cluster 2**	**Cluster 3**
Gujarat	Assam	Andhra Pradesh
Maharashtra	Chhattisgarh	Bihar
Tamil Nadu	Delhi	Haryana
Uttar Pradesh	Goa	Karnataka
	Himachal Pradesh	Kerala
	Jammu and Kashmir	Madhya Pradesh
	Manipur	Punjab
	Meghalaya	Rajasthan
	Mizoram	Telangana
	Nagaland	West Bengal
	Odisha	
	Puducherry	
	Sikkim	
	Tripura	
	Uttarakhand	
	Jharkhand	

The results show that the states within each cluster share similar economic profiles. Cluster 1 includes economically advanced states with high-development indicators. Cluster 2 represents moderately developed states, while cluster 3 consists of states with relatively lower economic performance. These groups highlight regional economic patterns and can guide more focused policy-making and resource allocation.

4.3.4. Cluster-wise Economic Profile of Indian States

Cluster 1: High-Performing States

Table 15: Summary of Economic Indicators and Interpretation

Variable	**Value**	**Interpretation**
GDP Growth	+0.254	Highest positive GDP growth
Inflation (CPI)	+0.483	Higher inflation (could be a concern)
Employment	+2.18	Much higher employment
Gross Fiscal Deficit	+1.42	High fiscal deficit (government spending)
Revenue Deficit	-0.36	Slightly negative (less revenue deficit)
Capital Expenditure	+1.76	High government capital spending
Agriculture, Manufacture, Construction, Industry, Banking, Services	All strongly positive (around +1.5 to +2.1)	Strong sectoral performance

As shown in Table 15, Cluster 1 includes states with the highest GDP growth, strong sectoral performance, high employment, capital spending, and fiscal deficit, alongside slightly higher inflation from strong demand.

Cluster 2: Moderate Growth with Weak Sectoral Support

Table 16: Summary of Economic Indicators and Interpretation

Variable	**Value**	**Interpretation**
GDP Growth	+0.14	Slightly above average
Inflation (CPI)	-0.42	Lower inflation
Employment	-0.59	Fewer employment
Gross Fiscal Deficit	-0.78	Lower deficit (could be good!)
Revenue Deficit	-0.39	Lower again (depends)
Capital Expenditure	-0.67	Less govt. spending
Agriculture, Manufacture, Construction, Industry, Banking, Services	All negative	Below average sectoral performance

As shown in Table 16, Cluster 2 comprises states with moderate GDP growth, lower sectoral performance, and investment, but low inflation and fiscal deficits, reflecting stable yet limited growth.

Cluster 3: Developing Economies with Sectoral Competence

Table 17: Economic Indicators with Values and Interpretations

Variable	Value	Interpretation
GDP Growth	-0.326	Negative GDP growth (below average)
Inflation (CPI)	+0.486	Slightly higher inflation
Employment	+0.079	Close to average employment
Gross Fiscal Deficit	+0.684	Higher fiscal deficit (more spending)
Revenue Deficit	+0.784	Higher revenue deficit
Capital Expenditure	+0.375	Moderate government spending
Agriculture	+0.524	Above average agriculture output
Manufacture	+0.066	Slightly above average manufacturing
Construction	+0.287	Above average construction
Industry	+0.112	Slightly above average industry
Banking and Insurance	+0.086	Slightly above average
Services	+0.342	Above average services

As shown in Table 17, Cluster 3 shows negative GDP growth but above-average performance in some sectors, with high deficits reflecting government efforts to boost recovery.

5. Summary and Conclusions

This study offers a detailed analysis of India's economic trends using advanced statistical methods. The ARIMA(2,2,2) model effectively captured the long-term GDP trajectory from 1965 to 2023, addressing data non-stationarity and validating model assumptions. The forecast projects steady GDP growth post the COVID-19 induced recession, reflecting resilience in the Indian economy. Regression analysis highlighted the significant contributions of the Trade, Agriculture, and Financial sectors to GDP growth during 2011–2022, with Trade showing the strongest impact. The study also identified and corrected multicollinearity issues to ensure reliable sectoral insights. Cluster analysis of Indian states and Union territories revealed three distinct economic groups, underscoring regional disparities and economic diversity. These findings can assist policymakers in designing targeted interventions to promote balanced regional development.

In general, this research integrates time series forecasting, regression, and clustering to provide a comprehensive view of India's economic dynamics. It contributes valuable knowledge to understanding sectoral influences and regional economic structures, especially in the context of recovery from recent economic shocks.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The authors received no financial support for the research, authorship and/or publication of this article.

ORCHID ID

Richu Rajesh https://orcid.org/0000-0003-2219-6190

References

Agrawal, V. (2018). GDP modelling and forecasting using ARIMA: An empirical study from India (Master's thesis). Central European University,33.

Basantwani, L. V., Sharma, S., Dagar, V., & Kharinta, K. (2021). Impact of national income and public expenditure on employment and its public-private sector composition in Indian economy. International Journal of Agricultural & Statistical Sciences, 17(1), 213.

Bhatt, R. K. (2011). Recent global recession and Indian economy: An analysis. International Journal of Trade, Economics and Finance, 2(3), 212–217.

Bhattacharya, R., Chakravartti, P., & Mundle, S. (2019). Forecasting India's economic growth: a time-varying parameter regression approach. Macroeconomics and Finance in Emerging Market Economies, 12(3), 205–228.

Daga, U. R., Das, R., & Maheshwari, B. (2004). Estimation, analysis and projection of India's GDP. Schoasticus, 1(2):1-13.

Gupta, A. S. (2009). India's Tryst with the Global Financial Crisis. Review of Market Integration, 1(2), 171-197.

Gupta, M., & Minai, M. H. (2019). An Empirical Analysis of Forecast Performance of the GDP Growth in India. Global Business Review, 20(2), 368-386.

Hassan, M. M., & Mirza, T. (2021). Using Time Series Forecasting for Analysis of GDP Growth in India. International Journal of Education and Management Engineering, 11(3), 40–49.

Jethwani, B., Dave, D., Ali, T., Phansalkar, S., & Ahhirao, S. (2021). Indian agriculture GDP and non performing assets: A regression model. In IOP Conference Series: Materials Science and Engineering (Vol.1042, No.1, p.012007).

Maiti, Dibyendu, Naveen Kumar, Debajit Jha, and Soumyadipta Sarkar. "Post-COVID recovery and long-run forecasting of Indian GDP with Factor-Augmented Error Correction Model (FECM)." Computational Economics 63, no. 3 (2024): 1095-1120.

Majumder, S., Kayal, P., Chowdhury, I. R., & Das, S. (2023). Regional disparities and development in India: evidence from Wroclow Taxonomy and K-means clustering. GeoJournal, 88(3), 3249-3282.

Malik, M. H., & Mirza, T. (2021). Using time series forecasting for analysis of GDP growth in India. International Journal of Education and Management Engineering (IJEME), 11(3), 40–49.

Manimala, M. J., & Kurian, C. (2024). Paradigms of people management: human resource strategies during an economic downturn. Human Resources Management and Services, 6(4), 3400-3400.

Micheal, V. A., & Abiodun, A. A. (2014). Estimation of regression coefficients in the presence of multicollinearity. Social and Basic Sciences Research Review, 2(10), 404–415.

Pandey, R., Chand, D. B., & Tolani, H. (2024). Bayesian multiple linear regression model for GDP in India. International Journal of System Assurance Engineering and Management, 15(6), 2170–2187.

Pervez, A., & Ali, I. (2024). Robust regression analysis in analyzing financial performance of public sector banks: A case study of India. Annals of Data Science, 11(2), 677-691.

Raj, A., Bhattacharyya, P. & Gupta, G.R. Clusters of COVID-19 Indicators in India: Characterization, Correspondence and Change Analysis. SN COMPUT. SCI. 3, 210 (2022).

Satpathi, S., & Hasan, M. R. (2021). Productivity Growth in the Indian Manufacturing Sector: A Way of Mitigating Recession. In Productivity Growth in the Manufacturing Sector: Mitigating Global Recession (pp. 155-165).

Singh, T. (2009). Services sector and economic growth in India. Applied Economics, 42(30), 3925–3941.

Singhania, M., & Gupta, A. (2011). Determinants of foreign direct investment in India. Journal of International Trade Law and Policy, 10(1), 64–82.

Sinha, J. K. (2024). Understanding economic growth in India: How it affects the everyday lives of common people. Journal of Comprehensive Business Administration Research. Advance online publication, 1–12.

Taneja, S., Bhatnagar, M., Kumar, P., & Rupeika-Apoga, R. (2023). India's Total Natural Resource Rents (NRR) and GDP: An Augmented Autoregressive Distributed Lag (ARDL) Bound Test. Journal of Risk and Financial Management, 16(2), 91.