Prediction of GDP in the BRICS Countries Based on the Proportions of Industry and Agriculture

Ziqi Guo[1st,1], Yonghao Li*[1st,2]

1764467235@qq.com[1],395665106@qq.com[2]

School of Economics and Management, Wuhan University, Wuhan, China

Abstract: In the realm of gross domestic product(GDP) research, the multiple regression model has garnered significant academic attention due to its ability to comprehensively consider various factors and its exceptional predictive performance. This study explores multidimensional GDP-related data and introduces a novel model to explain GDP, focusing on its relationship with the proportion of a country's industrial and agricultural output. To analyze the interaction between a nation's economic condition and its economic structure, we employ both static and dynamic research methods. We first decompose GDP into four components and select five countries with similar economic development stages for our analysis. We then use GDP as a measure of economic growth and conduct an in-depth analysis by creating scatter plots illustrating the relationship between the proportion of agriculture and industry in GDP. Finally, we construct a multivariate nonlinear regression model to examine the relationship between GDP and the proportion of industrial and agricultural output. The research findings reveal that our proposed GDP-industrial and agricultural proportion model has undergone rigorous significance testing and demonstrates strong explanatory capability. This study effectively elucidates the strong connection between GDP and a country's economic structure, providing a robust theoretical basis for shaping national policies aimed at adjusting economic structure and further fostering GDP development.

Keywords: Multiple regression analysis; GDP; BRICS countries; economic structure

1 Introduction

Gross Domestic Product (GDP) represents the total value of goods and services produced within a country or region and is a crucial indicator for assessing economic conditions, development, competitiveness, and growth potential.[1] Economic development varies among countries due to differences in economic scale, structure, and geographic location.[2] This study focuses on the BRICS countries to assess their economic development stages, aiming to provide targeted guidance for economic policies and resource allocation. [3]

Scholars have extensively used multiple linear regression models to study the factors influencing regional or national GDP. Cheng Huailin's research corrected multicollinearity in Guangdong Province's economic data and identified six major factors affecting GDP.[4]Li Jiajing used a VAR model and Granger causality tests to highlight the significant impact of per capita output and income growth on GDP. [5]

ICEMME 2023, November 17-19, Beijing, People's Republic of China

DOI 10.4108/eai.17-11-2023.2342697

This study offers a novel approach to analyze GDP factors, using economic data from the BRICS countries to explore the impact of industrial and agricultural output proportions on GDP. By combining dynamic and static analysis methods, it establishes a comprehensive model for a deeper understanding of factors influencing GDP in economically consistent countries, shedding light on macroeconomic trends and offering practical insights for policy adjustments.

2 Theoretical Analysis:

2.1 Time-Series Dynamic Analysis of GDP:

This paper uses two primary methods for this analysis: static and dynamic methods. We've gathered 2020 GDP data from five countries to assess their economic development status.Within the context of GDP analysis, we not only observe overall GDP time-series trends but also investigate how agriculture and industry influence GDP. Our aim is to predict future economic trends in BRICS countries and explore the impact of economic structure on GDP. This section provides a foundational overview of comparing the economic development of different nations and factors affecting GDP.

Our dataset covers 1970 to 2020, decomposing GDP into three sources: agriculture, industry, and craftsmanship. Due to limited craftsmanship data and a relatively small sample size, this paper specifically examines the influence of agricultural and industrial structures on GDP values. We break down GDP sources into agriculture and industry, using the GDP of the five nations as the dependent variable and the proportions of industrial and agricultural output as the independent variables.Then we observe their impact on GDP values.

2.2 Constructing Models for GDP and Industrial and Agricultural Output Proportions:

We generated a scatter plot, where the X-axis represented industrial output proportion, and the Y-axis represented GDP for the respective years. Similar scatter plots were created for the agricultural sector.By examining these scatter plots, we made initial model estimates and selections. Our initial hypothesis for the GDP-Industrial regression model was as follows(1):

$$Y = AX + B + \varepsilon \tag{1}$$

Where Y stands for GDP, A is the coefficient, ε represents the error term, B is the constant term, and X is the independent variable denoting the industrial output proportion.Following the same approach, we formulated the Agricultural-GDP regression model as follows(2):

$$Y = CX + D + \varepsilon \tag{2}$$

In this equation, Y represents GDP, C is the coefficient, ε denotes the error term, D is the constant term, and X is the independent variable representing the agricultural output proportion.

Regression analysis is a common technique to examine the relationship between one or more independent variables and a dependent variable. In this instance, our aim is to understand how GDP is affected by changes in agricultural and industrial output proportions.To enhance prediction accuracy, we utilized the least squares method to estimate model parameters, aiming to best fit the observed data points. We also conducted hypothesis tests to establish the statistical significance of the model. In this model, we test the following four hypotheses:

H0: A = 0, suggesting no linear relationship between industrial output proportion and GDP.

H1: A ≠ 0, indicating a linear relationship between industrial output proportion and GDP.

H2: C = 0, signifying no linear relationship between agricultural output proportion and GDP.

H3: C ≠ 0, revealing a linear relationship between agricultural output proportion and GDP.

We employed t-tests to assess the validity of the model and recorded the p-values for A and C in the model. If the p-value for A is significantly less than the 0.05 significance level, it implies a significant linear impact of industrial output proportion on GDP. Similarly, if the p-value for C is significantly less than the 0.05 significance level, it suggests a significant impact of agricultural output proportion on GDP, rendering the model predictive. A higher Multiple R-squared value indicates that the model better explains the data's variation.

2.3 Constructing a Multivariate Nonlinear Regression Model:

Based on the analysis of data and simple linear regression equations for Industrial-GDP and Agricultural-GDP, which showed linear relationships with GDP, we have developed a multivariate nonlinear regression model to better understand the relationship between GDP and the proportions of industrial and agricultural output. The equation is as follows(3):

$$\mathrm{GDP} = \mathrm{AX1} + \frac{\mathrm{B}}{\mathrm{X2}} + \mathrm{C} + \varepsilon \tag{3}$$

Here X1 represents the proportion of industrial output and X2 represents the proportion of agricultural output, A and B are the parameters for the respective variables, C is the constant term, and ε is the error term. [6]

This model aims to explore the relationship between GDP and industrial and agricultural output proportions. We used the least squares method to estimate the model's parameters and conducted significance tests, including t-tests and an F-test, to validate the model. [7] We also examined p-values and the multiple R-squared to assess the model's performance.

3 Model Construction and Testing

3.1 Data Source and Experimental Setup:

In our experiment, data was sourced from a dataset provided by the World Bank. After an initial selection process from the dataset, we chose three variables related to GDP: agriculture, industry, and craftsmanship. Additionally, we selected consumption, investment, government purchases, and net exports as the four components that constitute GDP. Our data covers the time period from 1970 to 2020.Before commencing the experiment, we performed data cleaning and organization. Linear interpolation was used to fill in missing values to ensure data reliability and authenticity. In this paper, the expenditure approach(4) was utilized to calculate a nation's total output:

$$\mathrm{GDP} = \mathrm{C} + \mathrm{I} + \mathrm{G} + \mathrm{NX} \tag{4}$$

3.2 Experimental Process:

The experimental process can be broken down into three main steps:

Step 1: Data Preprocessing

In this step, we utilized R Studio as a statistical analysis tool. We organized the data into a format suitable for statistical analysis and performed data cleaning to eliminate missing values (NA values) to make the dataset more suitable for subsequent research and experiments.

Step 2: Dynamic Analysis

In this step, we extracted data for consumption, investment, government purchases, net exports, and total GDP for the BRICS countries in the year 2020. We calculated the proportions of these data components within the total GDP for each country and created stacked bar charts Fig. 1to visually demonstrate the macroeconomic structure of each country.

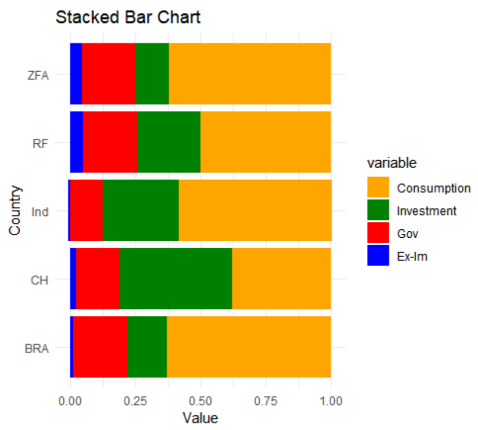

Fig. 1. Structure of Expenditure Approach Proportions for BRICS Countries

Continuing with GDP Trend Analysis, we observed the changing trends of GDP over time for the five countries to infer their temporal patterns. Time series is a sequence of numerical values or observations of a statistical variable or indicator arranged in chronological order with equal time intervals between them. In this context, the GDP values for each country constitute time series data.[8]We individually extracted the annual GDP values for each country from the dataset and plotted them on the same line chartFig. 2 to observe the evolving trends in GDP for these different countries over time. It is evident that within the available data timeframe, China's GDP significantly outpaces that of the other four nations. However, the overall consistency in the changing trends of GDP for the five countries further validates the similarity in the economic development stages of the selected nations.[9] This lays the foundation for our subsequent selection of a similar type of multivariate linear model.

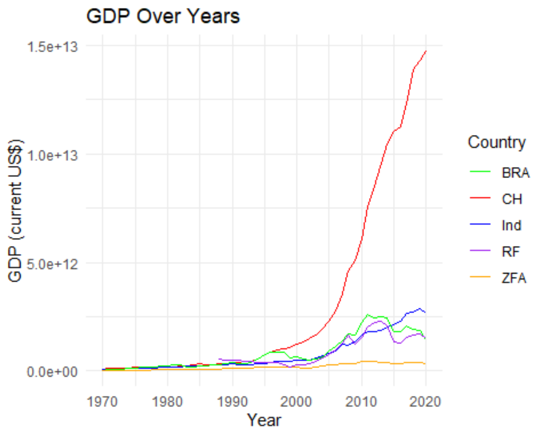

Fig. 2.GDP Trends for BRICS Countries

Step 3: Static Analysis and Modeling:

We extracted the annual industrial output, agricultural output, and total GDP values for each country from the dataset. Subsequently, we created relevant scatter plots Fig. 3. The following regression analysis is presented using charts and data for China as an example.

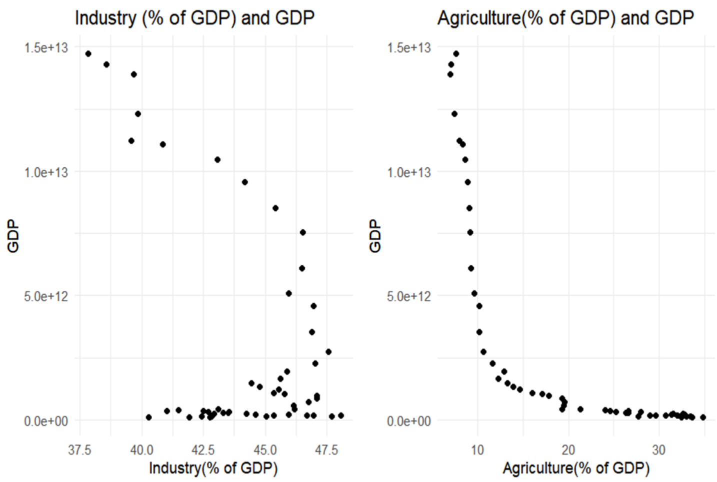

Fig. 3.GDP-Industrial and Agricultural Proportion Distribution Scatter Plot

Based on the distribution shape of the scatter plots above, we preliminarily assessed that the regressions of industry (5)and agriculture on GDP(6) can be represented by the models:

$$Y = \beta 0 + \beta 1 * X1 \quad (5)$$

$$Y = \beta 3 + {}^{\beta 4}/_{X2} \quad (6)$$

Here, Y represents China's Gross Domestic Product (GDP), β is the coefficient, X1 stands for the proportion of industrial output, and X2 represents the proportion of agricultural output.

Subsequently, we took the reciprocal of agricultural output as the independent variable and estimated the parameters for the GDP-Agricultural Proportion model. We conducted tests and calculations using the least squares method, resulting in a regression line Fig. 4represented as (7)

$$Y = -4.422e^{+12} + {1.134e^{+14}}/{X2} \quad (7)$$

In this model, the multiple R-squared (R^2) is 0.8534, indicating that the model can explain approximately 85.34% of the variation in GDP values. The p-value for 1/X2 is less than 2.2e-16, significantly lower than 0.05, suggesting that 1/X2 has a significant linear impact on the dependent variable Y. This implies that the proportion of agricultural output significantly influences GDP. Overall, this model exhibits strong explanatory power for the data, and it fits the data very well. It can be considered effective for predicting the relationship between Y and X2.

Using the same steps, we estimated the parameters for the GDP-Industrial Proportion model.(8)

$$Y = 4.116e^{+13} - 8.609e^{+11} * X1 \quad (8)$$

The regression equation obtained through regression analysis is as follows Fig. 5.

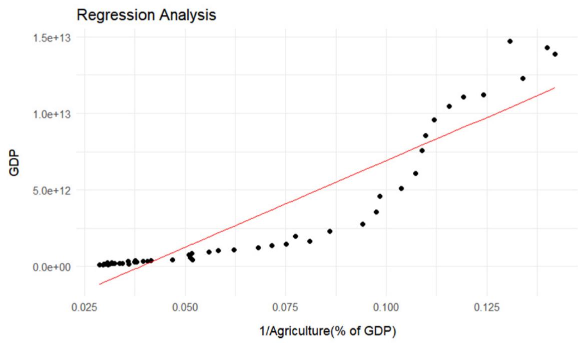

Fig. 4.GDP-Agricultural Output Proportion Regression Line

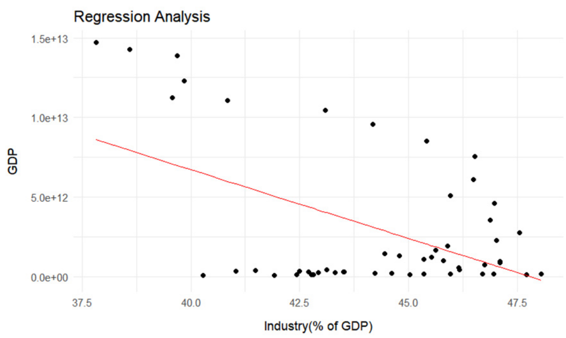

Fig. 5.GDP-Industrial Output Proportion Regression Line

In the examination of the regression coefficients for this model, the p-value for X1 is 0.000173, which is less than 0.05, indicating that X1 has a significant linear impact on the dependent variable Y. However, this model has relatively low goodness of fit, explaining only a small portion of the variance in the dependent variable. It might be necessary to consider other independent variables or more complex models to improve its explanatory power. Therefore, we constructed the GDP-Industrial and Agricultural Proportion model, comprehensively analyzing the influence of both variables on GDP.

Based on the existing models, we formulated the model [10]as follows (9):

$$Y = \beta0 + \beta1 * X1 + {}^{\beta2}/_{X2} + \varepsilon \quad (9)$$

Where Y represents GDP, and X1 and X2 are the explanatory variables. β0 is the constant term, and β1 and β2 are the regression coefficients for the independent variables X1 and X2. ε is the random disturbance term. Using the R software, we conducted regression analyses on the mentioned variables and obtained the preliminary regression equation for China as follows (10):

$$Y = 1.722e^{+13} - 4.75\ e^{+11} * X1 + {}^{1.044e^{+14}}/_{X2} \quad (10)$$

For India (11), Brazil(12), South Africa(13), and Russia(14), the regression equations are as follows:

$$Y = 2.352e^{+12} - 1.898\ e^{+11} * X1 + {}^{8.085e^{+13}}/_{X2} \quad (11)$$

$$Y = -8.434e^{+11} - 2.622\ e^{+9} * X1 + {}^{1.094e^{+13}}/_{X2} \quad (12)$$

$$Y = -6.648e^{+10} - 1.263\ e^{+9} * X1 + {}^{9.537e^{+11}}/_{X2} \quad (13)$$

$$Y = -3.611e^{+12} + 6.527\ e^{+10} * X1 + {}^{1.12e^{+14}}/_{X2} \quad (14)$$

3.3 Experimental Verification

We conducted significance tests on the model we constructed for the industrial and agricultural proportions with respect to GDP. We performed T-tests and P-tests on the coefficients of its independent variables and intercept. The final test results are presented in the following tableTable 1.Results tested by multiple regression models:

Table 1. Results tested by multiple regression models

Country	Coefficients	Estimate	Std.Error	T-value	Pr(>|t|)	Significance
China	Intercept	1.722e+13	3.216e+12	5.354	2.39e-06	Very High
	X_1	-4.757e+11	7.024e+10	-6.773	1.64e-08	Very High
	X_2	1.044e+14	5.030e+12	20.752	<2e-16	Very High
India	Intercept	2.352e+12	5.158e+11	4.560	3.54e-05	Very High
	X_1	-1.898e+11	2.338e+10	-8.118	1.46e-10	Very High
	X_2	8.085e+13	4.428e+12	18.260	<2e-16	Very High
Brazil	Intercept	-8.434e+11	7.630e+11	-1.105	0.0275	----
	X_1	-2.622e+09	1.477e+10	-0.177	0.860	----
	X_2	1.094e+13	2.192e+12	4.989	8.38e-06	Very High
South Africa	Intercept	-6.648e+10	1.165e+11	-0.571	0.571	----
	X_1	-1.263e+09	2.460e+09	-0.514	0.610	----
	X_2	9.537e+11	1.310e+11	7.281	2.73e-09	Very High
Russia	Intercept	-3.611e+12	7.516e+11	-4.805	4.36e-05	Very High
	X_1	6.527e+10	1.611e+10	4.052	0.000348	Very High
	X_2	1.120e+13	1.112e+12	10.073	5.59e-11	Very High

According to the results of the tests, it is evident that the coefficients of the models for each of the BRICS countries have all passed the significance tests and are at a high level of significance. Taking the model constructed based on Chinese data as an example, let's analyze its predictive capability. It can be observed that the p-values for the intercept term, as well as the coefficients for x1 and x2, are all significantly below 0.05, indicating their substantial linear impact on the dependent variable y. Furthermore, the corresponding coefficients have relatively small standard deviations, suggesting that the model we constructed has a good fit and a high level of explanatory power.

To further illustrate the explanatory power of our constructed models, we calculated the values of Multiple R-squared and Adjusted R-squared for the BRICS countries. Additionally, we conducted an F-test to verify the significance, calculated the corresponding degrees of freedom, and further examined the model's significance. The specific values are shown in the following tableTable 2:

Table 2. Significance test results of the model

Country	Residual Standard Error	Multiple R-squared	Adjusted R-squared	F-statistic	
China	1.245e+12 on 48 degrees of freedom	0.925	0.9219	296.1 on 2 and 48 DF	p-value:<2.2e-16
India	2.912e+11 on 48 degrees of freedom	0.8848	0.88	184.3 on 2 and 48 DF	p-value:<2.2e-16
Brazil	4.403e+11 on 48 degrees of freedom	0.6974	0.6848	55.31 on 2 and 48 DF	p-value:3.47e-13
South Africa	4.726e+10 on 48 degrees of freedom	0.8835	0.8786	182 on 2 and 48 DF	p-value:<2,2e-16

Russia	2.711e+11 on 29 degrees of freedom	0.8458	0.8352	79.56 on 2 and 29 DF	p-value:1.68e-12

From the values of Multiple R-squared and Adjusted R-squared, it is evident that the models we constructed all have a good level of explanatory power. Furthermore, our models have all passed the F-test, using China as an example. The p-value obtained in the F-test is less than 2.2e-16, indicating an exceptionally high level of significance for the model.[11]

4 Conclusion and Summary

In summary, an increase in the industrial value-added as a percentage of GDP accelerates economic growth, while an increase in the agricultural value-added as a percentage of GDP suppresses GDP growth. Utilizing a combined approach of dynamic and static analysis, along with modeling and testing based on economic data from the BRICS countries, this conclusion demonstrates a degree of universality. This research, based on existing analytical methods and regression model construction and testing, has observed the annual variations in the consumption, expenditure, government purchases, and net exports as a percentage of GDP, leading to the inference of the similarity in economic development stages among the five countries. Additionally, the impact of industrial and agricultural ratios on GDP has been explored, and a preliminary analysis has been conducted on how the distinct economic structural characteristics of countries affect GDP.[12]

In the future, as this model continues to evolve and improve, we will further expand the model's sample set to explore its applicability in countries with varying economic conditions. On the other hand, we will endeavor to optimize the model by introducing more relevant factors to enhance its fitting capability, making it more valuable for forecasting national GDP and policy adjustments.[13]

We believe that as the model continues to develop, improve, and expand its sample set, the guidance role of industrial and agricultural ratios in national economic development will be further reinforced. This study provides valuable insights for understanding the macroeconomic structures and GDP fluctuations of the BRICS countries, offering robust support for future policy formulation and economic development strategy planning. In the future, nations should pay more attention to macroeconomic regulation of industrial and agricultural ratios to stimulate sustained economic growth effectively.

References

[1] Liang, H. N. (2021). Empirical Research on the Influencing Factors of Anhui Province's GDP: Based on Multivariate Regression Analysis. Times Finance, 24, 73-75.

[2] Deng, C. Y. (2023). Analysis of the Influencing Factors of Hainan Province's GDP Based on the STATA Model. Heilongjiang Science, 18, 93-96.

[3] Karine HAJI,E-commerce development in rural and remote areas of BRICS countries,Journal of Integrative Agriculture,Volume 20, Issue 4,2021,Pages 979-997,ISSN 2095-3119,https://doi.org/10.1016/S2095-3119(20)63451-7.

[4] Cheng, H. L. (2023). Empirical Analysis of the Influencing Factors of Guangdong Province's GDP. Cooperative Economy and Technology, 23, 30-32. doi: 10.13665/j.cnki.hzjjykj.2023.23.045.
[5] Li, J. J., & Wei, Q. Y. (2023). Empirical Analysis of Factors Affecting GDP Growth Rate Based on the VAR Model. China Business Review, 6, 15-17. doi: 10.19699/j.cnki.issn2096-0298.2023.06.015.
[6] Advanced Statistics in Criminology and Criminal Justice，ISBN : 978-3-030-67737-4，David Weisburd, David B. Wilson, Alese Wooditch, Chester Britt；Journal: Economía y Negocios, 2022, Volume 13, Number 2, Page 30；DOI: 10.29019/eyn.v13i2.996
[7] Sureiman, Onchiri; Mangera, Callen Moraa1. F-Test of Overall Significance in Regression Analysis Simplified. Journal of the Practice of Cardiovascular Sciences 6(2):p 116-122, May–Aug 2020. | DOI: 10.4103/jpcs.jpcs_18_20
[8] Ju, Y. H., & Zhang, H. W. (2020). Overview and Selection of Economic Forecasting Methods. In Financial Observation (2020 Edition) (pp. 159-168).
[9] Wu, L., & Zheng, J. (2022). Comparative Analysis of Development Potential in BRICS Countries. Economic World, 1, 70-81.
[10] Gao, J. W. (2023). Estimation of Parameters in Multivariate Linear Regression. Journal of Xi'an University of Arts and Sciences (Natural Science Edition), 3, 1-5+12.
[11] Wang, H. (2020). Analysis of Factors Affecting Economic Development Based on Multivariate Linear Regression. Business News, 36, 146-147.
[12] Han, M. L. (2022). Research on the Upgrading of China's Economic Industrial Structure under the Background of High-Quality Development (Doctoral Dissertation, Jilin University).
[13] Xue, Y. X., & Wang, P. A. (2023). Private Investment, Upgrading of Industrial Structure, and Economic Growth: An Analysis Based on Provincial Panel Data Using the PVAR Model. Modern Business, 16, 55-59. doi: 10.14097/j.cnki.5392/2023.16.011.