Journal of Global Trends in Social Science

https://doi.org/10.70731/3ercnp44

Research on The Correlation between Investments in Various Industries Based on Regression Model

Jia Gao [a], Chenxi Ye [b,*], Fujiang Yuan [b], Qiang Xiong [a]

[a] *Taiyuan Normal University, Taiyuan, China*
[b] *Hong Kong Baptist University, Hong Kong, China*

KEYWORDS

Industrial Structure Adjustment;
Linear Regression;
Nonlinear Regression;
GDP

ABSTRACT

As China transitions to high-quality economic development, industrial structure adjustment has become crucial. Government investment, as a key tool of macroeconomic regulation, plays an important role in guiding industrial upgrading, optimizing resource allocation, promoting growth, and expanding employment. This study analyzes the relationship between industry-specific investment and GDP to provide a scientific basis for investment allocation. Using industry GDP data from 1990 to 2023, issues such as outliers, missing values, and unit inconsistencies were addressed to ensure data integrity. Pearson correlation analysis revealed that sectors such as industry, construction, manufacturing, finance, IT services, education, and healthcare are strongly linked to GDP growth. Based on the characteristics of each sector, linear or nonlinear regression models were developed to quantify investment impact. Under a total government investment constraint of 1 trillion yuan, an optimized investment plan was proposed using optimization algorithms like fmincon to maximize GDP. Analysis of historical investment proportions showed rising shares for IT services and finance, while agriculture and textiles declined. The findings demonstrate that scientific planning and optimized investment can effectively support industrial upgrading, improve resource efficiency, and promote sustainable economic growth and employment.

Introduction

With the continuous evolution of the global economic structure [1] and the transformation of the domestic economic development stage, China faces the urgent task of adjusting the industrial structure [2]. In this process, government investment, as an important means of macroeconomic regulation [3], plays an irreplaceable role in guiding industrial upgrading, optimizing resource allocation, promoting economic growth and expanding employment. However, how to scientifically plan and effectively utilize government investment to maximize GDP while promoting full employment has become a key issue that needs to be solved urgently. At present, China's economy has shifted from a stage of high-speed growth to a stage of high-quality development, and its industrial structure is undergoing profound changes. On the one hand, traditional industries are

* Corresponding author. E-mail address: u430233106@mail.uic.edu.cn

Received 30 March 2025; Received in revised form 18 April 2025; Accepted 28 April 2025; Available online 30 April 2025.
2759-7830 / © 2025 The Author(s). Published by Jandoo Press. This is an open access article under the CC BY 4.0 license.

facing pressure of transformation and upgrading, and need to improve their competitiveness through technological innovation and model innovation; on the other hand, emerging industries are rising rapidly and becoming a new driving force for economic growth. In this context, the direction, scale and structure of government investment directly affect the adjustment path and effect of the industrial structure.

Government investment [4] plays an important role in promoting economic growth and expanding employment. Through measures such as investing in infrastructure construction, supporting scientific and technological innovation, and cultivating emerging industries, the government can effectively stimulate domestic demand, promote the coordinated development of upstream and downstream enterprises in the industrial chain, and thus create more employment opportunities. However, government investment also faces challenges such as resource allocation efficiency and return on investment. How to achieve the maximum benefits of government investment under limited resources is a question that policy makers need to think deeply about. At the critical period when China is facing a critical period of industrial structure adjustment and upgrading, advanced learning methods such as artificial intelligence [5] and deep learning [6] are used to solve the problems of government investment and industrial structure adjustment in China's future industries. Artificial intelligence and deep learning have significantly improved decision-making efficiency and economic benefits through efficient processing of data, optimizing resource allocation, adapting to market changes, promoting industrial innovation, enhancing policy transparency, improving employment quality and quantity, and supporting the Sustainable Development Goals in government investment decisions [7].

Related Work

Yao Zhanqi et al. [8] used the transcendent logarithmic production function and Cobb-Douglas (C-D) production function method to calculate the total factor productivity (TFP) of China's overall economy and manufacturing industry, quantify the impact of capital transfer and labor flow on TFP, and test the structural dividend hypothesis. Pan Huifeng et al. [9] collected relevant data from ten western provinces and regions from 1985 to 2008, established a panel data model, and explored the impact of agricultural industrial structure on the income of rural farmers in the west and the relationship between agricultural structure and energy efficiency. Optimizing agricultural structure will help improve the income and energy efficiency of rural farmers in the west. He Yonggui et al. [10] used the gray comprehensive evaluation method, hierarchical analysis method and fuzzy evaluation method to comprehensively evaluate the internal structure adjustment of the primary industry. The research emphasizes that in the adjustment of regional industrial structure, it is necessary to achieve the transformation of the industrial structure from "two, one, three" to "three, two, one", and it is necessary to ensure that the second and third industries provide support and a good environment for the adjustment of the primary industry, so as to promote the coordinated development of the third industries and achieve sustainable economic growth. Jun Han et al. [11] used an empirical approach to explore the impact of China's industrial structure adjustment and labor mobility on urban and rural income gaps between 1990 and 2019. Ding Qiang Duan et al. [12] studied the relationship between industrial structure adjustment, technological innovation and energy efficiency in Hubei Province from 2001 to 2010 by reviewing and analyzing relevant literature. It was found that industrial structure adjustment and technological innovation can promote energy efficiency. In response to subsequent energy conservation and other issues, it is recommended that the government encourage citizens to participate, pay attention to the application of market mechanisms, and use preferential tax policies to promote technological development and application. Tianhe Jiang et al. [13] first analyzes the impact of industrial structure adjustment and economic development quality on the city-agricultural income gap theoretically, and then uses spatial dynamic models to explore the relationship and threshold impact of the three through complete sample analysis and regional analysis. The problem of clarifying the impact of industrial structure adjustment scope, quality and economic development quality on urban agricultural income gap was solved, and policy recommendations for more equitable development transition were put forward.

This study focuses on multiple economic issues. Through the analysis of data set processing and regression model, it provides a scientific basis for government investment allocation. The main contributions are as follows:

1) Process the GDP data of various industries from 1990 to 2023 to solve the problems of outliers, missing values and unit conversion, and ensure data integrity.
2) This paper analyzes the contribution of each industry to GDP in different years, and reveals the potential correlation between industries through regression models.
3) This paper quantifies the impact of investment in each industry on GDP, establishes linear or nonlinear regression models for each industry, and concludes which industries have the highest return on investment, providing theoretical support for subsequent industrial adjustments.
4) Based on the regression model, under the constraint of the total government investment of 1 trillion yuan, an industry investment allocation plan

was proposed using an optimization algorithm (such as ‘fmincon’) to maximize the total GDP.

5) By analyzing the rationality of investment in various industries and visually demonstrating the contribution of various industries to GDP,Provides a quantitative basis for investment decisions..

Data Description

Dataset Source

The current data set comes from the National Bureau of Statistics and the China Economic and Social Big Data Research Platform. The data set covers multiple fields such as agriculture, industry, and industrial manufacturing. There are complex interrelated relationships between industries in this data set, and the data set designs the added value of GDP in various industries in China, reflecting the contribution of each industry to the total GDP. Table 1.2 is an illustration of the dataset fields used in this article [14].

Table 1 | Description of the fields of contribution of each industry to GDP in different years

Field Name	Illustrate
YEAR	Years
Total GDP	Measurement of national economic scale and economic activity
Output value of each industry	Total output value of each industry, added value

Table 2 | Description of fields of investment and corresponding output value indicators for each industry

Field name	Illustrate
YEAR	years
Investment value of each industry	Investment costs for each industry every year
Output value of each industry	The output value corresponding to the investment of each industry every year

Data Preprocessing

In the above data set, in order to unify the units and ensure that all data use the same dimensions, we assume that the default unit of all GDP is "10,000 yuan", but for the sake of easy analysis and comparison, all industries' GDP data are converted into "10,000 yuan", which ensures that the comparison between industries is consistent. The unit conversion formula is the following formula 1.

$$\mathrm{GDP}_{\mathrm{converted}} = \frac{\mathrm{GDP}_{\mathrm{original}}}{10000} \tag{1}$$

And for the exact value, considering that this part of the data set is a numerical data set, we use linear interpolation to estimate based on the existing data points before and after, and fill in the missing values. The linear interpolation formula is shown in Equation 2 below.

$$y_k = y_i + \frac{x_k - x_i}{x_j - x_i}(y_j - y_i) \tag{2}$$

Pearson Correlation Analysis

In this problem, this paper uses Pearson's correlation coefficient method [14] to analyze the correlation between each index parameter in the dataset and the total GDP. Pearson's correlation coefficient method is a classic statistical analysis method used to quantify the intensity and trend of linear correlation between continuous variables. The core of Pearson correlation is to calculate the Pearson Correlation Coefficient (R) of investment and GDP, which is a value between [-1, 1]. This value not only reveals whether there is a linear correlation between variables, but also points out the direction of the correlation (if R>0, it indicates that the two variables are positively correlated, if R<0, it indicates that the two variables are negatively correlated) and the strength of the correlation (nearly 1 or -1 means strong correlation, close to 0 means weak correlation or unrelated). The formula is:

$$R = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}} \tag{3}$$

Thermal graph 1 This graph shows the thermal graph of correlation between various industries. The horizontal and vertical axis lists multiple industrial and economic indicators, such as the total GDP, such as the chemical industry, finance industry, construction industry, etc. The depth of the color in the figure indicates the strength of the correlation, and the color gradually changes from yellow (negative or weak correlation) to blue (strong positive correlation). It can be seen that the correlation between total GDP and industry, construction, manufacturing, finance, insurance, education and medical care is very strong (nearly 1.0, dark blue), indicating that these two major industries have made great contributions to GDP growth. Some industries are highly correlated (such as manufacturing and construction, finance and IT services) between specific industries, which may be because there are synergies or complementary relationships between these industries.

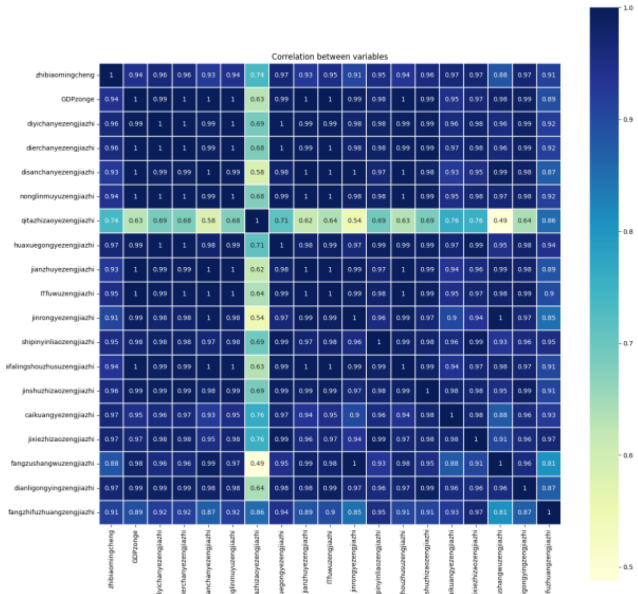

Figure 1 | Heatmap of correlation between various industrial indicators and total GDP

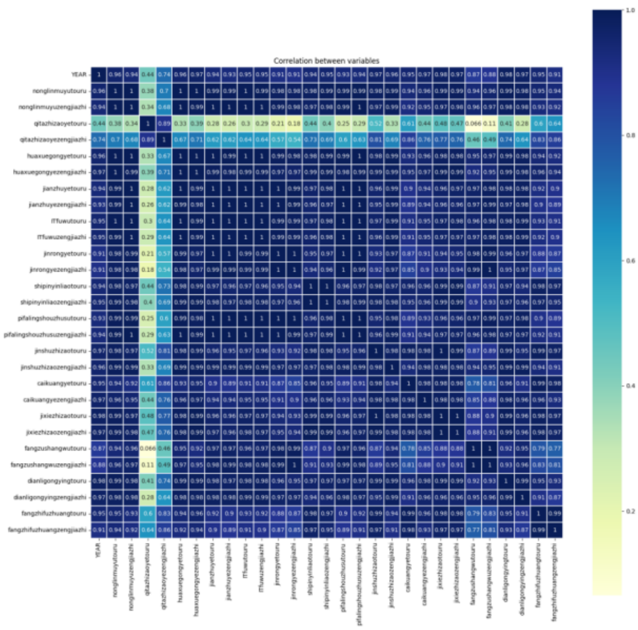

Figure 2 | Heat map of investment and corresponding output value indicators in various industries

The correlation between most indicators of Heat Figure 2 is very high, while the correlation between the two indicators of other manufacturing input and output value is low. When we make investment adjustments, we will temporarily account for a small proportion.

Regression Model

This article will study the relationship between investment in each industry and its corresponding GDP increase, analyze the impact of investment on GDP by constructing a regression model, and use linear models and nonlinear models to describe the relationship between investment adjustment and GDP increase according to the strength of correlation.

Linear Regression Model

Linear Regression is a statistical method that models the relationship between dependent variables (objectives) and independent variables (features) through linear combinations (weighted sums). Its core assumption is that the relationship between variables is linear and the error term is distributed from normal.

Formula: Suppose there are n independent variables x1, x2,..., xn, the mathematical expression of the linear regression model is:

$$y = \beta_0 + \beta_1\chi_1 + \beta_2\chi_2 + .. + \beta_n\chi_n + \epsilon \qquad (4)$$

y is the dependent variable; β_0 is an intercept term; $\beta_0,\beta_1,\ldots\beta_n$ is the regression coefficient; ϵ is the random $\mathcal{N}(0,\sigma^2)$ error of obedience.

Parameter estimation: Estimate parameters by least squares (OLS), with the goal of minimizing the residual sum of squares (RSS):

$$\min_{\beta} \sum_{i=1}^{N} (y_i - (\beta_0 + \beta_1\chi_{i1} + \beta_2\chi_{i2} + .. + \beta_n\chi_{in}))^2 \qquad (5)$$

In matrix form, the parameter solution is:

$$\hat{\beta} = (X^T X)^{-1} X^T y \qquad (6)$$

In the case where the correlation coefficient R exceeds the preset threshold of 0.90, we believe that there is a significant linear relationship between investment volume and GDP industry returns. Therefore, this study uses a linear regression model to describe this relationship, and the model expression is as follows:

$$\mathrm{GDP\_income}_{\mathrm{industry}} = \mathrm{k} \, . \, \mathrm{Invest}_{\mathrm{industry}} + \mathrm{b} \qquad (7)$$

The linear regression model architecture diagram is as Figure 3.

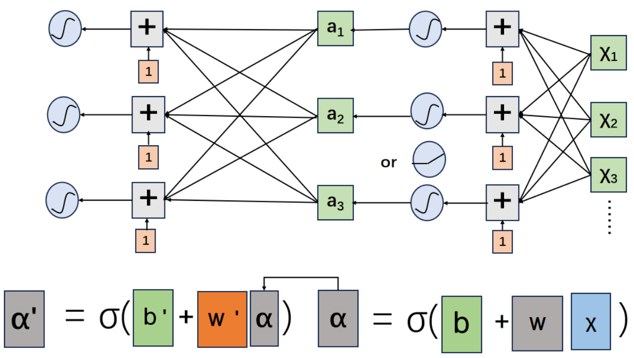

Figure 3 | Linear regression model architecture diagram

Nonlinear Regression Model

Nonlinear Regression is used to model the nonlinear relationship between dependent variables and independent variables. Its regression coefficients do not appear in the form of linear combinations and may involve tricks, logarithms, polynomials, or other complex functions.

The general form of nonlinear models is:

$$y = f(x, \beta) + \epsilon \tag{8}$$

$f(x, \beta)$ is a nonlinear function; $\beta = [\beta_0, \beta_1 \ldots\ldots \beta_k,]$ is the parameters to be estimated; ϵ is the error term.

Example

1) Exponential model:

$$y = \beta_0 e^{\beta_1 \chi} + \epsilon \tag{9}$$

2) Polynomial model (parameters are linear but features are nonlinear):

$$y = \beta_0 + \beta_1 \chi + \beta_2 \chi^2 + \epsilon \tag{10}$$

3) Logical Sty Growth Model (parametric nonlinear):

$$y = \frac{\beta_0}{1 + e^{\beta_1(\chi - \beta_2)}} + \epsilon \tag{11}$$

Parameter estimation Nonlinear regression usually uses numerical optimization methods (such as gradient descent, Newtonian method, or nonlinear least squares method) to estimate parameters.

$$\min_{\beta} \sum_{i=1}^{m} \left(y_i - f(x_i - \beta)\right)^2 \tag{12}$$

Table 3 | The core differences between linear regression and nonlinear regression

Features	linear regression	Nonlinear regression
Model form	Parameter linear (eg $\beta_1 x_1$)	Parameter nonlinear (eg $\beta_1 e^{\beta_2 \chi}$)
Interpretability	Intuitiveness, and coefficients directly reflect the influence of variables.	Complex, need to be analyzed in combination with functional form
Parameter estimation	Analytical solution (closed solution)	Numerical optimization (iterative approximation)
Application scenario	Linear relationships, low-dimensional data	Nonlinear relationships, complex pattern fitting

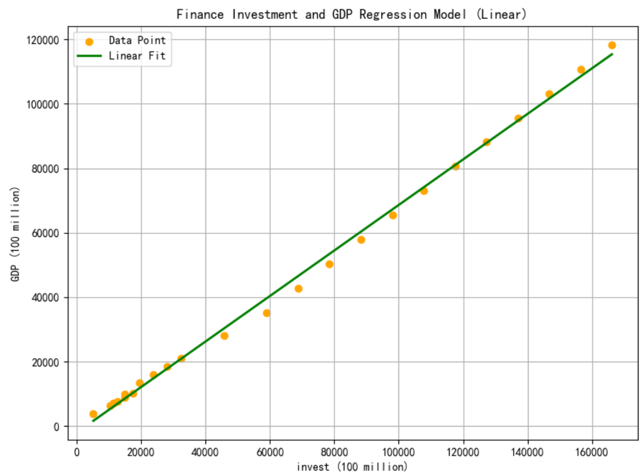

Figure 4 | Finance_Investment and GDP regression fitting diagram

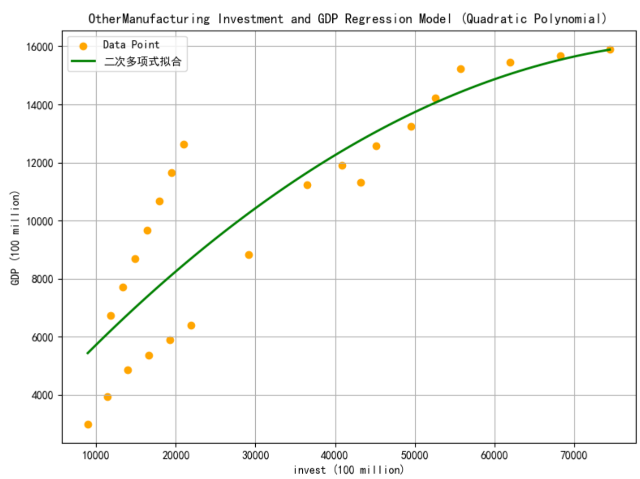

Figure 5 | OtherManufacturing_Investment and GDP regression fitting (quadratic term) diagram

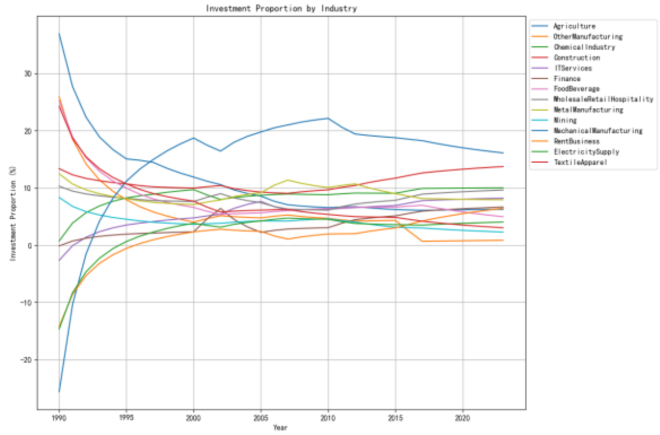

Figure 6 l Investment situation in various industries

The core differences between linear regression and nonlinear regression are as Table 3.

When the correlation coefficipent R of investment volume and GDP industry returns is lower than the threshold of 0.90, it indicates that the linear relationship between the two is not significant enough, so this study uses a nonlinear regression model for analysis. The formula for the nonlinear regression model is as follows:

$$GDP\_income_{industry} = \alpha + \beta * Invest_{industry} + \gamma * (Invest_{industry})^2 \quad (13)$$

Figure 4 shows the linear fit between investment and GDP. The orange point represents the actual data, and the green line is the regression fitting curve. The data points are closely surrounding the regression line, indicating that the linear model has good fit and explanatory power.

Figure 5 shows the nonlinear fitting relationship between investment and GDP in an industry, where the orange point is the actual data and the green curve is a quadratic polynomial regression fit. The model effectively captures the nonlinear trend of investment and GDP, showing the accelerated GDP growth brought about by investment growth and the slowdown in growth at high investment levels. The fitting curve is highly consistent with the data points, confirming the applicability of the model in explaining nonlinear relationships.

Investment Proportion Analysis

This article aims to analyze the proportion of various industries in investment over the years and provide reference for subsequent investment optimization.

Figure 6 shows the changing trend of investment share in various industries from 2000 to 2025. Different fold lines represent changes in investment proportion in various industries. It can be seen from the figure that the proportion of investment in the agriculture and financial industry is relatively stable, while the proportion of construction and IT services has changed significantly, among which the proportion of IT services and financial industry has increased, while the proportion of agriculture and textile and clothing industry has decreased. Overall, the figure reveals industry changes in the economic investment structure and provides a visual reference for industrial transformation and investment strategy analysis.

Optimization Model Establishment

The goal of this section is to calculate the optimal allocation plan for 1 trillion yuan of government total investment among industries by building an optimization model to maximize the total GDP. During the optimization process, the constraint optimization algorithm fmincon is used, using the investment amount of each industry as the decision variable, with the goal of maximizing the total GDP.

Objective function (maximizes total GDP):

$$\max GDP_{total} = \sum_{i=1}^{n} GDP_i(x_i) \quad (14)$$

Model Solution

This article will use the Fmincon optimization algorithm to solve it. By setting the target function and constraints, Fmincon can effectively find the optimal solution that satisfies all constraints. The target function is a negative value of total GDP:

$$\min -\sum_{i=1}^{n} GDP_I(x_i) \quad (15)$$

Table 4 | Fixed investment quota list for each restricted industry

Industry Names	in 100 million yuan
Agriculture	1238.00
OtherManufacturing	517.00
ChemicalIndustry	747.00
Construction	853.00
ITServices	871.00
Finance	714.00
FoodBeverage	446.00
WholesaleRetailHospitality	1012.00
MetalManufacturing	656.00
Mining	178.00
MechanicalManufacturing	1560.00
RentBusiness	692.00
ElectricitySupply	263.00

Table 5 | Fixed investment ratio table for each industry under restricted conditions

Optimal Investment Amount of Each Industry	Investment Proportion	GDP Proportion GDP
Agriculture	12.38%	6.68%
OtherManufacturing	5.17%	13.13%
ChemicalIndustry	7.47%	14.59%
Construction	8.53%	0.64%
ITServices	8.71%	7.04%
Finance	7.14%	-6.80%
FoodBeverage	4.46%	2.93%
WholesaleRetailHospitality	10.12%	5.18%
MetalManufacturing	6.56%	18.29%
Mining	1.78%	17.86%
MechanicalManufacturing	15.60%	7.64%
RentBusiness	6.92%	9.32%
ElectricitySupply	2.63%	12.96%
TextileApparel	2.53%	5.19%

Figure 7 I Investment situation in various industries

Figure 8 I The optimal investment allocation ratio of each industry under the maximization of GDP

The sequence planning (SQP) algorithm is used to iteratively solve the optimal investment solution. The formula for calculating the investment ratio to GDP ratio:

$$\text{InvestShare}_i = \frac{x_i}{\text{Total Investment}} \times 100\% \qquad (16)$$

$$\text{GDPShare}_i = \frac{\text{GDP}_i(x_i)}{\text{GDP}_{\text{total}}} \times 100\% \qquad (17)$$

Table 5 reveals the investment allocation ratio of each industry under the optimized resource allocation strategy. The financial industry is leading the market. It shows the importance of investment in the construction and chemical industries. The investment ratio between mining and metal manufacturing is also significant. Relatively speaking, the proportion of investment in traditional industries such as agriculture, food and beverage, textile and clothing is relatively low. The chart generally shows that investment allocation tends to high-return industries, while taking into account the moderate support from other industries.

Figure 7 shows the ideal investment distribution for each industry under the goal of achieving GDP maximization. Each sector represents the investment proportion of different industries, and the colors distinguish industries. It can be seen from the figure that the metal manufacturing and mining industries have the largest share of investment, highlighting their contribution to GDP. The machinery manufacturing industry and RENTBusiness list the importance of its investment.

Figure 8 shows the optimal investment ratio for each industry when achieving maximum GDP. In the figure, mining and metal manufacturing have the highest investment share, highlighting their key role in economic growth. Agriculture ranks an important position with an 11.4%, showing the importance of its basic industries. The proportion of investment in industries such as construction, chemical industry and leasing commercial services is also relatively high, while the lower proportion of investment in other industries reflects its lower sensitivity to economic output.

Increase Industry Constraints

In order to maximize the total GDP, this paper introduces industry constraints and uses regression models to establish linear or nonlinear regression functions of investment and GDP for each industry to evaluate the contribution of investment to GDP. On the premise of meeting the constraints, investment in some industries is fixed at the minimum limit, while the remaining funds are optimized to be allocated in three high-return variable industries, including the finance industry, information technology services industry and agriculture, to maximize the total investment benefits.

Conclusions

This article aims to explore how to promote China's industrial structure adjustment and high-quality economic development through scientific planning and optimization of government investment. First, by processing and analyzing the GDP data of various industries from 1990 to 2023, the outliers, missing values and unit conversion problems in the data are solved, and the relationship between GDP contribution between industries is revealed. Secondly, the relationship between investment and GDP in various industries was studied, a regression model was constructed, and the return on investment was quantified, providing a quantitative basis for investment decisions. Finally, based on the regression model, the optimization algorithm is used to allocate the total government investment to maximize the total GDP, and visually demonstrate the contribution and rationality of each industry. Through data analysis and optimization modeling, this study provides a scientific basis for government investment decisions, aiming to improve resource allocation efficiency, promote economic growth, promote employment and sustainable development in the process of industrial structure adjustment and high-quality development.

1. Ravallion M .An Exploration of the International Comparison Program's New Global Economic Landscape[J].World Development, 2014, 105:201-216. DOI:10.1016/j.worlddev. 2017.12.035.
2. Korpi T , Mertens A .Training and industrial restructuring: Structural change and labour mobility in West Germany and Sweden[J].International Journal of Manpower, 2004. DOI:10.1108/01437720410525018.
3. Traxler F .Bargaining (De)centralization, Macroeconomic Performance and Control over the Employment Relationship[J].British Journal of Industrial Relations, 2003, 41(1):1–27. DOI:10.1111/1467-8543.00259.
4. Chen C , Yao S , Hu P ,et al.Optimal government investment and public debt in an economic growth model[J].China Economic Review, 2016, 45.DOI:10.1016 /j .chieco. 2016.08.005.
5. Ramos C , Augusto J C , Shapiro D .Ambient Intelligence—the Next Step for Artificial Intelligence[J].IEEE Intelligent Systems, 2008, 23(2):15-18.DOI:10.1109/MIS.2008.19.
6. Seonwoo M , Byunghan L , Sungroh Y .Deep Learning in Bioinformatics[J].Briefings in Bioinformatics, 2017(5):851.DOI:10.1093/bib/bbw068.
7. Aremu O , Fabry A , Meemken E M .Farm size and the quality and quantity of jobs—Insights from Nigeria[J].Food Policy, 2024, 128.DOI:10.1016/j.foodpol.2024.102731.
8. Yiran Chen, Yidong Li, George Lodorfos & Junjie Wu. (2023) Impact of Structural Distortions on Resource Allocation in China: Evidence from an Innovative Empirical Model. The Chinese Economy 56:5, pages 366-383.
9. Pan, H., Liu, Y. & Gao, H. Impact of agricultural industrial structure adjustment on energy conservation and income growth in Western China: a statistical study. Ann Oper Res 228, 23–33 (2015).
10. Y. He and G. Zhang, "Policies-oriented research of the first industrial adjustment based on the adjustment of the regional industrial structure," 2013 6th International Conference on Information Management, Innovation Management and Industrial Engineering, Xi'an, China, 2013, pp. 461-463
11. Han, Jun and Jiang, Shicheng. "Does the Adjustment of Industrial Structure Restrain the Income Gap between Urban and Rural Areas" Economics, vol. 18, no. 1, 2024, pp. 20220112.
12. Duan, Ding Qiang, and Qiu Hong Hu. "Industrial Structure Adjustment, Technology Innovation and Energy Efficiency Improvement - An Empirical Analysis Based on the Statistical Data of Hubei Province." Advanced Materials Research, vol. 734–737, Trans Tech Publications, Ltd., 16 Aug. 2013, pp. 1688–1692.
13. Yuan, F.; Huang, X.; Zheng, L.; Wang, L.; Wang, Y.; Yan, X.; Gu, S.; Peng, Y. The Evolution and Optimization Strategies of a PBFT Consensus Algorithm for Consortium Blockchains. Information 2025, 16, 268. doi: 10.3390/info16040268
14. Jiang Tianhe , Han Jun "The effect of industrial structure adjustment and economic development quality on transitional China's urban–rural income inequity",JOURNAL=Frontiers in Environmental Science,vol.11,2023.1084605.