Investment output decision problem solved based on multi-dynamic combination algorithm

Zhenyu Hu, Pengbin Wang*

Jiangsu Normal University KeWen College, JiangSu, China

ABSTRACT

This article examined the development trends of various industries amid economic growth, analyzing the correlations between sectors, and constructed mathematical models to promote the upgrading of industries in China. The analysis focused on 9 core industry indicators and builds a Multi-Factor Industry Interaction Correlation Model (MFIICM). The Shapiro-Wilk test was first employed to examine the data distribution types for 2017 and 2019, calculating the significance level (P-value). It was found that P-value<0. 05 for all cases, indicating a normal distribution. Subsequently, the Spearman coefficients were calculated, and a heatmap was plotted. The correlation coefficient between manufacturing and construction industries was the highest at 0.97, while other industries showed trends of mutual promotion. Causal analysis was then introduced for cross-validation, with an average bar error rate of 0.04%. A topological relationship network was constructed, revealing that industries such as Agriculture, Accommodation, and Retail occupy central positions and exhibit significant systemic influence. To study the quantitative relationship between investment and the GDP of various industries, an Input-Output Economic Investment Efficiency Model (IEIEM) based on Leontief's theory was constructed, with the goal of maximizing GDP growth. Using factor analysis, 17 indicators were extracted as evaluation criteria to calculate the output value of each industry, among which other service industries had the highest output value of 379, 099.9. The deviation value Δ_{ratio} was then introduced for model evaluation, yielding a deviation of Δ_{ratio} =0.38 between the actual and theoretical values, thereby verifying the model's validity. Based on this, an analysis without restrictions by industry was conducted. Bi-Objective Programming Model (BOPM) was constructed, adding the objective of minimizing investment ratio changes and introducing the investment return to parameter β_i. The GA algorithm was used to BOPM, yielding β_{i1}=1.26, β_{i2}=1.09. The iterative update curves for the bi-objective model were plotted, showing consistent trends and high synergy. Specific allocation plans are detailed in the main text.

Keywords: Shapiro-Wilk test; MFIICM; IEIEM; BOPM; GA algorithm

1 INTRODUCTION

China's industrial structure is experiencing a gradual upward development trend. With the advancement of artificial intelligence and the progress of industrialization, various industries are continuously iterating and upgrading, driving China's economy. These industries cover a wide range of sectors, from basic resource development to high-end services, including agriculture, forestry, animal husbandry, fisheries, industry, construction, wholesale and retail, transportation, warehousing and postal services, accommodation and catering, finance, real estate, and more. Together, these industries form the diversified and integrated foundation of China's economy, reflecting the overall and balanced development of the national economy [1].

The relationships between industries are intricate, encompassing both the potential for mutual promotion and the risk of mutual restriction. Policymakers must identify the issues within these relationships, analyze their causes, and strive to achieve economic balance and sustainable development. At the same time, they should leverage the government's proactive role to increase employment rates and improve people's living standards, ensuring continued economic growth and optimizing the future industrial structure of the country.

2 RELEATED WORK AND ASSUMPTION

The analysis examines the interrelationships between industries and their impact on economic development, identifying key industries like agriculture, accommodation, and catering, and the strong synergy between construction and industry. This supports understanding the industry linkage mechanism.

Based on the input-output model, this study explores the quantitative impact of investment in different industries on GDP. Using Leontief's input-output theory and factor analysis, it quantifies the driving effects of different investment configurations on GDP, highlighting that while other service industries have the largest output, traditional industries play a crucial role in economic stability.

The study uses a bi-objective programming model and genetic algorithm to optimize government investment allocation, showing that investments in manufacturing, services, and information technology significantly drive GDP growth. This improves the scientific allocation of investments and the sustainability of economic growth, providing data and theoretical support for resource optimization and policy-making.

Therefore, this paper makes the following assumptions.

(1) The industries considered in the model are independent and do not have overlapping investment requirements.

(2) The investment efficiency for each industry remains constant over the analysis period.

(3) External factors such as market demand, government policies, and global economic conditions are stable during the investment period.

3 MODEL ESTABLISHMENT AND SOLUTION

3.1 Data sources

Our main indicators and data sources are shown in Table 1

Table 1: Sources and Format Types of Various Data

Website	Data format
https://www.stats.gov.cn	CSV

https://www.cssn.cn	CSV
https://www.cnenergynews.cn	CSV
http://www.sasac.gov.cn	CSV
https://unstats.un.org	CSV

3.2 Multivariate Factors Industry Interaction Correlation Model

3.2.1 Data Distribution Test

Given that this study examines nine types of indicators, including data from eight sectors and one additional variable, with a limited sample size, the Shapiro-Wilk test is adopted to evaluate the data distribution [2].

(1) Definition of Test Statistic W

The test statistic W is expressed as:

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{1}$$

Where: $x_{(i)}$: The i-th value of the sample data arranged in ascending order. a_i: Weight coefficients determined by the theory of normal distribution, related to the sample size n (obtained via table lookup or computation). $\bar{x}$: Sample mean, calculated as:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n=9} x_i \tag{2}$$

Where: n: Sample size, which in this paper is 9 (corresponding to the 9 indicators).

(2) Calculation of Weight Coefficients a_i

The weight coefficients a_i are calculated based on the expected values and covariance matrix of order statistics under the normal distribution [3].

$$\mathrm{a} = \frac{\mathrm{m}^{\mathsf{T}}\mathrm{V}^{-1}}{\sqrt{\mathrm{m}^{\mathsf{T}}\mathrm{V}^{-1}\mathrm{m}}} \tag{3}$$

Where m: The expected values (mean vector) of order statistics under the normal distribution. V: The covariance matrix of order statistics under the normal distribution. The values of a_i vary with sample size n and are computed using Matlab.

(3) Hypothesis Testing

Null Hypothesis (H_0) : The data follow a normal distribution.

Alternative Hypothesis (H_1): The data do not follow a normal distribution.

(4) Standardized Statistic Transformation

The test statistic W is transformed into a standardized statistic z using an empirical formula:

$$z = f(W, n) \tag{4}$$

$f(W, n)$represents a complex relationship derived through data enhancement techniques.

(5) Significance Calculation

Under the standardized distribution, the p-value is calculated using the cumulative distribution function (CDF) of the standard normal distribution:

$$p = \Phi(z) \tag{5}$$

Where $\Phi(z)$ is the CDF of the standard normal distribution:

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} dt \tag{6}$$

To determine the distribution type, the significance results are as follows:

Table 2: Normal Distribution Test Results

Indicator	Significance (p-value)	Indicator	Significance (p-value)
Agriculture, Forestry, Animal Husbandry, and Fishery	0. 000017	Transportation, Warehousing, and Postal Services	0. 000035
Manufacturing Industry	0. 000005	Accommodation and Catering Industry	0. 000044
Construction Industry	0. 000043	Financial Industry	0. 000021
Wholesale and Retail Industry	0. 000004	Real Estate Industry	0. 000028
Others Industry	0. 000004		

Based on the above calculations, the significance values were all found to be less than 0. 05, indicating that the data follows a normal distribution. Therefore, this study introduced the Spearman correlation coefficient for further analysis.

3.2.2 Spearman Rank Correlation Coefficient

The Spearman rank correlation coefficient is a method used to study the correlation within ranked data. It measures the strength and direction of the monotonic relationship between two variables. The calculation formula is as follows [4]:

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \tag{7}$$

Where ρ: Spearman rank correlation coefficient; d_i: The difference between the ranks of each pair of observations; n: The number of observations

The correlation graph based on the above calculation formula is shown below.

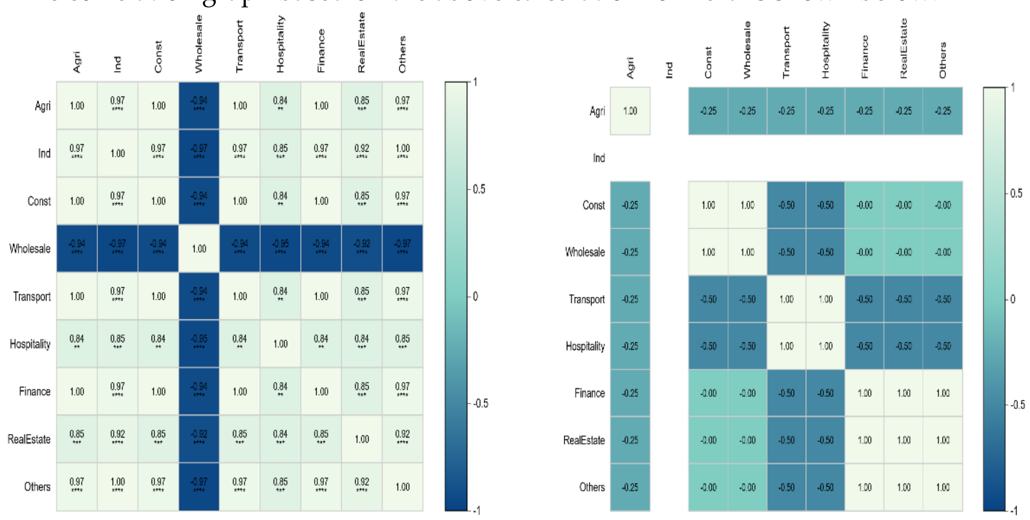

Fig. 1: Correlation Analysis Between Industries in 2017 and 2020

The left chart represents data from 2017. The correlation between Ind (Industry) and Const (Construction) is 0. 97, indicating a strong linkage between the two due to the high demand for construction by the industrial sector. On the other hand, the correlation between Wholesale and Finance is -0.57, showing a strong negative relationship, reflecting competition between these two sectors in terms of resource allocation or economic objectives. Additionally, the correlation between Transport and Wholesale is 0.84, highlighting the close collaboration between logistics and wholesale/retail in the supply chain.

The right chart represents data from 2020, revealing that Ind (Industry) and Const (Construction) exhibit high coordination within the group. However, inter-group correlations generally show negative relationships, such as a correlation of -0.25 between Ind (Industry) and Finance, and -0. 50 between Transport and Real Estate.

3.2.3 Correlation Analysis Based on Causal Cross-Improvement

Based on the above analysis, it is observed that industries mutually promote each other, but there is no specific indication of how one indicator affects another. Therefore, causal analysis is adopted to conduct an in-depth discussion. The foundation of causal analysis originates from counterfactual theory and Structural Causal Models (SCM), which reveal causal relationships between variables through methods such as intervention, hypothesis, and inference.

Causal relationships indicate that changes in indicator X lead to changes in another indicator Y, whereas correlations only suggest a statistical association between X and Y. The core question is: under the given intervention condition $do(X = x)$, how does the outcome variable Y change?

(1) The causal effect is defined as follows:

$$E[Y \mid do(X = x_1)] - E[Y \mid do(X = x_0)] \tag{8}$$

Where: $do(X = x)$: Represents an external intervention on X, rather than natural observation. $E[Y \mid do(X = x)]$: Represents the expected value of Y when X is forcibly set to x. Causal effects differ from conditional expectations; $E[Y \mid X = x]$, which represent the expected value of Y when X isobserved to equal x and include the influence of confounding factors.

(2) Connection Method of Causal Topology Graph

A causal graph $G = (V, E)$ is a Directed Acyclic Graph (DAG). Where: $V = \{X_1, X_2, \ldots, X_n\}$: Represents the set of variables (nodes in the graph). $E \subseteq V \times V$: Represents the set of causal relationships (directed edges), where $(X_i \to X_j) \in E$ indicates that X_i is a direct cause of X_j. The Structural Causal Model (SCM) of the graph describes the generation mechanism of each node X_i:

$$X_i = f_i(PA(X_i), U_i),\ i = 1, 2, \ldots, n \tag{9}$$

Where: $PA(X_i)$: The set of direct parent nodes of X_i (i. e., $\{X_j : X_j \to X_i\}$). U_i: The unobserved variables (noise), where $U_1, U_2, \ldots, U_n$ are mutually independent. f_i: The causal function of node X_i, which describes how X_i is generated from its parent nodes and noise variables.

(3) Joint Distribution

Based on the topology graph mentioned above, a causal decomposition of the joint distribution is proposed. According to the structure of the causal graph G, the joint distribution can be decomposed as follows:

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^{n} P\big(X_i \mid PA(X_i)\big) \tag{10}$$

This indicates that the conditional probability of any variable X_i depends only on its parent nodes $PA(X_i)$.

The joint distribution can be decomposed as:

$$P(X_1, X_2, X_3, X_4) = P(X_1)P(X_2 \mid X_1)P(X_3 \mid X_1)P(X_4 \mid X_2, X_3) \tag{11}$$

(4) Selection of Criteria

a. Backdoor Criterion

The backdoor criterion is used to adjust for the effects of confounding variables. If there are confounding variables Z along the causal path $X \rightarrow Y$, we can eliminate the confounding effect by controlling Z:

$$P\big(Y \mid do(X = x)\big) = \sum_{z} P(Y \mid X = x, Z = z)P(Z = z) \tag{12}$$

Conditions: Z blocks all backdoor paths connecting X and Y. Z is not a descendant of X.

b. Frontdoor Criterion

The frontdoor criterion is applicable in cases involving mediator variables. For instance, if $X \rightarrow M \rightarrow Y$, we can infer the causal effect of X on Y through the mediator variable M :

$$P\big(Y \mid do(X = x)\big) = \sum_{m} P(Y \mid M = m)P\big(M \mid do(X = x)\big) \tag{13}$$

(5) Calculation of Causal Effect

The causal effect is usually expressed as:

$$T = E[Y \mid do(X = x_1)] - E[Y \mid do(X = x_0)] \tag{14}$$

Under the backdoor criterion:

$$E[Y \mid do(X = x)] = \sum_{z} E[Y \mid X = x, Z = z]P(Z = z) \tag{15}$$

Under the frontdoor criterion:

$$E[Y \mid do(X = x)] = \sum_{m} E[Y \mid M = m]P\big(M \mid do(X = x)\big) \tag{16}$$

3.2.4 Causal Analysis Results

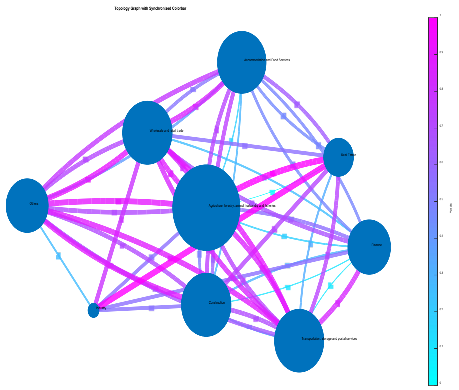

Fig. 2: Topology Map & Bar Error Chart of Various Industries

The topology map shows the causal relationships and association strengths among industries, with node size and connection weight reflecting industry importance and interrelations. Agriculture, forestry, animal husbandry, and fisheries are core industries, strongly linked to accommodation and food services (0.92), retail trade (0.87), industry (0.78), and transportation (0.85). In contrast, industry and real estate have a weaker connection (0.32), while retail trade connects moderately with accommodation services (0.81). Overall, agriculture, accommodation, and retail hold central positions with significant systemic influence. The bar error chart highlights the average correlation and variability across industries, showcasing differences and stability in their relationships.

3.3 Input-Output Economic Investment Efficiency Model

3.3.1 Model Construction Based on Leontief Theory

To analyze the quantitative relationship between government investment and the gross domestic product (GDP) of various industries and to formulate a scientific investment allocation plan, this study employs the Leontief input-output model. Widely used in analyzing the complexity of economic systems, this model precisely characterizes the interdependencies among industries and reveals the investment multiplier effect [5]. It uses a technical coefficient matrix to describe the input-output relationships across industries, providing a means to measure the direct and indirect dependency effects among different sectors.

(1) The constructed equation is expressed as:

$$X = AX + Y \tag{17}$$

Where X: An $n \times 1$ total output vector, representing the total output level of n industries in the economic system. A: An $n \times n$ input-output technical coefficient matrix, where A_{ij} denotes the proportion of intermediate input required from industry i for each unit of total output in industry j. Y: An $n \times 1$ final demand vector, representing the final demand for each industry, including consumption, investment, exports, and other demands.

Based on the above equation, the total output of the economic system can be expressed as:

$$X_t = \left(I - A_0 \cdot (I + \Gamma_t)\right)^{-1} \cdot (Y_t + \Delta Y_t) \tag{18}$$

Where I: An $n \times n$ identity matrix. $(I-A)^{-1}$: The Leontief inverse matrix, which captures the direct and indirect effects of final demand on total output.

(2) Incorporating Investment-Induced Demand

Government investment in various industries drives the growth of final demand through industrial chain effects. Let the investment vector be F, and the relationship between investment and final demand can be expressed as:

$$\Delta Y = BF \tag{19}$$

Where F: An $n \times 1$ investment vector, representing the scale of government investment in each industry. B: An $n \times n$ investment-to-demand conversion matrix, describing the impact of unit investment on the final demand of each industry.

(3) Calculating the Total Output Change Induced by Investment

The change in total output induced by investment can be expressed as:

$$\Delta X = (I-A)^{-1}\Delta Y \tag{20}$$

Combining the investment-demand relationship, we have:

$$\Delta X = (I-A)^{-1}BF \tag{21}$$

Where: ΔX: The change in total output. $(I-A)^{-1}$: The Leontief inverse matrix, capturing the direct and indirect effects of input output relationships within the industrial chain.

(4) Mechanism of Investment on GDP Growth

To quantify the contribution of investment to GDP, a value-added coefficient vector v is introduced, where v_i represents the contribution rate of industry i to GDP per unit of total output. The change in GDP can be expressed as:

$$\Delta GDP = v^{\top}\Delta X \tag{22}$$

Expanding further:

$$\Delta GDP = v^{\top}(I-A)^{-1}BF \tag{23}$$

This formula shows that the change in GDP is influenced by the following three c omponents: Leontief inverse matrix $(I-A)^{-1}$: Measures the direct and indirect interconn ections among industries. Investment-to-demand conversion matrix B: Reflects the impa ct of investment on final demand. vector v: Represents the contribution of each industr y's output to GDP.

(5) Optimization Objective

To maximize the GDP increment, the objective function for investment optimization is:

$$\max_{F} \Delta GDP = v^{\top}(I-A)^{-1}BF \tag{24}$$

(6) Constraints

a. Total Investment Constraints:

$$L_i \le F_i \le U_i,\ F_i \ge \beta_{ij}F_j,\ \forall i,j \tag{25}$$

Where: $L_i$and U_i: Represent the lower and upper bounds of investment in industry $i. F_i \ge \beta_{ij}F_j$: Represents proportional constraints between the investments in industries i j.

b. Non-Negativity Constraints:

$$F_i = \sum_{t=1}^{T} F_{i,t}, \ \sum_{i=1}^{n} F_{i,t} \leq I_t, \ \forall t \tag{26}$$

This indicates that the investment amount for each industry must be non-negative.

3.3.2 Input-Output Indicator Analysis

This study focuses on GDP, with agricultural, forestry, animal husbandry, and fishery products as foundational industries supplying raw materials for manufacturing sectors like food, textiles, and leather products [6]. Extractive and chemical products are critical for machinery, transportation, and construction, while non-metallic and metal processing support manufacturing and construction.

These industries drive upstream and downstream linkages, boosting electricity, heat, gas, and water supply, while services like wholesale, retail, and logistics ensure production and distribution. Digitalization through IT services enhances efficiency, finance and real estate provide capital, and R&D drives innovation. Together, these sectors contribute to GDP through consumption, investment, and exports, forming a complete economic cycle. The formula is as follows:

$$GDP = \sum_{i=1}^{n} VA_i + C + I + G + NX \tag{27}$$

Where VA_i: The value-added of the i-th industry (total output minus intermediate input). C: Final consumption (including household and government consumption). I: Gross capital formation (fixed asset investment). G: Government expenditure. NX: Net exports (exports minus imports).

Integrating industry indicators into the formula:

$$GDP = VA_a + VA_b + VA_c + \cdots + VA_Q + C + I + G + NX \tag{28}$$

Through input-output analysis, this can be further detailed as:

$$VA_i = S_i - \sum_{j=1}^{n} Z_{ij} \tag{29}$$

Where S_i is the total output of industry i, and Z_{ij} represents the intermediate input from industry j to i.

Based on this, Spearman's rank correlation coefficient is calculated, and a comparison heatmap of input and output before and after the analysis is drawn using MATLAB's built-in functions, as shown in Figure 6.

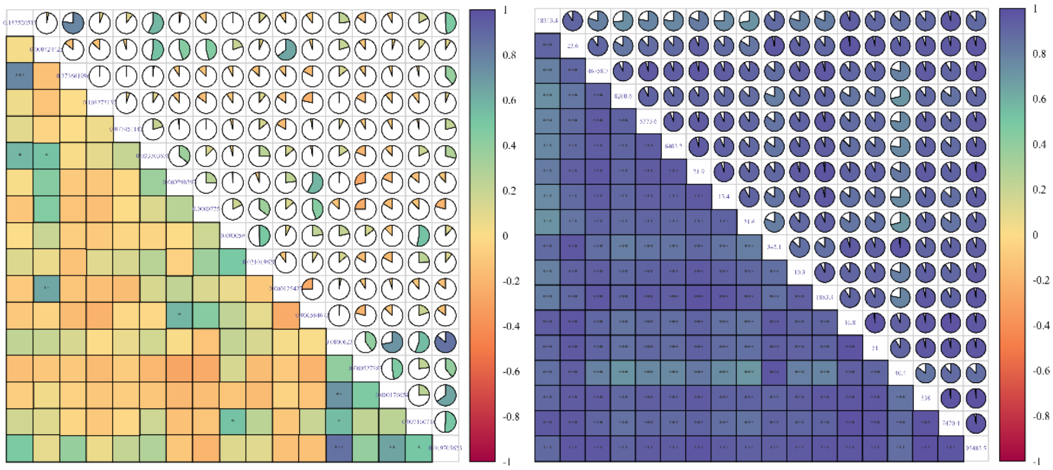

Fig. 3: Heatmap Analysis of the Indicator Matrix Before and After Projection

The heatmap comparison before and after the projection reveals significant changes in the correlations among 17 system indicators. Before projection, correlations were dispersed with weak overall dependencies, while after projection, stronger positive **(e. g. , indicator 8 and 11: 0.3 → 0.9)** and negative correlations **(e. g. , indicator 2 and 16: 0.02 → -0.5)** emerged, highlighting enhanced system synergy and opposition. Previously unrelated indicators **(e. g. , 11 and 15: 0.05 → 0.7)** became significantly correlated, while some dependencies weakened **(e. g. , indicator 4 and 7: -0.4 → -0.05)**, reflecting reduced reliance. These changes demonstrate the projection's deep impact on system relationships, strengthening synergies and adjusting internal balances, offering key insights for future optimization.

3.3.3 Model Solution Results

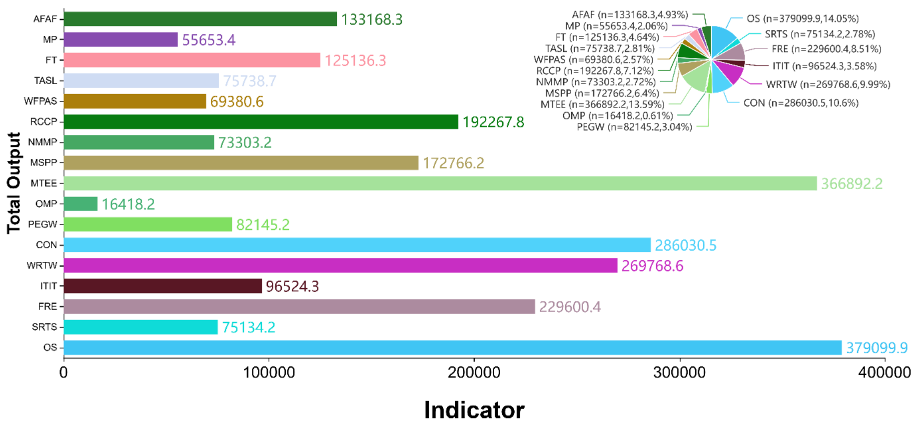

Fig. 4: Visualization of Total Investment by Industry

This table presents the total output of various industries, covering the output value of major sectors in the Chinese economy. Among them, the total output of machinery, transportation equipment, electronics, and other equipment is the highest, reaching 366, 892. 2 billion yuan, reflecting the significant position of this sector. The construction industry follows closely with 286, 030. 5 billion yuan, while other service industries also show substantial output at 379, 099. 9 billion yuan. The output of industries such as finance and real estate, wholesale and retail, transportation and warehousing, as well as refining and chemical products, is also notable. Additionally, traditional industries like agriculture, forestry, animal husbandry, and fisheries (133, 168. 3 billion yuan) and mining products (55, 653. 4 billion yuan) still hold a certain share in the economy. These figures reflect the diversity of China's industrial structure and the contributions of different sectors to the economy.

3.3.4 Model Evaluation

To evaluate the changes in the heatmap before and after adjustments, the differences in the correlation matrix can be quantified using multiple indicators to measure the dynamic adjustments in the relationships between the system's variables. By constructing a difference matrix between the pre- and post-correlation matrices, the overall trend and specific degree of changes in the interrelationships among the variables can be observed. This study uses variation intensity deviation values for analysis [7].

$$\Delta_{\text{ratio}} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\left|R_{\text{after},ij}-R_{\text{before},ij}\right|}{\sum_{i=1}^{n}\sum_{j=1}^{n}\left|R_{\text{before},ij}\right|} \tag{30}$$

Where $R_{\text{before},ij}$ and $R_{\text{after},ij}$: Represent the correlation between the i-th and j-th indicators before and after projection, respectively. Δ_{ratio} Deviation value.

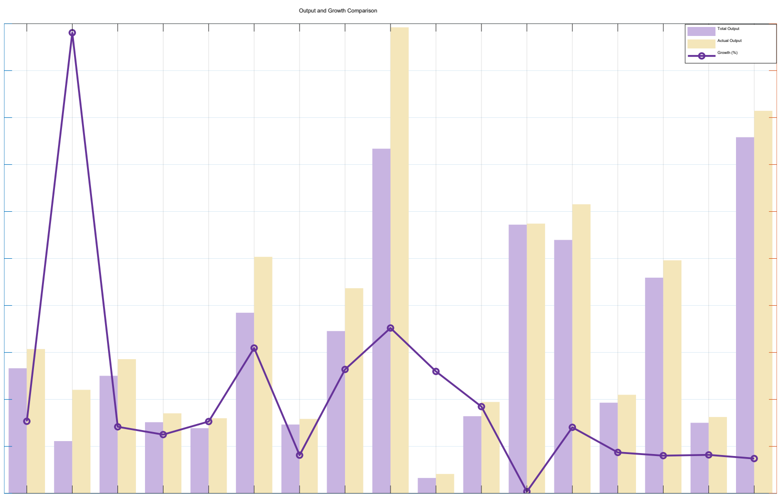

Fig. 5: Comparison of Actual Output and Original Output

This study evaluated the predictive performance of the model for total output and actual total output, and the results indicate significant differences in the model's performance across various industries. Actual total output increased by 20.67% compared to the predicted values, with notable growth in industries such as extractive products (98.07% increase), refining and chemical products (30.95% increase), and machinery manufacturing (35.2% increase), suggesting that the model failed to fully capture external driving factors in high-growth sectors. In contrast, the deviation between predicted and actual values in the construction industry was only 0.38%, demonstrating that the model has high predictive accuracy in relatively stable industries.

3.4 Bi-objective Programming Model

3.4.1 Model Development

On the one hand, it is necessary to maximize the Gross Domestic Product (GDP), and on the other hand, to ensure stable economic development. If the optimization focuses solely on maximizing GDP, the solution would allocate all investments to the industries with the highest current profitability based on the current Leontief inverse matrix. However, there are mutual promotion and constraints among industries. Ignoring balanced development among industries during the investment process may lead to distortions in future industry growth. Therefore, this paper establishes a bi-objective optimization model, simultaneously considering the maximization of GDP and the minimization of changes in investment proportions [8].

(1) Objective 1: Maximize GDP growth

In the input-output model, it is assumed that changes in GDP resulting from government investments in different industries are driven by the investment multipliers of each industry. Let the government's investments in various industries be $X_1, X_2, \ldots, X_n$ (where n represents the total number of industries) [9].

Based on the Leontief inverse matrix L, the change in GDP can be expressed as:

$$\Delta Y = A_Y \cdot \left[(L \cdot X)^{\beta_Y} \cdot (1 - \delta_Y \cdot D_Y) \cdot \exp\left(-\frac{\Delta R}{R_{\max}}\right)\right] \tag{31}$$

Where ΔY: Vector of GDP changes for each industry. A_Y: Production efficiency coefficient for GDP growth. L: Leontief inverse matrix, describing the input-output relationships among industries. X: Investment vector for each industry. β_Y: Investment multiplier parameter. δ_Y: GDP suppression factor. $R_{\max}$: Maximum threshold for resource utilization.

The constraints are:

$$\sum_{i=1}^{n} X_i = 10^{12},\ X_i \geq 0,\ \forall i = 1, 2, \ldots, n \tag{32}$$

(2) Objective 2: Minimize the variation in investment proportions

To ensure stable economic development and avoid excessive and unbalanced investments in certain industries, the objective is to minimize changes in investment proportions.

Specifically, this can be measured by the fluctuation of the investment proportions for each industry. Assume the initial investment proportions are $\alpha_1, \alpha_2, \ldots, \alpha_n$.

$$\alpha_i^{\mathrm{I}} = \frac{X_i}{\sum_{i=1}^{n} X_i} \tag{34}$$

The variation in investment proportions for each industry can be defined as the difference between the actual investment proportion β_i and the initial investment proportion α_i. The objective is to minimize this difference. The variation in investment proportions can be measured using the mean squared error (MSE), represented by the following formula:

$$\text{Minimize} \sum_{i=1}^{n} \left[\frac{X_i}{\sum_{i=1}^{n} X_i} \cdot \left(1 - \frac{\Delta R_i}{R_{\max}} \cdot \exp\left(-\kappa_i \cdot I_i^{\text{tech}}\right)\right) - \alpha_i\right]^2 \tag{35}$$

Introduction of the Investment Return Parameter

$$\beta_i = \frac{X_i}{\sum_{i=1}^{n} X_i} \tag{36}$$

Through this bi-objective optimization model, the government can find an optimal investment allocation plan by balancing the maximization of GDP and the minimization of changes in investment proportions. This approach will help achieve long-term and stable economic growth, avoiding economic imbalances caused by over-investment in certain industries, while simultaneously promoting an overall increase in GDP.

3.4.2 Solving Based on the Improved Genetic Algorithm (NSGA-II)

(1) Principle of the NSGA-II Algorithm

Genetic algorithms, as a global optimization method, possess the ability to quickly find global optimal solutions in complex search spaces. They consist of three fundamental operations: selection, crossover, and mutation. In the selection operation, the fitness of everyone is evaluated based on a fitness function, and individuals with higher fitness are selected for the next generation. The crossover operation simulates the process of genetic recombination, combining the genes of two individuals to form a new individual. The mutation operation mimics the random process of genetic mutation, randomly altering an individual's genes. Genetic algorithms are characterized by adaptability, parallelism, and robustness, and are widely used in fields such as machine learning, artificial intelligence, and optimization problems.

(2) Establishing a Multi-Matrix Chromosome Encoding and Decoding Scheme

Decision variables involve the integration of multiple indicators, and these high-dimensional variables are challenging to represent using traditional encoding methods [10]. Therefore, this paper introduces multi-type matrices from linear algebra, where different types of indicators are naturally mapped to complex factor variables in the form of 3×3 matrices. Simultaneously, cases where certain data overlap are considered, and partially overlapping matrices are visualized. As a result, the following encoding method is proposed:

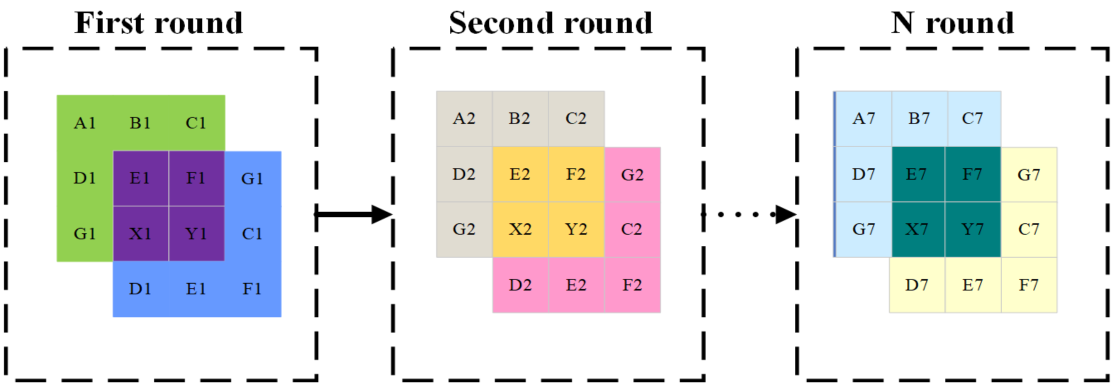

Fig. 6: Genetic Encoding Scheme

(3) Customized Chromosome Crossover Strategy

In the solution designed in this paper, the original data is divided into paternal and maternal parts in a 1: 1 ratio. Through crossover in the first generation, genetic information is passed to offspring 1 and offspring 2. Chromosome crossover is performed for different years to obtain the optimal solution. The crossover method is illustrated in the figure below.

Fig. 7: Chromosome Crossover and Mutation

3.4.3 Model Parameter Settings

The parameters for the genetic algorithm designed in this study are detailed in the following table.

Table 3: Genetic Algorithm Parameter Settings

Detailed Parameter Settings	Simulated Data Division
Number of Iterations	60
Crossover Rate	0. 7
Mutation Rate	0. 3
Population Size	10000

3.4.4 Model Solution Results

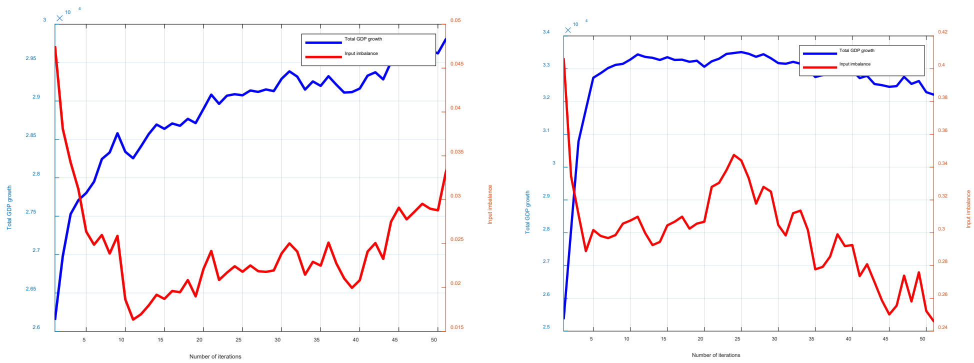

Fig. 8: Convergence Curves of Returns and Proportion Deviation for Scenario 1 -2

Left Chart: Returns show a gradual increase and stabilization. In the early stage (26, 147. 71–28, 999.03), returns grew by 10. 9%, and the proportion deviation dropped by 45. 2% (0. 0475 to 0.0260). In the mid-stage (29, 003.08–29, 999.48), growth slowed to 3.4%, and the proportion deviation dropped to 0.0224 with slight fluctuations. In the late stage (above 30, 004. 83), returns stabilized with less than 1% growth, and the proportion deviation remained between 0.035 and 0.037. Overall, returns grew 15.7% (26, 147. 71–30, 258. 71), with a significant drop in proportion deviation, demonstrating good convergence.

Right Chart: Returns gradually increased and stabilized. In the early stage (25, 360. 68–32, 865.70), returns grew by 29.6%, and the proportion deviation dropped by 26.7% (0.4065 to 0. 2980). In the mid-stage (32, 865.70–31, 621.56), growth fluctuated slightly, and the proportion deviation oscillated between 0. 2967 and 0. 2079. In the late stage (31,621. 56–32, 011.35), returns stabilized with less than 2% growth, and the proportion deviation dropped to around 0. 1910. Overall, returns grew 26.2% (25, 360.68–32, 011.35), with a significant drop in proportion deviation, showing strong convergence.

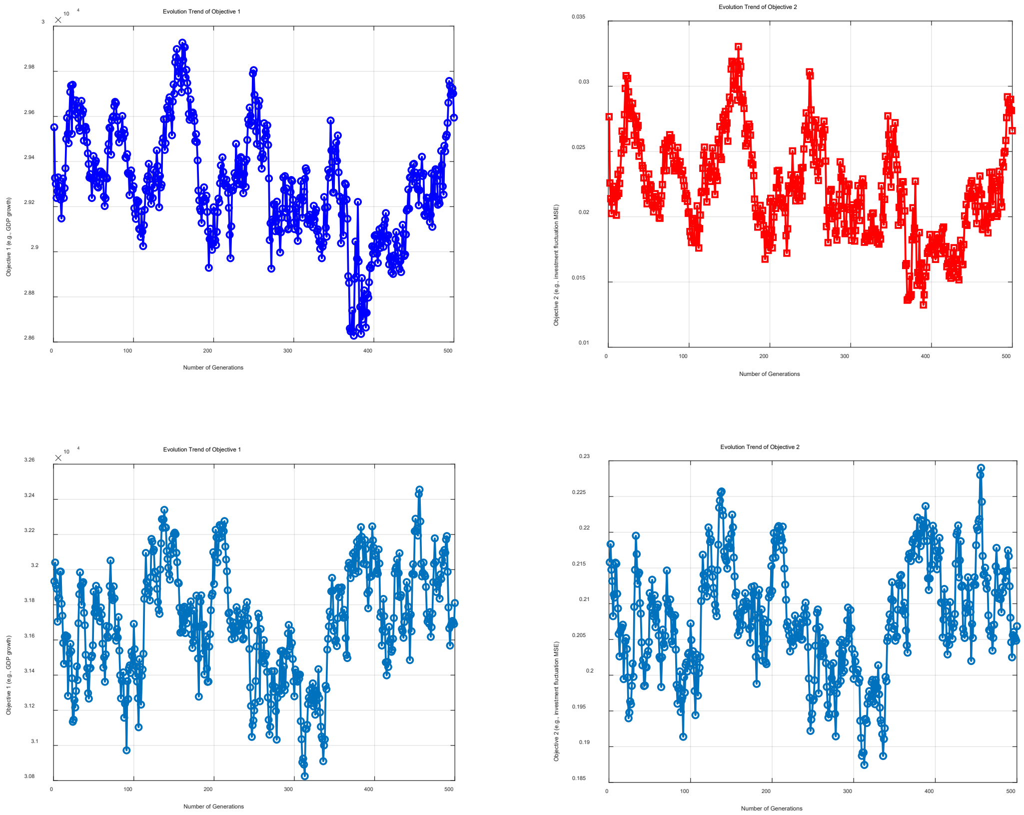

Fig. 9: Iterative Updates of Objective 1、2 Populations Under Two Scenarios

For Scenario 1, Objective 1 (GDP increment) and Objective 2 (mean squared error of investment proportion changes) exhibit characteristics of synergy and trade-off during the multi-objective optimization process. In the early stage, Objective 1 rises rapidly to nearly 3. 2×10^4, while Objective 2 remains between 0.035 and 0.04, indicating a focus on increasing GDP increment during this phase. In the mid-stage, Objective 2 continues to improve, dropping from 0.035 to 0.02, with significant fluctuations between the two objectives, reflecting the trade-off relationship. In the late stage, Objective 1 stabilizes, and Objective 2 converges, with fluctuations narrowing to 0.02–0.025, showing that the optimization process achieves a synergistic balance, simultaneously improving GDP increment and investment proportion balance. Scenario 2 exhibits a similar situation, characterized by synergistic balance with identical trends, differing only in numerical values.

4 CONCLUSIONS

This study addresses the critical issue of optimizing investment decisions in industries to maximize GDP growth while maintaining economic stability. By applying a multi-dynamic combination of methods, we proposed a Multi-Factor Industry Interaction Correlation Model (MFIICM), which successfully analyzed inter-industry relationships, revealing critical insights into their mutual influences.

Further, we developed an Input-Output Economic Investment Efficiency Model (IEIEM) based on Leontief's theory to quantify the relationship between investments and GDP growth. This model enabled a detailed assessment of the economic impact of sectoral investments. To ensure balanced development, we introduced a Bi-objective Programming Model (BOPM), which simultaneously aimed to maximize GDP growth and minimize the variation in investment proportions across industries.

Through the application of a Genetic Algorithm (GA), specifically the NSGA-II method, the study provided optimized investment strategies that promote both high economic returns and long-term sector stability. The results demonstrate the robustness and applicability of the proposed models, offering valuable decision-making tools for policymakers and investors.

The key advantage of this study lies in its ability to balance the maximization of economic output with the strategic allocation of investments across industries. It provides an in-depth analysis of the model's flexibility and adaptability under changing economic conditions.

REFERENCES

[1] Wu, Q., Yang, Y., Liu, L., Hu, X., Bo, L., & Yang, J. (2024). Lithium battery health state estimation method based on GA-SA-BP neural network. *Power System Protection and Control, 19*, 74-84. doi:10.19783/j.cnki.pspc.240248.

[2] Qiu, X., Zhang, H., & Zhao, R. (2024). Warehouse location allocation optimization for rear warehouse based on PSO-GA algorithm. *Journal of Information Engineering University, 04*, 423-427.

[3] Liu, H., & Jian, H. (2024). Water resource allocation model and application based on co-evolutionary particle swarm optimization algorithm. *People's Yellow River, 11*, 74-79.

[4] Yang, J., Zhang, X., Zhou, H., Zhou, Y., & Wang, J. (2024). Optimal measurement point selection and inversion calculation for transmission line non-contact voltage measurement based on genetic particle swarm algorithm. *Science and Technology and Engineering, 19*, 8135-8141.

[5] Liu, Y., Zhang, S., Wang, X., & Bai, W. (2023). Design of PI controller for wind turbine variable speed and pitch system based on genetic particle swarm algorithm. *Science and Technology and Engineering, 31*, 13375-13386.

[6] Liang, Q., & Zhang, L. (2024). Yarn production process parameter inversion based on particle swarm genetic algorithm. *Cotton Textile Technology, 06*, 1-7.

[7] Li, Z., & He, Y. (2024). UAV takeoff and landing site selection in complex mountainous environments based on GA-SA combined algorithm. *Science and Technology and Engineering, 02*, 850-857.

[8] Liu, J., Wang, H., & Zhou, W. (2018). Low Earth orbit constellation sensor resource scheduling algorithm based on GA-SA. *System Engineering and Electronics Technology,*

11, 2476-2481.

[9] Li, S., Fang, T., & Ye, J. (2024). Suppression of low-frequency oscillation modes in DFIG grid-connected power systems using controllable load sensitivity. *Proceedings of the CSEE, 07*, 2641-2655. doi:10.13334/j.0258-8013.pcsee.222084.

[10] Song, Z., Ji, S., Hu, W., Ren, Z., & Wang, L. (2024). Amphibious robot 3D path planning research based on improved particle swarm algorithm. *Science and Technology and Innovation, 04*, 59-61+65. doi:10.15913/j.cnki.kjycx.2024.04.015.