A Study on the Correlation of China's Major Industries Development Based on Spearman Correlation Coefficient

Xin Wang[#, 1], Zhiyang Liu[#, 1], Peng Li[#, 1], Huaming Yu[*, #, 2]

[1]School of Information and Control Engineering, Liaoning Petrochemical University, Fushun, China, 113001

[2]College of Sciences, Liaoning Petrochemical University, Fushun, China, 113001

*Corresponding author: h.m.yu@163.com

#These authors contributed equally.

Abstract. China's national economic system is large and diversified. Various industries cooperate and have complex connections. There are both incentives and constraints. These challenges to policymaking and government investment are crucial to economic development. It is about employment and social stability. To analyze the interrelationship between the major industries in China. Gross domestic product (GDP) data covers the period from 1978 to 2023 by industry, and the proportion of GDP in total GDP by industry. Visually, the line chart and the stack chart are used to show the changing trend of the total GDP and the proportion of each industry. It is found that there is a positive correlation between various industries. There has been no growth in GDP in one sector and a decline in GDP in another. Various industries are interdependent and influence each other to promote sustained economic growth. In addition, to conduct correlation analysis on the development of major industries in China, the normality test and Pearson correlation analysis were used for the GDP data of various industries. The former ensures that the data conforms to the normal distribution hypothesis to ensure the effectiveness of subsequent analysis, and the latter calculates the correlation coefficient among various factors. Measure the closeness of the relationship. The result displays that all industries have a strong positive correlation with China's GDP, and the lowest correlation coefficient is 0.975. This fully demonstrates the interdependence of various industries and their joint contribution to sustained and steady economic growth.

Keywords: Normality Test, Pearson, Visual Analysis, Gross Domestic Product.

1. Introduction

In the field of economic research, an extensive collection of authoritative data covering multiple dimensions and a long period is an important prerequisite for research. Through cooperation with government authorities such as the National Bureau of Statistics and the National Development and Reform Commission, we obtain macroeconomic data regularly. At the same time, the joint industry association collects first-hand information about various industries, uses the power of professional data research institutions, and also uses big data capture, field research, and other means to collect comprehensive and accurate data. These data cover economic indicators such as industry size, profit rate, and market share, as well as innovation factors such as research and development investment intensity and the number of patent applications, spanning from the early days of reform and opening up to now.

After obtaining the data, the input-output table is used to analyze the input-output ratio between industries. The Granger causality test is used to explore the causality of industries, and complex network analysis tools are used to study the correlation structure between industries. In-depth analysis of the internal links between agriculture, industry, service, and other major industries, to explore their interaction in the process of economic development. For example, the development of the manufacturing industry drives the rise of the logistics service industry, and the prosperity of the service industry provides a broader market for the manufacturing industry. This paper analyzes the promoting effect of these functions on economic growth and the problems that restrict economic development such as resource misallocation caused by unreasonable industrial structure.

2. Normality test and correlation analysis

2.1. Visual analysis of the gross domestic product of various industries

The data of this study, which is mainly derived from the China Statistical Yearbook, is used to analyze the interrelationships among major industries in China. The data cover GDP data for each sector over the longer period from 1978 to 2023, as well as the sector's share of overall GDP.

Visualizing GDP data by industry from 1978 to 2023 is shown in Figure 1.

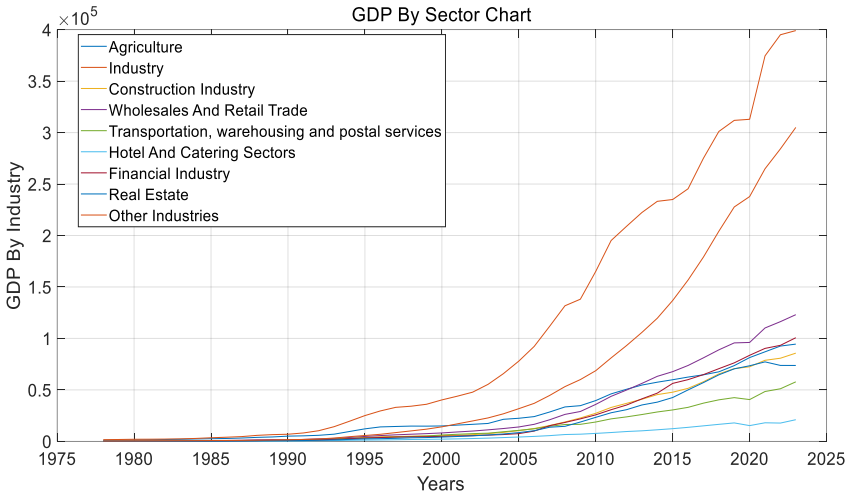

Figure 1. GDP by sector, 1978-2023

From 1978 to 2023, China's GDP of all industries showed a significant growth trend, reflecting the vigorous development of China's economy and the gradual optimization of industrial structure. Various industries are interdependent and influence each other to jointly promote sustained economic growth. The following select representative industries for in-depth analysis:

During 1978-2023, the GDP of industrial development jumped from 162.14 billion yuan in 1978 to 399.1031 billion yuan in 2023. Many factors contributed to this remarkable achievement. The implementation of the reform and opening up policy, such as the establishment of special economic zones and other measures[1], attracted a large number of foreign investment and advanced technology influx, injected a strong impetus to industrial development. The huge domestic market demand, from daily necessities to high-end equipment, provides a broad space for the development of the industry. Continuous technological innovation and industrial upgrading[2] have enabled China's industry to achieve breakthroughs in high-speed rail, 5G, and other fields, and move to the forefront of the world. At the same time, the perfect infrastructure construction reduces the logistics cost and ensures the stability of production. The rich and constantly improving quality of labor resources provide solid human support for industrial development and promote the leapfrog development of industry.

The GDP of the real estate industry was only 7.97 billion yuan in 1978 and reached 7.372.27 billion yuan in 2023. The acceleration of urbanization is an important driving force for its development. A large number of people have gathered in the city, and the demand for housing has increased sharply, providing a broad space for real estate development. The continuous expansion of urban scale and the continuous improvement of infrastructure have further promoted the prosperity of the real estate industry. As an important field of fixed asset investment, real estate has attracted a large amount of capital inflow. At the same time, the development of the real estate market is also guided and regulated by policies and plays an important role in promoting economic growth and improving residents' living conditions[3].

The wholesale and retail industry made a huge leap between 1978 and 2023, with a GDP of 24.24 billion yuan in 1978 and 12,307.24 billion yuan in 2023. As a key link between production and consumption, with the increasing abundance of industrial and agricultural products, it has built a huge sales network to transport goods to all parts of the country. The growth of residents' income and the change of consumption concepts have promoted the upgrading of the consumer market and prompted

the industry to constantly innovate its business model, such as the rise of e-commerce, breaking the limitations of time and space, and expanding the size of the market. The improvement of modern logistics and supply chain provides a solid guarantee for its development, and an efficient distribution system reduces costs and improves service quality. Logistics enterprises such as SF Express and Cainiao help the wholesale and retail industries flourish.

"Other industries" cover many subdivisions, such as science and technology services, education services, etc. The GDP started from 26.56 billion yuan in 1978, and grew to 30501.02 billion yuan in 2023, playing an increasingly important role in the national economic system. Take the education industry as an example. Since 1978, China's education has been developing and growing. The state pays more attention to and invests in education, which promotes the expansion of education scale and the improvement of education quality. The education industry has trained a large number of high-quality talents for other industries, provided professional and technical personnel and innovative talents for industry, and promoted industrial technology research and development and industrial upgrading; For the wholesale and retail industry, it has sent talents who know marketing and business management, helping the industry to innovate business models and improve service levels; It has trained professionals in architectural design, project management and other aspects for the real estate industry to ensure the high-quality construction of real estate projects. To intuitively understand the proportion of GDP of different industries in China from 1978 to 2023. Figure 2 is shown.

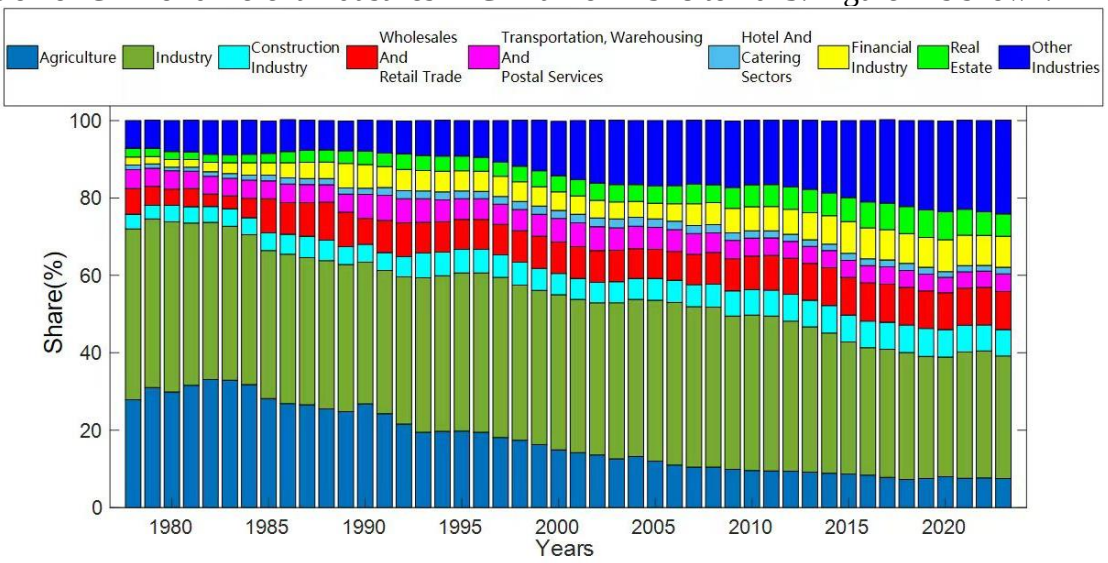

Figure 2. GDP by industry, 1978-2023

During the period 1978-2023, the share of GDP in China's GDP by industry has changed significantly. In 1978, agriculture, forestry, animal husbandry, and fishery accounted for 27.9%, industry accounted for 44.1%, construction accounted for 3.8%, wholesale and retail trade accounted for 6.6%, transportation, warehousing, and postal services accounted for 4.9%, accommodation, and catering industry accounted for 1.2%, financial industry accounted for 2.1%, real estate industry accounted for 2.2%[4], other industries accounted for 7.2%. By 2023, the proportion of agriculture, forestry, animal husbandry, and fishery will decrease to 7.5%, industry to 31.7%, construction to 6.8%, wholesale and retail to 9.8%, transportation, warehousing, and postal industry to 4.6%, accommodation and catering industry to 1.7%, financial industry to 8%, real estate to 5.8%[5], and other industries to 24.2%. These data clearly show the dynamic evolution of the position of various industries in the economic structure.

During this period, other sectors performed particularly well, climbing from 7.2 percent in 1978 to 24.2 percent in 2023, reflecting the growing strength of the many emerging sectors it covers and their continued contribution to economic growth. The financial sector's share has also increased significantly, from 2.1% in 1978 to 8% in 2023. With the development of the economy and the continuous improvement of the financial market, the financial industry is playing an increasingly important role in resource allocation and risk diversification[6]. In addition, the share of wholesale

and retail trade increased from 6.6% in 1978 to 9.8% in 2023, thanks to the continuous expansion of the consumer market and innovative development of business models.

In contrast, there are also some industries whose GDP share has changed relatively little. Transportation, warehousing, and postal services accounted for 4.9% in 1978 and 4.6% in 2023. Although there are slight fluctuations, the whole remains in a relatively stable range, which indicates that the industry is steadily moving forward with the overall development of the economy, and its basic position in the economic structure is relatively stable. In addition, the proportion of the construction industry has increased from 3.8% in 1978 to 6.8% in 2023, but compared with the changes of other industries, the range is not large, and it has been playing a relatively stable supporting role in economic construction[7].

From 1978 to 2023, the share of industry in GDP shows a complex change trajectory, but it still occupies an important position in the composition of China's GDP. In 1978, industry accounted for 44.1 percent and dominated the economic structure. In the 1980s, the proportion of industry as a whole showed a downward trend, and by 1984 it had dropped to 38.7%. From 1985 to 1997, the proportion of industry was relatively stable, basically maintaining between 38% and 41.4%, which benefited from the deepening of reform and opening up. The reform in the industrial field stimulates the vitality of enterprises, and the development of other industries has not yet had a major impact on the proportion of industry. In the 21st century, especially since 2013, China has paid more attention to economic quality and efficiency, accelerated economic restructuring, and the rapid rise of the service industry, while traditional industries are facing transformation pressure, some resources are transferred to emerging industries and service industries, and the proportion of industry continues to decline, accounting for 37.5% in 2013, to 31.7% in 2023.

2.2. Correlation analysis of major domestic industries

2.2.1 Data selection

To conduct a correlation analysis on the development of major industries in China, the GDP data of the National Bureau of Statistics of some industries are selected.

2.2.2 Normality test

Pearson correlation analysis is a parameter test method, and one of its important basic assumptions is that the two variables involved in the analysis should obey the normal distribution. Only when the data meet the distribution requirement, the correlation coefficient calculated based on this method and the corresponding statistical inference be effective and reliable.

The following table shows the normality test of the above-influencing factors by using Spsspro software:

Table.1. Normality test results

Variable name	Sample size	Skewness	Kurtosis	S-W test	K-S test
GDP (100 million yuan)	46	1.224	0.23	0.779(0.000***)	0.236(0.010***)
Agriculture, forestry, animal husbandry and fisheries	46	0.981	-0.29	0.84(0.000***)	0.212(0.028**)
Industry	46	1.035	-0.235	0.806(0.000***)	0.227(0.014**)
Building industry	46	1.235	0.149	0.758(0.000***)	0.258(0.004***)
Wholesale and retail trade	46	1.275	0.306	0.759(0.000***)	0.26(0.003***)
Transportation, warehousing, and postal services	46	1.167	0.235	0.812(0.000***)	0.21(0.030**)
Accommodation and catering	46	1.052	-0.218	0.81(0.000***)	0.209(0.031**)
Banking industry	46	1.363	0.516	0.732(0.000***)	0.291(0.001***)
Real estate industry	46	1.3	0.247	0.73(0.000***)	0.263(0.003***)
Other industries	46	1.505	1.102	0.73(0.000***)	0.243(0.007***)

Table 1 shows the descriptive statistics and normality test results of GDP (100 million yuan), agriculture, forestry, animal husbandry and fishery, industry, construction, wholesale and retail, transportation, storage and postal industries, accommodation and catering industries, financial industry, real estate industry and other industries[8], including median and average value, which are used to test the normality of data. Generally, there are two test methods for normal distribution. One is the Shapiro-Wilk test, which is suitable for small sample data (sample size ≤5000)[9]. The other is the Kolmogorov-Smirnov test, which is suitable for large samples (sample size >5000). If it is significant (P<0.05), it means that the null hypothesis (data conforms to normal distribution) is rejected and the data does not meet the normal distribution; otherwise, it means that the data meets the normal distribution[10].

Considering the N < 5000 of all samples, the S-W test was adopted, and the significance P value was 0.000***, indicating the level of significance, rejecting the null hypothesis[11]. Therefore, the data did not meet the normal distribution.

Since these data do not meet the normal distribution, the Pearson correlation coefficient method cannot be used to measure the relationship between different industries. Spearman's correlation coefficient rule can be used to measure the correlation when the normal distribution is not satisfied.

2.2.3 Spearman correlation coefficient

Spearman's rank correlation coefficient is a non-parametric measure of correlation in statistics that shows the statistical correlation between the grades of two variables. Spearman's correlation coefficient describes the degree to which the relationship between two variables can be described using a monotone function[12].

When the two variables being compared are monotonically correlated, the Spearman correlation coefficient is 1, that is, their relationship is not linear[13]. This means that all data points with an x value greater than a given data point will also have a larger y value. The Spearman correlation coefficient formula is as follows.

$$r_s = \rho_{R(X),R(Y)} = \frac{cov(R(X),R(Y))}{\sigma_{R(X)}\sigma_{R(Y)}} \tag{1}$$

Spearman correlation coefficients of GDP data of major domestic industries were visualized using thermal maps.

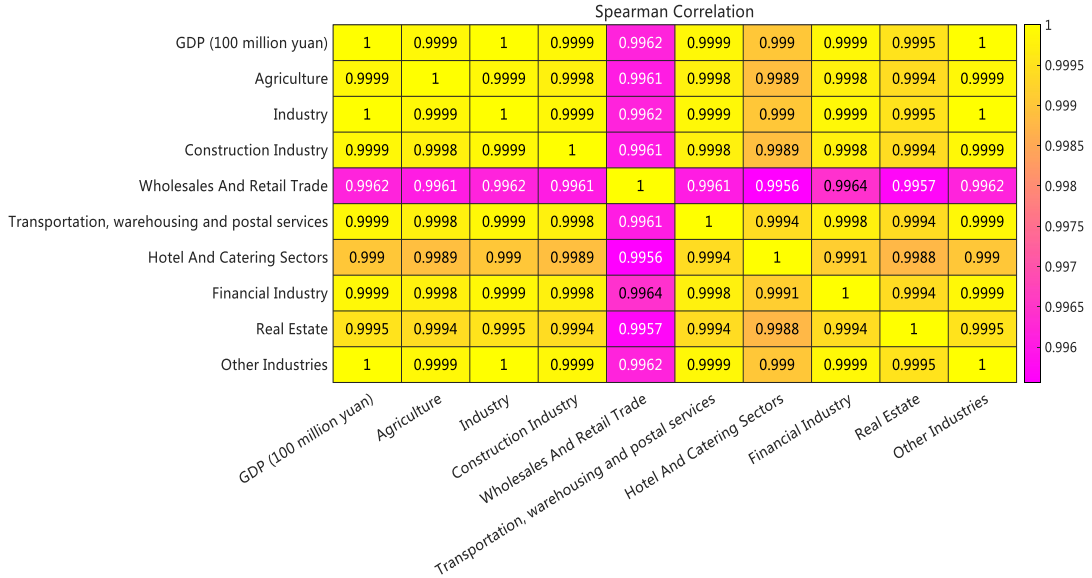

Figure 3. Spearman correlation coefficient visualization heat map of GDP data of major domestic industries

As can be seen from Figure 3, there is a strong positive correlation between industries, and the GDP of one industry does not increase while the GDP of another industry declines. Various industries are interdependent and influence each other to jointly promote sustained economic growth.

As a general rule, the Pearson correlation coefficient and Spearman correlation coefficient are usually calculated simultaneously when conducting data exploration. This is because any difference between the two can further illuminate the relationship between the variables, as well as any anomalies in the data itself. The Pearson correlation coefficient is more appropriate when the relationship between the guessing variables may be linear or when the data is measured on an interval scale. Spearman's correlation coefficient is more appropriate when the relationship between the guessed variables is monotonous, when the data is measured on an ordinal scale, or when the outlier is several orders of magnitude larger than most data points[14].

The pair-to-pair comparison of some industries shows a positive correlation with the lowest correlation coefficient of 0.975, and the GDP of one industry does not increase while the GDP of another industry declines. Various industries are interdependent and influence each other to jointly promote sustained economic growth.

3. Conclusions

Normality test and Pearson correlation analysis were carried out on GDP data of various industries. The former ensured that the data were in line with normal distribution and provided effective support for subsequent analysis; the latter calculated the correlation coefficient to measure industry relations. The results show that each industry has a strong positive correlation with China's GDP, with the lowest correlation coefficient reaching 0.975, indicating that each industry is interdependent and jointly promotes economic growth.

However, there are still some shortcomings in this study. At the data level, relying only on GDP data may not fully reflect the actual situation of the development of the industry and some unique indicators of the industry. For example, the number of R&D input and patent output in the scientific and technological innovation industry, as well as the influence of the communication of works in the cultural and creative industry were not included in the analysis, resulting in the lack of comprehensiveness of the research results. In terms of analysis methods, only the normality test and Pearson correlation analysis are used. For complex industry relations, the analysis means are slightly single, and it is difficult to deeply explore the complex nonlinear relationship between industries.

In the future, to further improve this research, we can start from many aspects. In terms of data collection, broaden the scope of data collection. In addition to GDP data, the representative core index data of various industries are widely collected to make the research data more comprehensive and representative. In terms of analysis methods, multivariate statistical analysis methods such as principal component analysis and gray correlation analysis are introduced to complement existing analysis methods, in-depth analysis of the internal relations between industries from different angles, mining complex association patterns between industries, to more accurately and comprehensively reveal the internal relations between the development of China's major industries.

References

[1] Zuo Xiuxia. ADF unit root test with higher order trend term [J]. Quantitative economy technical economy research, 2019,36(01):152-169.

[2] Jiang Feng, Zhu Jialing, Hu Kaiyong, et al. Application of Pearson correlation coefficient to evaluate the evaporator characteristics of ORC system [J]. Journal of Solar Energy, 2019,40(10):2732-2738.

[3] Zhang Wenhong. Analysis of the development of Hebei real estate market and real estate finance [J], 2007(S1):3.

[4] Ren Weidan, Liu Guiwei, Ye Haopeng, et al. Screening of key long non-coding RNAs of colorectal cancer and the mechanism of action of LINC00174 based on bioinformatics [J]. Chinese Journal of Experimental Surgery, 2023,40(12):2588-2591.

[5] Dong Yi, Tan Hao, Tao Xi, et al. Correlation analysis between coma recovery Scale and event-related potential in patients with chronic consciousness disorder [J]. Chinese Journal of Rehabilitation Medicine, 2024,39(5):663-668.

[6] Zheng Wei, Luo Danglun. Employment, social goals, and capital resource allocation [J]. Quarterly journal of finance, 2015, 9(1):31.

[7] Yang Xinbo. A note on discrete grey prediction model of inhomogeneous exponents and its application [J/OL]. The practice and understanding of mathematics,1-13[2024-11-22].

[8] Li Shuhua, Bei Jing, Wu Yaoye, et al. Lightweight design of integrated front hatch cover based on FAHP-TOPSIS method [J]. Vibration and shock, 2024,43(23):125-133.

[9] Pan Jie. Study on industry selection of comparative advantage in Kashgar City. [J]. 2020.

[10] Li Donghao, Management science and engineering. Research on pricing and causes of premium of convertible bonds [D]. [2025-02-06].

[11] Li Chengfeng, Fu Xiaofan, Xiong Guojiang, et al. Grid-connected inverter control of sliding mode photovoltaic based on grey Wolf parameter optimization [J]. Science, Technology and Engineering, 2024,24(36):15474-15482.

[12] Wu Yuhang, Cha Wenzhong, Ge Jianjun, et al. Task adaptive ranking model based on improved linear regression method [J]. Journal of Huazhong University of Science and Technology (Natural Science Edition), 2020,48(01):93-97+107.

[13] Lin Hu, Gu Dong, Kang Lin. Prediction of line loss rate in the low-pressure region based on gradient lifting decision tree [J]. Information technology, 2020, 44(8):7.

[14] Feng Jianxin, Li Haoyang, Gong Jianxiong, et al. Active disturbance rejection control of ultrasonic motor based on improved whale optimization algorithm[J/OL]. Systems engineering and electronics,1-9[2025-01-12].