ISSN 2706-6827 Vol. 3, Issue 8: 91-94, DOI: 10.25236/IJFS.2021.030814

# Research on the Influencing Factors of China’s Industrial Sector Gross Product Based on Multiple Linear Regression

**Yinglv Cao[1*], Bihui Huang[2]**

[1]*School of Mathematics, University of Jinan, Jinan, Shandong 250022, China*
[2]*Business School, University of Aberdeen, King’s College, Aberdeen, AB24 3FX, Britain*
*caoyinglv@outlook.com*
**Corresponding author*

***Abstract:*** *The development of industry is of vital importance to a country. The gross product of a country’s industrial sector can be used to measure the level of industrial development of the country. And the gross output value of a country's industrial sector is affected by many factors. This paper collects data on total industrial energy consumption, total industrial water consumption, fixed-asset investment in the manufacturing industry, and employment in the secondary industry from the National Bureau of Statistics of China. It uses method of econometrics to establish a multiple linear regression model, which has realistic meaning in economics and is statistically significant, and analyzes which factors have significance on industrial growth.*

***Keywords:*** *Gross Product of Industrial Sector, Influencing Factors, Multiple Linear Regression, Econometrics*

## 1. Introduction

Since the founding of New China, industry has always been a key factor for China’s development. After the reform and opening up, China has successively implemented policies such as giving priority to the development of the light and textile industry, focusing on strengthening basic industries, revitalizing pillar industries vigorously, developing high-tech industries actively. The industrial structure is continuously optimized and upgraded from then on. The development of the private economy has injected vitality into the development of industrialization, and the value of industrial output has continued to break new records. By the end of the 20th century, China had basically completed industrialization, the added value of industrial output accounted for half of the world's, and the development achievements have attracted the attention of the world[1].

## 2. Methodology

### *2.1. Influencing Factors*

In order to establish a multiple linear regression model, the gross domestic product (unit: billion RMB) of the industrial sector is chosen as the explained variable. The rapid development of industrialization is often accompanied by the huge consumption of resources, so the total amount of industrial energy consumption (unit: ten thousand tons of standard coal) and the total amount of industrial water (unit: one hundred million tons of water) are included in the explanatory variables. Industrial development cannot develop without a large amount of investment, because the construction of factories and the purchase of machines are quite expensive, so fixed asset investment (unit: one hundred million RMB) can be used as an explanatory variable. Finally, the explanatory variable--the number of people employed in the industrial sector (unit: ten thousand people)--can also measure the level of industrial development, since the industry that has started not so far in China to develop is low-end and labor-intensive[2].

ISSN 2706-6827 Vol. 3, Issue 8: 91-94, DOI: 10.25236/IJFS.2021.030814

### *2.2. Model Building and Estimation*

Based on the data, a model of the gross domestic product of the industrial sector is established:

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + u_1 \quad (1)$$

Table 1 shows the statistics on the industrial sector from 2011 to 2018, in which the explained variable Y is gross domestic product of the industrial sector, the explanatory variables $X_2$ is the total amount of industrial energy consumption, $X_3$ is the total amount of industrial water, $X_4$ is the fixed asset investment of the whole society of manufacturing industry, and $X_5$ is the employment figures of the secondary industry

*Table 1: Data from National Bureau of Statistics*

| Year | Y | $X_2$ | $X_3$ | $X_4$ | $X_5$ |
|---|---|---|---|---|---|
| 2011 | 28478.64 | 278048 | 1461.8 | 102712.85 | 22544 |
| 2012 | 31919.19 | 284712 | 1380.7 | 124550.04 | 23241 |
| 2013 | 34535.8 | 291130 | 1406.4 | 147704.96 | 23170 |
| 2014 | 38359.51 | 298449 | 1356.1 | 167025.29 | 23099 |
| 2015 | 39501.15 | 295953 | 1334.8 | 180370.38 | 22693 |
| 2016 | 40492.79 | 295615 | 1308 | 187962.12 | 22350 |
| 2017 | 42393.68 | 302308 | 1277 | 193709.97 | 21824 |
| 2018 | 43987.65 | 311151 | 1261.6 | 237899 | 21390 |

With the help of software EViews, the parameters of the regression model are estimated by the OLS estimator and the result is :

$$Y = 29362.49 + 0.07X_2 - 32.09X_3 + 0.07X_4 + 0.82X_5$$

$$t = \ (0.46) \ \ (0.32) \ \ (-1.46) \ \ (0.99) \ \ (0.70)$$

$$R^2 = 0.973 \quad \bar{R}^2 = 0.939 \quad F = 27.95 \quad (2)$$

### *2.3. Model Testing and Modifying*

The regression results show that 93.9% of gross domestic product of the industrial sector in 2011-2018 can be explained by the above four explanatory variables, and the sign of regression coefficient is reasonable, which is in line with economic significance. $R^2$ is more than 90%, and F=27.95, which indicates that the model has a high degree of fitting. But the four explanatory variables all failed to pass the $t$ tests, which mean the model needs to be modified.

*Table 2: Correlation coefficient matrix*

| | Ye | $X_2$e | $X_3$e | $X_4$e | $X_5$e |
|---|---|---|---|---|---|
| E | 1 | 0.959993 | -0.961148 | 0.974345 | -0.663937 |
| $X_2$ | 0.959993 | 1 | -0.913576 | 0.973473 | -0.661969 |
| $X_3$e | -0.961148 | -0.913576 | 1 | -0.940762 | 0.703 |
| $X_4$e | 0.974345 | 0.973473 | -0.940762 | 1 | -0.731151 |
| $X_5$e | -0.663937 | -0.661969 | 0.703 | -0.731151 | 1 |

It can be seen from the correlation coefficient matrix that the correlation coefficients of a majority of explanatory variables are above 0.9, which shows that the explanatory variables are highly correlated, indicating the existence of multicollinearity.

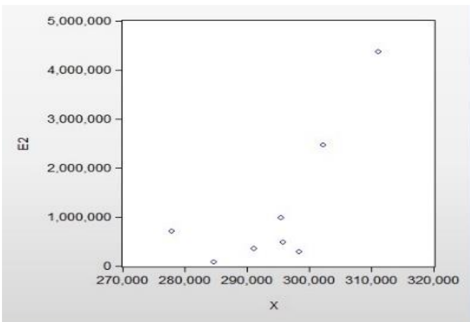


*Fig.1 Square of residual sequence of $X_2$*

ISSN 2706-6827 Vol. 3, Issue 8: 91-94, DOI: 10.25236/IJFS.2021.030814

From the scatter diagram of the square of residual sequence of $X_2$, we can see that the $e_i^2$ has an increasing trend with the growth of $X_2$, demostrating the existence of heteroscedasticity[3].

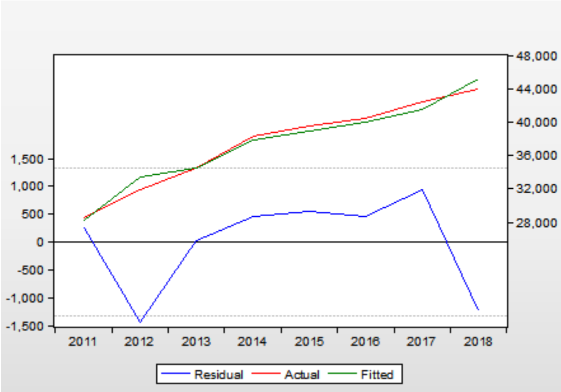


*Fig.2 Residual sequence*

The variation of the residual graph above shows an obvious system pattern: value of residual is continuous negative and then positive, and the residual has a first-order positive autocorrelation

To improve accuracy of this model, firstly we apply weighted least square method to modify heteroscedasticity, and we let the weight equals to $\omega_1 = 1 / |e_i|$, and the result is:

$$Y = 41487.94 + 0.06X_2 - 37.88X_3 + 0.06X_4 + 0.83X_5$$
$$t = (1.15) \quad (0.40) \quad (-2.30) \quad (1.21) \quad (1.09)$$
$$R^2 = 0.989 \quad \bar{R}^2 = 0.975 \quad F = 69.64 \tag{3}$$

The result shows that both $R^2$ and F are improved, the model fits better, and heteroscedasticity is eliminated, but some explanatory variables should be rejected at 5% significance level.

Then, we use stepwise regression method to eliminate multicollinearity, and finally find that abandoning variable $X_2$ can significantly improve the precision of model[4], and the result is :

$$Y = 80853.04 - 62.92X_3 + 0.04X_4 + 1.59X_5$$
$$t = (6.02) \quad (-7.39) \quad (2.29) \quad (9.14)$$
$$\bar{R}^2 = 0.998 \quad F = 1473.12 \quad DW = 1.97 \tag{4}$$

Since Durbin-Waston statistic is very close to 2, we consider there is no autocorrelation in the modified model.

## 3. Conclusion

Applying the econometric model, we can know that China’s Industrial Sector Gross Product is significantly affected by $X_2, X_3$, $X_4$, and has a negative correlation with $X_2$. Due to China's energy saving and emission reduction measures in recent ten years, the industrial sector's dependence on water resource has been significantly reduced, which is in line with China's national conditions and the economic significance of the model. We can also draw a conclusion from the model: China is the "factory of the world". At the present stage, China's industrial development relies heavily on sufficient labor supply, and cheap labor brings opportunities and growth to China's industrial development. Of course, this model has some weaknesses, which is partly different from the actual situation, because the model does not include all the influencing factors and the sample size is relatively small.

## References

*[1] Wang Yan, Su Yi, Gang Cui. Analysis of Quality Measurement and Factors of Industrial Growth in China. Productivity Research, 2020 (07): 77-81.*
*[2] Wang Peigang, Zhou Great Wall. Empirical Analysis and Dynamic Research on the Widening Income Gap of Chinese Residents- Interpretation based on Multiple Linear Regression Models. Management World, 2005 (11): 34-44-171-172.*

ISSN 2706-6827 Vol. 3, Issue 8: 91-94, DOI: 10.25236/IJFS.2021.030814

*[3] Tan Xin, Deng Guangming. Improvements in the Hetero-Difference White Test Methods, Statistics and Decision-Making, 2020, 36 (19): 42-46.*
*[4] Xinyi Shen, Yang Hong, Qiming Qin, Weilin Yuan, Sheng Chen, Trevor Grout, and Shaohua Zhao, (2011) Orientation Angle Calibration for Bare Soil Moisture Estimation Using Fully Polarimetric SAR Data. IEEE Transactions on Geoscience and Remote Sensing.vol. 49(12) pp.4987-4996.*
*[5] Yang Mei, Xiao Jing, Cai Hui. Multiple collinearity in multiple analysis and its treatment methods. China Health Statistics, 2012, 29 (04): 620-624.*