Research on the impact of industrial agglomeration on industrial green efficiency from the perspective of big data

Guanyi Wang
School of economics and management, Beijing Jiaotong University
Beijing, China

Abstract—This paper uses the data of labor force, capital, all kinds of energy input, industrial pollutant emissions and output value of industries in 30 provinces and autonomous regions of China from 2011 to 2015, and uses big data and data envelopment analysis to calculate the industrial green efficiency of provinces and cities in China. At the same time, in order to explore the relationship between industrial agglomeration and industrial green efficiency, the panel regression model is used to conduct quantitative analysis. The results show that: whether from the national level or the eastern, central and western regions, industrial agglomeration can significantly promote the improvement of industrial green efficiency. Environmental regulation, industrial enterprise scale and other control variables can also significantly promote the improvement of industrial green efficiency in some areas, but not in some areas.

1. INTRODUCTION
Since the reform and opening-up, the rapid development of China's industry has led to the recovery and development of other industrial sectors, making fundamental changes in China's economic and social outlook. The rapid rise of China's industry has made an important contribution to the rapid development of China's economy. However, the resources and environment problems caused by the extensive and rapid development path that only focus on quality have overshadowed the sustainable development. According to the 2018 global environmental performance index report, China's environmental performance index is only 50.74, ranking 120 out of the 180 countries and regions involved in the survey. In terms of air quality, it ranks the fourth from the bottom, far behind countries such as Europe, the United States, South Korea and Japan. Therefore, how to solve the resource and environmental problems caused by extensive development and how to promote the sustainability of industry Sustainable development has become an urgent problem.

Industrial green efficiency refers to a kind of comprehensive economic efficiency in which resources and environment costs are included in the accounting system. At present, scholars mainly study from the following aspects: From the perspective of methodology, Charnes, Cooper and Rhodes first proposed DEA model based on constant returns to scale to study efficiency. This method does not need to set the specific form of production function, and the evaluation results are more objective [1]. If DEA model is used to evaluate the industrial environmental efficiency, it is necessary to introduce environmental factors into the model, and the environmental factors do not include only the expected output, but the environmental pollution emission is usually called the unexpected output. In order to solve this problem, Fare and Grosskopf first proposed the DEA model including the unexpected output [2]. Anderson and Petersen established the super efficiency DEA model [3]. At the same time, there are many literatures focusing on the impact of enterprises and industrial agglomeration on green efficiency.
Taylor and Copeland found that, based on the externality theory of industrial agglomeration, the control cost of ecological pollution in the industrial agglomeration area decreases with the increase of returns to scale, and concludes that industrial agglomeration has a positive impact on green development[4]. Cingano and Schivardi selected enterprise level micro data, and empirical research shows that industrial agglomeration can improve the efficiency of green development of enterprises[5]. To sum up, this paper uses the sample data of 30 provinces and autonomous regions in China from 2011 to 2015, through the quantitative analysis of the relationship between industrial agglomeration and industrial green efficiency, trying to solve two problems: first, whether industrial agglomeration has an impact on industrial green efficiency? The second is whether there are regional differences in the impact of industrial agglomeration on industrial green efficiency?

2. VARIABLE SELECTION, RESEARCH METHODS AND DATA EXPLANATION

2.1 Variable Selection
Industrial green efficiency (IGE) is a kind of comprehensive economic efficiency which takes the cost of resources and environment into the accounting system. Based on the traditional calculation of industrial production efficiency with labor and capital as input and gross industrial output as output, the cost of resources and environment is included in the accounting system, and the industrial green efficiency with labor, capital, energy and environment as input and industrial gross output as output is constructed. Labor input is expressed by the average number of employees of industrial enterprises above designated size, and capital input is used as capital input. It is expressed by the net value of fixed assets of industrial enterprises above designated size, and energy input is expressed by the total energy consumption of industrial enterprises above designated size at the end of the year. It should be pointed out that the discharge of industrial waste water, industrial waste gas and industrial solid pollutants belong to the undesirable output. Here, drawing on the relevant research results, the undesirable output is treated as the input variable, hoping that the less the input, the better. The output variable is expressed by the total industrial output value. In terms of estimation method, the relative efficiency calculated by DEA method has many advantages, for example, it is not necessary to determine the specific form of the frontier production function, compensate for the subjectivity of the weight given in the study of multi input and multi output problems, and reduce the calculation error. However, the traditional DEA model cannot distinguish the evaluation units on the effective frontier and identify the differences between the units. Therefore, this paper uses the non radial super efficiency DEA model established by Anderson et al. To estimate the economic efficiency under the constraints of resources and environment.

Industrial agglomeration (IA). At present, the commonly used methods to express the degree of industrial agglomeration are: Concentration Ratio, Herfindahl index, spatial Gini coefficient and location entropy. Based on the availability of data and the need to use location entropy to measure the level of industrial agglomeration, the index can effectively eliminate the impact of regional scale differences, and more accurately reflect the spatial distribution of regional factors. Generally speaking, the greater the location entropy, the higher the degree of industrial agglomeration, otherwise the lower.

The main control variables in this paper are environmental regulation (ER), economic development level (EDL) and industrial enterprise scale (IES). Environmental regulation is expressed by the proportion of industrial pollution control investment in the main business cost of industrial enterprises above designated size. The level of economic development is expressed by real GDP per capita, and the scale of industrial enterprises is represented by the main business income of industrial enterprises above scale divided by the number of industrial enterprises above scale.

2.2 Data Sources
The empirical sample of this paper is the balanced panel data of China's provinces, autonomous regions and municipalities (Tibet, Taiwan, Hong Kong and Macao are not calculated due to lack of data). The data of GDP, gross industrial output value of the whole country, real GDP per capita of all provinces, autonomous regions and municipalities directly under the central government, and gross industrial
output value are derived from *China Statistical Yearbook*. Industrial wastewater discharge, industrial waste gas emission, industrial solid pollutant emission, average number of employees of industrial enterprises above designated size, net fixed assets value of industrial enterprises above designated size, main business income and cost of industrial enterprises above designated size and the number of industrial enterprises above designated size are from *China Industrial statistical yearbook*. The total energy consumption of industrial enterprises above designated size at the end of the year and industrial pollution control investment of industrial enterprises above designated scale are from *China Environmental Statistical Yearbook*.

2.3 Measurement Model

In order to investigate the impact of industrial agglomeration on industrial green efficiency, this paper sets three control variables: environmental regulation, economic development level and industrial enterprise scale. In order to eliminate heteroscedasticity, some variables are treated by logarithm when building the model. The basic model is as follows:

\[
\ln IG_E = \alpha_i + \beta_1 \ln IA + \beta_2 \ln ER + \beta_3 \ln DEL + \beta_4 \ln IES + \mu_i. \tag{1}
\]

Where i is province, t is year; IG is industrial green efficiency; IA is industrial agglomeration; ER is environmental regulation; DEL is economic development level; IES is industrial enterprise scale; \(\alpha_i\) is regional individual effect; \(\mu_i\) is random interference item.

3. EMPIRICAL RESULTS AND ANALYSIS

3.1 Descriptive Statistics

This paper will analyze the industrial green efficiency and industrial agglomeration of provinces and cities in China from two dimensions of time and space.

From the perspective of time, the industrial green efficiency in the eastern region is generally higher than the national level, and shows an upward trend year by year; the industrial green efficiency of the central region is ahead of the national level in 2011, but it is surpassed by the eastern region in 2012, showing a downward trend year by year; the western region is far behind the national level, showing a trend of growth first and then falling.

In terms of space, Guangdong (0.9837), Tianjin (0.9601), Inner Mongolia (0.9271) and Beijing (0.9053) exceeded 0.9; Hainan (0.4418), Hebei (0.4329), Yunnan (0.3968), Qinghai (0.3748), Xinjiang (0.3736), Guizhou (0.3726), Gansu (0.3498), Shanxi (0.3493) and Ningxia (0.2609) were lower than 0.5; the industries of other 17 provinces and cities were lower than 0.5 Green efficiency is between 0.5 and 0.9.

| Year | Whole Country | Eastern Region | Central Region | Western Region |
|------|---------------|----------------|----------------|---------------|
| 2011 | 0.5979        | 0.6373         | 0.6658         | 0.4776        |
| 2012 | 0.6130        | 0.6712         | 0.6603         | 0.4882        |
| 2013 | 0.6175        | 0.6858         | 0.6498         | 0.4941        |
| 2014 | 0.6024        | 0.7057         | 0.6290         | 0.4381        |
| 2015 | 0.5940        | 0.7270         | 0.5943         | 0.4165        |

| Year | Whole Country | Eastern Region | Central Region | Western Region |
|------|---------------|----------------|----------------|---------------|
| 2011 | 1.0686        | 1.0177         | 1.1739         | 1.0312        |
| 2012 | 1.0722        | 1.0232         | 1.1769         | 1.0327        |
| 2013 | 1.0788        | 1.0312         | 1.1848         | 1.0362        |
| 2014 | 1.0605        | 1.0257         | 1.1572         | 1.0101        |
| 2015 | 1.0294        | 1.0246         | 1.1313         | 0.9339        |
Figure 1. Annual average of green economic efficiency and industrial agglomeration level of provinces (autonomous regions and municipalities directly under the central government) from 2011 to 2015

From the perspective of time, the industrial green efficiency in the eastern region is generally higher than the national level, and shows an upward trend year by year; the industrial green efficiency of the central region is ahead of the national level in 2011, but it is surpassed by the eastern region in 2012, showing a downward trend year by year; the western region is far behind the national level, showing a trend of growth first and then falling.

In terms of space, Guangdong (0.9837), Tianjin (0.9601), Inner Mongolia (0.9271) and Beijing (0.9053) exceeded 0.9; Hainan (0.4418), Hebei (0.4329), Yunnan (0.3968), Qinghai (0.3748), Xinjiang (0.3736), Guizhou (0.3726), Gansu (0.3498), Shanxi (0.3493) and Ningxia (0.2609) were lower than 0.5; the industries of other 17 provinces and cities were lower than 0.5 Green efficiency is between 0.5 and 0.9.

3.2 Analysis of estimation results
According to the results of Wald and Hausman test, the fixed effect model is used to investigate the relationship between variables. During the period of industrial agglomeration, the industrial agglomeration coefficient in the eastern and western regions is positive, which shows that the industrial agglomeration coefficient in the eastern and western regions is significant.

Specifically, when the industrial enterprises are relatively dispersed, the cost of each enterprise to deal with pollution is high, so enterprises will choose to deal with pollution as little as possible. When industrial enterprises start to gather and form a certain scale, the industrial agglomeration dividend begins to appear. The entry of specialized pollution treatment enterprises will greatly reduce the cost of pollution treatment of enterprises, which is conducive to promoting industrial green Efficiency.

| Independent Variable | Explained Variable lnIGE |
|----------------------|--------------------------|
|                      | Whole Country | Eastern Region | Central Region | Western Region |
| lnIA                 | 1.1586***     | 0.4631**      | 1.7132***     | 1.0568***     |
|                      | (7.43)        | (2.21)        | (4.83)        | (3.95)        |
| lnER                 | 0.0108        | 0.062***      | -0.0016       | -0.0103       |
|                      | (0.53)        | (3.15)        | (-0.04)       | (-0.23)       |
| lnEDL                | 0.0322        | -0.0945       | -0.2086       | 0.0563        |
|                      | (0.43)        | (-0.92)       | (-1.09)       | (0.48)        |
| lnIES                | 0.2062*       | 0.5317***     | -0.2711       | -0.1628       |
|                      | (1.67)        | (4.34)        | (0.0427)      | (-0.69)       |
| _cons                | -1.1066       | 0.4855        | 1.4012        | -1.2955       |
|                      | (-1.39)       | (0.43)        | (0.74)        | (-1.02)       |
For the control variables of the model, the coefficient of environmental regulation on the improvement of industrial green efficiency in the eastern region is significantly positive, which indicates that the government's environmental policy can promote industrial enterprises to take the road of green development and improve industrial green efficiency; the coefficient of industrial enterprise scale in the whole country and the eastern region is significantly positive, which shows that higher industry scale will be beneficial to the improvement of industrial green efficiency If the scale is small and the distribution of resources is scattered, it will not be conducive to the intensive use of energy and the promotion of industrial green efficiency.

4. CONCLUSION AND SUGGESTION

4.1 Conclusion
Based on the provincial panel data of 30 provinces and autonomous regions in China from 2011 to 2015, this paper mainly demonstrates the relationship between industrial agglomeration and industrial green efficiency by using fixed effect model estimation method. The main conclusions are as follows:

- Industrial green efficiency and industrial agglomeration level show a strong collaborative change trend, which shows that the agglomeration effect of industrial enterprise agglomeration can promote the improvement of industrial green efficiency. Therefore, we should guide the development of industrial agglomeration and promote the green development of industry.
- Control variables such as environmental regulation, economic development level and the scale of industrial enterprises have different impacts on industrial green efficiency. Government environmental regulation policies can improve industrial enterprises' willingness to improve technology level, energy conservation and emission reduction, thus promoting the improvement of industrial green efficiency.
- The scale expansion of industrial enterprises can effectively promote the promotion of industrial green efficiency. Therefore, at this stage, provinces should give full play to the scale economy of industrial enterprises, make enterprises bigger and stronger, and regard it as an effective way to improve industrial green efficiency in the future.

4.2 Suggestion
- As industrial agglomeration can significantly improve the green efficiency of industry, all provinces and cities should speed up the construction of industrial agglomeration area, improve the infrastructure of industrial agglomeration area, attract more industrial enterprises to settle in, at the same time, it should strengthen the division of labor and cooperation of enterprises in industrial agglomeration area, give full play to the advantages of scale economy, so as to promote the improvement of industrial green efficiency.
- Since environmental regulation policies can improve the green efficiency of industry, provinces and cities should formulate environmental policies according to local conditions, which should not only ensure the development speed of industrial enterprises, but also take into account environmental problems, so as to promote the high-quality development of industries in all provinces and cities.
- As the scale of industrial enterprises can improve the green efficiency of industry, the provinces and cities should encourage the industrial enterprises in high pollution industries to become bigger and stronger, and give them corresponding preferential policies to promote their rapid development.

REFERENCES
[1] Charnes. A, Cooper. W. W, Rhodes. E. “Measuring the efficiency of decision making units,” J. European Journal of Operational Research, 2(6):429-444, 1978.
[2] Fare. R, Grosskopf. S. “Multilateral productivity comparisons when some outputs are undesirable: a nonparametric approach,” J. The Review of Economics and Statistis, 71(1):90-98, 1989.
[3] Andersen, P., Petersen, N. C. “A procedure for ranking efficient units in data envelopment analysis,” J. Management Science, 39(10):1261-1264, 1993.

[4] Copeland, B. R., Taylor, M. S. “Trade and Environment: Theory and Evidence,” J. Canadian Public Policy, 06(03):339-365, 2013.

[5] Cingano, F., Schivardi, F. “IDENTIFYING THE SOURCES OF LOCAL PRODUCTIVITY GROWTH,” J. Journal of the European Economic Association, 2(4):720-742, 2004.