Measuring the capacity utilization of China’s coal industry based on latent class stochastic frontier model

Yanping Ju 1,2,3
1 School of Education & Psychology, University of Jinan, Jinan, 250022, China;
2 College of Economics & Management, Shandong University of Science & Technology, Qingdao, 266500, China
3 Email: 3064844749@qq.com

Abstract. Due to the regional differences of coal resources storage and mining in China, the latent class stochastic frontier model is used to calculate the capacity utilization of coal industry from 2001 to 2016. The results show that China's coal industry began to suffer from overcapacity in 2012, and capacity utilization dropped to 0.68 in 2016. The coal-producing provinces are grouped into four classes: abundant, moderate, poor and exhausted. The capacity utilization of coal industry in poor and exhausted classes are generally high and there is little room for improvement. After 2011, the capacity utilization of coal industry in rich and moderate classes showed an obvious trend of decline, falling to a low level in 2015~2016. This indicates that there was a serious waste of resources in coal production of these provinces during this period. Capacity utilization of coal industry is sensitive to the economic development variable, while the market fluctuation make inter-provincial capacity utilization performance similar.

1. Introduction
Due to the extensive development mode in the early stage, China's coal industry has suffered from overcapacity since 2012, and its development has encountered difficulties. In recent years, China's supply-side reform takes promoting quality power and total factor productivity as the main theme of economic development. It is pointed out that the extensive development model in the early stage has caused the waste of resources and restricted the high-quality development of coal industry.

Capacity utilization is an important index to evaluate the operation status of production capacity. Its value is the ratio of actual output to production capacity, that is, how much actual production capacity plays a role in operation. Capacity utilization can be used to indicate the degree of overcapacity. In theory, if the capacity utilization is lower than 1, there will be overcapacity. However, in actual production, the capacity utilization equal to 1 cannot be realized. The reason is that, on the one hand, the factory equipment and other means of production from construction to having products need a certain time delay. On the other hand, due to the influence of management level, market environment, weather and other factors in the production process, the invested production capacity cannot be fully utilized. Therefore, developed countries generally set the standard for the reasonable level of capacity utilization based on historical and empirical data. For example, the European Union considers the capacity utilization below 0.82 as overcapacity, while Japan sets a reasonable range of 0.83~0.86. According to the historical experience of its economic operation, the United States finds...
that when the capacity utilization is below 0.79, overcapacity will cause a negative impact on economic development that cannot be ignored.

The measurement methods of capacity utilization include micro survey and methods with macro data. Micro survey is not suitable to Chinese industries, which require large and accurate corporate databases, and China is a late starter. According to the measurement method with macro data, capacity is the maximum or optimal output of an enterprise under a certain specific circumstance. For example, the concept of technical capacity is the maximum output that an enterprise can achieve by inputting certain factors of production in the production process at a given technical level. Economic capacity is defined as the capacity when the enterprise's cost is minimum or profit is maximum. Thus, the ratio of actual output to capacity is used to get the value of capacity utilization. Based on the concept of technical capacity, Aigner and Chu proposed the frontier production function model, which divided the producer efficiency into two parts: the technology frontier and the technical efficiency. The former describes the boundary of the producer input-output function, while the latter describes the gap between the producer's actual production and the technical frontier[1]. Kirkley used the stochastic frontier analysis to measure the production capacity and capacity utilization of U.S. fisheries[2]. Papers of Garcia[3], Shamim[4] and others are also representative literatures. Fare et al. first introduced DEA method into the calculation of capacity utilization[5]. Based on the concept of economic capacity, Morrison[6], Carol and Mattey[7] have all made in-depth studies on the method of cost function.

Most of the methods used by scholars to calculate the capacity utilization of China's coal industry are based on macro data, and have been improved and perfected according to the particularity of China's coal industry. Such as, Zhang Y. et al. used the Cobb-Douglas production function method to measure the capacity utilization of China's coal industry from 1990 to 2014[8]. Feng et al. used data envelopment model to estimate the capacity utilization of China's coal industry from 1980 to 2013[9]. Du et al. constructed the production function model of China's coal capacity, and drew the conclusion that the amount of capacity is positively correlated with the amount of investment and negatively correlated with the number of employees[10]. Zhao and Huang calculated the capacity utilization of China's coal industry from 1994 to 2012 by using the logarithmic cost function, and found that the overall capacity utilization of China's coal industry was low, with an average capacity utilization rate of 79.96%, which had a large room for improvement[11].

Scholars' research on capacity utilization of China's coal industry is mostly based on the national level. There are few studies on the heterogeneity of capacity utilization in different production areas and its influencing factors. However, the mining conditions of China's coal-producing regions are quite different. For example, the central and western regions have abundant and shallow resources, while the eastern and northeastern regions are facing exhausted and buried deep on resources. These differences will affect the production efficiency and maximum output (optimal output) of coal production units, while the existing measurement methods of capacity utilization generally place all production units under the same output frontier, which will inevitably affect the objectivity of the measuring results. In view of this, this paper adopts the latent class stochastic frontier model to measure the coal capacity utilization in China, takes the mining conditions as the category screening variable that affects the capacity of each coal production unit, and calculates the coal capacity utilization of each coal-producing province. In addition, the evolution characteristics of capacity utilization in different provinces with different production conditions are compared and analyzed in order to understand the capacity utilization situation of China's coal industry more objectively.

2. Formulations for capacity utilization

2.1. Latent class stochastic frontier model

This paper calculates the capacity utilization of China's coal industry based on the concept of technical capacity. The concept of coal production capacity is the maximum output that can be achieved when the factors of production in a coal production unit are fully utilized at a given reserve resource and
technical level. Factors of production may include capital, labor and so on. The capacity utilization is the ratio of the actual output of the coal production unit to the production capacity.

Due to the differences in resource storage status and production conditions between China's coal-producing regions, the production capacities formed by a certain amount of production factors cannot be equal. The paper introduces the latent class stochastic frontier model (LCSFM) to estimate the capacity utilization of each coal province in China. This model is proposed by Orea & Kumbhakar [12] and Greene [13], which classifies the units according to the endogenous attribute, and the statistical frontier estimation is carried out for each class. The basic form of the model is:

$$ Y_{ijt} = f_j(A; X_{ijt}; \beta) e^{u_{ijt}} - u_{ijt}, \quad u_{ijt} \geq 0 \tag{1} $$

Where, $Y_{ijt}$ is the actual output of production unit $i$ in period $t$, $X_{ijt}$ is the amount of input factors, $f_j(\Theta)$ is the production function, $A$ and $\beta$ are production structure coefficients, $v_{ijt}$ and $u_{ijt}$ are the random error and productivity inefficiency terms respectively. $f_0$ is the class label, indicating that production unit $i$ belongs to class $j$. Production capacity can be expressed as:

$$ Y_{ijt}^* = f_j(A; X_{ijt}; \beta) e^{u_{ijt}} \tag{2} $$

Suppose that $v_{ijt}$ and $u_{ijt}$ obey a certain distribution, for example, $v_{ijt} \sim i.d. N(0, \sigma^2_{ij})$, $u_{ijt} \sim N_+(b_{ij}, \sigma^2_{ij})$. The heterogeneity of variance of $u_{ijt}$ can be set as follows:

$$ \sigma^2_{ijt} = \exp(b_{ij} + z_{it} \gamma_{ij}) \tag{3} $$

Among them, $b_{ij}$ is a constant, $z_{it}$ is an influential factor of heterogeneity of non-efficiency item, and $\gamma_{ij}$ is a coefficient. Then the conditional likelihood function of class $j$ in period $t$ of unit $i$ is:

$$ LF_{ijt} = \frac{1}{\sqrt{\sigma^2_{ijt} + \sigma^2_{ij}}} \phi \left( \frac{b_{ij} + \varepsilon_{ijt}}{\sigma^2_{ijt} + \sigma^2_{ij}} \right) \phi \left( \frac{b_{ij} + \varepsilon_{ijt} + \varepsilon_{ijt}}{\sigma^2_{ijt} + \sigma^2_{ij}} \right) / \Phi \left( \frac{b_{ij}}{\sigma^2_{ijt}} \right) \phi(\cdot) \text{ and } \Phi(\cdot) \text{ are the standard normal distribution density function and cumulative distribution function.} \varepsilon_{ijt} \text{ is the sum of random error term and productivity inefficiency term. Then the coefficient of formula (1) can be obtained by solving the following global likelihood function.} \tag{4} $$

$$ \log LF = \sum_{t=1}^{N} \log \sum_{j=1}^{J} P_i(\delta_{ij}) \prod_{t=1}^{T} LF_{ijt} \tag{5} $$

Where, $P_i(\delta_{ij})$ is the prior probability that the production unit $i$ belongs to class $j$, and $\delta_{ij}$ is the category screening variable.

2.2. Production function setting

Considering the substitution effect among production factors, transcendental logarithmic production function is selected to construct the $f_j(\Theta)$. Input factors include capital and labor. Substitute it into formula (1) and use the logarithmic form, as shown below:

$$ \ln Y_{ijt} = \beta_0 + \beta_1 \ln K_{it} + \beta_2 \ln L_{it} + \beta_3 \ln L_{it}^2 + \beta_4 \ln K_{it} \ln L_{it} + \beta_5 \ln K_{it}^2 + \beta_6 \ln L_{it}^2 + \beta_7 t^2 + \beta_8 \ln K_{it} + \beta_9 \ln L_{it} + \varepsilon_{ijt} - u_{ijt}, \quad u_{ijt} \geq 0 \tag{6} $$

Where, $K_{it}$ represents the capital input, $L_{it}$ represents the labor input, $t$ is a year, and $\beta_{0ij} \sim \beta_{6ij}$ is the parameter to be estimated. By substituting (6) into equation (5), and solving it, the following formula can be used to calculate the productivity utilization of each production unit.

$$ \bar{u}_{ijt} = \frac{Y_{ijt}}{Y_{ijt}^*} \tag{7} $$

2.3. Data sources and variable settings

Taking 24 coal-producing provinces in China (Chinese mainland) as the basic coal production units, the capacity utilization of China's coal industry from 2001 to 2016 are calculated. The actual annual output of coal industry in each province is the value of variable $Y_{ijt}$, while the capital stock and labor force quantity are the values of $K_{it}$ and $L_{it}$. According to the research of Feng[14], the economic development level of each province and the fluctuation of coal market are taken as exogenous
influencing factor variables. The production condition of coal industry in each province is taken as the variable for classification screening. The data are from China Energy Statistics Yearbook, China Labor Statistics Yearbook, China Industrial Economy Statistics Yearbook, China Statistics Yearbook and Wind Information.

The capital stock is calculated by using the perpetual inventory method[15], and the annual GDP of each province is taken as the indicator of economic development level, and the change rate of annual coal sales of each province is taken as the index of market fluctuation. The difference in production conditions will ultimately be reflected in the production cost. Therefore, the ratio of the cost of the main business to the income of the main business is taken as the indicator of production conditions, that is, the classification screening variable.

3. Calculation results of China's coal industry capacity utilization rate

3.1. The number of classes

The number of classes was set as 1 ~ 5, and the optimal number was determined by information criterion (AIC or BIC) or likelihood ratio comparison. According to Orea & Kumbhakar, the smaller the AIC or BIC value or the larger the likelihood ratio is, the more likely it is to be chosen as the optimal number of classes[12]. The data results show that when the number of class is 4, the AIC value is the smallest (233.4) and the likelihood ratio is the largest (-94.712), so the classes of coal producing provinces in China are determined as 4. According to the data results of prior probability and marginal mean of the classes, with the increase of production condition index value, the probability of production unit falling into the next class increases. Therefore, the four class of coal-producing provinces in China are named as abundant (I), moderate (II), poor (III) and exhausted (V). The classes are shown in Table 1.

| Classes | Province                          |
|---------|-----------------------------------|
| I       | Inner Mongolia, Guizhou, Shaanxi, Xinjiang |
| II      | Shanxi, Sichuan, Yunnan, Qinghai   |
| III     | Hebei, Fujian, Jiangxi, Shandong, Henan, Hunan, Hubei, Gansu, Ningxia |
| V       | Beijing, Liaoning, Jilin, Heilongjiang, Jiangsu, Anhui, Guangxi |

3.2. Result of parameter estimation

Maximum likelihood estimation was performed for formula (5), and coal capacity utilization of each province from 2001 to 2016 was calculated according to formula (7). From the results of parameter estimation, the variable of economic development level has a significant inhibitory effect (the coefficient of GDP is -0.019) on the heterogeneity of inefficiency of the abundant class, and a significant promoting effect on the other three classes (the coefficient of GDP is positive and significant). It indicates that economic development is more conducive to attracting excellent talents and introducing advanced management experience as well as technology in abundant areas, so that the utilization rate of production capacity shows a trend of convergence.

Market volatility has a significant inhibitory effect (negative and significant coefficient) on the heterogeneity of capacity utilization in abundant, moderate and poor classes. The reason is that the higher the prosperity index of the coal market, the coal enterprises will increase the intensity of the production and output one after another, and the capacity utilization will converge. The time term coefficients of the abundant and moderate classes were significantly positive, indicating that the capacity utilization rate of the two classes had a significant time effect.
3.3. Measurement results of capacity utilization

The calculation results of the capacity utilization of coal industry in each province are shown in the following table. In order to compare the results of other methods, the results of the stochastic frontier analysis (SFA) are also shown in the table.

| Prov. | SFA   | LCSFM | LCM  |
|-------|-------|-------|------|
| BF    | 0.64  | 0.57  | 0.45 |
| LC    | 0.67  | 0.60  | 0.54 |
| HB    | 0.65  | 0.63  | 0.55 |
| LCM   | 0.69  | 0.64  | 0.55 |
| SX    | 0.80  | 0.75  | 0.74 |
| LM    | 0.70  | 0.71  | 0.72 |
| LN    | 0.65  | 0.67  | 0.68 |
| JS    | 0.61  | 0.57  | 0.57 |
| LS    | 0.81  | 0.75  | 0.72 |
| JN    | 0.63  | 0.61  | 0.59 |
| HL    | 0.64  | 0.62  | 0.60 |
| JX    | 0.68  | 0.70  | 0.68 |
| SD    | 0.74  | 0.72  | 0.71 |
| HA    | 0.72  | 0.68  | 0.68 |
| HB    | 0.74  | 0.72  | 0.70 |
| HN    | 0.75  | 0.71  | 0.69 |
| GN    | 0.78  | 0.75  | 0.73 |
| GS    | 0.74  | 0.72  | 0.70 |
| GZ    | 0.72  | 0.64  | 0.63 |
| YN    | 0.71  | 0.69  | 0.67 |
| SN    | 0.76  | 0.72  | 0.70 |
| LH    | 0.70  | 0.68  | 0.68 |
| QH    | 0.60  | 0.57  | 0.55 |
| NX    | 0.61  | 0.59  | 0.58 |
| XI    | 0.70  | 0.67  | 0.65 |
| NT    | 0.68  | 0.66  | 0.64 |

It can be seen from Table 2 that, according to the calculation results of LCSFM, the average level of capacity utilization of China's coal industry from 2001 to 2016 was 0.81, with a trend of first rising and then declining. This is consistent with the development history of China's coal industry. At the beginning of the century, influenced by the industrial policy to accelerate the development of township coal mines, the mining of coal resources adopted an extensive mode with relatively low capacity utilization. From 2004 to 2011, the capacity utilization of China's coal industry was at a relatively high level. This stage was the rapid development stage of China's coal industry, with deepening marketization reform and strengthening resource integration. The construction of large bases and groups was accelerated, the production efficiency of the industry was improved, and the profit margin...
of the industry increased significantly. At the same time, the rise of coal price and the improvement of industry profit margin have largely stimulated the enthusiasm of the whole society to invest in coal construction, laying a foreshadow for the subsequent coal overcapacity. From 2012 to 2016, China's economy entered a low-speed growth stage, and the demand for coal consumption declined. In addition, under the influence of international coal prices, new energy and environmental policies, coal prices have fallen, coal enterprises have lost money and their capacity utilization has decreased.

Compared with the results of SFA, the fluctuation rules of the average capacity utilization of the coal industry calculated by the two methods are consistent. Both of them experienced a process of first rising and then falling, but the value of the result of LCSFM is larger than that of the SFA. This is because, in the SFA results, some provinces with low capacity utilization mainly exhausted and poor provinces, the capacity utilization generally increased after calculating by LCSFM. The traditional stochastic frontier method does not consider the difference of production conditions of production units, and puts all production units on the same frontier. For some provinces with poor production conditions and high production costs, their production frontier is overestimated, thus their capacity utilization is underestimated.

4. Capacity utilization comparation among different classes
The average value of capacity utilization of the four classes from 2001 to 2016 was calculated respectively, and the evolvement rule and differences of capacity utilization with each class was analyzed. From the data results, it can be seen that the coal capacity utilization rate of exhausted provinces such as the Three Northeastern provinces, Beijing and Jiangsu has been relatively stable and at a high level from 2001 to 2016, with little room for upward improvement. Poor provinces such as Shandong, Henan, Hunan and Hubei have been at a high level of capacity utilization since 2004. Before 2004, there were dramatic fluctuations, with a decline followed by a rise, in sync with the abundant and moderate classes. Provinces in this class have faced the same situation in recent years, such as Shandong and Henan, where the combined cost of coal production is relatively high. The main reason is that the mining time is longer, the resources are deeper, the total amount of employment is larger, and the mining cost is higher than that of other regions in the industry. The capacity utilization of the moderate class has a certain room for improvement, and the development of the coal industry in the provinces such as Shanxi and Sichuan is in the stage of replacing old coal mines with new ones. The capacity utilization of the abundant class is at a relatively reasonable level. This class has rich coal resources and good mining conditions, but it also has problems such as shortage of water resources, fragile ecological environment and relatively independent market.

5. Conclusions and suggestions
5.1. Conclusions
(1) The production capacity put into China's coal industry is underutilized.
The coal capacity utilization of 24 coal-producing provinces in China from 2001 to 2016 was calculated by the latent class stochastic frontier method. The calculation results show that the average level of capacity utilization in China's coal industry is 0.81, and shows a trend of first rising and then declining, dropping to 0.68 in 2016. By US and Western European standards, there is a overcapacity. This is related to the industrial environment of excess capacity in China's coal industry, which is not fully utilized. The capacity utilization value of LCSFM is generally larger than that of SFA. This is because, after considering the difference in production costs, some coal-producing provinces have reached a high level of capacity utilization under the current production conditions, with little room for improvement.

(2) There are differences in capacity utilization among groups.
The latent class model groups China's coal-producing provinces into four classes: abundant, moderate, poor and exhausted. By comparing and analyzing the coal capacity utilization of the four classes, the conclusion is that the coal capacity utilization of the poor and exhausted provinces is
generally at a high level and the room for improvement is little. The capacity utilization of moderate class dropped to about 42% in 2015～2016, indicating that production resources were not fully utilized.

(3) Coal capacity utilization is affected by related factors. The capacity utilization of moderate, poor and exhausted classes are sensitive to economic development variables. Market fluctuation cannot cause effective change of capacity utilization, which on the one hand indicates that the marketization level of China's coal industry is relatively low, on the other hand, it is also determined by the high input characteristics of the coal industry.

5.2. Suggestions
(1) The key to improve the capacity utilization rate of China's coal industry lies in the western and emerging coal production bases, so the investment in infrastructure construction and technological innovation should be increased. (2) Improve the marketization level of China's coal market, especially abundant coal groups, and improve the guiding role of the market for coal production under the premise of total quantity control. (3) Give play to regional advantages, rationally distribute, and optimize the allocation of production factors. (4) According to the constraints of the future environment on coal production and consumption, the investment in the coal industry should be rationally planned to avoid overcapacity or insufficient.

Acknowledgments
This research is supported by the National Science Foundation of China (Grant No. 51574157). Thanks to the editor and the anonymous reviewers for their constructive comments, which is much helpful for improving the manuscript.

References
[1] Aigner D J, and Chu S F 1968 On Estimating the Industry Production Function The American Economic Review 58 826
[2] Kirkley J, Morrison P C J, and Squires D 2002 Capacity and Capacity Utilization in Common-pool Resource Industries Environmental & Resource Economics 22 71
[3] Garcia A, Shen Z 2010 Equilibrium Capacity Expansion under Stochastic Demand growth Operations Research 58 30
[4] Shamim A 2016 Profit and Cost Efficiencies of Conventional Banking (CB) and Islamic Banking (IB) Using Stochastic Frontier Approach: A Comparative Study Journal of Islamic Economics, Banking and Finance 12 63
[5] Färe R, Grosskopf S, and Kokkelenberg E C 1989 Measuring Plant Capacity, Utilization and Technical Change: A Nonparametric Approach International Economic Review 30 655
[6] Morrison C J 1985 On the economic interpretation and measurement of optimal capacity utilization with anticipatory expectations Review of Economic Studies, 52 295
[7] Carol C, and Mattey J 1997 Capacity utilization Journal of Economic Perspectives 11 151
[8] Zhang Y, Nie R, Shi R, Zhang M 2018 Measuring the capacity utilization of the coal sector and its decoupling with economic growth in China’s supply-side reform Resources, Conservation and Recycling 129 314
[9] Feng D M, Wang S, Zhai C X 2015 Empirical Research for Coal Industry Estimation and Influencing Capacity Factors in Utilization Rate China Statistics & Information Forum 30 48 (in Chinese)
[10] Du G L, Yang W G, Wang D 2016 The construction of China's coal capacity function and Suggestions on the placement policy of coal miners -- based on the macro level Economic Research Guide 25 13 (in Chinese)
[11] Zhao B F, Huang Z G 2014 Study on capacity utilization estimation and evolution characteristics of China's coal industry Statistics & Information Forum 29 15 (in Chinese)
[12] Orea L, and Kumbhakar S C 2004 Efficiency measurement using a latent class stochastic
frontier model *Empirical Economics* **29** 169

[13] Greene 2005 Reconsidering Heterogeneity in Panel Data Estimators of the Stochastic Frontier Mode *Journal of Econometrics* **126** 269

[14] Feng W 2018 Coupling analysis of economic fluctuations to capacity utilization: an empirical study based on China provincial panel data *Economic Fabric* **35** 82 (in Chinese)

[15] Tian Y C 2016 Estimation of capital stock in different industries in China:1990–2014 *Research on quantitative economy and technology economy* **6** 3(in Chinese)