Research on influencing factors of coal price based on improved principal component analysis and grey correlation method

Zhaobin DUAN¹, Yanjun MENG²
¹Engineering Techniques Training Center, Civil Aviation University of China, Tianjin 300300, China
²College of Electronic Information and Automation, Civil Aviation University of China, Tianjin, 300300, China
E-mail: zbduan@cauc.edu.cn, 26428496@qq.com

Abstract. As an important basic energy in China, the price fluctuation of coal is related to the operation of China’s coal market and the whole energy market and coal price analysis has always been the focus of the industry. By establishing an improved principal component-grey correlation analysis (PCA-GRA) model, the influencing factors of coal price were analyzed and tested one by one, and the influence degree of the influencing factors such as total coal production, total coal consumption, coal import and export volume, and world coal price on China's coal price was studied. The results show that coal price fluctuation is greatly influenced by coal energy storage level and market environment level. In order to promote the development of coal industry, some suggestions are put forward.

1. Introduction
Coal is an important part of China’s energy structure [1]. The analysis of influencing factors of coal price is conducive to the in-depth study on the change trend of coal price, which provides a basic evaluation index system for the specific coal price forecast in China in the future. Reference [2] used analytic hierarchy process (AHP) to analyze influencing factors of coal price, emphasizing the influence of microscopic factors on coal price fluctuation. However, it relied heavily on expert scoring, which was highly subjective. Reference [3] selected eight influencing factors of coal price for stepwise regression analysis of variables, studying and screening the main influencing factors. But it only focused on retaining the most significant influencing factors and neglected the analysis of other variables.

Based on the above studies, this paper establishes an improved PCA-GRA model to conduct quantitative representation and ranking of each influencing factor. Through empirical research, a more scientific index system affecting coal price can be obtained, which makes policy recommendations and the prediction of future coal price fluctuations more reasonable.

2. Improved PCA-GRA model

2.1. Improved PCA model
The process of the model is as follows:

Step 1: Standardized treatment of indicators. Suppose the observation matrix of the original data of the index is:
Then standardized processing of the data is

\[ x'_j = \frac{x_j}{\max_{i=1}^n x_{ij}}, 1 \leq j \leq p \]  

(2)

In formula (1), \( X \) represents the \( n \times p \) matrix formed by the \( n \) sets of raw data for \( p \) indicators; \( x_{ij} \) is the \( i \)-th raw data of index \( j \); \( x'_j \) is the \( i \)-th standardized data of index \( j \).

Step 2: Optimize index data. Because the influence degree of each index on the system is different, the influence degree variable of data is introduced to optimize the index data before using the data to analyze the system. In order to increase the objectivity of the data, the entropy weight method is introduced to calculate the weight \( \omega_j \) of each index.

- Calculate the information entropy \( E_j \) of the index \( j \). Notes: \( E_j = -\frac{1}{\ln n} \sum_{i=1}^n (x'_j \ln x'_j) \)
- Calculate the difference coefficient \( g_j \) of the index \( j \). Notes: \( g_j = 1 - E_j \)
- Calculate the weight \( \omega_j \) of index \( j \). Notes: \( \omega_j = \frac{g_j}{\sum_{j=1}^n g_j} \)
- The \( i \)-th optimization data after the introducing influence degree of the index \( j \) is

\[ y'_i = x'_i \times \omega_j \]  

(3)

Step 3: Determine correlation coefficient matrix \( R \) of optimized data. Notes:

\[ R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{np} \end{bmatrix} \]

(4)

\[ r_{ij} = \frac{\text{Cov}(y_i, y_j)}{\sqrt{\text{Var}(y_i)} \sqrt{\text{Var}(y_j)}} \]  

(5)

In formula (5), \( r_{ij} \) is the \( i \)-th correlation coefficient of the index \( j \).

Step 4: Extract principal component factors. The eigenvalue \( \lambda_i \) of the correlation coefficient matrix and the corresponding eigenvector \( F_i \) are calculated to determine the contribution rate and the cumulative contribution rate. Determine the number \( m \) of principal components according to the cumulative contribution rate \( \sum_{i=1}^m \lambda_i \geq 0.85 \), so that the utilization rate of information can reach more than 85%.

2.2. The GRA model

The process of the model is as follows:

Step 1: Determination of principal component reference sequence and comparative sequences. The reference sequence is recorded as \( X_0 = \{x_{0}(1), x_{0}(2), \ldots, x_{0}(n)\} \), and the coal price is taken as the reference sequence. Comparison sequence is recorded as \( X_j = \{x_j(1), x_j(2), \ldots, x_j(n)\}, j = 1, 2, \ldots, m \) [5].
Step 2: Calculate the absolute difference sequence $\Delta_j(k)$ of the index $j$. Notes:

$$\Delta_j(k) = \left| x_j(k) - x_k(k) \right|, \quad k = 1, 2, \ldots, n; j = 1, 2, \ldots, m$$  \hspace{1cm} (6)

In formula (6), $x_j(k)$ is the number of the $k$-th node of the reference sequence; $x_k(k)$ is the number of the $k$-th node of the comparison sequence.

Step 3: Find the maximum difference $\Delta_{\text{max}}$ and minimum difference $\Delta_{\text{min}}$. Notes:

$$\Delta_{\text{max}} = \max_j \max_k \Delta_j(k)$$ \hspace{1cm} (7)

$$\Delta_{\text{min}} = \min_j \min_k \Delta_j(k)$$ \hspace{1cm} (8)

Step 4: Calculate correlation coefficient $\gamma_j(k)$ of the $k$-th node of index $j$. Notes:

$$\gamma_j(k) = \frac{\Delta_{\text{min}} + \varepsilon \Delta_{\text{max}}}{\Delta_j(k) + \varepsilon \Delta_{\text{max}}}, \quad k = 1, 2, \ldots, n; j = 1, 2, \ldots, m$$ \hspace{1cm} (9)

In formula (9), $\varepsilon$ is the resolution coefficient, and $\varepsilon \in (0, 1)$, usually taking $\varepsilon = 0.5$.

Step 5: Determine the comprehensive grey correlation degree $\gamma_j$ of the index $j$. Notes:

$$\gamma_j = \frac{1}{n} \sum_{k=1}^{n} \gamma_j(k), \quad j = 1, 2, \ldots, m$$ \hspace{1cm} (10)

3. Empirical Analysis

3.1. Determination of coal price impact index

In essence, the extraction and determination of any system index depend on the qualitative analysis of the system itself and the process of transforming qualitative analysis into quantitative expression, in which reducing the subjectivity of qualitative analysis is a key step. This paper makes a systematic analysis of the influencing factors of coal price with the help of energy manual and literature. By comprehensively considering the international market price of coal, macro economy, domestic coal market, coal demand of Chinese downstream enterprises, and the price of coal substitute products [6-8], a qualitative analysis of influencing factors of coal price is carried.

According to the analysis, influencing factors of coal price can be selected. The following principles should be followed in the selection of influencing factors. First, the principle of strong correlation. Second, the principle of quantity should not be too much, not more than 15. Based on the above principles, ten influencing factors are selected from five aspects, as shown in Fig. 1.
3.2. Application of improved PCA in studying influencing factors of coal price

In view of the impact indicators listed in Fig. 1, the coal index data from 2008 to 2018 are selected as the data required by the system. And the entropy weight method was used to calculate the weight of each index. The original index data are standardized and the influence degree of the index is introduced to obtain the optimized index data. Then through the improved PCA component matrix, two principal component analysis models can be obtained, as shown below:

\[
F_1 = -0.017y_1 + 0.048y_2 + 0.284y_3 + 0.227y_4 + 0.270y_5 - 0.258y_6 - 0.082y_7 - 0.010y_8 + 0.216y_9 - 0.120y_{10}
\]

\[
F_2 = 0.134y_1 + 0.235y_2 + 0.139y_3 + 0.049y_4 + 0.108y_5 - 0.096y_6 - 0.275y_7 - 0.204y_8 + 0.371y_9 + 0.059y_{10}
\]

We know from the model that in the first principal component, the scores of total coal production, coal import volume and total coal consumption are higher, so the first principal component can be defined as the coal reserve level. In the second principal component, the scores of international market oil price and GDP growth rate are higher, which can represent the market environment level. Therefore, the fluctuation of coal price is influenced by the level of coal reserves and market environment.

The optimized data are added to the principal component analysis model and the respective scores and comprehensive scores of the two principal components can be obtained by using formula (11), as shown in Table 1.

\[
F_Z = \frac{F_1 \times 68.282\% + F_2 \times 17.543\%}{85.824\%} 
\]

Table 1. Principal component score results

| Year | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 |
|------|------|------|------|------|------|------|------|------|------|------|------|
| $F_1$ | -0.082 | -0.005 | 0.001 | 0.013 | 0.033 | 0.036 | 0.024 | 0.002 | 0.003 | 0.010 | 0.013 |
| Ranking | 11 | 10 | 9 | 4 | 2 | 1 | 3 | 8 | 7 | 6 | 5 |
| $F_2$ | 0.020 | 0.030 | 0.035 | 0.037 | 0.038 | 0.040 | 0.033 | 0.004 | 0.002 | 0.009 | 0.014 |
| Ranking | 7 | 6 | 4 | 3 | 2 | 1 | 5 | 10 | 11 | 9 | 8 |
| $F_Z$ | -0.061 | 0.002 | 0.008 | 0.018 | 0.034 | 0.037 | 0.026 | 0.002 | 0.003 | 0.009 | 0.013 |
| Ranking | 11 | 10 | 7 | 4 | 2 | 1 | 3 | 9 | 8 | 6 | 5 |

In order to facilitate observation and analysis, the scores of principal components in Table 1 are plotted as graphs. And making a comparison with the fluctuation of coal price, as shown in Fig. 2. The comparison shows that with the change of coal reserve level $F_1$ and market environment level $F_2$, the coal price changes accordingly. It can be found that 2009, 2011 and 2015 are the major years of change.
Coal reserves rose sharply in 2009, leading to oversupply, which made the price of coal fall. With the slow growth of coal reserves, the price of coal had been rebounded. Until 2011, the growth rate of coal reserves accelerated, making the price fall again. By 2013, although the level of coal reserves had been slowed down, the level of market environment began to decline. Due to economic constraints, coal prices continued to decline, which was in line with the market price mechanism. With the steady growth of coal reserve level and market environment level in 2015, the coal price gradually and steadily rose. It further explains that with the implementation of the 13th Five-Year plan, the reform of the coal industry mechanism has achieved certain results.

3.3. Correlation degree analysis of main influencing factors of coal price
According to the grey correlation analysis model, the coal price in China is taken as the reference series. And world coal price, GDP growth rate, total coal production, total coal consumption, China’s coal import, China’s coal export, steel production, thermal power generation, international market oil price and international market natural gas price are the comparative data series.

| Relevance | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ | $x_9$ | $x_{10}$ |
|-----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|---------|
| Size      | 0.666 | 0.789 | 0.707 | 0.696 | 0.691 | 0.475 | 0.614 | 0.642 | 0.750 | 0.533   |
| Ranking   | 6     | 1     | 3     | 4     | 5     | 10    | 8     | 7     | 2     | 9       |

Taking the standardized data of indicators into the grey correlation analysis model we can get the size and change trend of the correlation coefficient between coal price and influencing factors during 2008-2018, as well as the grey correlation analysis results of each index, as shown in Fig. 3 and Table 2.

It can be seen from the chart that GDP growth rate, international market oil price, total coal production, total coal consumption and China’s coal import are highly correlated with coal price, which
is consistent with the analysis results of improved PCA. The grey correlation analysis can further obtain the ranking result of the main influencing factors of coal price.

4. Conclusions
To sum up, through the qualitative and quantitative analysis of the influencing factors of coal price, it can be found that the main influencing factors of price fluctuation are GDP growth rate, oil price in the international market, total coal production, total coal consumption and China's coal imports. Among them, on the one hand, GDP growth rate and oil price in international market represent the market environment level. On the other hand, the total coal production and coal import determine the level of coal supply and total coal consumption determines the level of coal demand. The change of supply and demand represents the change of coal reserve level. Therefore, the influencing factors mainly affect the coal price from the two aspects of coal reserve level and market environment level, and then the evaluation index system of the coal price analysis system can be obtained. In the process of regulating the coal market and reforming the energy system, it is particularly important to focus on the influencing factors of coal price from the perspectives of coal supply and demand and market environment.

Acknowledgments
Thanks for the support of Civil Aviation University of China special fund for the basic scientific research service fee project of central universities(3122018S001).

References
[1] Zhao X M, Zhang T, Zou S H. Research on the impact of thermal coal futures price volatility on China's coal economy[J]. China mining, 2020, 29(1): 34-40.
[2] Wang W, Li G D. Research on influencing Factors of Coal price in China based on analytic Hierarchy Process[J]. Operation and Management, 2016(12): 87-90.
[3] Zhang T G, Zhao D Z. Empirical Study on influencing Factors of Coal price Fluctuation in China[J]. China Energy, 2018, 40(3): 16-21.
[4] Wang X L. Application of Grey Relation Analysis Theory to Choose High Reliability of the Network Node. 2019, 1237(3).
[5] Wang X L. Application of Grey Relation Analysis Theory to Choose High Reliability of the Network Node. 2019, 1237(3).
[6] Yang J. Multi-factor Analysis of coal price Fluctuation[J]. Knowledge Economy, 2015(14): 81-82.
[7] Gao D, Nie R. Construction of evaluation Index System for coordinated Development of coal enterprise economy and energy environment in the Context of new Normal Economy[J]. China Coal, 2017, 43(11): 25-29.
[8] Li B. China's Coal price Formation mechanism and influencing factors of Coal price[J]. China's Strategic Emerging Industries, 2018(12): 82-83.