Influencing Factors and Projections of Coal Price

Wenlong Zhang*
School of Energy and Mechanical Engineering, North China Electric Power University, Baoding 130600, China;
*Corresponding author e-mail: wenlongzhang@ncepu.edu.cn

Abstract. The price of coal is related to the healthy development of China's coal industry and the national economy and people's livelihood, and has important research significance. The coal price is not only regulated by the relevant departments of the state, but also affected by the domestic coal market. In this paper, we systematically analyze the internal and external factors of coal price, and study the formation mechanism of domestic coal price and the main problems existing in the current coal price.

Keywords: Coal industry, Multiple regression analysis, Neural networks.

1. Introduction
In 2010, China consumed more than 3.2 billion tons of coal, accounting for more than 70 percent of primary energy. Coal is the grain of Chinese industry, coal price is not only related to the healthy development of coal industry, but also related to power plant, steel, building materials, chemical industry and other basic industries. The rise of coal price will inevitably cause the rise of PPI and CPI, and then affect the national economy and people's livelihood. Taking the price of coal as the research object, it is of great practical significance to analyze the problems existing in China's coal industry.

2. Analysis of Factors Influencing Coal Price

2.1. Analysis of Factors
By means of quantitative analysis, the main factors affecting coal prices are given, relevant data are found, and the coefficient is compared by SPSS multivariate regression analysis. The absolute value of the coefficient reflects the influence of the influencing factors, and the positive and negative coefficients reflect the positive and negative factors. The theory of supply and demand in economics holds that under the market mechanism, the size of commodity price and its direction of change are determined by the quantity of supply and demand in the market, while the change of price level and its direction in turn affects the adjustment of supply and demand quantity in the market, and further adjusts the allocation of resources in the market. In the market mechanism, when the supply is greater than the demand, the buyer's market is formed and finally formed; in the market mechanism, when the demand is greater than the supply, the seller's market is formed, at this time, finally. Therefore, the factors that affect the fluctuation of coal price are mainly those that affect coal supply and coal demand.
2.2. Multiple regression analysis
Through the analysis of coal price data from 2003 to 2017, we get the following figure 1

![Figure 1. Coal price trends](image)

After analyzing the trend of price in the chart, we choose the index regression equation as the simulation of the change trend of price index.

2.2.1. Variables and data selection
First of all, the article to coal prices and six variables: production costs, coal transport capacity, raw coal production, coal net import volume, domestic GDP, national policies Granger causality test.

Then the linear relationship between coal price and supply and demand is tested. The formula is:

\[ PIN = C + \beta_1 SALE + \beta_2 SUP + \beta_3 NIM + \mu \]  

(1)

The linear relationship between coal price and national economic development is tested. The formula is:

\[ PIN = C + \beta_1 GDP + \beta_2 ROC + \mu \]  

(2)

The linear relationship between domestic coal price and foreign coal price is tested. The formula is:

\[ PIN = C + \beta_1 BJP + \mu \]  

(3)

The data selected here are monthly data from January 2006 to May 2017, and coal prices are measured by the national composite coal price index, measured as 100 in January 2006. Finally, the linear regression test of domestic coal price and supply and demand, national economic development and foreign coal price is carried out. The formula is:

\[ PIN = C + \beta_1 GDP + \beta_2 ROC + \beta_3 BJP + \beta_4 CHA + \mu \]  

(4)

Supply and demand gap (CHA)= output-volume-consuming net imports are used to describe the variable supply and demand.

2.3. Unit root test
Because the sample data is time series data, and the vast majority of economic time series variables are non-stationary, before regression analysis of the time series of economic variables, the unit root test should be carried out first to judge the stability of the variables. This paper will carry on the unit root
test to the variable, examine the national economy development, the coal transportation capacity, the coal production quantity, the production cost, the import and export quantity and the national policy. The following Figure 2 shows that there may be a single first-order difference between nim, pin, sale, bip, roc and sup. The root hypothesis was rejected at a significant level of 1%. All of these variables have first order single integrality. And GDP second-order differences reject the hypothesis at the same significant level, indicating that the variable is second-order single.

| Variables | Type of inspection | Inspection value | 1% threshold | 5% threshold |
|-----------|-------------------|-----------------|--------------|--------------|
| nim       | Original value    | -1.18           | -3.51        | -2.9         |
|           | First order differential | -11.6          | -3.51        | -2.9         |
|           | Second order differential | -8.53          | -3.51        | -2.9         |
| pin       | Original value    | -3.85           | -3.51        | -2.9         |
|           | First order differential | -5.84          | -3.51        | -2.9         |
|           | Second order differential | -7.59          | -3.51        | -2.9         |
| sale      | Original value    | -2.51           | -3.51        | -2.9         |
|           | First order differential | -8.81          | -3.51        | -2.9         |
|           | Second order differential | -7.72          | -3.52        | -2.9         |
| sup       | Original value    | -1.67           | -3.51        | -2.9         |
|           | First order differential | -10.07         | -3.51        | -2.9         |
|           | Second order differential | -7.59          | -3.52        | -2.9         |
| gdp       | Original value    | 0.12            | -3.51        | -2.9         |
|           | First order differential | -2.19          | -3.72        | -2.99        |
|           | Second order differential | -13.18         | -3.51        | -2.9         |
| bjp       | Original value    | -2.32           | -3.51        | -2.9         |
|           | First order differential | -6.51          | -3.51        | -2.9         |
|           | Second order differential | -13.18         | -3.51        | -2.9         |
| roc       | Original value    | -2.2            | -3.67        | -2.96        |
|           | First order differential | -4.83          | -3.67        | -2.96        |
|           | Second order differential | -7.45          | -3.69        | -2.97        |

**Figure 2.** Unit root test

### 3. Coal price forecast

According to the main factors that affect the coal price, and combined with the coal price data, according to the mapping relationship between them, each factor is predicted by the neural network model, and substituted into the model, and finally the coal price can be obtained.

#### 3.1. Construction of BP Neural Network Model

In order to improve the accuracy of this model, feedforward neural network is chosen as the basis of modeling. The structure diagram is shown in the diagram: According to the model of BP neural network constructed by this paper, it is assumed that the input layer in the neural network is 6 neurons, which are national economic development, coal transportation capacity, coal production, production cost, import and export quantity and national policy, and the hidden layer is 31 neurons, and the output layer is coal price.
3.2. Genetic Algorithm-based Model Solutions
Calculation of input and output of

Step 1: Calculation of neuron input and output

The data of each input layer and output layer are defined separately. After calculating the input and output of neurons in each layer, we can get:

1) The input vector for the hidden layer is:

\[ h_{i_k}(k) = \sum_{i=0}^{n} w_{hi}x_i(k) \quad h = 1, 2, \cdots, 13 \]  

(5)

2) The output vector of the hidden layer is:

\[ h_{o_k}(k) = f(h_{i_k}(k)) \quad h = 1, 2, \cdots, 13 \]  

(6)

3) The output layer input vector is:

\[ y_{i_o}(k) = \sum_{h=0}^{12} w_{oh}h_{o_k}(k) \quad o = 1 \]  

(7)

4) The output vector of the output layer is:

\[ y_{o_o}(k) = f(y_{i_o}(k)) \quad o = 1 \]  

(8)

Step 2: Calculation of deviation of output layer neurons by error function

After getting the input and output function, we need to get the partial derivative of the error function to each neuron in the output layer by mathematical operation. The partial derivative can be obtained by calculation:

\[ \frac{\partial e}{\partial w_{oh}} = \frac{\partial e}{\partial y_{i_o}} \frac{\partial y_{i_o}}{\partial w_{oh}} \]  

(9)

Step 3: Calculation of deviation of error function to neurons in hidden layer

In order to calculate the partial derivation of each neuron in the hidden layer, we need to use the connection weight of the hidden layer to the output layer, the function of the output layer and the function of the hidden layer to solve the problem.

\[ \frac{\partial e}{\partial w_{oh}} = \frac{\partial e}{\partial y_{i_o}} \frac{\partial y_{i_o}}{\partial w_{oh}} \left( -\delta_o(k)h_{o_k}(k) \right) \]  

(10)

\[ \frac{\partial e}{\partial h_{i_k}(k)} = \frac{\partial e}{\partial h_{i_k}(k)} \frac{\partial h_{i_k}(k)}{\partial h_{i_k}(k)} \]  

(11)

\[ \frac{\partial h_{i_k}(k)}{\partial w_{hi}} = \frac{\partial (\sum_{i=0}^{7} w_{hi}x_i(k))}{\partial w_{hi}} = x_i(k) \]  

(12)
5 \frac{\partial e}{\partial h_i(k)} = \frac{\partial}{\partial h_o(k)} \sum_{o=1}^{1} \left( (d_o(k) - \left( \sum_{h=0}^{7} w_{ho} h_o(h(k))^2 \right) \right) \frac{\partial h_o(k)}{\partial h_i(k)} \tag{13}

Step 4: use the output layer to modify the connection weights
Correction of connection weights by using each neuron output. The specific solution is:

$$\Delta w_{oh}(k) = -\mu \frac{\partial e}{\partial w_{oh}} = \mu \delta_o(k) h_o(k)$$

$$w_{oh}^{N+1} = w_{oh}^{N} + \mu \delta_o(k) h_o(k) \tag{14}$$

Step 5: use input layer to modify connection weights
By using each neuron input to modify the connection weights, the specific solution is:

$$\Delta w_{hi}(k) = -\mu \frac{\partial e}{\partial w_{hi}} = \delta_h(k) x_i(k)$$

$$w_{hi}^{N+1} = w_{hi}^{N} + \mu \delta_h(k) x_i(k) \tag{15}$$

Step 6: global error

$$E = \frac{1}{2m} \sum_{k=1}^{m} \sum_{o=1}^{1} (d_o(k) - y_o(k))^2$$

Step 7: judgment
In the judgment, the accuracy needs to be compared with the maximum number of times of the budget, and when the precision and the number of times of the budget do not meet the requirements, the cycle of the above steps needs to be carried out until the requirements are met.

3.2.1. icon results
After adjusting the neural network model established in this paper, the neural network is returned Figure below:
The weights and thresholds obtained at this time are the relational values of each level of this neural network model under the current accuracy. Through the weights and thresholds between the levels, the remaining 20% of the data can be predicted, and the accuracy of this model can be tested at the same time. Final results.

3.3. Coal price fluctuations and projections
Firstly, the model is established to predict coal supply and demand, and then the fluctuation mechanism of coal price is analyzed. During the analysis of coal price fluctuation mechanism, the linear model is used as the benchmark model to simulate and depict the inherent fluctuation characteristics and behavior of coal price; then the established linear model is tested on the basis of the benchmark linear model, such as the stability test of parameters, the nonlinear test between variables, the nonlinear Granger causality test, and if there is a nonlinear relationship, the nonlinear model is constructed to analyze the coal price fluctuation mechanism and predict the coal price.

3.3.1. Linear VAR Model
Because economic theory can not strictly explain the dynamic relationship between variables, traditional econometric models are based on these economic theories, and the left and right end of the equation can appear endogenous variables. So this paper selects the vector autoregressive model to construct the benchmark linear model. The main principle of this model is to construct the model by using the lag value of each endogenous variable. VAR (P) model is expressed as:

\[ y_t = \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + Hx_t + \varepsilon_t = 1, 2, \ldots, T \]

(17)

Where \( y_t \) is the column vector of the d dimension exogenous variable, \( y_i \) is the column vector of the k dimension endogenous variable, \( T \) is the sample size, \( p \) is the lag order, the matrix \( \phi_1, \ldots, \phi_p \) and
matrix $H$ are the coefficient matrix to be estimated, and the $\varepsilon_t$ is the dimensionally scrambled column vector, which allows simultaneous correlation, but can not be correlated with its lag value.

3.3.2. Model stability test

The stability of the model is tested by the stability of the parameters, the stability of the model is the same structure, and the stability of the model parameters can improve the prediction and analysis function of the model. For the better use of the model for analysis and prediction, it is necessary to test whether the estimated parameters of the benchmark linear VAR model have stability. A number of methods can be used to test the stability of parameters, such as Chow structural stability test and recursive least square method.

Chow structure stability test includes two kinds: one is fracture point test, this method uses the least square method to carry on the regression, by judging whether the residuals of the two subsamples have a significant difference in statistics, and then according to the two structures to determine whether the parameters between the two subsamples are stable. another is Chow prediction failure test. The method divides the sample size into two subsamples: $T_1$ and $T_2$. After estimating the sample size by using the equation, the sample size is $T_2$. Then the difference between the actual value and the predicted value is compared. If the difference is large, the structure and the model estimation period are explained the expected structure is different. however, both methods have some defects and need to know the location of the fracture point of the model structure in advance. however, since most of the structural changes are gradual and non-outburst, in this case, the two methods of the Chow test are invalid. therefore, this paper uses cumulative and CUSUM and cumulative sum of squares CUSUMsq statistics to test the stability of the model parameters. because of the poor effectiveness of the stability of the CUSUMsq judgment coefficient, it is mainly used to test the stability of the regression residual variance. Turner think if the variance of the model parameter or error term exists at mutations, CUSUMsq tests are more effective.

3.3.3. Coal Price Forecast

The dynamic relationship between variables can be found by analyzing the pulse response function. This section mainly analyzes the response of exchange rate, economic growth, coal import and export and international coal price to domestic coal price shocks. as shown in figure 4 below, the horizontal axis represents the number of lag periods of impact action, the solid line represents the impulse response function, and the dashed line represents the deviation zone of positive and negative double standard deviation.

![Figure 4. Exchange rate, economic growth, coal exports impact on coal prices, respectively](image)
As can be seen from figures 4 and 5, given a positive impact on the exchange rate, the domestic coal price reaches the maximum negative response in the first phase, and then the negative response slowly weakens and continues until the sixth phase gradually disappears, which shows that giving the exchange rate a shock causes changes in international coal prices, affecting domestic coal demand, and then affecting changes in domestic coal prices. Give a positive shock to economic growth, domestic coal prices reach the largest positive response in the third phase, this response continues until the seventh phase, indicating that the acceleration of economic growth will cause domestic coal prices to rise, this shock has a significant promotion effect and a longer lasting effect. Export of coal a positive shock, domestic coal prices reached the maximum negative response in the first phase, followed by smaller fluctuations, and reached the maximum positive response in the third phase, but the magnitude of the shock was not very large. Give international coal prices a positive impact, domestic coal prices in the second phase reached the largest positive response, and then slowly weakened until disappeared, this impact is larger, international coal prices are affecting domestic coal prices. Less response of domestic coal prices to the impact of coal imports.

3.4. Model Improvement Based on Levenberg-Marquardt Algorithm

For improving the accuracy of model prediction and the speed of operation, this part uses another algorithm to solve the BP neural network model, which aims to reduce the residual error and reduce the number of iterations. Hence, this paper introduces Levenberg-Marquardt algorithm for model solving. An improved BP neural network model is solved according to the above steps. According to the corresponding weights and thresholds with higher accuracy, we can know that the residual error is more convergent than before, and the residual value is smaller, which indicates that the BP neural network model based on Levenberg-Marquardt algorithm in this question has higher accuracy than the BP neural network model based on genetic algorithm.

4. Conclusions

Ranking of the main factors affecting coal prices:
National economic development > coal transport capacity > production costs > national policies > coal production > import and export. According to the main factors that affect the coal price, and combined with the coal price data, according to the mapping relationship between them, each factor is predicted by the neural network model, and substituted into the model, and finally the coal price can be obtained. The BP neural network model based on Levenberg-Marquardt algorithm is more accurate than the BP neural network model based on genetic algorithm.

References
[1] Xie Shouxiang, Tan Qinghua, Song Yang. Correlation Analysis and Testing of Factors Affecting Coal Price [J].] Statistics and Decision Making, 2006, 000 (022):57-60.
[2] Yuan Guiqiu, Zhang Lingdan. Analysis on Influencing Factors of China's Coal Price [J]. Price monthly, 2009 (02):47-49.

[3] Pu Hongjiu, Wu Rongkang, Liu Caiying, et al. Study on the Mechanism of Coal Price Formation [C]// China Coal Economic Research (2005~2008).

[4] Zou Shaohui ,& Zhang Jinlock. (2010). Empirical study The empirical study on variable models of coal price in china% coal price change model in china. Journal of Coal ,035 (003),525-528.