Original Research Article

Prediction of Thermal Coal Prices in Qinhuangdao Port Based on ARIMA-AHP Model①

Abstract:Aiming at the price of commodity coal, the ARIMA-AHP combination model is constructed logically and comprehensively using factor analysis dimensionality reduction method, time series analysis and prediction method, and analytic hierarchy process decision method. This paper also uses R, PYTHON and other software programming to solve, and gives a comprehensive method for short-term accurate prediction of thermal coal prices in Qinhuangdao. The study draws conclusions: the main influencing factors of coal price, the internal structure of coal price time series, the weight ranking of uncertain influencing factors, etc.

Key Words:forecast; coal price; factor analysis; analytic hierarchy process; ARIMA model; uncertainty factors

Introduction

From 2012 till now, problems such as structural overcapacity, falling average prices, and declining efficiency have plagued the coal market industry. In order to promote the sustainable and healthy development of the coal industry, the General Office of the State Council of China has repeatedly issued emphasis on the urgent need to promote the smooth operation of the coal industry. Moreover, the report of the 19th National Congress of the Communist Party of China has repeatedly emphasized the strengthening of industry development to protect my country's traditional resources. Therefore, regardless of the macro or micro level, the reasonable inheritance and strategic structural reform of the coal resource industry are inevitable, which further invisibly forms the driving factor for all walks of life to pay attention to the price forecast of the important core economic indicators of the coal industry. As a bulk commodity, coal resources are one of the most important strategic energy sources in my country. They are the upstream industries of the basic industries of the national economy such as the power industry, building materials industry, and metallurgical industry. The price level orientation of its resource status is extremely sensitive, and the magnitude of changes directly affects the national economy. Whether the long-term development of the economy is stable or not, the research and prediction of coal prices are of extraordinary significance to the strategic formulation of relevant enterprises and the policy formulation of government departments.

Ning Hui[1],etc.Through the support vector regression machine prediction model of rolling time window, the change rule is accumulated as time goes on, so as to achieve the prediction result；Liu Bo[2],etc.Start with the coal market price data indicators in the first half of 2019, and then predict the coal price in the second half of 2019 through visual trend prediction；And Chang Lidan[3],etc.Reasonable use of artificial neural network to build a nonlinear model to train coal price fluctuation law, predict the later coal price trend and predict its model accuracy through error comparison analysis.However, none of these studies take into account the uncertainties in the actual fluctuations of the coal market, such as the unpredictable COVID-19, so that the problems are not considered comprehensively.

1 Data sources and model assumptions

The data in this paper comes from the Qinhuangdao Port thermal coal price real-time database of China Coal Information Network. In order to facilitate the research question, the following assumptio

UNDER PEER REVIEW

ns are made:(1) The unanimous assumption of the freight trains is that the internal combustion engine drives the combustion of diesel oil; (2) A month is 30 days; (3) It is assumed that the policy issued by the relevant state departments will affect the coal price with a delay and no suddenness; (4) It is assumed that the official data has been obtained. The error is small and there is no falsehood.

2 Main influencing factors of coal price based on factor analysis

2.1Research ideas

After standardizing and preprocessing the data in advance, it was imported through SPSS data processing software for factor analysis and dimension reduction processing. Then collect panel data corresponding to time series prices, including average temperature, coal substitutes (crude oil), upstream and downstream products (natural gas), industrial concentration, inventory-to-consumption ratio, cost, net international coal import and export, economic growth, national government related The department supervises these 9 main factors affecting coal prices. And use MATLAB software to get the correlation coefficient, and finally use the correlation coefficient in the model to sort the main influencing factors of thermal coal in Qinhuangdao Port.

2.2Research method

Determine the factors affecting coal prices by consulting relevant websites and papers and consulting front-line market staff. One of the first-level indicators and quantification is interpreted as:cost factor, climate change, mode of travel (month-on-month growth rate of rail passenger traffic),international coal import and export net value, domestic coal market (stock-to-consumption ratio),supervision by relevant state departments, complementary alternative resources (crude oil, natural gas production),industrial concentration($CR4$)、competition in the industry (single-digit month-on-month growth rate of coal enterprises)、economic growth (Consumer Price Increase Index).The secondary indicators and quantitative explanations are as follows:production cost, operating cost of washing, transportation cost ((average daily coal loading * saverage loading volume per vehicle * average wholesale price of diesel) * 30), humidity, temperature, total import (average import price * import volume), total export (average export price * export volume), coal inventory in key power plants (days available for power generation group inventory), demand (coal consumption in steel, electricity, fertilizer, and cement industries), environmental protection policy (membership range (0,1), satisfactory degree range (-1,1)).

After the standardized data is imported into the data processing software, it needs to undergo factor analysis to reduce dimensionality, but before that, it is necessary to perform KMO test and $Bartlett$ sphericity fitness test on the sample data in order to make a rational judgment on whether the collected data is suitable for factor analysis.Get an KMO value of 0.727 and $Bartlett-p-value$ is 0.0002 that is less than the significance level of 0.05.Therefore, the null hypothesis can be rejected at the significance level.According to the test results, the sample data in this paper meet the preconditions for principal component analysis or factor analysis.The factor loading matrix obtained from the analysis is shown in Table 1 and the factor analysis score coefficient matrix is shown in Table 2.

See Table 1,Based on the premise that the eigenvalue is greater than 1, three main components are extracted, namely the three first index categories.Based on the premise that the eigenvalue is greater than 1, three main components, namely the three major categories of the first indicators, are extracted, which are: X_2 Temperature, X_7 Supervision by relevant state departments, X_{11} Natural gas production；In the second category of ingredients are: X_1 Cost, X_2 Temperature, X_5 International coal import and export net, X_8 Industry concentration；In the third category of components, there are: X_6 Domestic coal market (stock-to-consumption ratio), X_{10} Economic growth.

Table 1 Factor Loading Matrix

UNDER PEER REVIE

	1	2	3
X1 COST	-0.167	0.781	0.363
X2 TEMPERATURE	0.788	0.575	0.102
X3 HUMIDITY	-0.773	-0.522	0.229
X4 TRAVEL MODE	-0.139	0.391	-0.590
X5 INTERNATIONAL COAL IMPORT AND EXPORT NET	-0.186	0.849	-0.002
X6 DOMESTIC COAL MARKET	0.262	-0.366	0.797
X7 SUPERVISION BY RELEVANT STATE DEPARTMENTS	0.522	-0.636	-0.315
X8 INDUSTRIAL CONCENTRATION	0.471	0.791	0.024
X9 COMPETITION IN THE INDUSTRY	-0.558	0.118	0.415
X10 ECONOMIC GROWTH	-0.204	0.173	0.671
X11 NATURAL GAS PRODUCTION	0.902	-0.259	0.216
X12 CRUDE OIL PRODUCTION	0.934	-0.133	0.207

As shown in Table 2, the linear estimation function of the mathematical model of the total sc ore can be established by combining the component score coefficient matrix and the weighting of the variance contribution rate to the cumulative contribution rate, as shown below:

$$Y = 0.47(-0.043X_1 + \ldots + 0.239X_{12}) + 0.28(0.231X_1 + \ldots - 0.039X_{12}) + 0.25(0.182X_1 + \ldots 0.104X_{12})$$

Table 2 Factor Score Coefficient Matrix (extraction method: principal component analysis)

	Element		
	1	2	3
X1 COST	-0.043	0.231	0.182
X2 TEMPERATURE	0.202	0.170	0.051
X3 HUMIDITY	-0.198	-0.155	0.115
X4 TRAVEL MODE	-0.036	0.116	-0.297
X5 INTERNATIONAL COAL IMPORT AND EXPORT NET	-0.048	0.251	-0.001
X6 DOMESTIC COAL MARKET	0.067	-0.108	0.401

UNDER PEER REVIEW

X7 SUPERVISION BY RELEVANT STATE DEPARTMENTS	0.134	-0.188	-0.159
X8 INDUSTRIAL CONCENTRATION	0.121	0.234	0.012
X9 COMPETITION IN THE INDUSTRY	-0.143	0.035	0.209
X10 ECONOMIC GROWTH	-0.052	0.051	0.337
X11 NATURAL GAS PRODUCTION	0.231	-0.077	0.108
X12 CRUDE OIL PRODUCTION	0.239	-0.039	0.104

2.3Result analysis

Finally, the main influencing factors affecting coal prices are obtained by taking the factor loa d matrix union: average temperature, coal substitutes, upstream and downstream products, industrial concentration, inventory consumption ratio, cost, net international coal import and export, economi c growth, Supervision by relevant state departments.

3 Time series structure of coal price

3.1Research ideas

The R software is used to draw the Qinhuangdao thermal coal price time series chart, and th e difference method is adopted to smoothly pass the residual analysis stationarity test, and an ARI MA model is established. Combined with the principal component linear regression model before t his paper, the forward-looking short-term range price of thermal coal in Qinhuangdao Port is predi cted. Its model establishment framework is shown in Figure 1.

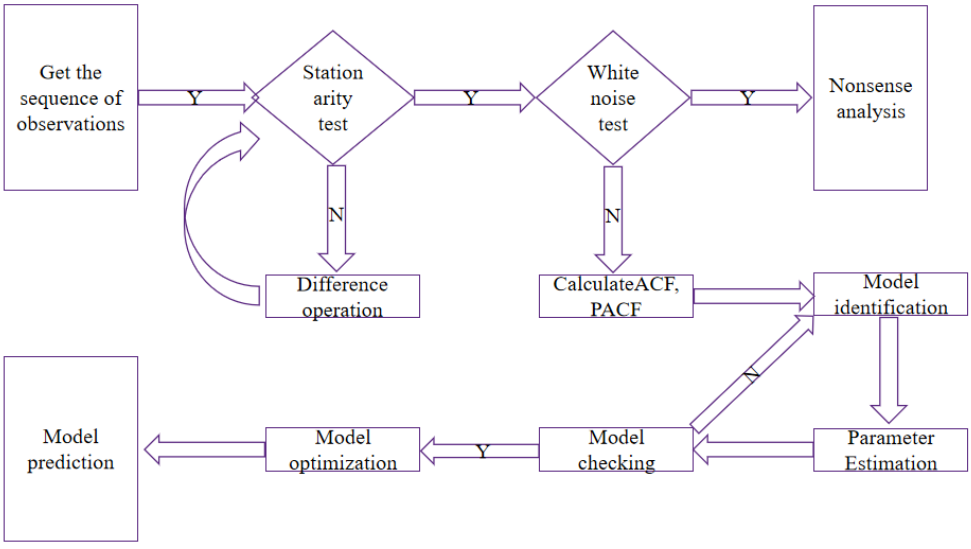

Figure 1 ARIMA Model Establishment Framework

3.2Research method

Use PYTHON to import the collected CSV time series data to draw the Qinhuangdao thermal coal weekly price time series chart to test the stability. It is observed that there is a clear down ward trend, so the difference method is adopted to obtain the expected relatively stable data, and

$d=1$, In order to further accurately check the non-white noise of the time series model, the residuals of the time series model are analyzed. The normal distribution verification diagram of the standard residual results is shown in Figure 2, and in the ADF unit root test, it can be considered that the null hypothesis is rejected at the significance level, that is, the difference time series is stationary.

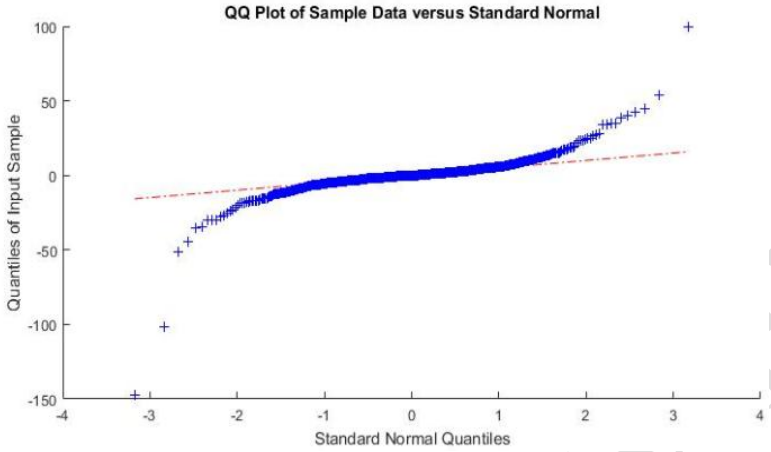

Figure 2 Residual Normal Distribution Plot

It can be seen from Figure 2 that the red line is the standard normal distribution curve, and the model data is stable and satisfies the normal distribution. After verification, it can be used for time series forecasting. The following is the identification and fitting of the ARIMA model for the coal price of Qinhuangdao Port with a time span of weeks. By observing Figure 3, it is analyzed that the coefficients of ACF and $PACF$ are stable within the marked difference after the lag period, which is a first-order tail, so $p=1,q=1$.

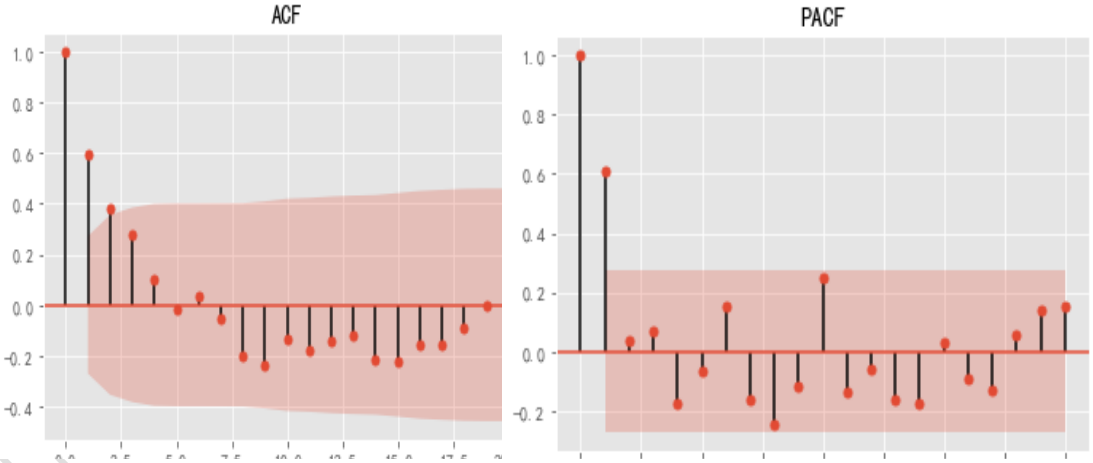

Figure 3 ACF and PACF Test

The model can be identified as ARIMA(1,1,1) as shown below. The parameters of the model were estimated using the $Arima$ function in the $forecast$ package in the R software, which passed the model $Box-Pierce$ significance test and the model parameter significance t test respectively. By comparing the possible AR(1) and MA(1) models, this model shows a small size on the AIC and BIC information criteria and has a good fit. Considering the influence of the main factors on the thermal coal of Qinhuangdao Port, the obtained multiple linear regression equation shown below was combined with the autoregressive model. The mean coefficient polynomial of the variable was obtained from the regression model and the ARIMA standard model. The improved ARIMA model looks like this:

$$y_t=\mu+\sum_{i=1}^{k}\frac{\Theta_i(B)}{\Phi_i(B)}B^iX_{it}+\varepsilon_t,\varepsilon_t=y_t-(\mu+\sum_{i=1}^{k}\frac{\Theta_i(B)}{\Phi_i(B)}B^iX_{it})$$

where is the μ constant term, k is the order, and ε_t is the regression residual. According t o y_t, the future thermal coal price of Qinhuangdao Port can be predicted.

4 Uncertainty factors affecting coal prices

4.1Research ideas

As the saying goes: "The sky is unpredictable, and people have good fortune." For these sudd en or silent hidden factors of uncertainty, it becomes unpredictable. Some uncertain factors affectin g some important secondary indicators of coal prices are sorted out by consulting the data. Using MATLAB to apply the AHP to the sample data, construct a suitable hierarchical structure model, obtain the comparative discriminant matrix and pass the consistency test, and then obtain the linear model of the influencing factors of uncertainty. Combined with the combination model above, a c omprehensive coal price prediction model is finally established.

4.2Research method

Therefore, by consulting relevant website information and communicating with front-line staff, our team has concluded that the general direction uncertainty influencing factors (quasi-measuremen t layer) corresponding to Qinhuangdao Port thermal coal price A (target layer) are: objective envir onmental factors $B1$, Market structure driving factor $B2$, subjective direct connection factor $B3$. The standard layer corresponds to their own unique plan layers: market economic fluctuations $P1$ (inflation, deflation, tax policy, etc.), major macro policy adjustments $P2$, sudden natural or man-made disasters $P3$, population mobility $P4$ (structural unemployment, Excessive generation element s, etc.), related technological breakthroughs $P5$. Establish a hierarchical structure model for the ev aluation of uncertainty influencing factors (as shown in Figure 4).

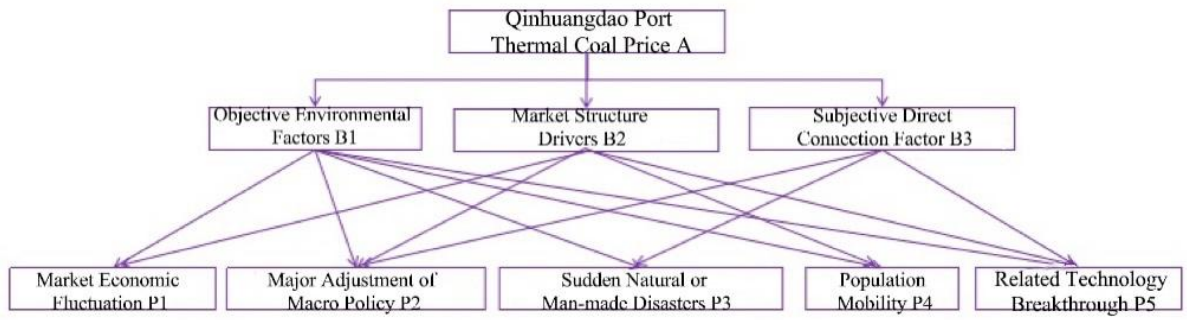

Figure 4 Hierarchical structure of uncertainty factors

When A is the comparison criterion, the mutual comparison judgment matrix of each uncertai nty influencing factor in the B layer is A, so as to transmit the next level. When B is the com parison criterion, the mutual comparison judgment matrix of the corresponding P layer under $B(i=1,2,3)$ is: $B-P(i=1,2,3)$. The team obtained 4 discriminant matrices by understanding the Q inhuangdao Port thermal coal related website and papers, communicating with the front-line work c onsultants, communicating with the local statistical bureau work consultants, and synthesizing the w eights of the three first-level indicators of the above problems, that is, one second The discriminan t matrix of the layer to the first layer, and the discriminant matrix of the three third layers to the second layer. Use MATLAB programming to solve, import the above discriminant matrix, and get the largest eigenvalue and the corresponding matching eigenvector. After standardizing it, you can get the relative importance weight vector of the corresponding level single ordering, consistency I ndicator CI, consistency ratio CR. See Table 3.

Table 3 Calculation Results of Uncertain Factors Affecting Thermal Coal Prices in Qinhuangdao Port

UNDER PEER REVIEW

matrix	Hierarchical single sorted weight vector	λ_{max}	CI	RI	CR
A-B	$(0.9161,0.1506,0.3715)^T$	3.0385	0.0193	0.58	0.0332
B_1-P	$(0.8535,0.3767,0.3210,0.0903,0.1355)^T$	5.4276	0.1069	1.12	0.0955
B_2-P	$(0.5478, 0.7984, 0.2008, 0.1492)^T$	4.1906	0.0635	0.90	0.0706
B_3-P	$(0.9161,0.1506,3715)^T$	3.0385	0.0193	0.58	0.0332

It can be seen from the observation that the CR results of the single-ranking of the four lev els are all less than 0.1, which is in line with the satisfactory consistency test.There is a sorted ve ctor of layer B relative to layer A: $W^{(2)}=(0.9161,0.1506,0.3715)^T$,And the sorting vector of the P layer when the factor B_1 is the criterion:

$$p_1^{(3)}=(0.8535,0.3767,0.3210,0.0903,0.1355)^T$$
$$p_2^{(3)}=(0.5478,0.7984,0.2008,0.1492)^T$$
$$p_3^{(3)}=(0.9161,0.1506,0.3715)^T$$

Combining the above, the sorting vector of the P layer relative to the A layer is:

$$W^{(3)}=(P_1^{(3)},P_2^{(3)},P_3^{(3)})W^{(2)}=\begin{pmatrix}0.8535 & 0.5478 & 0\\ 0.3767 & 0.7984 & 0.9161\\ 0.3210 & 0 & 0\\ 0.0903 & 0.2008 & 0.1506\\ 0.1355 & 0.1492 & 0.3715\end{pmatrix}\begin{pmatrix}0.9161\\ 0.1506\\ 0.3715\end{pmatrix}$$
$$=(0.8644,0.8057,0.2941,0.1689,0.2846)^T$$

The consistency check results are as follows:

$$CI^{(2)}=(CI_1^{(2)},CI_2^{(2)},CI_3^{(2)})=(0.0193,0.1069,0.0635,0.0193)$$
$$RI^{(2)}=(RI_1^{(2)},RI_2^{(2)},RI_3^{(2)})=(0.58,0.12,0.90,0.58)$$

therefore:

$$CI^{(3)}=CI^{(2)}W^{(2)}=(0.1069,0.0635,0.0193)(0.9161,0.1506,0.3715)^T=0.1147$$
$$RI^{(3)}=RI^{(2)}\bullet w^{(2)}=(0.12,0.90,0.58)(0.9161,0.1506,0.3715)^T=0.4609$$
$$CR^{(3)}=CI^{(2)}+CI^{(3)}/RI^{(3)}=0.0332+0.1147/0.4609=0.0282$$

Conclusion: Hierarchical total ranking also passed the consistency test. Therefore, the compreh ensive prediction model of coal price: this uncertainty influencing factor model is a multivariate li near model:

$$Y=\gamma_0+\gamma_1P_1+\gamma_2P_2\ldots+\gamma_iP_i$$

Combined with the synthetic mathematical model given by the ARIMA model, the final comp rehensive mathematical model is obtained:

$$y_t=\mu+\sum_{i=1}^{k}\frac{\theta_i(B)}{\varphi_i(B)}B^iX_{it}+\gamma_0+\gamma_1P_1+\ldots+\gamma_iP_i+\varepsilon_t$$

UNDER PEER REVIEW

4.3Result analysis

By normalizing the previous comprehensive weights, the weights of the five schemes to the target are 0.3577, 0.3334, 0.1217, 0.0695, and 0.1178. From this it is concluded that the five scenarios are ranked 1, 2, 3, 5, 4 on the target impact. The fluctuation of the market economy involves all aspects of the commodity economy, and it has a repeated cycle. Hyperinflation will directly affect the long-term, universal, and irreversible rise of prices, and the devaluation of currency. The coal price market has both the production supply chain and the consumer demand side. great hindrance. Austerity has led to a decrease in social investment, and coal enterprises are unable to produce in large quantities due to the shortage of investment ports. Although prices have fallen, it is due to the nature of human beings. People will delay consumption as much as possible, thus reducing the scale of consumption. Therefore, the market economic fluctuations that affect the uncertainty factors of coal prices have a high weight; and the major adjustment of macro policies can directly or indirectly affect the strategic development direction of coal enterprises, which has a constructive and guiding influence, so the weight is high; Uncertain factors of natural or man-made disasters will affect the external market environment more. Sudden disasters will inevitably lead to the take-off of a few industries, and the slump of most industries will have a wave effect; for related technological breakthroughs Uncertainty factors in the labor force bring more liberation of productivity and a qualitative leap in the production chain, but in the long run, its periodicity is too long; and although population mobility is indeed a reasonable factor affecting labor prices, However, with the development of related technologies and the optimization of management, the price of productivity affected by labor is almost the same in the automated production chain that prevails today.

5 Concluding remarks

Constructive suggestions can be made to relevant departments based on the price predicted by the combination model based on historical coal price data. It is suggested that enterprise risk management departments should always be vigilant against hidden uncertain fluctuation factors, and do a good job of early prediction and later avoidance; government management departments also need to mobilize macro policies in a timely manner to keep pace with development. The advantage of this paper is that it fits the real market environment and explores the impact of uncertain behavioral factors on coal prices. The established model effectively and reasonably links the internal and external mechanisms of normalized and abnormal coal prices; and the direction of mechanism visualization is also written. narration, so that information users can see the essential problems from the outside to the inside through the phenomenon. Of course, there are also some shortcomings. For the subjective comparison of experts in AHP, the weight sampling is more accidental. Due to the lack of resources, the average and reasonable judgment results of most experts are not obtained. For the establishment of the ARIMA model, due to the limitations of its own stochastic process rational spectrum, As a result, the model cannot make long-term predictions. In this paper, the ARIMA-AHP model solves the short-term price forecast of the coal market, and there are many resources in the same bulk energy status. The commonality of these resources can be attributed to the fact that they can enter the circulation market, but they are not retailed. Therefore, whether in industry, agriculture or service industries, they are all material commodities that are produced and consumed in large quantities. Therefore, this model can be used by analogy, involving other commodities. Price forecast. For example: wood, steel, minerals and other material resources. Of course, when using this model, it should be combined with the actual situation of the relevant departments, and it is necessary to satisfy the scientific nature of the determination indicators and the authenticity of the indicator weights.

UNDER PEER REVIEW

References

[1] Ning Hui, Zhou Wenwen. Research on ε-SVR Coal Price Forecast Model Based on Rolling Time Window[J].Coal Economics Research,2020,40(03):12-18.DOI:10.13202/j.cnki.cer.2020.03.003.

[2] Liu Bo, Gong Dayong. Coal Market Analysis in the First Half of 2019 and Forecast for the Second Half of the Year[J].Coal Economics Research,2019,39(07):29-33.DOI:10.13202/j.cnki.cer.20190801.005.

[3] Chang Lidan, Liang Zhijian. Research on Coal Cost Prediction Based on BP Neural Network[J]. Coal Technology,2019,38(04):180-182.DOI:10.13301/j.cnki.ct.2019.04.062.

[4] Zhu Jiaming, Duan Hanbing, Wang Zijian, Zhang Junling. Quantitative Analysis of Waste Classification Strategies in Beijing Based on ARIMA Model[J].Journal of Minzu University of China (Natural Science Edition),2020,29(02):49-56.

[5] Gao Xinwei, Yang Ao, Han Yukai. Analysis and Countermeasures of Energy Price Fluctuations during the New Crown Epidemic[J/OL].Price Theory and Practice:1-5[2022-03-08].DOI:10.19851/j.cnki.CN11-1010/F.2021.11.353.

[6] Zou Shaohui, Li Mengjia. Research on China's Coal Price Forecast under the Carbon Peak Target[J].Price Monthly,2022(02):20-27.DOI:10.14076/j.issn.1006-2025.2022.02.03.

[7] Chen Jingru. Research on Coal Price Forecast Based on Modified ARIMA Model of Intervention Items[J].Logistics Engineering and Management,2021,43(09):132-135.

[8] Gao Luhui,Wang Guoqing. Research on the Improved Neural Network of Coal Price Forecast Based on Co-integration Theory[J]. IOP Conference Series: Earth and Environmental Science,2021,769(4).

[9] Yin Zheng,Kun Yuan,Lei Dai. Research on Coal Price Forecast based on ARIMA and SVM combination Model[J]. E3S Web of Conferences,2021,257.

[10] Xiaofan Zhang,Chao Liu,Yuhang Qian. Coal Price Forecast Based on ARIMA Model[J]. Financial Forum,2020,9(4).

[11] Jiale Xu,Zhen Wang,Dongxiao Li. Coal Price Forecast Based on ARIMA-BP Combination Model[P]. 2020 International Conference on Economic Management and Corporate Governance,2020.

UNDER PEER REVIEW