Algorithm research and empirical analysis of container transportation production prosperity index in China

Weiwei Qiu[a] , Jishuang Zhu*[a,1] , Xinzi Wang[b]

[a] China Waterborne Transport Research Institute, Beijing 100088, China
[b] Research Institute of Highway Ministry of Transport, Beijing100088, China
Weiwei Qiu and Jishuang Zhu contributed equally to this paper

*Corresponding author: Zhujishuang2004@163.com

Abstract. Shipping is an important channel for the international trade transport goods and one of the basic industries of the national economy. With the gradual transformation of China's water transport industry to high-quality development, the development goals of the port container transport need to be guided by more scientific and accurate indicators. Based on AIS data and big data methods, this research uses the Vessels arriving at port, ship type and tonnage as the measurement indicators of port container production prosperity, constructing the container transportation production prosperity index (CTPPI), and use time difference correlation and vector auto-regressive mode to do empirical analysis. The results will provide quantitatively support for the industry authorities to study and judge the development form of coastal ports.

Keywords: container transportation production, shipping, AIS, vector auto-regression model, index, time difference correlation analysis method

1 Introduction

Shipping is an important channel for the international trade transport goods and one of the basic industries of the national economy. As a barometer of national economy, port can reflect the development trend of society, economy and trade to a certain extent. With the rapid development of macro-economy, the fluctuation of port production activities will not only affect the national strategy, but also have a direct impact on the industrial chain and supply chain. Among them, port container transport with high transport efficiency, high degree of cooperation, high economic benefits in China's port enterprises and logistics industry have been rapid development. According to statistics, the national port container throughput totaled 295.87 million TEU in 2022, with 4.7% growth rate. With the gradual transformation of China's water transport industry to high-quality development, the development goals of the port container transport need to be guided by more scientific and accurate indicators. However, most of the indicators reflect the prosperity of the port container industry through turnover and price. Due to the relevant regulations on pricing, the price factor is difficult to reflect the prosperity of the port to a certain extent, and the port operation monitoring is based on traditional

Y. Jiao et al. (eds.), *Proceedings of the 3rd International Conference on Internet Finance and Digital Economy (ICIFDE 2023)*, Atlantis Highlights in Economics, Business and Management 1,
https://doi.org/10.2991/978-94-6463-270-5_19

statistical data, mostly based on the statistical results of the port. The statistical data is not time-sensitive and is mainly based on scale indicators such as throughput. Therefore, the establishment of reasonable, effective and timely index to realize the dynamic tracking of China's port and the evaluation of economic fluctuations, to assist the analysis of macroeconomic operation gradually become the focus of research.

Based on AIS data and big data methods, this research uses the Vessels arriving at port, ship type and tonnage as the measurement indicators of port container production prosperity, constructing the container transportation production prosperity index (CTPPI). Cargo throughput, import and export amount, GDP, and added value of the primary industry are selected to do the empirical study and the fluctuation analysis of influencing factors by using the time difference correlation and vector auto-regressive mode. The research provides quantitatively support for the industry authorities to study and judge the development form of coastal ports.

2 Literature review

2.1 Port container index research

Nowadays, studies on port container index mainly include Container Port Performance Index (CCPI), container port comprehensive service time efficiency measured by Shanghai Shipping Exchange, coastal container port comprehensive service evaluation Index, and Port Congestion index Index), China Export Container Freight Index (CCFI), etc. in the industry. It can reflect the prosperity of port production to a certain extent, but there are also shortcomings. For example, CPPI tends to use rankings of different port areas, and the efficiency level cannot be compared uniformly from port caliber, and the setting of weights may lead to large differences in rankings. Due to the difference in the number of containers loaded and unloaded at the port of departure, transit port and destination port, it is difficult to compare the service efficiency of container ships between different ports.

2.2 Analysis of influencing factors related research

The fluctuation analysis of index influencing factors mainly uses the correlation model to analyze index and macroeconomic indicators. Wang (2016) analyzed the influencing factors of coal price fluctuation in China from macro, industry, international market and micro factors by VAR model[9]. Tsioumas et al. (2017) ananlyzed a multi-variate VAR model includes steel production, dry bulk fleet and dry bulk ecnomoic index to predict BDI, which has higher accuracy than ARIMA[11]. Wang (2018) adopted the methods of Johannes test, VAR model and pulse response, and found that the current price was balanced in the long run, and the feedback of futures prices to new information was timelier and more flexible in the short run, but the price fluctuation of futures was relatively large, so the suggestion is to adjust the product structure of iron ore futures[3]. Jafari (2018) used ARIMA model to predict BDTI in the past 5 years and found that it is less value for weekly data prediction[12]. Ye (2020) combined manufacturing PMI and CCFI and found that there was a one-way long-term co-integration relationship between PMI and CCFI with a lag time of about half a month, while CCFI had little impact on PMI[4]. Mao and Wang (2020) found that import and export value, BDI, oil price and economic situation has impact on CCFI[5]. Zhang (2020) explored freight volume fluctuation by using X-12-ARIMA and HP Filtering decomposition method, and analyzed the freight price elasticity of different types of goods based on co-integration theory[10]. Wang (2021) use VAR model to study the external influencing factors of WTI crude oil price, Brent crude oil price and other variables on the container freight rate index of the Yangtze River, and further found that WTI and PMI had stronger explanatory ability to the index, and found that the internal influence relationship of the container of the Yangtze River showed the influence sequence from upstream to midstream and then to downstream[6]. Fan (2021) explore the equilibrium relationship between China iron ore futures and spot, established the VAR model, and found that the influence of iron ore futures prices on domestic spot and Platts index prices was stronger than that of domestic and foreign spot prices on China's futures through pulse impact analysis[7]. Chen (2022) explained the main influencing factors of CCFI through the coefficient estimation of multiple linear regression equation, and found that container transportation was the shipping capacity provided by the market and the Shanghai Composite Index had a negative impact on CCFI, while container export value and WTI crude oil price had a significant positive effect on CCFI[8].

3 CTPPI calculation method

In order to break through the limitation of freight rate index and reflect the supply and demand state of the entire water transportation industry more comprehensively and systematically, this paper adopts the core indicators of production scale of the port, including the arrival of ships, tonnage and other core indicators, and establishes the container transportation production prosperity index based on AIS data, which can more sensitively reflect the changes in the container production scale of China's coastal ports.

3.1 Data source

The ship operation data used in this paper is mainly from Automatic identification System (AIS), which is mainly received by shore-based equipment and satellite equipment. Inland ships and some domestic trade ships are not loaded with the equipment, or the use is not standardized, resulting in the loss ship data. Therefore, container ships of 1000TEU and above are selected as sample ship types, and different weights are set according to the differences in operation efficiency and operation volume of ships in different tonnage in ports.

3.2 CTPPI calculation method

According to the AIS data and the ship static database, the ship type and tonnage information of each arriving ship can be obtained. If, the number of container ships with s tonnage calling at Port i in the n time period is S_{in}^{s}, then the weighted value of the CTPPI T_n in the n time period is:

$$T_n = \sum_{i=1}^{23} \sum_s S_{in}^{s} \quad * E_s \tag{1}$$

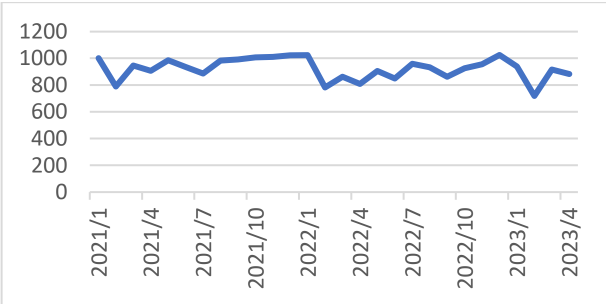
Among, E_s represents the weight of the tonnage. Since the operation efficiency and operation volume of ships with different tonnage in the port are quite different, different weights are set respectively according to the operation efficiency of ships with different tonnage of different cargo classes and weighted to obtain the index.

Let T_0 as the base period is 100 points in January 2021, CTPPI (C_n) in the n time period is calculated as follows:

$$C_n = 100 * \frac{T_n}{T_0} \tag{2}$$

3.3 CTPPI results

Overall, CTPPT fluctuates greatly. The index fluctuated at the beginning of 2021 due to COVID-19 and returned to the level in the second half of the year. With the gradual control of the COVID-19, the CTPPI began to rise after May, and by July it had risen to about 950 points as shown in Figure 1.



Source: Authors

Fig. 1. CTPPI results.

4 Empirical analysis

4.1 Data source

The data published by the National Bureau of Statistics are mainly used for indicators such as cargo throughput, import and export value, GDP and the value-added of the primary industry[1-2]. All data is selected from January 2021 to March 2023, a total of 27 monthly data.

4.2 Time difference correlation analysis method

The time difference correlation analysis method is a method to analyze the correlation between indicators in economic analysis. Generally, a correlation coefficient between 0.8-1 indicates a strong correlation between the two, 0.6-0.7 indicates a strong correlation between the two, and 0.4-0.6 indicates a moderate correlation between the two.

Table 1. the correlation between variables

Factors	CTPPI
cargo throughput (Ten thousand tons)	0.7
import and export value (Hundred million yuan)	0.6
GDP	0.5
the value-added of the primary industry	0.7

Source: Authors

From the Table 1, it can be seen that the cargo throughput and the value-added of the primary industry have strong correlation with CTPPI, and GDP has moderate correlation. From the view of lag period, CTPPI is ahead of the GDP, the added value of the first industry, cargo throughput and China's import and export value are synchronized.

4.3 VAR model development

In order to eliminate the possible heterosexuality problem of time series variables, it is necessary to logarithmically transform the variables indicators. Let CTPPI as LNY, cargo throughput as LNX1, China imports and export value LNX2, GDP as LNX3, the added value of the first industry index as LNX4.

4.3.1 Stationarity test.

ADF unit root test is adopted. When the unit root test statistic value is less than the critical value under the significance level of 5%, the null hypothesis can be rejected, and there is no unit root in the original sequence. Otherwise, it is considered that the original sequence is not stable, and differential processing is needed to ensure the stationarity. The specific test results are shown in Table 2.

Table 2. Stationarity test results

Variables	Difference	t-statistic	Prob.	1% significant level	5% significant level	10% significant level	Results
LNY	0	-4.548	0.0062	-4.339	-3.587	-3.229	stationary
LNX_1	0	-4.944	0.0005	-3.711	-2.981	-2.630	stationary
LNX_2	0	-3.841	0.0296	-4.339	-3.587	-3.229	stationary

LNX_3	0	-3.403	0.075	-4.394	-3.612	-3.243	Not stable
	1	-4.978	0.0026	-4.374	-3.603	-3.238	stationary
LNX_4	0	-2.274	0.1872	-3.711	-2.981	-2.630	Not stable
	1	-4.913	0.003	-4.374	-3.603	-3.238	stationary

Source: Authors

It can be seen from Table 2 that the P-values of the original sequence of CTPPI, cargo throughput, and China import and export value are all less than 0.05, indicating that the null hypothesis is rejected at 95% confidence level and the sequence is stable. In the case of first-order difference, the P-value of GDP and value-added of primary industry is less than 0.01, indicating that the null hypothesis is rejected at 99% confidence level, and the series is stable at this time, which can be further analyzed by VAR model.

4.3.2 Cointegration test.

If the original sequence variable has no unit root, the VAR model can be carried out directly. If the original sequence variable has a unit root and satisfies the homogeneity of order, the VAR model can be built only after the system has long-term stability. In this research, the Johansen test method is adopted, and the variables are tested based on trace statistics and maximum eigenvalue statistics. The output results are shown in Table 3.

Table 3. Johansen test results

Unrestricted Cointegration Rank Test (Trace)				
Hypothesized No. of CE(s)	**Eigenvalue**	**Trace Statistic**	**0.05 Critical Value**	**Prob.** Critical Value**
None *	0.839	82.977	69.819	0.003
At most 1	0.416	37.323	47.856	0.332
At most 2	0.389	23.858	29.797	0.207
At most 3	0.284	11.522	15.494	0.181
At most 4	0.119	3.1672	3.8415	0.075
Unrestricted Cointegration Rank Test (Max-eigenvalue)				
Hypothesized No. of CE(s)	**Eigenvalue**	**Max-Eigen Statistic**	**0.05 Critical Value**	**Prob.** Critical Value**
None *	0.839	45.653	33.877	0.001
At most 1	0.416	13.466	27.584	0.858
At most 2	0.389	12.336	21.132	0.515
At most 3	0.284	8.3549	14.265	0.344
At most 4	0.119	3.167	3.841	0.075

Source: Authors

According to trace statistics and Max-eigenvalue, when the null hypothesis is "at most one", p=0.3324>0.05, the null hypothesis is accepted. It indicates that there is at most one co-integration relationship, and the relationship between variables is stable for a long time.

4.3.3 VAR model development.

Vector autoregressive model (VAR) uses the current variables in the model to regression the lagging variables, and establishes the model by taking each exogenous variable as a function of the lagging value of the endogenous variable. The model is used to estimate the dynamic relationship between endogenous variables. The equation to build the model are as follows.

VAR model contains n endogenous variables, each endogenous variable regressed several lag values of itself and other endogenous variables. When the lag order is assumed to be m, the VAR model is generally expressed as follows.

$$Y_t = \sum_{i=1}^{m} A_i Y_{t-1} + B_t \tag{3}$$

Where, represents the dimensional column vector composed of the t phase observations. is matrix of series. is a dimensional column vector composed of random error terms, and the random error term satisfies the white noise process.

4.3.3.1 The selection of the optimal order.

Before constructing the VAR model, the optimal lag order of the model should be determined. Generally, the methods to determine the best lag order mainly include LR test, FPE, AIC, SC and HQ. By using this theoretical method, the output of the optimal lag order is obtained as shown in Table 4.

Table 4. Optional lag order of VAE model

Lag	LogL	LR	FPE	AIC	SC	HQ
0	142.4051	NA	7.33e-12	-11.45042	-11.20499	-11.38531
1	199.2832	85.31726	5.43e-13	-14.10694	-12.63437	-13.71626
2	223.3590	26.08210	8.03e-13	-14.02992	-11.33021	-13.31368
3	287.8461	42.99137*	8.22e-14*	-17.32050*	-13.39366*	-16.27871*

Source: Authors

The results show that the lag order of the 5 criteria is all 3, and the optimal lag order of the VAR model can be determined to be 3. Therefore, VAR (3) model is established according to the lag order.

CTPPI and cargo throughput model expression:

$$\begin{pmatrix} LnY \\ Lnx1 \end{pmatrix} = \begin{pmatrix} 0.067 & -0.449 \\ 0.062 & -0.301 \end{pmatrix} \begin{pmatrix} LnY_{t-3} \\ Lnx1_{t-3} \end{pmatrix} + \begin{pmatrix} 10.628 \\ 16.968 \end{pmatrix} \tag{4}$$

CTPPI and China import and export value expression:

$$\begin{pmatrix} LnY \\ Lnx2 \end{pmatrix} = \begin{pmatrix} 0.221 & -0.395 \\ -0.128 & 0.026 \end{pmatrix} \begin{pmatrix} LnY_{t-3} \\ Lnx2_{t-3} \end{pmatrix} + \begin{pmatrix} 5.234 \\ 10.185 \end{pmatrix} \tag{5}$$

CTPPI and GDP expression:

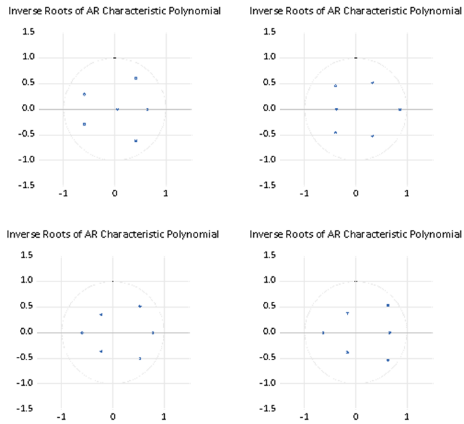
$$\begin{pmatrix} LnY \\ Lnx3 \end{pmatrix} = \begin{pmatrix} 0.147 & -0.198 \\ 0.022 & -0.348 \end{pmatrix} \begin{pmatrix} LnY_{t-3} \\ Lnx3_{t-3} \end{pmatrix} + \begin{pmatrix} 0.839 \\ 7.889 \end{pmatrix} \tag{6}$$

CTPPI and the added value of the first industry expression:

$$\begin{pmatrix} LnY \\ Lnx4 \end{pmatrix} = \begin{pmatrix} 0.062 & 0.031 \\ 0.392 & -0.581 \end{pmatrix} \begin{pmatrix} LnY_{t-3} \\ Lnx4_{t-3} \end{pmatrix} + \begin{pmatrix} 2.671 \\ 13.593 \end{pmatrix} \tag{7}$$

4.3.3.2 AR Roots text.

Generally, it is necessary to check the stability of the VAR model. If the reciprocal values of all roots fall within the unit circle, the VAR model is in a stable state. The test results show that the VAR model is stable and effective as shown in Figure 2.

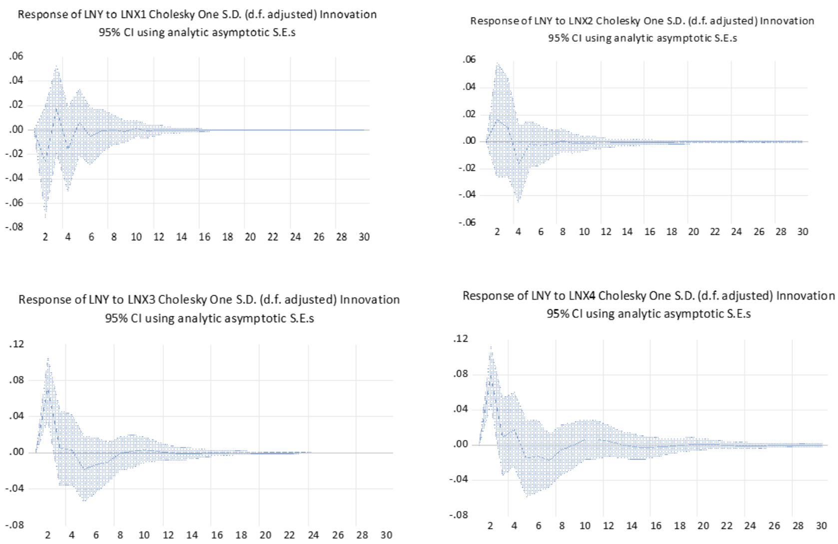


Source: Authors

Fig. 2. VAR (3) Modal AR root diagram of CTPPI and variables (cargo throughput, China import and export value, GDP, value-added of the primary industry)

4.3.4 Impulse response.

Impulse response function describes the impact of one of the most endogenous changes in the VAR model on other endogenous variables, and is often used to reflect the impact effect of policies or uncertainties on the economy or other aspects. In order to further analyze the corresponding direction and magnitude of the index unit shock and the prosperity index of container transportation production, pulse response analysis and function diagram are adopted, and Cholesky decomposition was selected to identify the impulse response function of variable. The results were shown in the Figure 3.



Source: Authors

Fig. 3. Impulse response of port container transport to Lnx_1, Lnx_2, Lnx_3 and Lnx_4

As shown in the Figure 3, when a unit of lateral impact is given to cargo throughput, the CTPPI response turns to positive after negative response, and finally becomes stable in the 8th period. It indicates that the increase of cargo throughput will bring short-term fluctuations to the index, and the influence will gradually weaken and become stable in the later period. It has a positive impact on the amount of China's import and export. After a positive response, CTPPI has a negative response in a short period of time, and finally becomes stable in the 10th period. The GDP and the added value of the primary industry had a positive impact respectively. After the positive response, the index showed a negative response in the short term, and finally stabilized in the 12th and 16th periods respectively. It indicates that the changes of these indicators will bring liquidity to CTPPI, which is beneficial to the market's healthy development.

4.3.5 Variance Decomposition.

The basic idea of variance decomposition is that after a variable is affected by exogenous variables, the impact degree of different variables can be quantified in the form of predicted variance percentage, and the most important variable for the corresponding variable can be obtained. Then, the most important influencing factor can be used to take corresponding measures to affect the corresponding variable, so as to achieve the purpose of research.

Table 5. Variance Decomposition

Variance Decomposition of LNY:						
Period	S.E.	LNY	LNX1	LNX2	LNX3	LNX4
1	0.042300	100.0000	0.000000	0.000000	0.000000	0.000000
2	0.087834	24.11894	65.22619	1.325325	8.394552	0.934986
3	0.097502	34.65243	56.42916	1.130955	7.020398	0.767058
4	0.104990	33.79917	54.37210	1.556291	6.929256	3.343181
5	0.124654	26.94743	45.94670	10.94920	12.61389	3.542781
6	0.125058	26.81418	45.69448	11.42051	12.53602	3.534807
7	0.130055	27.11641	45.24269	11.73880	12.21165	3.690445
8	0.137496	24.65310	48.26586	11.49768	11.58100	4.002351
9	0.139713	23.97769	47.55851	12.98070	11.21637	4.266727
10	0.146062	23.77281	44.57538	17.43735	10.26455	3.949911
11	0.149593	22.75257	42.54034	20.89729	10.03840	3.771393
12	0.150772	22.39791	42.34788	21.37283	9.944080	3.937297
13	0.154335	21.46188	44.14695	20.40175	9.969945	4.019473
14	0.160968	20.43672	46.22566	19.76794	9.355829	4.213854
15	0.166032	19.93792	45.79725	21.12963	8.797025	4.338172
16	0.170198	19.69712	44.27781	23.27736	8.439644	4.308062
17	0.173315	19.25044	43.09194	24.89607	8.604345	4.157201
18	0.176504	18.59977	43.75623	24.59167	8.958334	4.094000
19	0.181747	17.56200	45.95698	23.20027	9.114627	4.166122

Source: Authors

In the Table 5, the CTPPI is only affected by itself in the first period, and the contribution rate of other variables was 0. From the second period, the impact remained at about 25%. From the 2nd period, CTPPI is mainly affected by the cargo throughput, and the contribution rate can reach 65.23%, but with the increase of the number of periods, the contribution rate gradually decreases and remains at about 45% after the 5th period. The contribution rate of China's import and export value to the CTPPI ranks second, and since the 11th period, the contribution rate has gradually increased and maintained at about 20%. The contribution rate of GDP to CTPPI ranks third, and since the 15th period, the contribution rate has remained at about 8%. The contribution rate of the added value of the primary industry to CTPPI is relatively low, and the contribution rate has remained at about 4% since the 8th period. Overall, cargo volume fluctuation could give better explanation to the change of CTPPI.

5 Conclusion

In summary, through the times series correlation analysis, it is concluded that the cargo throughput, the added value of the primary industry, etc. are strongly correlated with CTPPI. While the import and export value and GDP are moderately correlated with the index. Through the co-integration test analysis, it is concluded that there is a long-term stable and mutually reinforcing relationship between cargo throughput, China's import and export amount, GDP and the added value of the primary industry. Through the VAR model, impulse response analysis and variance analysis, it can be seen that when CTPPI is taken as the corresponding variable, the cargo throughput and the import and export amount of China have more influence on it. Compared with GDP and the added value of the primary industry, it is more suitable to be taken as the explanatory variable of CTPPI.

CTPPI can provide a foundation for other researches. In the further study, the index calculation method could be improved by Laspeyres chain index method, which is internationally adopted such as CPI, PPI, etc. Besides, the base period should be rotated than fixed, otherwise it will lead to greater volatility in the index. Therefore, these aspects should be considered for the improvement.

References

1. National Bureau of Statistics (2023). National Data. Available at: http://www.stats.gov.cn/sj/. (Accessed: 20 May 2023).
2. Ministry of Transport of the People's Republic of China. (2023). Port cargo and container throughput. Available at: https://xxgk.mot.gov.cn/2020/jigou/zhghs/202205/t20220524_3656659.html (Accessed:12 May 2023).
3. Wang, Y. and He, H.D. Multifractal study of Baltic Crude oil Freight Index and Sino-US Oil Index under the influence of financial crisis [J]. *Journal of Dalian Maritime University*, ***44*** (2018):104-112.
4. Ye, S. C. Research on the correlation between China's Export container Freight Index and manufacturing PMI [J]. *Price Theory and Practice*, ***09*** (2019), 79-82.
5. Mao. H.D. and Wang A.R. Study on the influence of economic situation on container freight rate in China after 2020[J]. *Pearl River transport* ***24*** (2022), 72-74.
6. Wang, Y.F. Research on the Influencing Factors Analysis and Forecast on the Yangtze River Container Freight Index[D]. Nanjing Audit. (2021).
7. Fan, M.L. The co-integration relationship between iron ore futures and current prices. [D]. Shanghai University of Finance and Economics. (2016).
8. Chen, P.K. Analysis of influencing factors of China's export container freight rate index[J]. *Cooperative economy and technology* 6 (2022), 74-77.
9. Wang. W.Q. Study on influencing factors of coal price fluctuation in China under the new normal: Cointegration analysis based on VAR model [J]. *Economic Forum 550(05)*, 83-88.
10. Zhang, S.S. Research on the price elasticity of railway freight transportation [D]. Southwest Jiaotong University. (2020).
11. Tsioumas V, Papadimitriou S, Smirlis Y, et al. A novel approach to forecasting the bulk freight market[J]. The Asian Journal of Shipping and Logistics, 2017,33(1):33-41.
12. Jafari H, Rahimi G. Forecasting dirty tanker greight rate index by using stochastic differential equations[J]. International Journal of Financial Engineering, 2018, 05(04):637-659.