International Core Journal of Engineering Volume 6 Issue 3, 2020
ISSN: 2414-1895 DOI: 10.6919/ICJE.202003_6(3).0024

Short-term Forecast of Tubarao-Ningbo Iron Ore Freight Rate Based on ARMA Model

Jiawei Wang[1, a], Ming Yin[1, 2, b]

[1]College of Transport and Communications, Shanghai Maritime University, Shanghai 201306, China;

[2]Shanghai International Shipping Institute, Shanghai Maritime University, Shanghai 200082, China.

[a] jwellswang@sina.com, [b]yinm@shmtu.edu.cn

Abstract

Impact of iron ore freight rate fluctuation on mining industry, steel enterprises and ship chartering are important. Using ARIMA and taking the iron ore freight rate from port Tubarao to port Ningbo as an example, with the help of Eviews 10, the historical time series data from January 2005 to October 2019 was preprocessed, modeled and tested, then obtain the ARMA (1,1), a freight rate model with good fitting effect. And the iron ore freight rate of the route was predicted in a short term, the results shows that half year rate will show a slow growth trend in the range of $21.29-23.35 per ton.

Keywords

Forecast; ARIMA; Iron ore; Freight rate.

1. Research Background

With the change of supply and demand in iron ore market, steel market and ship chartering market, the freight rate of iron ore also changes. However, freight rates are of great significance to these enterprises, which may affect their short-term development plans or even their medium-term market strategies. So, it's necessary to forecast freight rates. Generally speaking, it can be predicted from two dimensions of influencing factors or time series. There are many factors influencing the freight rate of iron ore. In addition to the changes in supply and demand in the three typical markets mentioned above, there are also national reserve policies and international trade environment, and even random factors such as the mining accident and fire in the world's three iron ore giants in 2019, as well as bad weather and strikes, this is a very complicated system. In addition, although the freight rate fluctuates in the short term, it is generally stable in a long time, thus, faced the uncertain supply and demand, it is meaningful to adopt the time series method and make short-term forecast, so ARIMA model was considered.

2. Introduction

2.1 ARIMA

ARIMA model is the abbreviation of autoregressive integrated moving average model. It is a common and high precision short-term time series prediction method. It was put forward by George Box and Gwilym Jenkins in the 1970s. It is also called box Jenkins model. The main idea is to arrange the historical data of the predicted target into time series according to the order of time, and to form a group of time-dependent random variables. Although the single value of the sequence is uncertain, the change of the whole sequence has certain regularity and thus the future value of the target can be predicted by using this law which can approximately express the change with time.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202003_6(3).0024

There are many researches on ARIMA model. Zhu and Wei applied ARIMA and SVM to predict the price of carbon dioxide. It is not only the prediction in the economic field, but also used in many fields such as medicine [2], stock market [3], meteorology [4], agriculture [5], etc. In recent years, in order to get a more accurate result or solve some problems more appropriately, scholars gradually combine the new technology or other disciplines with the traditional time series model, such as BP neural network [6]. In addition, some people apply the idea of signal analysis in ARIMA.

2.2 Modeling Principle

ARIMA consists of AR, I and MA, which represent auto expression, integration and moving average. It can be seen that ARIMA model is actually a combination of AR model and MA model.

The formula of AR (p) is

$$x_t = c + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \cdots + \alpha_p x_{t-p} + \varepsilon_t, \tag{1}$$

$x_{.}$ is the value of freight rate at each time, c is a constant, $\alpha_{.}$ is the coefficient of each lag, p is the lag of autoregression, ε_t is a white noise sequence with mean value of zero and variance of constant.

The formula of MA (q) is

$$x_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}, \tag{2}$$

I is the lag of the moving average, and $\theta_{.}$ is also the coefficient of each lag. The rest is the same as above.

The formula of ARMA (p, q) is

$$x_t = c + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \cdots + \alpha_p x_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}. \tag{3}$$

ARIMA and ARMA are essentially the same that both of them use present value and historical value to predict the value of future, meanwhile eliminate the accumulation of random errors in the prediction process. The difference is whether the original time series is stable or not. ARMA model requires the sequence to be stable, so it is necessary to do differential processing for the unstable time sequence to convert it into a stable one, and then use ARMA to model. Therefore, the model established according to the unstable time sequence is ARIMA (p, d, q), d represents the number of differences.

2.3 Stationarity Test

When applying time series models, the time series are required to be stable. Therefore, the first step is to test the stability of the time series. In addition to the intuitive feeling of the eyes, the more commonly used and strict statistical test method is ADF test, that is, to judge whether there is unit root in the series. The original hypothesis of ADF test is that there is a unit root. Generally, if the value of obtained statistic is less than the critical value of the *5%* confidence level, then there is no unit root corresponding to *95%* confidence to reject the original hypothesis, which means that the sequence is stable.

2.4 Establish the Model

The key of ARIMA modeling is to determine the lags, that is, to determine p and q. Generally, the possible values of p and q are initially determined by observing the tailing and truncation features of *PACF* and ACF obtained by correlation test, and then the model is basically determined by comparing the model features under various parameter combinations, including *AIC*, *SC*, *BIC* and the *Adjusted* R^2. Finally, in order to ensure that information is extracted as fully as possible, it is also required that the residual sequence passes the correlation test, and the modeling of ARIMA is not finished until this step.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202003_6(3).0024

3. Example Analysis

3.1 Data Presentation

The historical data of iron ore shipping freight rate of the Tubarao-Ningbo route from January 2005 to October 2019 were collected and a complete time series was formed. The data came from *Xiben New Line Database*.

3.2 Data Preprocessing

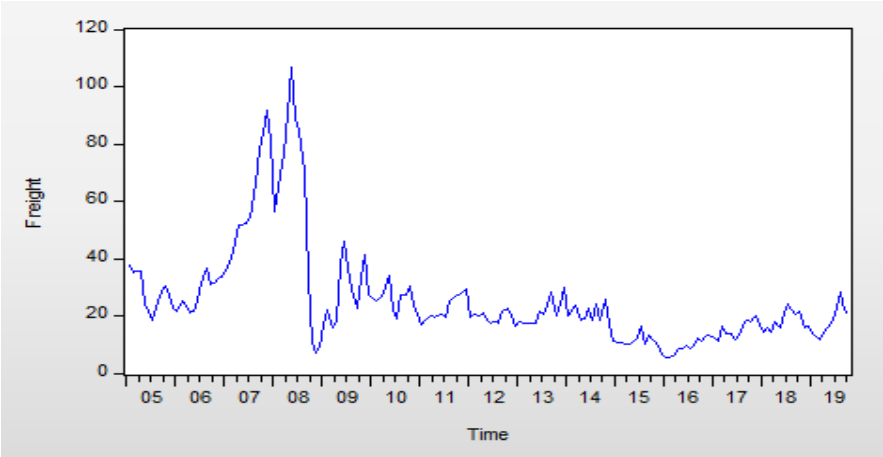

Fig. 1 Time series plot

It can be known from the *Fig.1* made from the time series data that the original series does not have seasonality, and the mean value is relatively stable, so it can be preliminarily considered that the series is stable. For the sake of caution, it is essential to judge its stability, then used Eviews to carry out ADF test. *Fig. 2* shows that the value of t-statistic is *-3.139*, which is less than the critical value of *-2.878* at the level of *5%*, and the value of prob.* is less than *0.05*, so the time series can be considered as a stationary one, so there is no need for difference.

Lag Length: 1 (Automatic - based on SIC, maxlag=14)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-3.148531	0.0249
Test critical values:	1% level	-3.467633	
	5% level	-2.877823	
	10% level	-2.575530	

Fig. 2 ADF test result

3.3 Modeling

3.3.1 Model Recognition

The autocorrelation test of the sequence shows that the original sequence is not a white noise sequence, which shows that there is a law of extractable information in the sequence, thus the original sequence is a stable and non-white noise sequence, which is suitable for building ARMA model.

According to *Fig. 3*, the *ACF* of the time series is tailed. After the first step, the *PACF* rapidly drops to the vicinity of the double standard deviation, and then falls to within the borderline from the second step, but the fifth step slightly exceeds it and then steadily falls back. According to the general experience and caution, it is reasonable that the possible value of *p* may be [*1, 2, 5*], and the *q* may be *0* or *1*, thus the model may be AR (*1*), ARMA (*1, 1*), AR (*2*), ARMA (*2, 1*), AR (*5*) and ARMA (*5, 1*).

ISSN: 2414-1895 DOI: 10.6919/ICJE.202003_6(3).0024

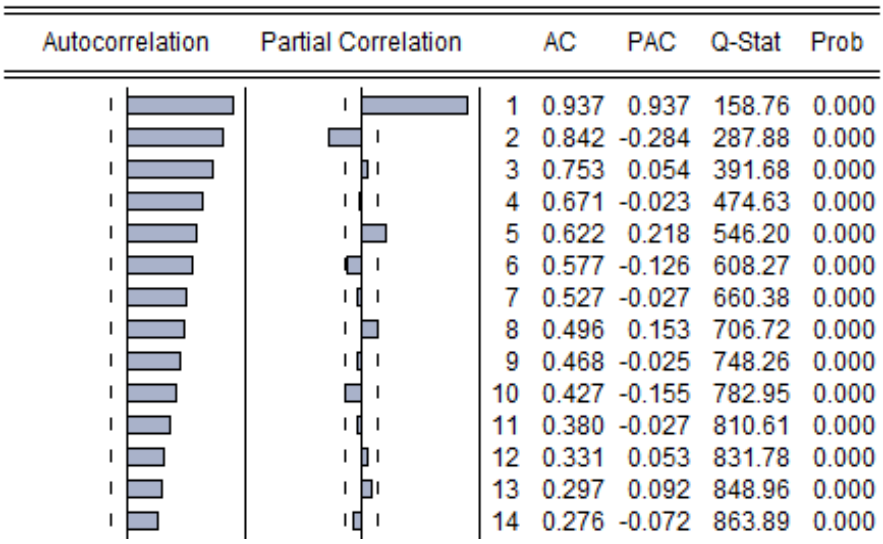

Fig. 3 Correlation test result

In addition, it can be proved according to *Fig. 2* that the series can be considered stable at *95%* confidence level, but it can be considered unstable at *1%* confidence level. The effect of ARIMA established after one difference is not as good as the ARMA (*1, 1)*.

Table 1 Evaluation indicators of alternative models

Model	Adjusted R^2	AIC	SC
(1, 0)	0.878	6.565	6.618
(1, 1)	0.889	6.48	6.552
(2, 0)	0.709	7.436	7.49
(2, 1)	0.88	6.554	6.625
(5, 0)	0.383	8.188	8.242
(5, 1)	0.745	7.312	7.383

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	26.15304	7.423438	3.523036	0.0005
AR(1)	0.895985	0.022210	40.34179	0.0000
MA(1)	0.345391	0.046588	7.413764	0.0000
SIGMASQ	36.03658	2.392002	15.06545	0.0000

R-squared	0.890609	Mean dependent var	25.82045
Adjusted R-squared	0.888723	S.D. dependent var	18.20137
S.E. of regression	6.071656	Akaike info criterion	6.480219
Sum squared resid	6414.511	Schwarz criterion	6.551720
Log likelihood	-572.7395	Hannan-Quinn criter.	6.509215
F-statistic	472.2077	Durbin-Watson stat	2.005537
Prob(F-statistic)	0.000000		

Fig. 4 The modeling results

According to the adjusted R^2 and information criterions, and combined with the habit of model simplification as much as possible, it can be seen from *Table 1* that ARMA (*1, 1*) has the best fitting effect and *AIC* and *SC* are the smallest among the alternative models. In addition, according to *Fig. 4*, the *DW* statistic of the model is very close to *2*, and the probability of *F statistic* is almost *zero*, which meets the accuracy requirements. Therefore, it can be determined that this example is suitable for the ARMA (*1, 1*) model. Therefore, the formula of the model was obtained:

$$Freight_t = 26.15304 + 0.895985Freight_{t-1} + \varepsilon_t + 0.345391\varepsilon_{t-1} \quad (4)$$

ISSN: 2414-1895 DOI: 10.6919/ICJE.202003_6(3).0024

3.3.2 Residual Test

The residual test of ARMA (*1*, *1*) was carried out and the test result in *Fig. 5* shows that most of the lags are within the double standard deviation, and the residual is a white noise sequence, indicating that ARMA (*1*, *1*) model has fully extracted the data information.

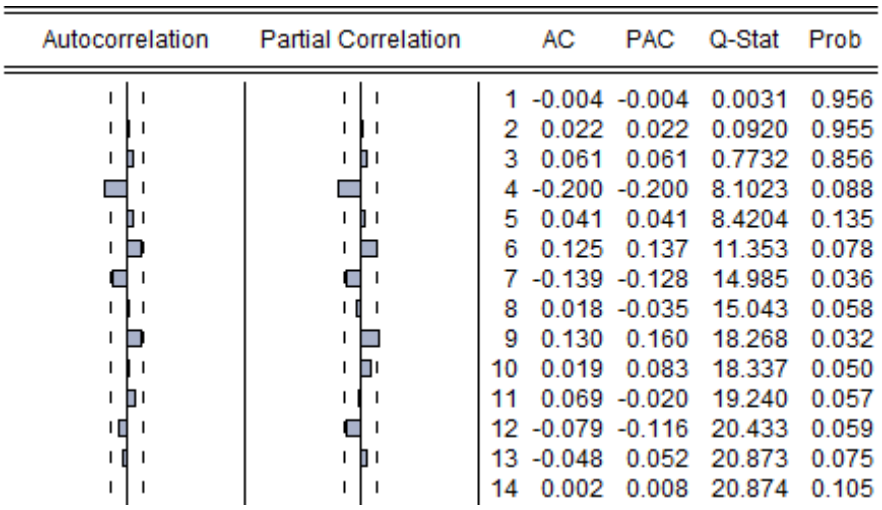

Fig. 5 Residual correlation test

3.3.3 Hypothesis Test

Table 2 shows that the three common statistical error index values of ARMA (*1*, *1*) are very small which means the model could hold error to a minimum, and according to *Fig. 6*, the two curves representing the actual observation value and the predicted value are basically consistent in the whole duration, which proves the high accuracy of ARMA (*1*, *1*) in this case.

Table 2 Statistical error index

Error Index	Value
MAE	4.103
MAPE	17.424
Theil Inequality Coefficient	0.096

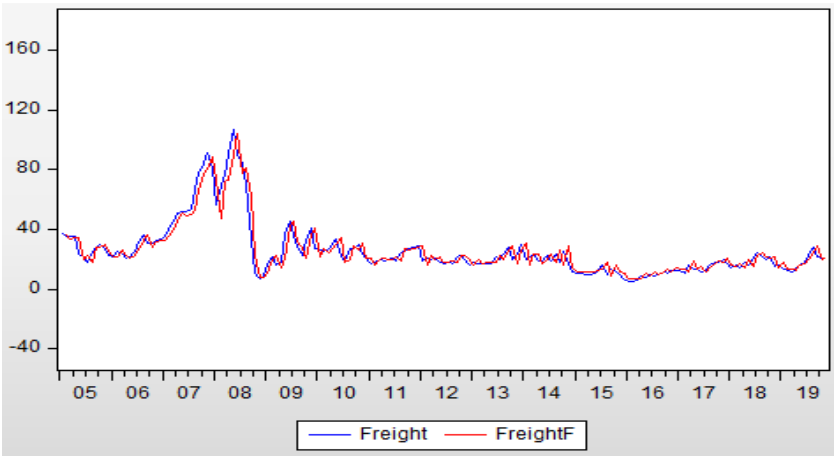

Fig. 6 Fitting effect

3.4 Forecast and Analysis

According to formula (*4*) and time series data, the sea freight rate of iron ore from Tubarao to Ningbo after October 2019 could be predicted. Considering ARMA *(1*, *1*) is suitable for short-term prediction,

ISSN: 2414-1895 DOI: 10.6919/ICJE.202003_6(3).0024

it was decided to set the prediction period from November 2019 to April 2020. The results is shown in *Table 3*:

Table 3 Predictive value from November 2019 to April 2020

Month	Predictive value
Nov-19	21.29
Dec-19	21.79
Jan-20	22.24
Feb-20	22.65
Mar-20	23.02
Apr-20	23.35

It is obvious that the iron ore freight rate of the route will show a slow growth trend in the short term, probably in the range of *$21.29 ~ 23.35* per ton, with an average growth of *61.4%* compared with the same period last year. It is far away from the historical high level caused by irrational market behavior and other factors, while the segment shipping market will show signs of slow recovery, combined with the overall freight rate level of more than ten years.

4. Conclusion

China is the largest iron ore importer in the world, while Brazil is the only iron ore exporter after Australia. It is of great significance for the global iron ore trade market, steel supply market and Capesize ship chartering market to predict the iron ore shipping price between the two countries. ARIMA model, as a common time series prediction model, has the advantages of high precision and small amount of data.

According to the historical freight rate data of Tubarao-Ningbo iron ore from January 2005 to October 2019, the unit root test and autocorrelation and partial autocorrelation chart of Eviews were used to test the stability and identify the parameters of the time series, and then the optimal model was determined through the goodness of fit and information criteria, and finally ARMA (*1*,*1*) was established through the residual correlation test and fitting test. In addition, the effect of the error index of the model is good, and the observation value in the observation period is very close to the prediction value of the model, which shows that the model is very effective. Finally, the model was used to predict that the iron ore freight rate of the route from November 2019 to April 2020 will show a slow growth trend in the range of *$21.29-23.35* per ton, which has certain reference value for iron ore trade, ship chartering and steel market of China.

References

[1] B.Z. Zhu, Y.M. Wei: Carbon price forecasting with a novel hybrid ARIMA and least squares support vector machines methodology, Omega, vol. 41 (2013), 517-524.

[2] J. Yang, L. Li, Y.M. Shi, X.L. Xie: An ARIMA model with adaptive orders for predicting blood glucose concentrations and hypoglycemia, IEEE Journal Of Biomedical and Health Informatics, vol. 23 (2019), 1251-1260.

[3] J.J. Wang, J.Z. Wang, Z.G. Zhang, et al. Stock index forecasting based on a hybrid model, Omega, vol. 40 (2012), 758-766.

[4] L.Y. Zhang, J. Lin, R.Z. Qiu, et al. Trend analysis and forecast of PM2.5 in Fuzhou, China using the ARIMA model, Ecological Indicators, vol. 95(2018), 702-710.

[5] M. Ohyver, H. Pudjihastuti: ARIMA model for forecasting the price of medium quality rice to anticipate price fluctuations, Procedia Computer Science, vol. 135 (2018), 707-711.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202003_6(3).0024

[6] Q. Liu, Z.Q. Li, Y.Ji, et al. Forecasting the seasonality and trend of pulmonary tuberculosis in Jiangsu Province of China using advanced statistical time-series analyses, Infection and Drug Resistance, vol. 12 (2019), 2311-2322.