Industrial Engineering and Innovation Management (2021) 4: 56-60
Clausius Scientific Press, Canada
DOI: 10.23977/ieim.2021.040108
ISSN 2522-6924

Analysis and Forecast of China Railway Freight Volume based on ARIMA Model

Meng Xiao[1], Wenjun Li[2]

[1]*School of Mathematics and Big Data, Chongqing University of Science and Technology, Chongqing, 401331*

[2]*School of Engineering, Queen Mary University of London, Northwestern Polytechnical University, Xi'an, Shaanxi, 710129*

Keywords: ARIMA model, railway freight volume, time series, prediction

Abstract: In order to forecast China's railway freight volume scientifically, this paper uses ARIMA model and R-studio software to analyze China's railway freight volume from 1949 to 2008 in detail, and takes the data of China's railway freight volume from 1999 to 2008 as the test set and compares it with the predicted value of the model. The results show that the error between the predicted value and the real value of the ARIMA model with drift term is smaller, it can provide more accurate prediction results, has high feasibility and credibility, and can effectively make a reasonable forecast of China's railway freight volume in the future, and provide a reliable basis for its development.

1. Introduction

The forecast of railway freight volume plays an important role in the economic development of railway transportation industry and is beneficial to improve the operating efficiency of railway transportation system. Forecasting railway freight volume can provide reference for railway transport enterprises to formulate marketing plans, provide basis for railway related departments to designate railway related policies and provide solid and reliable practical basis for the future investment direction of railway construction of the country.

Before that, some scholars have done related research. In "Research on Railway Freight Volume Forecast and influencing factors" in 2016, Zhang Jie used several time series forecasting models to analyze the relative error [1]. And Tuan Jin uses Eviews to forecast railway freight volume for short-term forecasts [2]. Because of the high accuracy of ARIMA model in short-term trend prediction, ARIMA model has a good simulation effect when the mean square error of other single variable time series forecasting model is large, and in the process of railway freight volume prediction. The effect of quadratic exponential smoothing method is the best. It fits the time series model with the highest railway freight volume in different types of time series, which simplifies the prediction process and modeling speed [3].

2. Establishment of ARIMA model

ARIMA model, the full name of differential integrated moving average autoregressive model, is a non-stationary time series model, which is transformed into a stationary time series model through difference [4]. ARIMA is a model obtained by the combination of autoregressive model and moving average model and difference operation. Where *p* is the autoregressive coefficient, *q* is the difference order made to become a stationary time series, and *d* is the number of moving average terms.

A model with the following structure is called an autoregressive moving average model, abbreviated as *ARIMA(p,q)*.

$$\begin{cases} x_t = \phi_0 + \phi_1 x_{t-1} + \ldots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q} \\ \varphi_p \neq 0, \theta_q \neq 0 \\ E(\varepsilon_t) = 0, Var(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t, \varepsilon_s) = 0, s \neq t \\ E(x_s \varepsilon_t) = 0, \forall s < t \end{cases} \tag{1}$$

Then, the ARIMA model structure can be obtained by differential processing of *ARIMA(p,q)* as follows:

$$\begin{cases} \phi(B)\nabla^d x_t = f(B)\varepsilon_t \\ \nabla^d = (1-B)^d \\ E(\varepsilon_t) = 0, Var(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t, \varepsilon_s) = 0, s \neq t \\ \forall s < t, Ex_t \varepsilon_t = 0 \end{cases} \tag{2}$$

Among them,

$$\begin{cases} \phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_P B^P \\ f(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q \end{cases} \tag{3}$$

When the non-stationary sequence is encountered, the difference operation is needed to convert the non-stationary sequence into a stationary time series, and then the ARIMA model fitting is carried out for the sequence after the difference, and the prediction is made.

3. Application of ARIMA Model in China Railway Freight Volume Forecast

3.1 Stationary test

Before modeling time series, it is necessary to ensure the stationarity of the series. [5] Therefore, the first step is to draw a time sequence diagram, as shown in Figure 1.

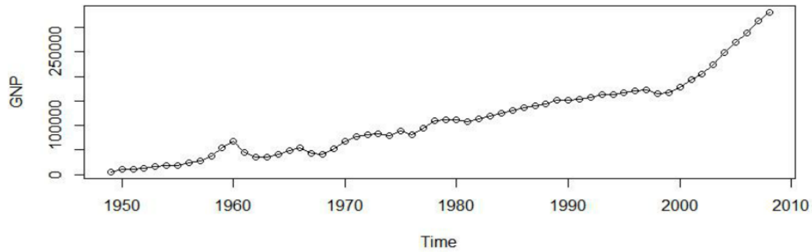

Figure 1: Sequence time sequence diagram

As can be seen from Figure 1, the sequence has an obvious upward trend, indicating that this sequence has a strong non-stationarity and needs to be stabilized. In this paper, the first-order difference method is used to stabilize the sequence, and the results are shown in Figure 2.

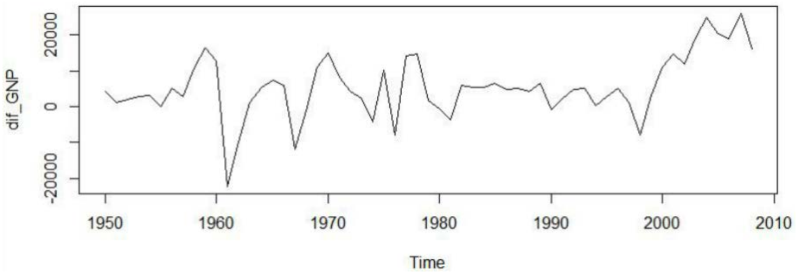

Figure 2: Sequence diagram after first-order differential processing

It can be found from the figure that the sequence after the first-order difference shows a relatively stable fluctuation. In order to further determine the stationarity of the post-sequence, the ADF test is carried out on the differential sequence, and it is found that the mean values of all ADF test statistics of the sequence are less than 0.05, and the sequence also passes the randomness test, so it can be determined that the sequence after the first-order difference is a stationary non-white noise sequence.

3.2 Model order determination and parameter estimation

In this paper, R-Studio software is used to draw the autocorrelation diagram (ACF) and partial autocorrelation diagram (PACF) of this sequence to determine the most appropriate P and Q in the *ARIMA(q, d, q)* model. As can be seen from Fig.3, after first-order difference, the autocorrelation value shows trailing phenomenon, and the partial autocorrelation graph shows trunking phenomenon of first-order, so p=1 is set in this paper[6].

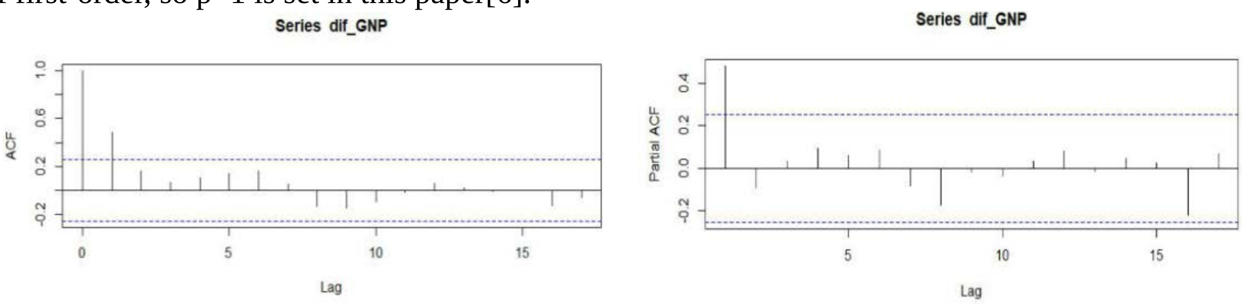

Figure 3: Autocorrelation and partial autocorrelation graphs

Because of the first order difference operation, this paper uses *ARIMA*(1,1,0) model to fit the sequence of railway freight volume in China.

4. ARIMA model prediction

Simulation tests were carried out on the *ARIMA*(1,1,0) fitting model with no drift term and with drift term (the mean value of the sequence after difference [7] [8]), and it was found that the fitting model could pass the significance test.

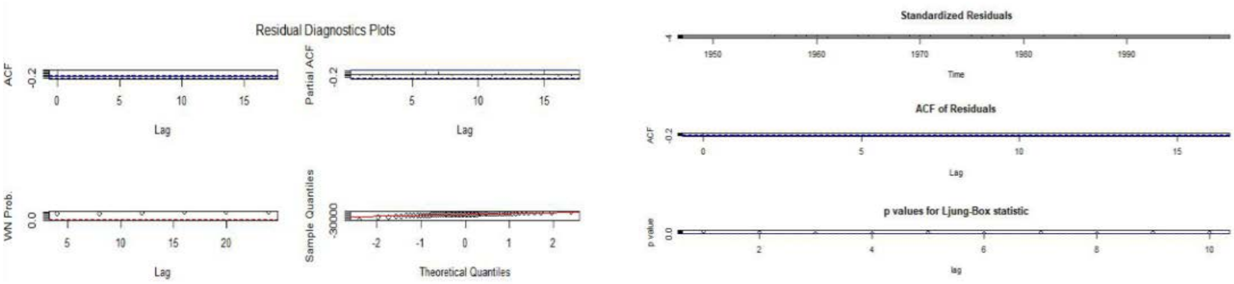

Figure 4: Test chart of significance of models with and without drift terms

The development trend from 1999 to 2008 is predicted, and the predicted results are shown in Figure 5 and Table 1.

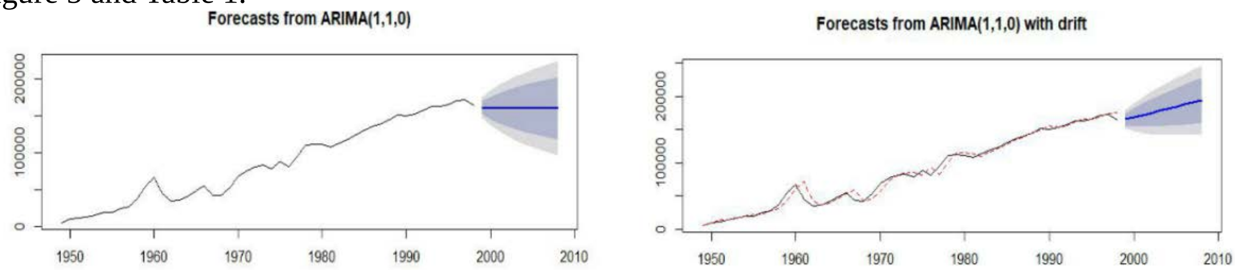

Figure 5: Comparison of trends predicted by models with and without drift terms

Table 1: Comparison of prediction results between models with and without drift terms

	No drift items			Drift items		
Year	fore1.mean	test1	error1	fore2.mean	test2	error2
1999	161770.7	161770.7	161770.7	161770.7	161770.7	161770.7
2000	160949.0	160949.0	160949.0	160949.0	160949.0	160949.0
2001	160682.9	160682.9	160682.9	160682.9	160682.9	160682.9
2002	160596.8	160596.8	160596.8	160596.8	160596.8	160596.8
2003	160568.9	160568.9	160568.9	160568.9	160568.9	160568.9
2004	160559.9	160559.9	160559.9	160559.9	160559.9	160559.9
2005	160556.9	160556.9	160556.9	160556.9	160556.9	160556.9
2006	160556.0	160556.0	160556.0	160556.0	160556.0	160556.0
2007	160555.7	160555.7	160555.7	160555.7	160555.7	160555.7
2008	160555.6	160555.6	160555.6	160555.6	160555.6	160555.6

From the table, it can be seen that the prediction value of the non-drift model shows a slow downward trend, but the prediction error is larger in the short term; the prediction value of the model with drift term shows an increasing trend, which basically continues the development trend of the original series, and the error is relatively small in the short term.

5. Conclusion

In this paper, the ARIMA model with and without drift term is fitted respectively to predict and analyze the railway freight volume in China. From this, it can be seen that the prediction effect of the ARIMA model with drift term is significantly better than that of the model without drift term. The fitting effect of the two fitting models for China's railway freight volume data is better, but the prediction effect is significantly different, and the prediction trend is also quite different, but the prediction trend of the ARIMA model with drift term is more consistent with the reality and has a higher degree of fit with the original data. To sum up, it is suggested that the ARIMA model with drift term should be used to predict China's railway freight volume, which can effectively make a reasonable forecast of China's railway freight volume in the future and provide a reliable basis for its development.

References

[1] Zhang yue. Research on railway freight volume prediction and influencing factors [D]. Beijing Jiaotong University, 2016.

[2] Zhao Naqing. Railway Freight Volume Forecast based on ARIMA Model [J]. The Age of think tanks, 2019 (22): 187: 190.

[3] Zhang Jun. Research on the Forecast of fixed assets Investment in China based on ARIMA Model [J]. Modern Commerce, 2019 (05): 77-78.

[4] Luo Mengjuan. Research on GDP Forecast of Anhui Province based on ARIMA Model and regression Analysis [D]. Anhui Agricultural University, 2015.

[5] Yan Jing, Yang Bei. Application of exponential smoothing method and ARIMA model in predicting the incidence of viral hepatitis C in Hubei Province [J]. China Vaccine and Immunization, 2017 Jol 23 (03): 292-297.

[6] Ni Ningning. Fitting and forecasting of fixed assets investment in the whole society based on extended ARIMA model [J]. Statistics and decision-making, 2016 (13): 72-75.

[7] Wang Yongbin, Chai Feng, Li Xiangwen, Yuan Juxiang, Wu Jianhui. Application of Arima model and residual autoregressive model in the prediction of hand-foot-mouth disease [J]. Chinese Journal of Disease Control, 2016J.20 (03): 303-306.

[8] Zhang Ke, Meng Xiujuan. Visualization method of numerical variables in R-Studio environment [J]. Southern Agricultural Machinery, 2020, 51 (15): 194-195.