ISSN 2616-5775 Vol. 6, Issue 8: 138-145, DOI: 10.25236/AJCIS.2023.060817

Research on Logistics Cargo Volume Forecasting Based on SETAR Model

Xingguo Xu[1,*], Qiaojun Chen[2], Yiqiang Xia[1]

[1]*College of Science, Liaoning Technical University, Fuxin, Liaoning, 123000, China*
[2]*College of Safety, Liaoning Technical University, Huludao, Liaoning, 125000, China*
[*]*Corresponding author*

Abstract: *With the rapid development of e-commerce platforms, logistics networks have ushered innew challenges. In order to predict the cargo volume of logistics routes over a period of time, this paper takes a transportation network as an example, and calibrates three routes as observation pointsbased on the cargo volume of historical logistics routes. Selects a SETAR model with more time considerations than time series, establishes an AR model in each subinterval by setting the upper limitof the number of model orders, delay steps and the number of thresholds, and then changes the threshold value to select the optimal threshold value is selected by the AIC criterion. Then the numberof delay steps is changed and the procedure is repeated to achieve the optimal prediction result. Finally, the prediction results of cargo transportation volume of the three routes are obtained. This study is important for the rational arrangement of logistics transportation resources.*

Keywords: *SETAR Model, Logistics Cargo Volume, AR Model*

1. Introduction

After the 21st century, e-commerce platforms have developed rapidly, providing us with convenient ways to shop for everything from simple hairpins and pencils to sophisticated computers,cell phones, and even cars and speedboats. These goods can be ordered through e-commerce platforms and delivered within a few days through the logistics network. According to statistics, China sends and receives more than 300 million pieces of express delivery every day, which requiresa large and systematic logistics network to support. The logistics network is interwoven by numerouslogistics transport routes and transit stations, each of which has its own regular transport mode and common carrying capacity. Logistics companies usually forecast future volumes based on past periods of shipments in order to arrange appropriate manpower and capacity. Therefore, it is importantfor logistics companies to accurately predict the volume of cargo transportation on logistics routes inthe coming period of time.

Time series forecasting models have become one of the preferred methods for many, with the autoregressive moving average model(ARMA)being one of the commonly used time series forecasting[1-2] methods. The ARMA model combines the characteristics of autoregressive (AR) and moving average(MA) models to capture trends and periodicity in time series data and is used to forecast the volume of traffic on logistics routes. Previous studies have focused on forecasting the volume of goods transported along logistics routes, using ARMA models [2-5] for predictive modeling. The researchers collected data on cargo traffic volumes on different logistics routes and selected appropriate ARMA models for parameter estimation and model fitting to derive cargo traffic forecasts for future time periods for each route. The research results show that the ARMA model can predict the cargo transportation volume on logistics routes more accurately, which provides a reference basis for logistics planning and resource allocation. However, ARMA models have difficulties in dealing with complex nonlinear relationships and may not capture features and variations well. At the same time, ARMA models may not accurately capture and predict data with complex seasonality or unconventional cyclicality. Although ARMA models can provide more accurate forecasts of cargo traffic in some cases, their prediction accuracy is still affected by various factors such as data quality, model parameter selection and observation errors. Therefore, these factors need to be considered together in practical applications to evaluate and revise the forecast results.

The purpose of this paper is to combine the advantages of the ARMA model and to use the SETAR model [6-7] with more time considerations than the time series. The SETAR model is a nonlinear time series model that better captures the nonlinear relationships and potential threshold effects in cargo traffic

ISSN 2616-5775 Vol. 6, Issue 8: 138-145, DOI: 10.25236/AJCIS.2023.060817

data compared to traditional linear models. By segmenting the data and applying different autoregressive models at different threshold points, the SETAR model is able to more accurately describe and predict the nonlinear dynamics of cargo traffic. In addition, the SETAR model is able toconsider the effect of threshold effects and select an appropriate autoregressive model at the time of forecasting, thus providing more accurate forecasting results on cargo traffic data with significant segmentation characteristics.

For the forecast of route cargo, a logistics network company is selected as a case study in this paper, and the historical data of daily cargo volume carried by each route for two years from January 1, 2021 to December 31, 2022 are fitted to the relationship in order to discover linear or nonlinear relationships and to find a suitable model. For the relationship fitting of these three routes, the SETAR model is considered for solving this paper. However, the simple time series model [6] may not be applicable to each route, so this paper also considers the improvement and optimization of the time series model to obtain a better model.

2. Prediction of route cargo volume based on nonlinear SETAR mode

In this paper, the route cargo volume of a company's logistics network is selected for prediction, and the route of this company's logistics network is specified in figure 1.

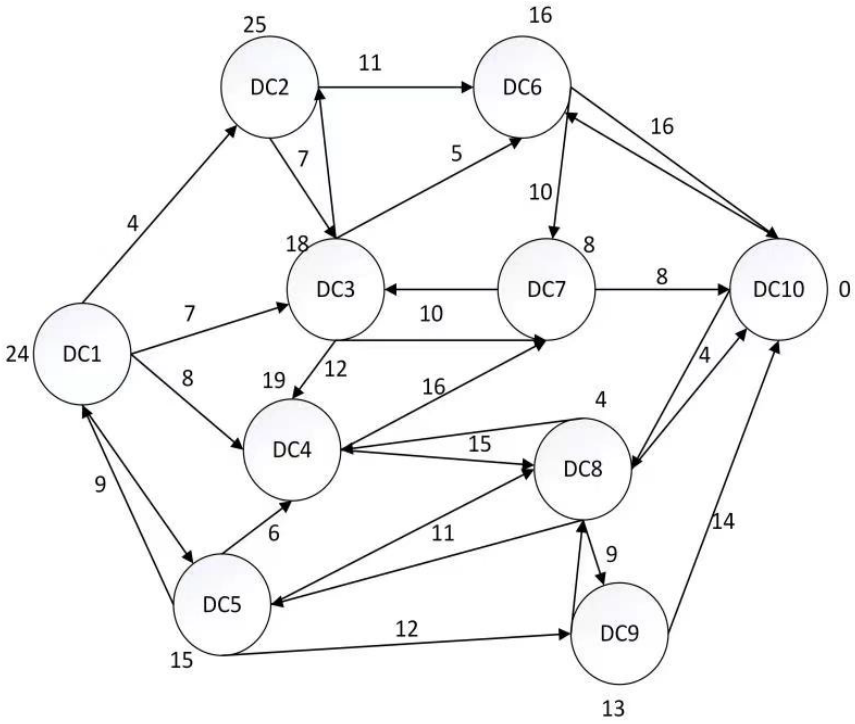

Figure 1: A company's logistics network route

2.1 Pre-processing and analysis of data

In this paper, based on the information reflected by the historical data of cargo transportation volume of a logistics network company in the past years, we find the data of the three routes to be predicted: DC14→DC10, DC20→DC35, DC25→DC62, and use Matlab software to make a scatter plot of the daily cargo volume following time.

ISSN 2616-5775 Vol. 6, Issue 8: 138-145, DOI: 10.25236/AJCIS.2023.060817

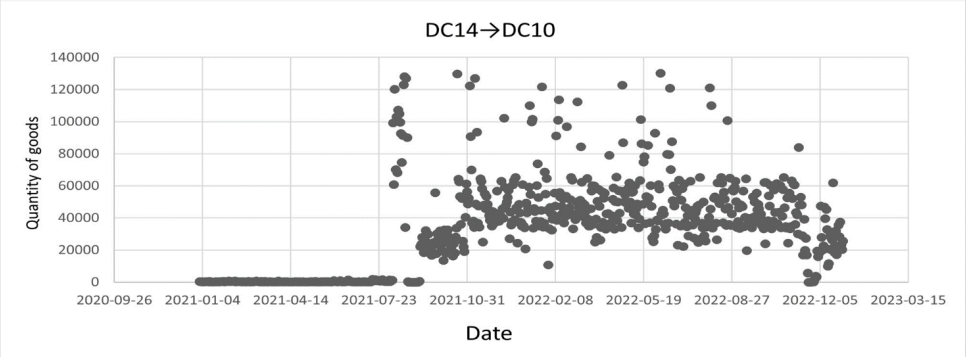

Figure 2: Scatter diagram of DC14→DC10 route

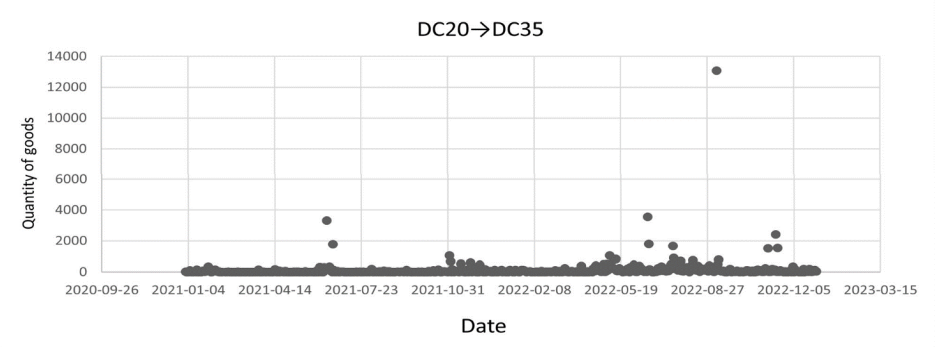

Figure 3: Scatter diagram of DC20→DC35 route

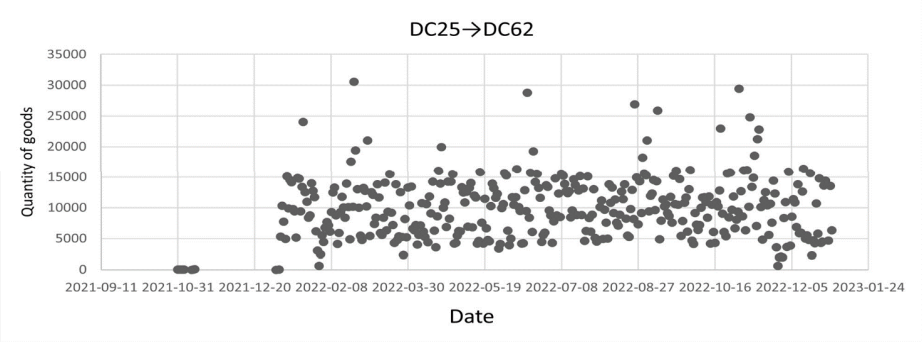

Figure 4: Scatter diagram of DC25→DC62 route

According to Figure 2, this paper observes the daily cargo volume of route DC14→DC10 from January 1, 2021 to December 31, 2022, and it is not difficult to find that it presents irregular state, which may be caused by holiday promotions and unexpected events, which presents serious nonlinearity. The ordinary linear model cannot satisfy this characteristic, so we can only consider other algorithms.

For Figure 3, this paper can observe that the daily shipments of route DC20→DC35 from January1, 2021 to December 31, 2022 present a smooth state. The influence of those anomalies can be removed when building the model, and the underlying time series can be used.

For Figure 4, this paper can observe that the daily cargo volume of route DC25→DC62 from January 1, 2021 to December 31, 2022, also presents a serious irregularity, which leads to the generation of nonlinearity, which also causes the ordinary time series to fail to solve this problem [8].

Lastly, Hansen's test is used to analyze the error in order to make a reasonable and accurate prediction.

2.2 Introduction and use of the SETAR model

The SETAR model, also known as the threshold autoregressive model, was first proposed by H.Tong [8] in 1978. Its basic idea is that the observation time series is divided into L subintervals according to the pre-defined (L-1) threshold values e_j(j=1, 2, 3. . . L-1), and the delay step a is set to assign different sizes of $\{y_t\}$ and $\{y_{t-a}\}$ values to different subintervals, and then continue to applydifferent AR model

ISSN 2616-5775 Vol. 6, Issue 8: 138-145, DOI: 10.25236/AJCIS.2023.060817

to describe the whole system. For the general form of the threshold autoregressivemodel as.

$$y_t = B_0^{(j)} + \sum_{i=1}^{m_{(j)}} B_i^{(j)} y_{t-a} + w_t^{(j)} \tag{1}$$

$$e_{j-1} < y_{t-a} < e_j (j = 1, 2 \ldots\ldots L) \tag{2}$$

The model notation is explained as follows.

Table 1: SETAR model symbol description

Symbols	Meaning
e_j	ndicates the threshold value
L	Indicates the number of subintervals
a	denotes the number of delay steps
$w_t^{(j)}$	denotes a white noise sequence with variance$\sigma^2{}_j$
$B_i^{(j)}$	denotes the autoregressive coefficient of the model in subinterval j
n_j	denotes the order of the model in subinterval j

Obsering Table 1 can also be represented by SETAR(L， a， $n_1, n_2, \ldots n_L$) to represent.

When L=1, a=0, the model is degenerated to the traditional autoregressive model, so the SETAR model is essentially an interval AR [2] model, where the threshold divides the time series {xt} into different intervals and builds the corresponding autoregressive model within that interval, describingthe whole nonlinear time series by means of a combination of multiple segmented linear autoregressive models [9].Flow chart of SETAR model usage as shown in Figure 5.

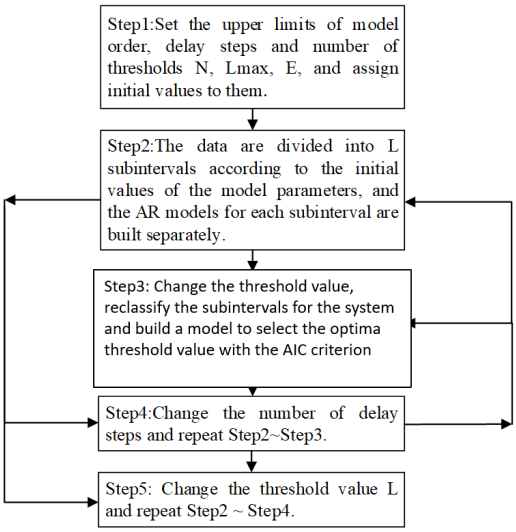

Figure 5: Flow chart of SETAR model usage

The key to the parameter identification of the threshold autoregressive model is how it goes to determine the threshold value e, the number of segment intervals L and the number of lag steps a, andthen model the AR model within its segmented intervals. Therefore, we combine the point-value map in SETAR and the AIC criterion in a hybrid method, and select the daily cargo volume from January 1, 2021 to December 31, 2022 in DC14→DC10 as the training set data to predict the daily cargo volume from January 1, 2023 to January 31, 2023 thereafter.

In this article, we first use these training data to plot the sequence point-value graph. In thispaper, we can use the point-value plot to test whether the series really has nonlinear characteristics, and this method is often called mathematical expectation estimation method in mathematics. We find the appropriate lag a for the time series $\{y_t\}$ by going to build the corresponding data pair (y_t, y_{t-a})， and we go to build the coordinate axes, taking y_{t-a} as the horizontal axis, and dividing it into m segments uniformly. We boldly make the assumption that with n_1 y_{t-a} falling within the first segment,then for this set of sequences we go for the mean, which is equivalent to going for the conditionalmathematical

ISSN 2616-5775 Vol. 6, Issue 8: 138-145, DOI: 10.25236/AJCIS.2023.060817

expectation within each segment.

$$E\left(\frac{y_t}{y_{t-a}}\right)_i = \frac{1}{n_i}\sum_{i=1}^{n_i} y_t^{(i)}, i = 1,2,\ldots m \tag{3}$$

In this paper, takingy_{t-a} as the horizontal axis and $E(\frac{y_t}{y_{t-a}})_i$ going as the vertical axis, we can obtain its relationship graph as shown in Fig.

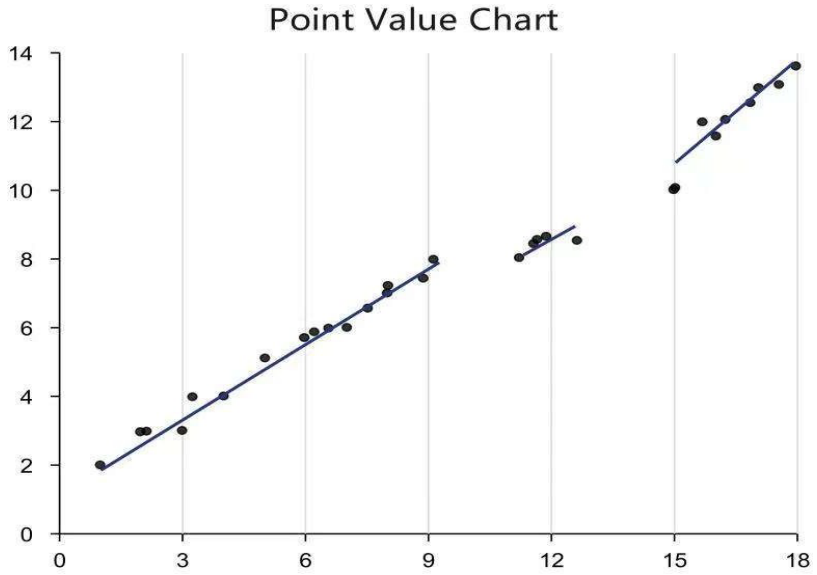

Figure 6: Graph of mathematical expectation values for problem 1

Observing Figure 6, we find that the points cannot be fitted with a straight line, so we can prove that it is a nonlinear sequence, and we can determine the number of segments L and the range of e according to the above figure, so as to reduce the complexity of the calculation.

After determining the specific range of *L* and *e*, we select$\{e_1, e_2, \ldots e_j\}$ as alternative threshold variables, and each time go to choose one of them as the required threshold value, thus go to divide the sequence into two segments, and build the AR model in each segment, and then start to calculate the correspondin.

$$\text{AIC(ej)=AIC(AR(n1)+AIC(AR(n2)} \tag{4}$$

Determine the number of lag steps *a*: Since considering that different lag steps change the numberof selected sample series, after determining *e*, the same method can be used to determine *a*.Choose$\{a_1, a_2, a_3, \ldots a_n\}$to calculate the value of the corresponding AIC.

$$AIC(a_n) = \frac{AIC(L;e_j)}{N-a} \tag{5}$$

Where: $a_n$is the number of lag steps for a given condition, *N* is the total number of samples, and the corresponding a_n is its lag step when $AIC(a_n)$obtains the minimum value. The above steps allow us to determine the number of lag steps required for SETAR(L;a； $n_1, n_2, \ldots n_j$)with the specific parameters required [9].

2.3 Solving with SETAR model

In this paper, according to the above table, the lag step number 15<a<20 is selected, and using the above modeling method, the final paper determines the threshold variable values $e_1, e_2$for the first route DC14→DC10 as follows: *L*=3, e_1=4. 873 e_2=16. 477, the lag step number is *a*=18, and the corresponding SETAR model parameters and expressions are obtained as follows:

ISSN 2616-5775 Vol. 6, Issue 8: 138-145, DOI: 10.25236/AJCIS.2023.060817

$$y_t = \begin{cases} -4.1\times10^{-7} + 2.23y_{t-1} - 1.98y_{t-2} + 0.7y_{t-3} - 3.14y_{t-4} \\ -0.29y_{t-5} - 0.66y_{t-6} \left(y_{t-18} \le 4.873\right) \\ -3.68\times10^{-7} + 4.43y_{t-1} - 2.21y_{t-2} + 0.13y_{t-3} - 3.13y_{t-4} \\ -033y_{t-5} - 0.37y_{t-6} \left(y_{t-18} \ge 16.477\right) \\ -7.77\times10^{-7} + 2.37y_{t-1} - 0.98y_{t-2} + 0.32y_{t-3} + 0.41y_{t-4} \\ +0.11y_{t-5} - 0.99y_{t-6} \left(4.873 \le y_{t-18} \le 16.477\right) \end{cases} \quad (6)$$

In this paper, we write the above model as SETAR (3;18;5;6;5) and calculate the daily cargo data from January 1 to 31, 2023 for the three routes as follows Table 2.

Table 2: 2023-01-01 to 2023-01-31 Daily three routes cargo data

Date	DC14→DC10	DC20→DC35	DC25→DC62
2023-01-01	20311	46	13362
2023-01-02	20567	8	1111
2023-01-03	28768	2	2435
2023-01-04	23456	4	4657
2023-01-05	35676	55	9876
•••	•••	•••	•••
2023-01-27	49687	25	25
2023-01-28	25688	9	9
2023-01-29	36887	4	4
2023-01-30	20311	9	9
2023-01-31	29679	34	34

In order to observe the cargo data on these routes more intuitively, this paper uses Matlab software to produce a scatter plot. The results are shown by Figure 7, Figure 8, Figure 9.

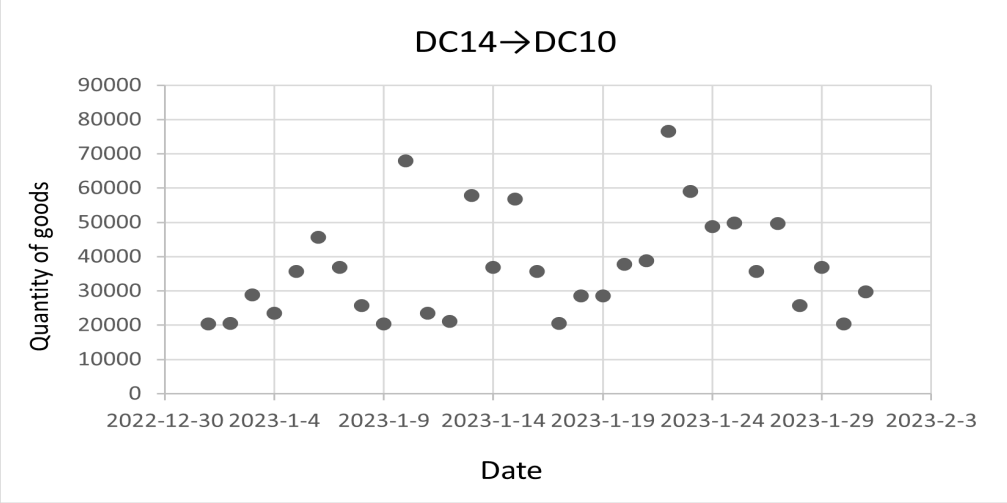

Figure 7: Daily DC14→DC10 route cargo data scatter plot from 2023-01-01 to 2023-01-31

ISSN 2616-5775 Vol. 6, Issue 8: 138-145, DOI: 10.25236/AJCIS.2023.060817

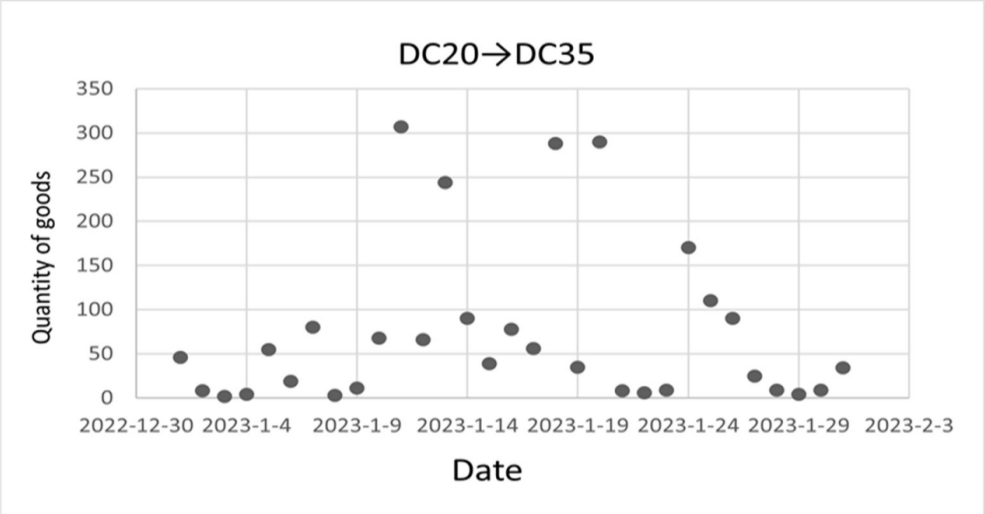

Figure 8: Daily DC20→DC35 route cargo data scatter plot from 2023-01-01 to 2023-01-31

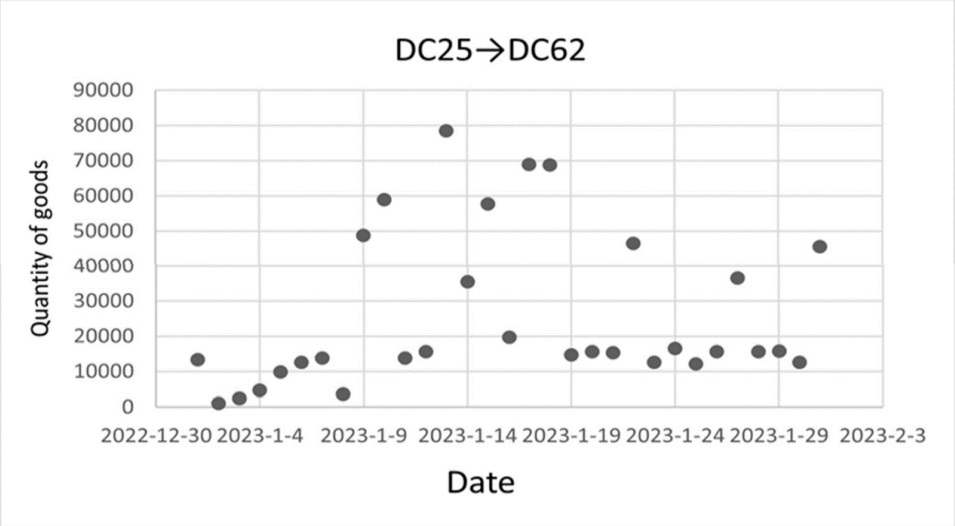

Figure 9: Daily DC25→DC62 route cargo data scatter plot from 2023-01-01 to 2023-01-31

2.4 Hansen's test for the prediction results of the SETAR model

In this paper, we use Hansen test for the non-linear SETAR model based on its previous use for the prediction of the results of the three routes [10]. Hansen is used to analyze the non-linear adjustment mechanism between economic variables and we can use it to test our prediction results, according to table 2 we use Matlab to implement Hansen test, to obtain Table 3.

Table 3: Hansen's test results

Route	5% critical value	10% Threshold	p-value
DC14→DC10	70. 5645	66. 5241	0. 001
DC20→DC35	71. 5549	63. 4478	0. 001
DC25→DC62	69. 3544	61. 6632	0. 000

From the analysis in Table 3: we can easily find that the p-values of all three routes are less than 0. 01, and we can conclude that the original hypothesis is rejected at 99% confidence level. Therefore,we determine that the error of the predicted results using the nonlinear SETAR model is small and acceptable.

3. Summary

In this paper, a large amount of historical data of a logistics network company was firstly screenedand cleaned, and three routes DC14→DC10, DC20→DC35 and DC25→DC62 were calibrated as

ISSN 2616-5775 Vol. 6, Issue 8: 138-145, DOI: 10.25236/AJCIS.2023.060817

observation points. From the results of MATLAB processing, some routes transported cargo volumeline graphs show non-linear and irregular relationships. So the SETAR model, which has more time considerations than the time series, is chosen in this paper to match with the data. By setting the upperlimit of the model order, the number of delay steps and the number of thresholds, the AR model is built in each subinterval, after which the threshold value is changed to select the optimal threshold value with the AIC criterion, and then the number of delay steps is changed and its steps are repeatedto achieve the optimal prediction results. We finally arrive at the prediction results of cargo traffic forthe three routes, which are shown in Table 2 of the main text.

References

[1] Yang H-M, Pan Z-Song, Bai Wei. A review of time series forecasting methods[J]. Computer Science, 2019, 46(01):21-28.
[2] Wang X, Wu J, Liu Chao, Yang H-Y, Du Y-L, Niu W-S. Fault time series prediction based on LSTM recurrent neural network [J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(04):772-784.
[3] Ruohui Zhang. A study of SSE Composite Index returns based on ARMA-GARCH model [J]. China Market, 2023(13): 43-46.
[4] Geng H, Wang W, Xing Chengbin. Research on tide level fitting and interpolation based on ARMA model [J]. Marine mapping, 2021, 41(05):17-20+25.
[5] Huang Shuiren, Liu Yuji, Hu J. Construction of robust ARMA residual control charts and their application in financial markets [J]. Mathematical Theory andApplications, 2022, 42(01):117- 129.
[6] Zhang Chen. A test of reversal and inertia effects in the Chinese stock market [D]. Northeast Normal University, 2016.
[7] Yu-Ping Y. Research on urban logistics demand forecasting based on the BP neural network method[J].Journal of Qinghai Normal University(Natural Science Edition), 2017.
[8] Chen Le, Prediction of plunger pump leakage based on time-series method [D]. Lanzhou University of Technology, 2022.
[9] W. J. Shi, L. S. Hu. Performance estimation of a class of nonlinear systems basedon SETAR model [J]. Control Engineering, 2010, 17(S2):97100.
[10] Wei J. Y. Research on the application of cross-period arbitrage of stock indexfutures based on GARCH-GED and SETAR models [D]. Henan University of Economics and Law, 2021.