Research on the Forecasting Model of Logistics Volume Based on the Theory of Large Transportation and Logistics: Taking Qingdao City as an Example

Hui Sun [*], Suwei Song[1]

College of Transportation Shandong University of Science and Technology Qingdao, Shandong,,China
[1]Corresponding author:Suwei Song, Tel:18765250776. Emil:18765250776@163.com

Abstract — In order to realize the effective forecast of logistics volume in Qingdao, this paper studied the forecasting method based on the combination of qualitative and quantitative analysis of the large transportation logistics theory. After analyzed the particularity of Qingdao logistics, this study selected and established a prediction model based on the multiple linear regression. Moreover, this study used the MATLAB software to calculate the equation of regression model according to the years of logistics-related data of Qingdao. Finally, it validated the feasibility of the forecasting model by system simulation in logistics.

Keywords - The theory of large transportation and logistics; Logistics volume of Qingdao; Model based on multiple linear regression; prediction

I. INTRODUCTION

Many factors which includes the development of network information, the adjustment of economic structure, the mainstream development of market competition and cooperation have been driving the port city to develop based on the important base of large transport and logistics,which plays the role of economic engine nowadays[1]. The purpose of the development of large transport is to build a modern land, sea and air transportation system; the purpose of the promotion of large logistics is to create well-developed modern logistics industry [2].Its essence is to realize the fusion development of comprehensive transportation and modern logistics, to optimize the transport and logistics system and to achieve the most optimal integrated allocation of resource.

As an eastern coastal metropolis , Qingdao plays an important role in the development of eastern China. From the past economic statistics, we can see that the added value of the tertiary industrial is increasing year by year in Qingdao, in which transport and logistics industry has developed rapidly. In the current context, Qingdao should speed up based on the large transportation and logistics system.The "Qingdao City's Comprehensive Transportation Planning (2008-2020)"made a plan which is called three-dimensional integrated transport system for Qingdao in the future , which simply put forward the conception of freight logistics system, described as "three parks, four centers, several distribution centers", and a detailed plan was not included on urban logistics. As a derived demand of economic development, the logistics demand's development has not been attached importance by Qingdao municipal government.

Qingdao, as a land, sea and air modes of transport , is a port and tourist city. Its logistics situation is complex and the logistics volume seasonal changes obviously[3].In fact, there are many problems with the logistics development of Qingdao. For example, items categories and numbers of logistics circulation are complex and varied. Moreover, the basic data of the collection is incomplete, inaccurate and historically limited. The relationship between logistics and the development of the variables is not sure. In addition, the relationship among the factors has two ways or more…Thus, all these problems are closely linked with the development of the logistics industry and will affect the long development of logistics demand.Soit is important and necessary to carry out reasonable analysis and accurate forecast of the logistics quantity in Qingdao.This has important theoretical significance and social economic significance to the development of Qingdao city logistics[4].

II. THE SELECTION AND FORECASTING PROCESS OF QINGDAO LOGISTICS VOLUME' FORECASTING MODEL

A. The Selection of Qingdao Logistics Volume's Forecasting Model

According to the characteristics of urban development and logistics quantity in Qingdao, the combination of qualitative and quantitative methods is used when predicting [5].

Qualitative analysis: the logistics demand in Qingdao is closely related to the economic situation; the changing market environment and consumption structure of Qingdao will make a change in the amount of logistics;Qingdao logistics volume is influenced by the industrial structure;the growth of Qingdao logistics demand and the development of transport conditions and infrastructure are closely related.

Quantitative analysis: according to the qualitative analysis,

we can see that the logistics quantity of Qingdao is closely related to the economic factors and the logistics factors. So it is necessary to choose the quantitative method to make the multi-factors analysis.

After reviewed the existing calendar year data about Qingdao City Statistical Yearbook and Statistical Bulletin, some of the logistics volume and economic indicators original data was collected, As shown in Table I.According to Table I, make the scatter plot of independent and dependent variables and observe its characteristics.

TABLE I. THE ORİGİNAL DATA OF QİNGDAO'S LOGİSTİCS INDİCATORS AND ECONOMİC INDİCATORS FROM 1997 YEAR TO 2014 YEAR

Year	Total Cargo Volume (Rail Way, High Way, Sea,Air)	Port Throughput	Cargo Turnover	The City's Total GDP	The City's Industrial Added Value	Retail Sales of Social Consumer Goods
1997	16242	6943	290	797	341	216.1
1998	20132.44	7043.5	1624.8	880	365	242
1999	20396.96	7282	1774.0	990	416	270.3
2000	24490.3	8661	3010.44	1191.25	500.32	307.7
2001	29118.4	10422.9	3560.12	1316	578.2	352.88
2002	32092	12251.6	1825.7	1518.2	681	400.5
2003	34946.6	14135.7	2005.87	1780.3	825.5	512.2
2004	38564.3	16302	3073.87	2163.8	1024.1	605.5
2005	37783.6	18727	3314.07	2687.46	1266.79	870.11
2006	39610.1	22438	3823.9	3183.18	1527.49	1016.35
2007	40141.9	26507	3458.21	3750.16	1785.31	1216.22
2008	42484.2	30029	4579.56	4401.56	2062	1492.22
2009	24408.19	31668	3912.03	4853.87	2338.13	1730.22
2010	26971.44	35012	3596.15	5666.19	2714.57	1961.13
2011	29133.93	37971	3819.45	6615.6	3081.04	2302.37
2012	29238.15	41465	2064.88	7302.11	3138.44	2635.62
2013	31308.31	45782	1113.93	8006.56	3248.4	2986.81
2014	26051.22	47701	1037.85	8692.1	3419.8	3361.72

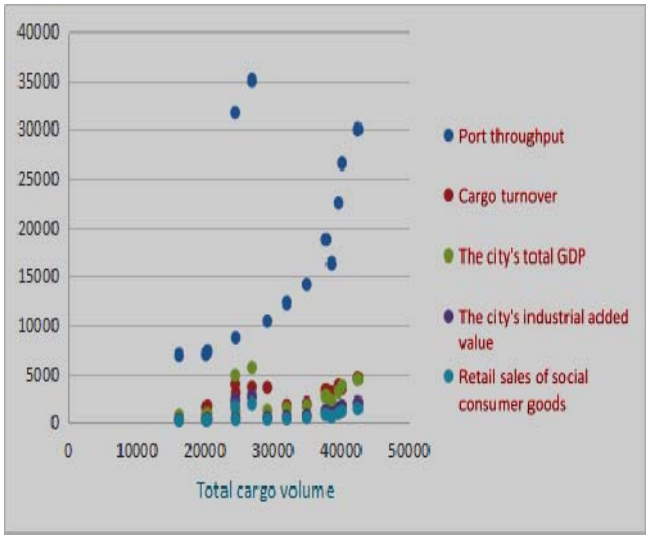

Fig.1 The Scatter Plot of the Independent Variable and Dependent Variable

According to the above scatter plot, there is a certain relationship between the total freight volume and the 5 variables.

With the analysis of the level of economic development, transportation logistics development[6]and scatter plot, we finally choose the multiple linear regression forecasting model to predict the logistic volume.

B. The Forecast Process of Logistic Volume[7]

The first step: collect the network and literature data and obtain the existing or historical data and related data of the influencing factors of logistics. Then the data will be standardized, the formula for the Standardization of extreme value is obtained for 2-1:

$$X_{pq} = \frac{X'_{pq} - X'_{q\min}}{X'_{q\max} - X'_{q\min}} \tag{1}$$

The Standardization formula for the original data is:

$$X'_{pq} = \frac{X_{pq} - X_q}{S_q} \tag{2}$$

Standard deviation is:

$$S_q = \sqrt{\sum_{p=1}^{m} \left(X_{pq} - X_q \right)^2} \tag{3}$$

The average value of the N factor of the M sample is:

$$X_q = \frac{1}{m} \sum_{p=1}^{m} X_{pq} \tag{4}$$

X_{pq} refers to the standardized value of the logistics demand. Respectively, $X'_{q\max}$ and $X'_{q\min}$ refer to the maximum and minimum value of the standardized sample X'_{mq}.

The second step: separate the standardized data and take the most reasonable means as the case might be to determine the coefficients of multivariate regression model equation so as to make sure the multiple regression equation. Then take the remaining standardized data into multiple regression equation to obtain the forecast value.

The third step: compare the actual value with the actual value ,then simulation fitting, and then carry out the relevant tests of the multiple regression model, such as: goodness of fit test, regression model significance test, finally analysis and concluded.

III. THE ESTABLISHMENT OF QINGDAO LOGISTICS PREDICTION MODEL

A. The Sketch of Multivariate Linear Regression Model

In practice, the factors that affect the results of random variable are often more than one.

$$\begin{cases} Y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + \varepsilon_i, i = 1,\ldots,n \\ \varepsilon_1,\ldots,\varepsilon_n iid\left(0,\sigma^2\right) \\ -\infty < \beta_0,\beta_1,\ldots,\beta_p < \infty, \sigma^2 > 0 \end{cases} \tag{5}$$

The model is called the multiple linear regression model.Unknown parameters $\beta_0, \beta_1, \ldots,$

β_p called regression coefficient, ε is a unobservable random error ,satisfy $E(\varepsilon) = 0$， $D(\varepsilon) = \sigma^2 < \infty$($\sigma^2$ keeps unknown).Argument x_i(i= 1, ..., p)is called regression factor or predictor, referred factor [8].

B. Data Standardization

Since the data units vary greatly, they will cause a serious error to the computed model relationship if we use them .So it need to standardize the data in the table before calculating the model, which means that the data will be standardized so that the eigenvalues are unified and mapped to the [0,1] range [9].Data standardization was calculated by the MATLAB software .Standardized data of Qingdao's logistics indicators and economic indicators from 1997 to 2010[10] are shown in Table II. The standardized program computing codes are as follows.

```
>> initial_fun=[16242 20132.44 20396.96 24490.3 29118.4 32092 34946.6 38564.3 37783.6 39610.1 40141.9 42484.2 24408.19 26971.44 ; 6943 7043.5 7282 8661 10422.9 12251.6 14135.7 16302 18727 22438 26507 30029 31668 35012 ; 290 1624.8 1774.0 3010.44 3560.12 1825.7 2005.87 3073.87 3314.07 3823.9 3458.21 4579.56 3912.03 3596.15 ; 797 880 990 1191.25 1316 1518.2 1780.3 2163.8 2687.46 3183.18 3750.16 4401.56 4853.87 5666.19 ; 41 365 416 500.32 578.2 681 825.5 1024.1 1266.79 1527.49 1785.31 2062 338.13 2714.57 ; 216.1 242 270.3
307.7 352.88 400.5 512.2 605.5 870.11 1016.35 1216.22 1492.22 1730.22 1961.13 ];
[ initial_fun_out,ps]=mapminmax( initial_fun,0,1)
initial_fun_out =
```

TABLE II. THE STANDARDİZED DATA OF QİNGDAO'S LOGİSTİCS INDİCATORS AND ECONOMİC INDİCATORS FROM 1997 YEAR TO 2010 YEAR

Year	Total Cargo Volume (Rail Way, High Way, Sea,Air)	Port Throughput	Cargo Turnover	The City's Total GDP	The City's Industrial Added Value	Retail Sales of Social Consumer Goods
1997	0	0	0	0	0	0
1998	0.1483	0.0036	0.3112	0.0170	0.0101	0.0148
1999	0.1583	0.0121	0.3460	0.0396	0.0316	0.0311
2000	0.3143	0.0612	0.6342	0.0810	0.0671	0.0525
2001	0.4907	0.1240	0.7623	0.1066	0.0999	0.0784
2002	0.6040	0.1891	0.3580	0.1481	0.1432	0.1057
2003	0.7128	0.2563	0.4000	0.2019	0.2041	0.1697
2004	0.8506	0.3334	0.6490	0.2807	0.2878	0.2231
2005	0.8209	0.4198	0.7050	0.3882	0.3900	0.3748
2006	0.8905	0.5520	0.8238	0.4901	0.4999	0.4586
2007	0.9107	0.6970	0.7386	0.6065	0.6085	0.5731
2008	1.0000	0.8225	1.0000	0.7403	0.7251	0.7313
2009	0.3112	0.8809	0.8444	0.8332	0.8414	0.8677
2010	0.4089	1.0000	0.7707	1.0000	1.0000	1.0000

C. The Determination of Regression Equation Formula

In the following multiple linear regression model,by treating the economic indicators such as Qingdao's total cargo volume and the city's industrial added value as a regression model case,the argument is the Qingdao economic and logistics index (X), the dependent variable is Qingdao total cargo (Y).

After MATLAB calculated:

B=regress(Y,X)， Y=[0； 0.1483； 0.1583； 0.3143； 0.4907； 0.6040； 0.7128； 0.8506； 0.8209； 0.8905； 0.9107； 1.0000； 0.3112； 0.4089]，X=[1 0 0 0 0 0 ;1 0.0036 0.3112 0.0170 0.0101 0.0148 ;1 0.0121 0.3460 0.0396 0.0316 0.0311 ;1 0.0612 0.6342 0.0810 0.0671 0.0525 ;1 0.1240 0.7623 0.1066 0.0999 0.0784;1 0.1891 0.3580 0.1481 0.1432 0.1057 ;1 0.2563 0.4000 0.2019 0.2041 0.1697 ;1 0.3334 0.6490 0.2807 0.2878 0.2231 ;1 0.4198 0.7050 0.3882 0.3900 0.3748 ;1 0.5520 0.8238 0.4901 0.4999 0.4586 ;1 0.6970 0.7386 0.6065 0.6085 0.5731 ;1 0.8225 1.000 0.7403 0.7251 0.7313 ;1 0.8809 0.8444 0.8332 0.8414 0.8677 ;1 1.000 0.7707 1.000 1.000 1.000];

Then the coefficients of multivariate regression model equation was get,

B=[0.0414;4.7331;0.3760;2.9794;-1.6545;-5.9826].

So the regression equation of multiple linear regression model is:

$$Y=0.0414+4.7331*X_1+0.3760*X_2+2.9794*X_3 -1.6545*X_4-5.9826*X_5 \quad (6)$$

Obviously, the model has a certain degree of correlation with the dependent variable, but the model has a certain degree of goodness of fit, but it needs to be recognized by the relevant test of the simulation.

D. Simulation Fitting and Test Analysis of Model

Use the MATLAB software to standardize the collected data about the Logistics and economic indicators from 2011 to 2014 and treat the data as the actual value.And then take the standardized independent variable data into the regression equation to get forecast value of the dependent variable[11].Contrast data and error, as shown in Table III.Finally, simulate systematically the actual value and forecast value and get the fitting graph [12], as shown in figure 2.

TABLE III. COMPARATİVE DATA AND THE ERROR TABLE OF LOGİSTİCS VOLUME FROM 2011 YEAR TO 2014 YEAR

Year	Actual Value	Predictive Value	Absolute Error Value
2011	0.5864	0.4174	0.169
2012	0.6062	0.7025	0.0963
2013	1	1.1647	0.1647
2014	0	0.1168	0.1168

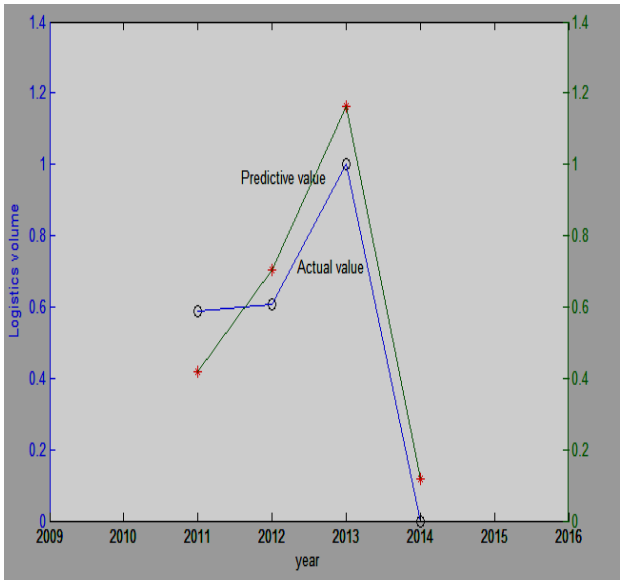

Fig.2 Fitting Chart of Predictive Value and Actual Value

From the fitting Figure 2, the comparison between the actual value of the logistic model and the forecast value is small, the distance of the two lines is close, which indicates that the model is in good fitting.But we also need to carry on the multivariate regression model correlation test to increase the persuasive power,which include standard error, goodness of fit test, significance test and so on [13].

(1) Standard deviation (the smaller the data, the stronger the degree of fit regression equation):

$$S_y = \sqrt{\frac{\sum (y-\hat{y})^2}{n-k-1}} = 0.099, \text{ (k denotes the number of independent variables)} \quad (7)$$

Estimated standard error is 0.099, it is described a high degree of regression equation model fit.

(2) The goodness of fit test (a measure of the independent variables and the dependent variable relevance):

$$\bar{R}^2 = 1 - \frac{RS/(n-k-1)}{TS/(n-1)} = 1 - (1-R^2)\frac{n-1}{n-k-1} = 0.839 \quad (8)$$

among them:

$$TS = \sum (y_2 - \bar{y})^2 \quad RS = \sum (\hat{y}_2 - \bar{y})^2$$
$$R^2 = \frac{ES}{TS} = 1 - \frac{RS}{TS} \quad (9)$$

Take a significant level 0.999, then look-up the F-distribution table which obey freedom (k,n-k-1) ,Fa = 8.89, therefore F> Fa, the regression equation is considered significant.

IV. SUMMARY

Model established in this paper can be seen, the logistics volume of Qingdao is not only closely related to the logistics index, but also affected by Qingdao's economy. Because Qingdao is a tourist city with high consumption, its coverage is very broad. This shows that the model and the reality fit well and further proves the feasibility of the model. So it can be used to predict the future of Qingdao logistics. Logistics demand forecasting is the foundation and precondition of logistics planning. It is also the key factor to meet the requirements of economic development. Furthermore, it is the critical step to be based on the theoretical basis of the large transport large logistics and build a modern logistics industry. Therefore, the government should pay more attention to the forecast of Qingdao logistics demand. So Whether the Qingdao government can pay more attention to the logistics demand forecasting has great significance for the development of Qingdao.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflicts of interest.

ACKNOWLEDGMENT

Research project supported by post doctoral research project of Qingdao City: A Study on Qingdao's Urban logistics planning based on the theory of large transportation and logistics.

REFERENCES

[1] Liya Ji. "Large logistics PK big traffic". Automotive observation,vol. 2, pp.48-49, 2014.
[2] Kejun Lao. "Baseon large transportation and develop large logistics". Ningbo Economy(Finance View), vol.7, pp.16-17, 2014.
[3] Wenming Wang, Wen Yang, Fengzhao Yin. "The operation status and development countermeasures of logistics enterprise in Qingdao development zone". Henan ScienceX, vol.3(32), pp.452-454, 2014.
[4] Leqin Wei. "A research review on regional logistics demand forecasting". Journal of Hubei University of Economics(Humanities and Social Sciences), vol. 11(11), pp. 57-58, 2014.
[5] Yu Liu, Yingxue Wu, Wenfeng Dang. "Regional logistics demand forecasting based on multiple linear regression". Logistics Engineering and Management, vol.3(36), pp. 52-53, 2014.
[6] Lei Li, Ye Liu. "Nonlinear regression prediction of the social logistics demand forecast in our country". Journal of Jiangnan University(Natural Science Edition), vol.3(13), pp. 375-377, 2014.
[7] Sanyong Niu. "Study on forecasting of sudden changes in logistics volume during e-commerce holiday discounts". Logistics Technology, vol.33(3), pp. 161-162, 2014.

[8] Changyin Zhou, Chun Yan et al. “Application of mathematical statistics”. Qingdao: School of Information Science and Engineering, Shandong University of Science and Technology, September,2011.
[9] Yangpei Cui, Qingxian Wu, Mou Chen. “Based on adaptive nervous fuzzzy inference system of air target intention prediction”. System Simulation Technology and Application, vol.15, pp. 277-278, 2014.
[10] Chuang Du. “Research on regional logistics development in Liaoning province”[D]. Liaoning: Dongbei University of Finance and Economics, 2011, pp.16-17.
[11] Lian Lin, Hua Lin. “The application of improved BP neural network in the prediction of port logistic”[J]. Traffic Information and Safety, 2009, vol.5, pp.163-164.
[12] Feng Zhang, Xiuqing Yin, Huizhong Dong. “Analysis on the effect of logistics demand forecast of combined gray forecast model”[J]. Logistics Technology, 2014, vol.9, pp. 222-223.
[13] Lin Zhong. “Forecast and analysis of regional logistics volume in Chengdu”[D]. Sichuan: University of Electronic Science and Technology, 2011, pp. 22-50.