International Journal of Engineering, Science and Mathematics
Vol. 7,Issue 12, December 2018,
ISSN: 2320-0294 Impact Factor: 6.765
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

LOGISTICS QUANTITY FORECAST BASED ON FACTOR ANALYSIS AND LOGISTIC MODEL

Ying Zhang*

Abstract

The factor analysis of Chinese logistics quantity and its influencing factors from 2006 to 2015 Using SPSS software, a multivariate statistical analysis model is established, which takes the object flow as the interpreted variable. The factor analysis method is adopted to reduce the dimension of the influence factors of the logistics quantity, reduction approach with the purpose of eliminating the multrcollinearity resulting from variables logistics quantity. Furthermore , a logisticbased forecast model is set up accordingly to further predict the future logistics quantity in China. The empirical results show that the longer-period Comprehensive economic Development Value value has less forecasting error.

Copyright © 2018 International Journals of Multidisciplinary Research Academy. *All rights reserved.*

Keywords:

Factor analysis;
Logistic Model;
Logistics quantity;
Forecast.

Author correspondence:

Ying Zhang,
College of Science, Binzhou University, Binzhou, China

1. Introduction

The flow of matter is an important economic maicator to measure the economic development of a country or region. China's internet economy rapid development stimulates the growing demand for l logistics quantity, studying of the main factors affecting the change of physical flow in China and prediction of logistics quantity by a scientific method is a great significance not only to analyze the present situation of China's logistics industry, to reveal the deep problems facing the development of logistics industry, but to describe the changing trend of market demand in the future. At the same time, it can provide reference basis for government departments to formulate macroscopic logistics measures, optimize resource allocation and reduce decision risk, and provide feasible suggestions for the development of related logistics quantity.

* College of Science, Binzhou University, Binzhou, China

International Journal of Engineering, Science and Mathematics
Vol. 7,Issue 12, December 2018,
ISSN: 2320-0294 Impact Factor: 6.765
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

At present, there are many methods to predict logistics quantity, The methods that have a wide application is neural network model, multivariate linear regression prediction model, time series prediction model, grey system grediction godel, etc[1-9].

In this paper, Using the logistics volume data in recent years analyzes and synthesizes the factors affecting the change of matter flow by factor analysis method, Interpreting variables is reduce ddimensions by factor analysis method. At the same time,it eliminate multiple collinearity between variablesand get the value of Comprehensive economic Development Value indicators reflecting different years. The nonlinear regression relationship between the value of Comprehensive economic Development Value index and the amount of logistics is determined by scatter plot, we Considering that the flow of matter is affected by many factors, we will forecast the logistics quantity in the next few years by logistic model.

2. The Factor Analysis Model

The core of factor analysis is reflect most of the information of the original variable with fewer independent factors. The idea of factor analysis can be expressed in the following mathematical model.Suppose original variables is $x_1, x_2, \mathrm{L}\ , x_q$, and each variable's mean value(be standardized treatment)is 0,it's standard deviation is 1,then, factor analysis model is

$$\begin{cases} x_1 = a_{11} f_1 + a_{12} f_2 + a_{13} f_3 + \ldots + a_{1k} f_k + \varepsilon_1 \\ x_2 = a_{21} f_1 + a_{22} f_2 + a_{23} f_3 + \ldots + a_{2k} f_k + \varepsilon_2 \\ x_3 = a_{31} f_1 + a_{32} f_2 + a_{33} f_3 + \ldots + a_{3k} f_k + \varepsilon_3 \\ \mathrm{M} \\ x_q = a_{q1} f_1 + a_{q2} f_2 + a_{q3} f_3 + \ldots + a_{qk} f_k + \varepsilon_k \end{cases} \quad (1)$$

Where $f_1, f_2, f_3, \ldots, f_k$ is factor variables, It can also be expressed as a matrix in the form of :

$$X=AF+\varepsilon$$

Where A is malled mactor load matrix , $a_{ij}(i=1,2,\mathrm{L}\ ,\ \mathrm{q};\ j=1,2,\mathrm{L}\ ,k)$ is malled mactor load, it is the load of the i original variables on the j factor. If variable x_i is seen as a factor of k dimension factor space,then a_{ij} is the projection on axis f_j of x_i ,it is equal to Standardized regression coefficient in multivariate linear regression model. ε is specific factors, it is expressed the part of the original variable that cannot be interpreted by the factor. It's mean value is 0 , it is equal to residuals in multivariate linear regression model[10].

3. Empirical Research

3.1. Selection of Indicators

There are many factors that affect the logistics quantity, In this paper, 6 main factors affecting the flow of matter are selected, They are Real GDP per capita, Total Retail Sales of Consumer Goods, turnover of freight traffic, Consumption level of residents,

International Journal of Engineering, Science and Mathematics
Vol. 7,Issue 12, December 2018,
ISSN: 2320-0294 Impact Factor: 6.765
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

Operational mileage of railways and Highway mileage, represented by x_1 , x_2 , x_3 , x_4 , x_5 and x_6 .Units are YUAN, Billion Yuan, Billion tons, Kilometers,yuan, Million km and Million km. Logistics quantity(freight) is represented by y, Unit is ton.Year is represented by t, In this paper, interpreting variables is reduce dimensions by factor analysis method,at same time,it will eliminate multiple collinearity between variables. The data(2006～2015) are shown in table 1.

Table 1. The Data of Dreight and its Influencing Factors

t	y	x_1	x_2	x_3	x_4	x_5	x_6
2007	2037060	16738	79145	88839.9	6416	7.71	345.7
2008	2275822	20505	93572	101419	7572	7.8	358.37
2009	2585937	24121	114830	110300	8707	7.97	373.02
2010	2825222	26222	133048	122133	9514	8.55	386.08
2011	3241807	30876	158008	141837	10919	9.12	400.82
2012	3696961	36403	187206	159324	13134	9.32	410.64
2013	4100436	40007	214433	173804	14699	9.76	423.75
2014	4098900	43852	242843	168014	16190	10.31	435.62
2015	4167296	47203	271896	181668	17778	11.18	446.39
2016	4175886	50251	300931	178356	19397	12.1	457.73

Data sources: statistics bureau

Table 2. Correlation coefficient matrix

	x_1	x_2	x_3	x_4	x_5	x_6
x_1	1.000	.996	.976	.997	.971	.996
x_2	.996	1.000	.955	.999	.988	.991
x_3	.976	.955	1.000	.960	.912	.976
x_4	.997	.999	.960	1.000	.982	.990
x_5	.971	.988	.912	.982	1.000	.969
x_6	.996	.991	.976	.990	.969	1.000

It can be seen from the correlation coefficient matrix of Table 2 that the simple correlation coefficients of the independent variables have reached 0.9 or more, and there are serious multicollinearity between the independent variables. A description of the public factors that can synthesize the common characteristics of the original variables from it, These 6 variables are suitable for factor analysis.

3.2. Factor Analysis Results

3.2.1. KMO and Bartlett's Test

International Journal of Engineering, Science and Mathematics
Vol. 7,Issue 12, December 2018,
ISSN: 2320-0294 Impact Factor: 6.765
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

Table 3. KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		.793
Bartlett's Test of Sphericity	**Approx. Chi-Square**	168.016
	df	15
	Sig.	.000

It can be seen from the Table 2 that the Bartlett Tes' Observation values of Approximate Chi-Square is 168.016, If Significance level is equal to 0.05, Since the probability p-value is close to 0, It is less than the Significance level, The original hypothesis should be rejected,the alternative hypothesis should be Acceptanced, it is considered that the correlation coefficient matrix is significantly different from the unit array.at same time, , the value of the KMO test statistics is 0.793, It's close to 0.8. So the original variable is suitable for factor analysis.

3.2.2. Communalities of Factor Analysis

Table4. Communalities

	Initial	**Extration**
x_1	1.000	.997
x_2	1.000	.995
x_3	1.000	.946
x_4	1.000	.995
x_5	1.000	.960
x_6	1.000	.993

Table 4 shows the Communalities' data for all variables, The first column in the right column is the Initial Communalities' data of factor analysis, It shows: It is shown that if extract 6 factor from the original 6 variables by principal component analysis,then, all the Communalities of the original variable can be interpreted,and all the Communalities of those variable is 1. The second column is the Extration Communalities' data when the Characteristic root is greater than 1. It can be seen from data that the proposed factor explain most of the information for 6 variables(All are greater than94%), information that loss is very little. Therefore, it can be considered that the effect of this factor extraction is very perfect.

3.2.3. Communalities of factor analysis

It can be seen from the Table 5 that The characteristic of the first factor is 5.886, A 98.108% of the original 6 variables' variance can be explained by it. Cumulative Variance contribution Rate is 98.108% .The characteristic of other factors are 0.096,0.010,0.007,0.000,0.000.Therefore, the first factor can be selected as the main factor.

International Journal of Engineering, Science and Mathematics
Vol. 7,Issue 12, December 2018,
ISSN: 2320-0294 Impact Factor: 6.765
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

Table5.Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of squared Loadings		
	Total	% of Variance	Cumulative%	Total	% of Variance	Cumulative%
x_1	5.886	98.108	98.108	5.886	98.108	98.108
x_2	.096	1.594	99.702			
x_3	.010	.172	99.874			
x_4	.007	.119	99.993			
x_5	.000	.005	99.998			
x_6	.000	.002	100.000			

3.2.4. Component Matrix

Table6. Component Matrix

	Component
	1
x_1	.999
x_2	.998
x_3	.972
x_4	.998
x_5	.980
x_6	.997

Table 6 is the factor load matrix, It is the core content of factor analysis. It can be seen that 6 variables have a high load on the first factor. It means that the correlation between Real GDP per capita, Total Retail Sales of Consumer Goods, turnover of freight traffic, Consumption level of residents, Operational mileage of railways,Highway mileage and the 1th factor is higher(All are greater than97%), it show that the 1th factor is important. It has well explaine the original variables. It can be viewed as a Comprehensive economic Development Value value,it represented by F_1.

3.2.5. Component Score Coefficient

It can be seen from the Table 7 that The Component Score Coefficient of Real GDP per capita, Total Retail Sales of Consumer Goods, turnover of freight traffic, Consumption level of residents, Operational mileage of railways,Highway mileage are 0.170,0.169,0.165,0.169,0.166,0.169.So you can write out factor scoring functions :

$$F_1 = 0.170x_1 + 0.169x_2 + 0.165x_3 - 0.169x_4 + 0.166x_5 + 0.169x_6 \quad (2)$$

International Journal of Engineering, Science and Mathematics
Vol. 7,Issue 12, December 2018,
ISSN: 2320-0294 Impact Factor: 6.765
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

Table7. Component Score Coefficient Matrix

	Component
	1
x_1	.170
x_2	.169
x_3	.165
x_4	.169
x_5	.166
x_6	.169

3.3. Improved Logistic Model

3.3.1. Component Score Coefficient

From formula (2),we got the factor score for each year from 2006 to 2015, These factor scores represent the comprehensive development value of the economy in different years. As a result, factor scores can be used to predict the volume of logistics in the next few years.

First,we make a scatter plot between the Comprehensive economic Development Value value F_1 and the logistics quantity Y in different years, See in Figure 1 :

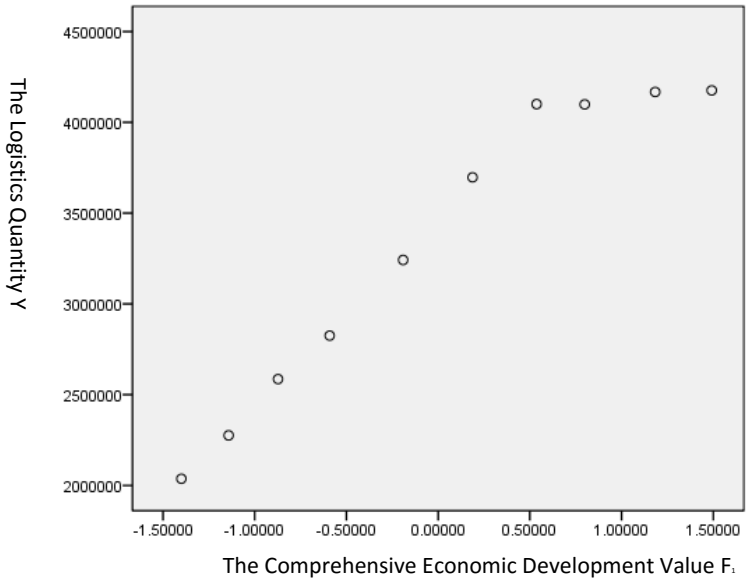


Figure 1 . The Comprehensive economic Development Value Value and The logistics quantity

It can seen from Figure 1 that the logistics quantity of grows rapidly at beginning, then slightly lower than in previous years, and then the pace of growth slowed.there are obvious

International Journal of Engineering, Science and Mathematics
Vol. 7,Issue 12, December 2018,
ISSN: 2320-0294 Impact Factor: 6.765
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

nonlinear regression relationship between The Comprehensive economic Development Value value and the logistics quantity , Combining with the practical significance of this paper. As the flow of things will not only be restricted by the industry's own level of development, but also by Economic growth rate, changes in residents ' consumption level, development status of transportation industry, resource environment and other objective factors, It is impossible to grow indefinitely, so it is appropriate to choose the logistic model for analysis.

The curve of logistic model is roughly S-type, which shows a steady growth trend in the early stage, then rises rapidly and tends to stabilize after reaching a certain scale. Its function model can be represented as:

$$y = \frac{1}{1/\mu + \beta_0 \beta_1^{F_1}} \qquad (3)$$

y is the logistics quantity, F_1 is the Comprehensive economic Development Value , The upper limit value in the logistic function must be greater than the maximum value of the due variable,Different μ gets different results. According to the actual situation of this paper, and the results obtained from different upper limit values are compared and synthesized, It is considered appropriate to take 100000000 of the upper limit value in this paper[11].

3.3.2. Model Summary and Parameter Estimates

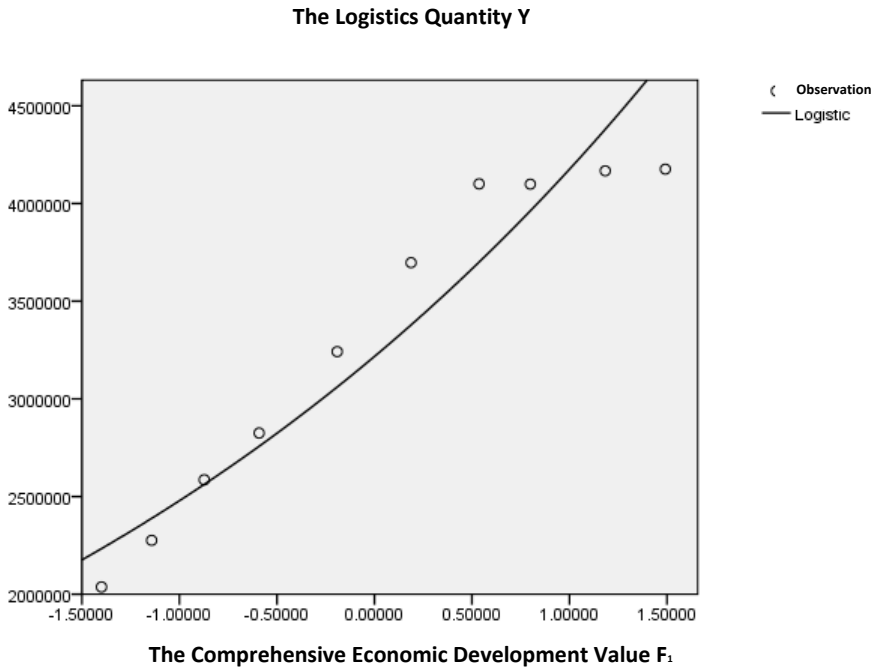


Figure2 . Fitting Diagram of Actual Data andLogistic Regression Equation

The model summary and the Parameter estimates are shown in table 8.

Table8. Model Summary and Parameter Estimates

International Journal of Engineering, Science and Mathematics
Vol. 7,Issue 12, December 2018,
ISSN: 2320-0294 Impact Factor: 6.765
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

Equation	Model Summary					Parameter Estimates	
	R Square	F	df1	df2	Sig.	b0	b1
Logistic	.920	91.539	1	8	.000	3.108E-7	.771

The sample determination coefficient of this regression model $R^2 = 0.920$,Test statistic Quantity F=91.539, The p-value is close to 0, less than the significant level 0.05, which indicates that the model has statistical significance. The estimation equation of t the logistics quantity can then be obtained:

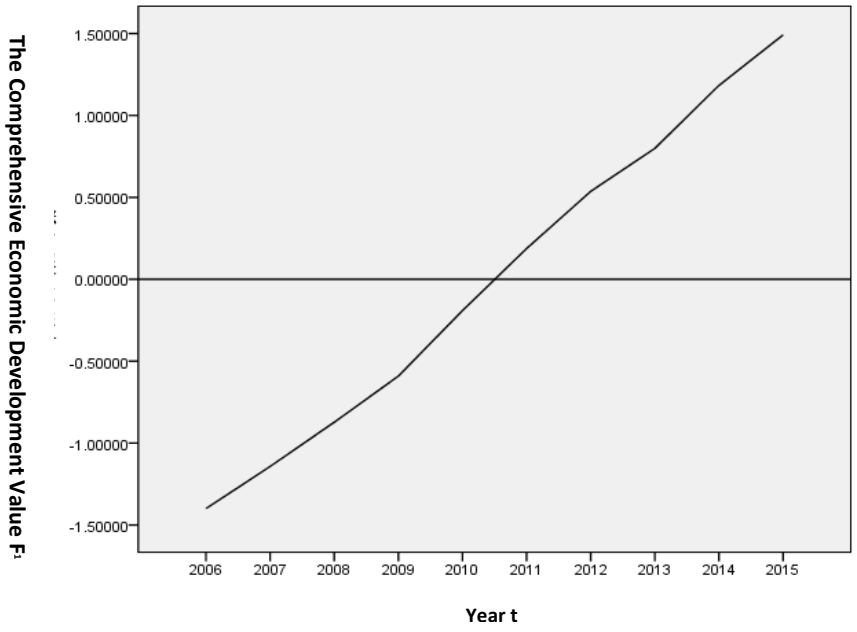
$$y = \frac{1}{1/10^8 + 3.108 \times 10^{-7} \times 0.771^{F_1}} \tag{4}$$

3.4. Prediction of Logistics Quantity

3.4.1. Prediction of Comprehensive economic Development Value

First,We draw a time series diagram of the Comprehensive economic development value F_1 .Then From Figure 3,We can see that the time series diagram of the Comprehensive economic development value Overall showing an upward trend from 2007 to 2016 year ,it is roughly linearly related to time t. Therefore, the comprehensive economic development prediction equation is:

$$F_1 = -663.181 + 0.330t \tag{5}$$



International Journal of Engineering, Science and Mathematics
Vol. 7,Issue 12, December 2018,
ISSN: 2320-0294 Impact Factor: 6.765
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

Figure3. Time Series Diagram of The Coprehensive Economic Development value（2007-2016）

3.4.2. Comparative analysis of forecast results

According to Formula (5), the Comprehensive economic development value of each year can be predicted. By Compared the prediction results and Actual Passenger volume from 2007 to 2016 ,we can calculate the relative error of each year. By calculation, we know that the Average relative error is 7.05%.

Table 9. Comparison of prediction results and ActualLogistics quantity volume

Year	Actual Logistics quantity volume	Forecast value	Error	Year	Actual Logistics quantity volume	Forecast value	Error
2007	2037060	2300197	12.92%	2012	3696961	3489833	5.60%
2008	2275822	2501167	9.90%	2013	4100436	3790711	7.55%
2009	2585937	2719208	5.15%	2014	4098900	4116422	0.43%
2010	2825222	2955681	4.62%	2015	4167296	4468820	7.24%
2011	3241807	3212039	0.92%	2016	4175886	4849860	16.14%

3.4.2. Forecast Value of the Logistics Quantity

The comprehensive economic development value of the future years can be calculated by the formula (5),it broughted into formula (4),we forecast the value of the Logistics Quantity , the result is shown in table 10.From table 10,we can see that the Logistics Quantity will continue to grow rapdely in the next few years. , it will 5261601 million tons by 2016, 7260163 million tons by 2020, 107733224 million tons by 2026, It's a double more than it is now. The rapid growth of Logistics Quantity is bound to pose a serious challenge to the development of transportation and related logistics industry,we need to be prepared in advance.

Table 10. Forecast Value of the Logistics Quantity(2017-2025)

Year t	Comprehensive economic Development Value F_1	Logistics quantity y
2017	2.099	5261601
2018	2.429	5706201
2019	2.759	6185916
2020	3.089	6703093
2021	3.419	7260163
2022	3.749	7859630
2023	4.079	8504054
2024	4.409	9196043
2025	4.739	9938224
2026	5.069	10733224

International Journal of Engineering, Science and Mathematics
Vol. 7,Issue 12, December 2018,
ISSN: 2320-0294 Impact Factor: 6.765
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

4. Conclusion

For the easy collection of data and statistical reasons, this paper uses the amount of cargo transport instead of logistics volume. In this paper, 6 main factors affecting the flow of matter are selected, They are Real GDP per capita, Total Retail Sales of Consumer Goods, turnover of freight traffic, Consumption level of residents, Operational mileage of railways and Highway mileage, interpreting variables is reduce dimensions by factor analysis method,at same time,it will eliminate multiple collinearity between variables. At the same time, the scoring coefficients of the main factors are estimated by regression analysis, and the scores of these factors fully reflect the comprehensive economic development values of different years. Then the scatter chart between the comprehensive economic development value and the logistics quantity is made, fitted with the logistic model in the curve estimation, the fitting equation is written according to the estimation values of each parameter, and then the logistics quantity is predicted in the next few years. Based on the above studies, it is concluded that:

(1) In this paper, many factors affecting the flow of matter are considered comprehensively, and the dimension is reduced by factor analysis, which avoids the multiple collinearity effects caused by multiple regression analysis alone, and the phenomenon that the predicted data is not in conformity with the actual economic significance appears.

(2) In this paper, the factor score of physical flow in each year is obtained by using the method of factor analysis in SPSS, and these factors represent the comprehensive economic development value of the flow rate of each year. The scatter plot between the flow rate and the comprehensive economic development value can be determined that there is a clear nonlinear regression relationship between the two. Considering that the traffic flow will not only be restricted by the development of their own industry, but also by the economic growth rate, the development of the transportation industry, the consumption level of residents, the social resources and environment and many other objective factors, it is impossible to increase the expansion of the infinite, so it is appropriate to choose the logistic model for regression analysis. And the prediction results show that the relative error is small in a short period of time.

(3) With the continuous growth of freight volume, the transportation industry will face more severe challenges, the country should stabilize economic growth, improve people's income level, increase investment in fixed assets, stimulate consumption, speed up the construction of basic transportation networks such as roads, railways and aviation, optimize the allocation of resources, and rationally plan logistics parks The government departments should step up the formulation of relevant policies and laws and regulations of the logistics industry, continuously promote and standardize the development of the logistics industry, and form a reasonable control mechanism for the price of goods; The relevant logistics enterprises should also start from a long-term perspective, formulate reasonable development plans, and continuously improve the service level and logistics and distribution capacity.

zAcknowledgments

This work is supported by Science and Technology Project of Shandong Province under grant J14li54,BYHXKC201505. Also the authors would like to thank the editor and the anonymous referee for their valuable comments and suggestions, which helped us to improve this paper greatly.

References

[1] Wang Xiaozhong. "Study on Prediction Method of Matter Flow," *Wuhan University*, pp. 170-202, 2005.

[2] Wei Lianyun,Pang mingbao. "Prediction of Object Flow Based on Neural Network," *Journal of Chang'an University(Social Science Edition),* vol. 06,pp. 55-59, 2004.

[3] Chen Zixia,Li Dong. "Prediction of urban physical flow based on improved grey theory," *China Management Informationization,* vol. 08,pp. 72-74, 2009.

[4] Hu Baoyua,Wang Xiaokun,Chen Yan. Capacity prediction of logistics park based on double improved gray Markov model," *Journal of Dalian Maritime University,* vol. 01,pp. 91-94, 2011.

[5] Guan Chunyan,Su Jianmei. "Prediction of regional traffic flow in Wuhan based on stepwise regression analysis," *Guide to Business*, vol. 16,pp. 133-135, 2012.

[6] Zhang Chunping,Yin Ying. "Based on the Optimal Weighting Combination Model in Logistics Volume Forecasting," *Logistics Engineering and Management* , vol. 01,pp. 85-88+91, 2013.

[7] Yue Hui,Wu Bo,Shan Cui,Chen Yanru. "Logistics Forecasting Based on Support Vector Machine: A Case Study of Chengdu," *Communications Standardization,* vol. 15,pp. 87-91, 2014.

[8] Li Wanchun. "An empirical study on the prediction of physical flow based on time series and regression analysis model," *Business Era* , vol. 20,pp.22-23, 2014.

[9] Guo Haihua,Li Ruirui. Study on the prediction of physical flow in China based on clustering analysis and time series method," *Journal of Commercial Economics,* vol. 20,pp.29-31, 2015.

[10] Ren Da,Wang Dongping. "Synthetic Evaluation on the Income Level in Urban Areas Based on Factor Analysis Method," *Journal of Chongqing Institute of Technologyt,* vol. 04,pp.127-132, 2014.

[11] Li Xiaohui,Tan Qingmei,Zhao Aimei. "China's Passenger Traffic Forecast based on Factor Analysis and Logistic Model," *Journal of Systems & Management,* vol. 03,pp.444-450, 2014.