

Analysis of Logistics Demand of Kashagar Administrative Offices Based on Multiple Linear Regression Model

Aimin Deng, Hancheng Yang

Abstract— In this paper, selecting the cargo transportation volume as the index measuring the logistics demand level, we analyzed empirically the economic data of Kashagar Administrative Offices for the period between 2000 and 2016, used the Eviews program to build a multiple linear regression model and found that the Industrial output value and the Total wages of employed persons in urban non-private units had marked impact on the cargo transportation volume of the city.

Index Terms— Logistics demand; Multiple linear regression model; Empirical analysis; Kashagar Administrative Offices.

I. INTRODUCTION

Logistics is the foundation of ensuring the normal operation of the city, and it promote the stable development of the urban economy and others. The continuous transformation of the economy and the continuous optimization of the industrial structure means that our logistics industry needs to be professionalized and socialized.

Kashagar Administrative Offices is an important node in the strategy of "The Belt and Road". The starting point of China-Pakistan Economic Corridor(CPEC)is in Kashagar Administrative Offices, which is the city can not be replaced by others in this plan. Kashagar Administrative Offices has become the "main hub of the west", the "main gateway" to the west, and the "important growth pole" in the economic development of the western region of China. The good development of Kashagar Administrative Offices is related to China's external security. This area is the strategic node and the throat of China's western international environment, and the strategic throat and strategic reserve of the national land energy transmission line. CPEC strengthened the economic status of Kashagar Administrative Offices. The construction of CPEC is the first large multinational engineering of "The Belt and Road" strategy, it has a very strong demonstration effect. The Central Asian countries near the corridor are also paying attention to this construction process. Kashagar Administrative Offices ,in the new round of development, carrying west and east two big market of material flow of energy and science and education exchanges, these market will have new opportunities for its development.

Rapid economic growth has led to the increase in the demand of logistics services, and it also poses a severe challenge to the development of the logistics industry. Therefore, the

analysis of logistics demand in Kashagar Administrative Offices can regulate and guide people's logistics management activities, so as to take appropriate strategies and measures to maximize profits and further promote the overall development of local logistics industry. Predicting the volume of logistics demand by the established model in this paper will help us to see what factors in the Kashagar Administrative Offices have a greater impact on the logistics demand in Kashagar Administrative Offices.

II. INDEX AND DATA SELECTION

Based on the data of the product of the three industry ,total retail sales of consumer goods and other index in the Kashagar Administrative Offices for 2000-2016 years, this paper makes an empirical analysis on the factors affecting the logistics demand in this region.

A. Variable selection

This paper selects freight traffic (Y) to reflect a regional logistics demand, the freight volume consists of three blocks, the completion of the railway freight volume, the completion of the highway freight volume, civil aviation cargo throughput.

The increase of urbanization level leads to the diversification and individuation of urban residents' consumption demand, which makes the demand for commodities for urban residents grow. The index of the Total retail sales of consumer goods(X_1) in the regional society reflects the size of the urban commodity demand in the regional logistics planning. Therefore, the total retail sales of consumer goods are also the main factors that affect the demand of the logistics market. The three industry output value reflects the development situation of the first industry , the second industry and the third industry .The contrast between these three output value reflect the industrial structure of a regional economy. We set the primary industry product to X_2 , the secondary industry product to X_3 , the tertiary industry product to X_4 . Government expenditure(X_5), a part of the government expenditure of a region will be invested in the construction of the transportation system in this area, increasing the mileage of the area and improving the transportation environment. A good transportation environment will improve the speed of transportation and thus expand the volume of freight .The logistics is responsible for providing the material produced in the industry to the people for consumption. A high industrial output value(X_6) will put forward new requirements to the logistics industry. Gross output value of agriculture, forestry, animal husbandry and fishery(X_7), The development of this industry will also increase the demand for cargo transport. In

Aimin Deng, the School of Economics and Trade, Hunan University, Changsha, China, (+86)18975184867;

Hancheng Yang, the School of Economics and Trade, Hunan University, Changsha, China, (+86)18374889942

particular, the development of animal husbandry and fisheries, the transport of products in these industries, we use cold chain logistics to transport. The total wages of employed persons in urban non-private units (X_8), it reflects the level of the total wage in a region, and the economic development of the region will bring about the rise of wages. With the increase of wages, people's disposable income will rise, people will have more money to buy consumer goods, and the increase of demand for consumer goods will increase the volume of freight.

B. Data selection and processing

The natural logarithm of the data of each sequence does not change the Cointegration and causality of each sequence. In order to make data easy to get a stationary sequence and eliminate the heteroscedasticity phenomenon in time series, we do logarithmic processing to the data before analyzing it. Add the Ln to the correlation sequence after the logarithm process. The processed data sequences are listed in Table 1.

Table 1 The processed data sequence of the variables

Year	LnY	LnX ₁	LnX ₂	LnX ₃	LnX ₄	LnX ₅	LnX ₆	LnX ₇	LnX ₈
2016	8.066	5.168	5.561	5.254	5.732	6.238	4.260	6.238	14.291
2015	8.029	5.125	5.424	5.484	5.745	6.086	4.890	6.172	14.278
2014	8.025	5.000	5.330	5.348	5.605	5.937	4.780	6.081	14.006
2013	7.610	4.877	5.254	5.220	5.485	5.848	4.653	5.963	13.911
2012	7.543	4.705	5.166	4.964	5.293	5.748	4.398	5.886	13.708
2011	7.440	4.494	5.015	4.675	5.089	5.586	4.075	5.729	13.435
2010	7.292	4.296	5.022	4.175	4.964	5.291	4.175	5.560	13.221
2009	7.204	4.250	4.748	3.889	4.788	4.996	3.995	5.480	12.997
2008	6.951	4.057	4.640	3.759	4.631	4.594	3.789	5.340	12.869
2007	6.931	3.839	4.511	3.919	4.319	4.143	3.356	5.098	12.842
2006	6.997	3.629	4.234	3.526	4.111	3.851	2.862	4.862	12.700
2005	7.003	3.437	4.094	3.332	3.871	3.642	2.407	4.676	12.571
2004	6.811	3.271	3.988	3.096	3.773	3.405	2.054	4.557	12.470
2003	6.633	3.104	3.914	2.918	3.597	3.345	1.421	4.476	12.406
2002	5.841	2.983	3.852	2.660	3.434	3.160	1.306	4.355	12.210
2001	5.969	2.914	3.718	2.832	3.366	2.978	1.418	4.248	12.071
2000	6.220	2.833	3.722	2.312	3.129	2.603	1.539	4.146	11.844

III. ECONOMETRIC ANALYSIS AND EMPIRICAL RESULTS

A. Data test

We first make the Stationary test and test the smoothness of the data. The stability test is the basis test for the analysis of time series data. The stationarity of the time series, refers to the statistical law of time series does not change with time, if there are two column time data showed the same variation trend, then the data is non-stationary, even if they do not have any meaningful relationship, but in regression analysis these two column time data may also show higher the coefficient of

determination. Therefore, it is necessary to analyze the stationarity of the time series data of the observed values.

First of all, unit root test (ADF) is performed on LnY, LnX₁, LnX₂, LnX₃, LnX₄, LnX₅, LnX₆, LnX₇ and LnX₈ respectively, and the trend of each variable sequence diagram is observed, and then according to the graph determine whether we select the intercept and trend items. After imply the unit root test of the original sequence, it is found that many variables are non-stationary. Then the first order difference and the two order difference are taken for each variables, and we carried out the unit root test of the differential sequence again. The results of the stability test are shown in Table 2.

Table 2 The result of the stability analysis (α=0.05)

	Result	LnY	LnX ₁	LnX ₂	LnX ₃	LnX ₄	LnX ₅	LnX ₆	LnX ₇	LnX ₈
level	P-value	0.9800	0.9277	1.0000	0.9969	1.0000	0.9304	0.5644	0.9338	1.0000
	Result	Non-stati onary	Non-stati onary	Non-stati onary	Non-stati onary	Non-stati onary	Non-stati onary	Non-stati onary	Non-stati onary	Non-stati onary
First order difference	P-value	0.0062	0.1922	0.0018	0.0072	0.0671	0.2640	0.5740	0.3159	0.0228
	Result	stationary	Non-stati onary	stationary	stationary	Non-stati onary	Non-stati onary	Non-stati onary	Non-stati onary	stationary
Two order difference	P-value	0.0048	0.0116	0.0065	0.0003	0.0003	0.0334	0.0292	0.0398	0.0034
	Result	stationary	stationary	stationary	stationary	stationary	stationary	stationary	stationary	stationary

We first make the Co-integration test. Since the original time series selected in this paper is non-stationary, it becomes a stationary sequence after the two order difference, it lead to that the total amount will lose the long term information.

The best method of multiple co-integration test is JJ test. However, due to the limited number of observations in this paper, the co-integration test we selected is based on regression residuals, we call it co-integration test of single equation.

First, the regression equation is established to generate the sequence of residual, and then the ADF unit root test is carried out on the residual sequence. The result is shown in Table 3. The output results show that the probability is $P=0.0006$, so at the 0.05 level, the residual does not exist in the unit root and it is stationary. There is a long-term co-integration relationship between the freight volume (Y) and the 8 independent variables.

Table 3 The result of the Co-integration test

Null Hypothesis: RESID has a unit root				
Exogenous: None				
Lag Length: 0 (Automatic based on SIC, MAXLAG=3)				
Augmented Dickey-Fuller test statistic			t-Statistic	Prob.*
			-3.987698	0.0006
Test critical values:	1% level		-2.717511	
	5% level		-1.964418	
	10% level		-1.605603	

B. Establishing a preliminary model

After the above analysis and test, we set up a preliminary multivariate linear regression model. Freight traffic (Y), Total retail sales of consumer goods (X_1), Primary industry product (X_2), Secondary industry product (X_3), Tertiary industry product (X_4), Government expenditure (X_5), Industrial output value (X_6), Gross output value of agriculture,

forestry, animal husbandry and fishery (X_7), Total wages of employed persons in urban non-private units (X_8), considering above index, we set up the regression equation, the multivariate element linear regression model as follows:

$$Y=C+\beta_1\text{Ln}X_1+\beta_2\text{Ln}X_2+\dots+\beta_8\text{Ln}X_8+\mu(1)$$

In this paper, using the regression function of Eviews, the coefficients of linear regression equation are estimated by the least square method. The results are shown in Table 4.

Table 4 The result of preliminary linear regression

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LnX ₁	0.623209	2.227378	0.279795	0.7867
LnX ₂	-1.046234	1.367216	-0.765230	0.4661
LnX ₃	-0.656520	0.611013	-1.074478	0.3139
LnX ₄	-1.407957	2.476219	-0.568592	0.5852
LnX ₅	0.249073	0.841674	0.295926	0.7748
LnX ₆	0.631570	0.540026	1.169518	0.2759
LnX ₇	0.009487	3.109067	0.003051	0.9976
LnX ₈	2.115150	1.630601	1.297160	0.2307
C	-12.45688	15.83172	-0.786831	0.4541
R-squared	0.934542	Mean dependent var		7.092018
Adjusted R-squared	0.869084	S.D. dependent var		0.674631
S.E. of regression	0.244097	Akaike info criterion		0.322548
Sum squared resid	0.476666	Schwarz criterion		0.763661
Log likelihood	6.258338	Hannan-Quinn criter.		0.366396
F-statistic	14.27700	Durbin-Watson stat		2.142002
Prob(F-statistic)	0.000547			

IV. UNITS

From the output results, we can see that when a significant level $\alpha=0.05$, for every explanatory variable, the critical value of the variable is $t_{\alpha/2}(n-k-1)=t_{0.025}(8)=2.306$. As shown in Table 4, The absolute value of the T value of variables LnX₁, LnX₂, LnX₃, LnX₄, LnX₅, LnX₆, LnX₇ and LnX₈ is less than 2.306, that is, these variables are failed to

pass the t test. Therefore, we considering that multiple collinearity exist among various explanatory variables, it is necessary to modify the model.

C. Model correction

We first carry out multiple collinear tests on the model. If there is a correlation between two or more explanatory variables, it is called the existence of multiple collinearity. In this paper, the correlation coefficient matrix is used to test the multiple collinearity. As a result, see Table 5.

Table 5 The result of multiple collinearity test

Correlation	LnX ₁	LnX ₂	LnX ₃	LnX ₄	LnX ₅	LnX ₆	LnX ₇	LnX ₈
LnX ₁	1.000000							
LnX ₂	0.994969	1.000000						
LnX ₃	0.985751	0.976107	1.000000					
LnX ₄	0.998277	0.995175	0.984858	1.000000				
LnX ₅	0.992325	0.994121	0.977076	0.995569	1.000000			
LnX ₆	0.967870	0.964742	0.944440	0.969702	0.960989	1.000000		
LnX ₇	0.998131	0.996780	0.980619	0.998678	0.996312	0.970526	1.000000	
LnX ₈	0.984781	0.977420	0.988608	0.982909	0.974913	0.918213	0.978418	1.000000

Because the correlation coefficients of each explanatory variable in Table 5 are more than 0.8, we think there is a serious multiple collinearity in the model. Therefore, we use a stepwise regression method to screen variables. This method is used to determine the final independent variable.

In this paper, the backward screening method in stepwise regression is selected to eliminate the variables that do not meet the requirements.

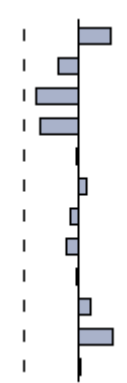
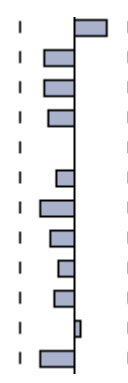
First of all, all the explanatory variables are introduced into the regression equation, and then test the regression equation.

Then, in the significance test of regression coefficient of the variables, if there are one or more variables is not significant, the least significant variable is removed, and re-established the regression equation again after the test. The above steps are repeated, until the regression coefficients of all the variables in the regression equation are all through the test, finally the equation is set up. By stepwise regression, the variables LnX₁, LnX₂, LnX₃, LnX₄, LnX₅, LnX₇ were eliminated in turn, and the regression results is in Table 6 .

Table 6 The result of model stepwise regression

Variable	Coefficient	Std. Error	t-Statistic	Prob.*
LnX ₈	0.493384	0.012259	40.24648	0.0000
LnX ₆	0.201032	0.045828	4.386665	0.0005
R-squared	0.911181	Mean dependent var		7.092018
Adjusted R-squared	0.905260	S.D. dependent var		0.674631
S.E. of regression	0.207650	Akaike info criterion		-0.195791
Sum squared resid	0.646780	Schwarz criterion		-0.097766
Log likelihood	3.664225	Hannan-Quinn criter.		-0.186047
Durbin-Watson stat	1.395261			

Table 7 The result of Sequence correlation test

Auto-correlation	Partial correlation		AC	PAC	Q-stat	Prob
		1	0.279	0.279	1.5707	0.210
		2	-0.173	-0.272	2.2157	0.330
		3	-0.371	-0.273	5.3868	0.146
		4	-0.339	-0.239	8.2358	0.083
		5	-0.020	0.005	8.2461	0.143
		6	0.068	-0.161	8.3819	0.211
		7	-0.080	-0.317	8.5864	0.284
		8	-0.101	-0.213	8.9538	0.346
		9	-0.019	-0.152	8.9679	0.440
		10	0.099	-0.179	9.4229	0.492
		11	0.306	0.045	14.462	0.208
		12	0.024	-0.315	14.498	0.270

It can be seen from the output results that R²=0.911, the model has a high degree of fitting to the sample. There is no F value because there is no intercept item. At the significant

level of $\alpha = 0.05$, and a critical value $t_{\alpha/2}(n-k-1)=t_{0.025}(14)=2.145$, when a significant test of each variable is carried out

for each explanatory variable, it is known from Table 6 that both $\text{Ln}X_6$ and $\text{Ln}X_8$ have passed the t test.

Secondly, we make the sequence correlation test. One of the basic assumptions of multivariate linear regression model is that the random interference terms of models are independent or irrelevant. If the model's random interference term violates the independent basic assumptions, it is called the sequence correlation, so we need to test the auto-correlation of the models. See the Q statistics in the regression equation window about the residuals, as shown in Table 7. The residual sequence is located within the dotted line, that is, there is no sequence correlation in the model.

A. Model meaning interpretation

We can get the final model after the preliminary model is passed through the above test and the correction of the model. The multiple regression models of the volume of freight ($\text{Ln}Y$) and industrial added value ($\text{Ln}X_6$) and the total wages of employed persons in urban non-private units ($\text{Ln}X_8$) are as follows.

$$\text{Ln}Y = 0.201\text{Ln}X_6 + 0.493\text{Ln}X_8 \quad (2)$$

The economic meaning of the two slope coefficients of the above equation are explained as follows.

If the total wages of employed persons in urban non-private units ($\text{Ln}X_8$) remains unchanged, the industrial added value ($\text{Ln}X_6$) increases by 10 thousand yuan, and the demand for freight volume ($\text{Ln}Y$) will increase by 49.3% on average. On the other hand, the industrial added value ($\text{Ln}X_6$) remains unchanged, the total wages of employed persons in urban non-private units ($\text{Ln}X_8$) will increase by 100 million yuan, and the freight volume demand ($\text{Ln}Y$) will increase 20.1% on average.

V. Conclusions and Suggestions

According to the predicted regression model, the logistics demand in Kashagar Administrative Offices in the next few years can be predicted. In order to expand the demand for logistics in Kashagar Administrative Offices, it is necessary to improve the development of industry and improve the people's income level. If we can strengthen the adjustment of the industrial structure, then this will provide better opportunities for the development of the logistics industry. The rapid growth of industrial output will promote the development of wholesale and retail trade and transportation industry and storage industry, which will provide a good industrial foundation for the development of local logistics industry.

At the same time, it is necessary to raise the people's income level, especially the wages of the local rural residents in this area. Data show that in 2016, the per capita disposable income of urban residents in Kashagar Administrative Offices was 22732 yuan, and the disposable income of rural residents was 7918 yuan, the former was 2.87 times the latter. When the salary increases, people have enough income to buy the products they need, and the local logistics industry will be further developed.

Acknowledgment

This research was financially supported by following projects from China and Germany: the National Science Foundation of China (71172194); 2015 Hunan's University Collaborative Innovation Project (No.2015/69); German BMBF (No.APR10/810); Social Science Funds Project in

Hunan Province (No.2012ZDB13); Hunan international cooperation projects (No.2010WK3041); Chongqing Major Social Science Foundation (No.2004ZD03).

REFERENCES

- [1] Chen B, Hu L · Forecast of Logistics Demand in Hercynian Region Based on BP Neural Network[C]//International Conference of Logistics Engineering and Management · 2015:4137-4143 ·
- [2] Du Y, Zhang J, Tang Z, et al · Application and Result Analysis Comparison of Based on MLP Neural Network and Econometric Models in Logistics Demand Forecasting[C]// International Conference of Logistics Engineering and Management · 2015:2768-2774 ·
- [3] Hu W, University T C · A study of regional logistics demand forecasting based on the genetic BP neural network[J] · Journal of Chongqing Three Gorges University, 2014 ·
- [4] Cao P, Chen F · Grey BP Neural Networks Model based on GA and Its Application in Regional Logistics Demand Forecasting[J] Journal of Beijing Institute of Technology, 2012 ·
- [5] LIU Yu · WU Ying xue · DANG Wen feng. Regional Logistics Demand Forecasting Based on Multiple Linear Regression[J] Logistics Engineering And Management. 2014, 36(03):52-54.
- [6] Mao Min, Liu Jian. Method for Combination Forecasting of Regional Logistics Demand Based on Multiple Linear Regression. [J]. Logistics Technology, 2015, 34(07):163-165.
- [7] Zhang Cheng, Feng Yaping. Logistic Demand Forecasting in Jiangxi Province Based on Grey Correlation. [J] Journal of East China Jiaotong University, 2014, 31(04):26-32.
- [8] HU Yun-qing. Application of Improved Genetic BP Algorithm in Logistics Demand Forecasting. [J] Logistics Sci-Tech, 2015, 38(11):107-109.
- [9] Dou Jin. Research and Demonstration of Gansu Province Logistics Demand Forecast "Belt and Road Initiative" [D]. Lanzhou University of Finance and Economics, 2017.
- [10] Tan Chengwei, Zhu He, Zhao Xiaomin. Analysis of Logistics Demand of Shanghai Based on Multiple Linear Regression Model. [J]. Logistics Technology, 2016, 35(01):34-37.
- [11] CHEN Yan · WANG Lu · SUN Feng yan. Research on the Port Logistics Demand Forecast Based on System Dynamics [J]. Logistics Engineering And Management. 2017, 39(08):76-79+75
- [12] Kashagar Bureau of statistics, National Bureau of statistics, Kashagari investigation team. (2017, May, 19). Statistical bulletin of 2016 national economic and social development in Kashagar Administrative Offices. <http://www.kashi.gov.cn/Item/41122.aspx>.



Dr. Aimin Deng is a full time professor and doctoral supervisor at the School of Economics and Trade of Hunan University; the founder and Director of the Institute of transport and logistics (www.tl-hnu.cn, 2006); the founder and Director of Sino-German Center for transportation and logistics of Hunan University (www.tl-cnde.org, 2008). Research areas: logistics and supply chain management (including finance) systems, transportation systems, e-commerce and IOT systems. She studied in Germany, has a

certificate of advanced English to study abroad by the Ministry of Education in China, and was the head of E-Commerce and Logistics Management Department and Information Management and Information System. She has a wide range of exchanges and cooperation with national and international enterprises, institutions and Governments, and social and academic positions; hosted and completed nearly more than 40 domestic and international projects, published more than 100 papers and guide more than 100 graduate students.



Hancheng Yang is now studying in the School of Economics and Trade of Hunan University, majoring in International Business. His major research direction is logistics and supply chain finance under the guidance of tutor Professor Aimin Deng. After one year studying, he has got the basic research quality and understanding of his research field. He had previously worked in rural commercial banks and had a preliminary understanding of the

banking system. He also has a keen interest in economics, usually reading a lot of economics books, learning and life are also using economic thinking to do things. He is very independent and creative.