

# Jilin Traffic Volume Impact Analysis on Economic Development

Yuanyuan Huang, Zhe Yin\*, Jinghu Shen

Mathematics Department, Yanbian University, Yanji, China  
Email: [\\*yinzhe@ybu.edu.cn](mailto:yinzhe@ybu.edu.cn)

Received 3 January 2015; accepted 23 January 2015; published 29 January 2015

Copyright © 2015 by authors and Scientific Research Publishing Inc.  
This work is licensed under the Creative Commons Attribution International License (CC BY).  
<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

Since the reform and opening up, China's transport development has made brilliant achievements and has a strong support for the rapid development of economy and society. In this paper, we collate, screen and analyze a total of 32-year data from 1980 to 2011 on traffic volume in Jilin Province, then we build a partial least squares regression model to the quantitatively predict and analyze the relations of transportation construction and economic development.

## Keywords

Factors, Gray Correlation, Partial Least Squares Regression, Forecast

---

## 1. Introduction

In this paper, according to the data that shown during the Eighth Five-Year Plan period, the correlation coefficient of road density, growth rate and the growth rate of GDP was 0.57, the correlation coefficient of rail, road density and per capita GDP was up to 0.8. In recent years, Jilin Province's economy has seen rapid development momentum, thanks to the first vanguard role of the railway.

Academics have done many researches in theory and practice on traffic volume due to the increasing importance of traffic volume. Such as Yang Zhengxiang and some people [1] use generalized regression neural network to predict traffic volume, Li Yongmei and some people [2] use BP neural network to predict traffic volume in Zhejiang Province, Gao Wei and some people [3] make a GM-Markov general prediction model in order to research on traffic volume, Cao Xuping and some people [4] use grey correlation to make an analysis on the relation of transportation industry and local economical development. These methods can predict the future development of transportation industry in some degree, but sometimes they can't overcome the problem about the existence of multiple relevance. Based on this, in this paper, we model Jilin traffic volume on the economic de-

---

\*Corresponding author.

velopment impact by partial least squares regression method.

In recent years, for practice needs, we develop the partial least squares (PLS) regression analysis, which is a kind of multivariate statistical analysis method with wide applicability. The type of data modeling and forecasting analysis methods and data analysis methods of non-understanding of patterns get together organically. This method can be easier to analyze the factors associated with the dependent variable, so that the independent variable regression coefficients are easier to interpret [5]. In this paper, a second-generation regression, known as partial least squares regression (PLS) method, analyzed the factors affected the economic development of Jilin Province. Because of the locality, not so many references are included in this paper.

## 2. PLS Regression Basic Ideas and Concrete Steps

### 2.1. Basic Ideas

Successive independent variable  $x$  to extract the principal component  $t_h$ ,  $h = 1, 2, \dots, A$  and as much as possible from the set of variables  $x$  summarize the information while the correlation between the dependent variable  $y$  can be maximized.

### 2.2. Specific Steps

- 1) The  $x$  and  $y$  are normalized to obtain the normalized matrix independent variable  $E_0$ , variable matrix  $F_0$ .
- 2) Component extraction: First, an ingredient extracted from  $E_0$ ,  $t_1 = E_0 W_1$ ,  $\left(W_1 = \frac{E_0' F_0}{\|E_0' F_0\|}\right)$ . Then  $E_1$ ,  $F_1$  substituted  $E_0$ ,  $F_0$ , can be extracted a second PLS components of  $t_2$ .
- 3) Effectiveness: For the determination of the number of component extraction, we use cross validation to

determine the coefficient  $Q_h^2$ : 
$$Q_h^2 = 1 - \frac{\sum_{k=1}^q PRESS_{hk}}{\sum_{k=1}^q SS_{(h-1)k}} = 1 - \frac{PRESS_h}{SS_{(h-1)}}$$

- (1) When  $Q_h^2 \geq (1 - 0.95^2) = 0.0975$ , the marginal contribution  $t_h$  component is significant.
- (2) For  $k = 1, 2, \dots, q$  at least there exists one  $k$ , so that  $Q_h^2 \geq 0.0975$ , with the increase of the component  $t_h$ , one dependent variable  $y_k$  has been significantly improved, and therefore, can be considered as a significant increase in component  $t_h$  beneficial.

## 3. Empirical Analysis

According to reports, with regard to the impact of economic factors, many scholars have done research on this. There are many factors that affect economy, on the basis of least squares regression model, considering the impact of transport on the economy's capacity. Data from 1980-2011 statistics in Jilin Province, according to the theory of economics, has made a number of conclusions, and after screening, there are six independent variables in the model were elected: railroad mileage (RM), highway mileage (HM), passenger traffic (PN), cargo (CN), passenger turnover (PTN), freight turnover (CTN), with the index of per capita GDP to represent economy of Jilin Province. The data can be seen from Statistical Yearbook 2013 in Jilin Province [6].

By the timing diagram, we found that per capita GDP has a parabolic trend growth and instability. After taking a second radical timing diagram of per capita GDP, it stabilizes state.

By the dependent variable and the independent variables scatter, taking the square root of the variable, there is a good linear relationship which is conducive to fit the following linear regression equation, therefore, the variables were first pretreated.

Variables were treated as follows:

$$\begin{aligned} x_1 &= \text{sqrt}(\text{RM}), x_2 = \text{sqrt}(\text{HM}), x_3 = \text{sqrt}(\text{PN}), x_4 = \text{sqrt}(\text{CN}), \\ x_5 &= \text{sqrt}(\text{PTN}), x_6 = \text{sqrt}(\text{CTN}), y = \text{sqrt}(\text{GDP}). \end{aligned}$$

Due to the level of GDP and economic value of units of each variable transportation unit volume is not uni-

form, there is a dimension, so before doing data analysis, the variables do standardized, eliminating the effects of dimensionless.

3.1. Data Analysis

3.1.1. The Association Test

Gray correlation analysis is a factor analysis. In practice, a lot of factors affect the economy, the effect of various factors on the total economy is not able to see at once, and the impact is different. The relationship is “primary and secondary”, or “pros and cons”. We classify these factors between the primary and secondary by the gray correlation relationship, and quantify the degree of each influential factor. Because distinguish coefficient  $P$  is a real number given in (0, 1), and generally  $P = 0.5$ . Correlation degree is the average of the n-dimensional space after each correlation coefficient f (k) accumulates [7]. When the resolution factor  $P = 0.5$ , considering correlation degree greater than 0.6 is acceptable, that is to say passing association test, otherwise the degree of association is not good.

As shown in **Form 1**, we can get that all the variables pass the test of association, that is to say there is a strong influence of independent variables to the dependent variable, using these variables to establish the model has a good effect. As can be seen from the associated values, the railway has the greatest impact on the economy.

3.1.2. Cointegration Tests

Non-stationary series is likely to occur spurious regression [8], cointegration is to test whether the significance of their causal relationship described by the regression equation is spurious regression. That is to say between the test variables whether exist a stable relationship.

As shown in **Procedures 1**, we know: when take unit root test, test level  $\alpha = 0.01$ , p-values of the 0-order zero-mean sequence, single whole mean sequences and trend sequences were less than 0.01. So under the level of significance  $\alpha = 0.01$ , the original hypothesis are refused, the results do not accept the existence of a unit root. We can determine that the sequence is a stationary series and spurious regression equation does not exist.

From the **Form 2**, we know that the variance of each variable inflation factor (VIF) are large ( $VIF > 10$ ), described the presence of severe multicollinearity among these variables. Using ordinary regression analysis has been unable to resolve, so we use the partial least squares regression analysis to overcome this shortcoming.

Form 1. Correlative degree of independent variable.

Independent Variable	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$
Correlative Degree	1.00	0.7530	0.7520	0.6103	0.7217	0.7169

Procedures 1. Unit root tests.

Type	Lags	Augmented	Dickey	Fuller	Unit Root	Tests	
		Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-33.3486	<0.0001	-4.09	<0.0001		
	1	-3.3316	0.2083	-1.04	0.2696		
	2	0.0090	0.6841	0.01	0.6838		
Single Mean	0	-237.550	0.0001	-17.13	<0.0001	146.71	0.0010
	1	-96.0969	0.0013	-6.46	<0.0001	21.00	0.0010
	2	-24.4549	0.0032	-2.93	0.0448	4.49	0.0584
Trend	0	-311.265	0.0001	-27.67	<0.0001	382.81	0.0010
	1	-471.337	0.0001	-14.43	<0.0001	104.34	0.0010
	2	-316.031	0.0001	-8.61	<0.0001	37.34	0.0010

### 3.2. PLS Regression Model

#### 3.2.1. PLS Components Extracted

According to the principles and practices of PLS extracts [9] [10], we use the  $Q_k^2$  value to determine the PLS components  $t_1, t_2$ . From the output of the program, select two PLS components to meet the accuracy requirements results in table 3.

As shown in Form 3: Because  $Q_k^2 = -0.2465 < 0.0975$ , So choose  $n = 2$ , namely with  $t_1, t_2$  two constituents partial least squares regression analysis, the model works best.

#### 3.2.2. Cumulative Explanatory Power

As can be seen from Form 4, composition  $t_1, t_2$  whose cumulative explanatory power of the independent variable X is 0.958556, the dependent variable explained cumulative capacity is 0.981631, thus  $t_1, t_2$  are very effective to achieve the requirements of the modeling.

#### 3.2.3. PLS Model Results

The Form 5 shows that the regression coefficients of the variables of the modeling which is established by partial least squares regression are positive numbers, which indicates they have a positive correlation between traffic volume and the economy. So enhancing the transportation industry can help raise the economic level.

**Form 2.** Colinearity test table of independent variable.

Model	Non-standardized coefficients		A coefficient			Correlation			Collinearity statistics	
	B	Standard error	Standard coefficient Trial Version	t	Sig.	Zero-order	Partial	Section	Tolerance	VIF
$x_1$	1.571	0.309	0.141	5.085	0.000	0.529	0.713	0.073	0.269	3.717
$x_2$	0.107	0.013	0.376	8.082	0.000	0.963	0.850	0.116	0.095	10.498
1 $x_3$	-0.125	0.056	-0.319	-2.237	0.034	0.921	-0.408	-0.032	0.010	98.623
$x_4$	0.114	0.025	0.281	4.554	0.000	0.897	0.673	0.065	0.054	18.396
$x_5$	0.082	0.072	0.199	1.137	0.267	0.972	0.222	0.016	0.007	148.980
$x_6$	0.103	0.029	0.419	3.488	0.002	0.971	0.572	0.050	0.014	70.032

a. The dependent variable: y

**Form 3.** Composition tables.

Number of components	$Q^2$	Thresholds
1	0.3498	0.0975
2	0.5577	0.0975
3	-0.2465	0.0975

**Form 4.** Cumulative explanatory power (%).

RD	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x$	$y$
$t_1$	35.9221	90.6519	94.7066	73.0094	98.6205	97.9766	81.8145	96.9074
$t_2$	94.2015	90.7931	96.9316	96.3384	98.7710	98.0978	14.0411	1.2557
Cumulative (RD)							95.8556	98.1631

**Form 5.** Regression coefficients of partial least squares regression.

Variable	Regression coefficients	Standardized regression coefficients
Constant term	-44.752	0
$x_1$	0.263	0.024
$x_2$	0.067	0.235
$x_3$	0.055	0.139
$x_4$	0.108	0.264
$x_5$	0.079	0.193
$x_6$	0.048	0.195

And the development of the railway has the biggest impact on economy, which is consistent with the gray correlation judgment. As we known, Jilin Province locates in the northeastern hinterland, borders on Russia and North Korea, there are many scenic areas. If traffic develops rapidly, it will be able to attract many tourists at home and abroad, and stimulate the local economy. Therefore, we should accelerate the development of railways mainly and develop the transport sector vigorously.

### 3.2.4. PLS Regression Equation

Standardized regression equation:

$$y^* = 0.024x_1^* + 0.235x_2^* + 0.139x_3^* + 0.264x_4^* + 0.193x_5^* + 0.195x_6^*$$

Original regression equation:

$$y = -44.752 + 0.263x_1 + 0.067x_2 + 0.055x_3 + 0.108x_4 + 0.079x_5 + 0.048x_6$$

## 3.3. Model Predictions

PLS regression models were created by using the above process. Then we forecast 2012 economic situation in Jilin Province (per capita GDP), compare with the actual economic situation and test the model's reliability.

Predictive value:

Put the values of **Form 6** into the equation, we can get the predictive value of the following as shown in **Form 7**.

The transformed data is squared to obtain the predicted values of the original data.

**Form 8** shows that predictive value of 2012 per capita GDP is relatively close to the actual value, slightly larger than the actual value, which is related to the development of the transportation industry in Jilin Province in recent years, also because we open the trade channel with the surrounding areas. After making predictions, the error of the predicted value is 4.83%, within the forecast error in the range of 5%, Once again shows the model is very significant.

## 4. Conclusions

Nowadays for improving levels of economic issues of Jilin Province on the transportation industry, the railway development should be enhanced mainly, great importance should be attached to railway construction, operation, management, and the development of railway construction should be vigorously promote. Opening the east of the Northeast corridor rail and road as quickly as possible built a new international channel of Northeast China gradually. Currently, the railway development in Jilin Province is mainly based on passengers and freight situation is not as good as passenger. But after modeling analysis, the cargo has a greater impact on the economy. For enhancing passenger traffic, we should seize the opportunity to carry out freight marketing vigorously and organize effective supply to stimulate economic development actively.

Currently, highway mileage is less, specific gravity is low, and debts are many. And for more far-reaching

**Form 6.** Actual value of transport volume indicators.

Year	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$y$
2012	4383 (66.204)	93208 (305.3)	72677 (269.587)	59768 (244.475)	57232 (239.232)	173129 (416.088)	5113 (71.5087)

Note: Value () is taking a second after radical.

**Form 7.** Predictive value of the transport volume indicators.

Year	Predictive value	The true value	Error
2012	73.2169	71.509	0.024

Note: The value after taking a second radical.

**Form 8.** Predictive value of the transport volume indicators.

Year	Predictive value	The true value	Error
2012	5360.7128	5113.5	0.0483

Note: The original value.

impact on the development of highways in the economy, we should speed up the sea entry and communicate the inter-provincial channels construction, and speed up the construction of the capital to city hall location, and the construction of Changbai Mountain Expressway, we should also adjust and improve highway development plan, increase the projects, increase the investment, to achieve that every county has highway. Additionally, in order to promote economic development, we should compressing passenger and cargo turnaround time and increase turnover volume and capacity.

**References**

- [1] Yang, Z.X. (2009) Transport Volume Forecast Based on GRNN. *Technology Economy in Area Communications*, **4**, 20-21.
- [2] Gao, W. (2009) Research on the Gray-Markov Forecast Model for Traffic Volume of Highway Transport. *China Market*, **15**, 95-98.
- [3] Li, Y.M. and Liang, Q.O. (2010) The Prediction of Traffic Transportation Volume of Zhejiang Province with BP Neural Network. *Journal of Lishui University*, **32**, 21-25.
- [4] Cao, X.P., He, Y.H. and Chen, G.P. (2010) Analysis of Grey Relatedness between Transportation and Regional Economy in Jiangsu from 1999 to 2008. *Journal of Nantong Vocational & Technical Shipping College*, **9**, 1-4.
- [5] Wang, H.W. (1999) Partial Least Squares Regression Method and Its Application. National Defense Industry Press, Beijing, 201-206.
- [6] (2013) Jilin Bureau of Statistics Yearbook 2013. Statistics Press, Beijing.
- [7] Yi, D.S. (1992) Grey Theory and Methods: Summary Program Application Problem Solution. Petroleum Industry Press, Beijing.
- [8] Li, Y.K. and Yao, C.Z. (2009) Analyzing the Impact of China's Economic Growth Factors. *Statistics and Decision*, **19**, 104-106.
- [9] Wold, H. (1975) Soft Modelling by Latent Variables: The Non-Linear Iterative Partial Least Squares (NIPALS) Approach. *Perspectives in Probability and Statistics: Papers in Honour of M. S. Bartlett*. Academic Press, London, 117-142.
- [10] Ma, M.D., Ma, X.J., Xie, Y.Z. and Ma, T. (2014) Analysis the Relationship between Ecological Footprint of Ningxia and Influencing Factors: Partial Least-Squares Regression. *Acta Ecologica Sinica*, **34**, 682-689.

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

**Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either [submit@scirp.org](mailto:submit@scirp.org) or [Online Submission Portal](#).**

