Econometric Analysis of the Influencing Factors of the Transportation Industry on GDP-Based on the Data of China Statistical Yearbook 2018

Min Huang[1, a], Xiaodan Zhang[2, b], Ruiting Zhang[1, c]

[1]Faculty of International Trade, Shanxi University of Finance and Economics, Taiyuan, Shanxi, China

[2]School of Management Science & Engineering, Shanxi University of Finance and Economics, Taiyuan, Shanxi, China

[a]hm197346285@163.com, [b]18134925399@163.com, [c]z15535021861@163.com

Keywords: GDP, Transportation Industry, Multiple Linear Regression, Multiple Collinearity

Abstract: As a carrier of the national economy, transportation communicates production and consumption, and plays an extremely important role in economic development. Looking at the interrelationships between transportation and economic development over the centuries, the higher the level of production, the more advanced the infrastructure is required to develop. Entering a modern society, the economic and social requirements for transportation are inherently advanced. Transportation is the forerunner of the national economy, and economic development and transportation first are the inherent laws of economic development. This paper uses statistical data from 2000 to 2016, and uses econometrics to study the value-added of transportation, warehousing, and postal industries, the number of employees on railways and highways, the mileage of railways and highways, passenger traffic, and cargo transportation. The impact of China's gross domestic product (GDP) is analyzed from four perspectives: determining variables and mathematical relationships, analyzing specific quantitative relationships between variables, testing the reliability of conclusions obtained, making economic analysis and economic forecasts.

1. Introduction

The main purpose of developing the transportation industry is to promote the economy. It involves the economic resources of a region and the output of local industries, improves the profitability of the economy, and guarantees economic revenue. With the continuous improvement of people's living standards, the requirements for the transportation industry have also been continuously improved, and the scale of economic development has been achieved [1]. In general, in some regions with relatively high economic conditions, the transportation industry is also developing faster and supporting facilities are more complete. The development of the transportation industry has also stimulated the scale of local economic development, which has formed a circular nature. Development has achieved a win-win situation. In some areas with insufficient economic conditions, the capital investment in the transportation industry has been reduced for development, resulting in the stagnation of the transportation industry, limiting the development of the local economy, and leading to a situation in which both have failed. Therefore, it is important to explore the impact of the transportation industry on China's GDP [3].

2. Theoretical Model Construction

2.1 Data Sources

The data in this article are from China Statistical Yearbook 2018. The selected variables are transportation, storage and postal value added (100 million yuan), employment of railways and highways (persons), mileage of railways and highways (10,000 kilometers), passengers The volume of transportation (10,000 people), the volume of cargo transportation (10,000 tons), and China's GDP (100 million yuan), of which China's GDP exists as an explanatory variable, and the rest are treated as explanatory variables. Data from 2000 to 2016 for each variable were selected, and the sample size was 17.

2.2 The Reason of Choosing Explanatory Variables

Explanatory variables mainly analyze rail transportation and road transportation in transportation. Railway is one of the important carriers of social and economic development, and at the same time has created prerequisites for social and economic development. Since the reform and opening up, China's road transport has entered a new period of development, and road mileage, road transport volume and civilian car ownership have all increased significantly. The national highway network with national highways as main trunk lines and provincial and county or township roads as branch lines has initially formed. The rapid development of highway construction has laid a good foundation for the basic role of highway transportation in the comprehensive transportation system. Therefore, railways and highways have a certain representation in the transportation industry.

3. Descriptive Analysis

3.1 Total Analysis

Table.1. Influencing Factors of the Development of Transportation Industry on GDP

variable	index
X1	Transportation, storage and postal added value (100 million yuan)
X2	Railway and highway employment (person)
X3	Railway and highway mileage (10,000 kilometers)
X4	Passenger traffic (10,000 people)
X5	Cargo transportation volume (10,000 tons)
Y	China's GDP (100 million yuan)

3.2 Modeling

From the scatter diagram, it can be seen that the value-added of transportation, warehousing and postal services, the number of employed persons on railways and highways, the mileage of railways and highways, the volume of passenger transportation, and the volume of freight transportation are basically linear with the GDP Therefore, a linear regression model can be initially established.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \mu 9$$

$\hat{\beta}_1$ represents the impact of added value in transportation, warehousing, and postal industries on GDP; $\hat{\beta}_2$ represents the impact of railway and highway employment on GDP; $\hat{\beta}_3$ represents the impact of railway and highway mileage on GDP; $\hat{\beta}_4$ represents the impact of passenger traffic on GDP; $\hat{\beta}_5$ represents the impact of cargo transportation on GDP; μ is a random perturbation term.

4. Model Estimation and Testing

4.1 Model Estimation

Table.2. Model Regression Results

	Coefficient	Std. Error	t-Statistic	Prob.
C	-59941.21	20675.90	-2.899086	0.0145
X1	17.51588	1.709148	10.24831	0.0000
X2	-0.006379	0.007478	-0.853068	0.4118
X3	-121.0342	41.99619	-2.882027	0.0149
X4	-0.014351	0.008415	-1.705417	0.1162
X5	0.078906	0.017206	4.586059	0.0008
R-squared	0.999477	Mean dependent var		358368.5
Adjusted R-squared	0.999239	S.D. dependent var		220498.8
S.E. of regression	6084.241	Akaike info criterion		20.53536
Sum squared resid	4.07E+08	Schwarz criterion		20.82943
Log likelihood	-168.5505	Hannan-Quinn criter.		20.56459
F-statistic	4200.705	Durbin-Watson stat		1.629794
Prob (F-statistic)	0.000000			

$$\hat{Y} = -59941.21 + 17.51588X_1 - 0.006379X_2 - 121.0342X_3 - 0.014351X_4 + 0.078906X_5$$

$$(20675.90) \quad (1.709148) \quad (0.007478) \quad (41.99619) \quad (0.008415) \quad (0.017206)$$

$$t = (-2.899086) \quad (10.24831) \quad (-0.853068) \quad (-2.882027) \quad (-1.705417) \quad (4.586059)$$

$$R = 0.999477 \quad \bar{R}^2 = 0.999239 \quad F = 4200.75 \quad D.W. = 1.629794$$

4.2 Economic Significance Test

The estimated parameter indicate that for every 100 million yuan in added value of transportation, warehousing, and postal services, GDP will increase by 1.751588 billion yuan; for each additional person employed by railways and highways, GDP will decrease by 0.00637.9 billion yuan; For every 10,000 kilometers of highway mileage, GDP will decrease by 12.103342 billion yuan; for every 10,000 passenger traffic increases, GDP will decrease by 0.014351 billion yuan; for every 10,000 tons of freight traffic increase, GDP will increase by 0.078906 million yuan. This is inconsistent with the expected economic significance, and it can be seen that there is a serious multicollinearity among the variables.

4.3 Statistical Inference Test

(1) Goodness-of-fit Test

As can be seen from Table 2, the coefficient of determination $R^2 = 0.999477, \bar{R}^2 = 0.999239$ indicates that the model as a whole fits the sample data well, that is, the explanatory variables "added value of transportation, warehousing and postal industry", "number of railway and highway employment", "Railway, highway mileage", "passenger traffic", and "cargo traffic" explained a larger part of the explanatory variable "China's gross domestic product (GDP)".

(2) F Test

As can be seen from Table 2, F = 4200.705, falling into the rejection domain, indicates that the regression equation is significant, that is, the explanatory variables "added value of transportation, storage and postal industry", "number of railway and highway employment", "railway and highway mileage" The combination of "passenger traffic" and "cargo traffic" has an impact on the explanatory variable "China's gross domestic product (GDP)".

(3) t Test

It can be known from Table 2 that the absolute value of t statistics with $\hat{\beta}_1$, $\hat{\beta}_3$, $\hat{\beta}_5$ and is greater than 2.201, so the null hypothesis is rejected, and the explanatory variables "added value of transportation, warehousing and postal industry", "railway, highway mileage", "cargo transportation volume" are rejected Significant impact on the explanatory variable "China's Gross Domestic Product (GDP)"; the absolute values of the other t statistics are all less than 2.201, accepting the null hypothesis, explaining the explanatory variables "number of railway and highway employees" and "passenger traffic" The effect on the explanatory variable "China's gross domestic product (GDP)" is not significant, which may be caused by the existence of multicollinearity in the model.

4.4 Econometric Significance Test

(1) Multicollinearity Test and Correction

- Multiple Collinearity Test

Table.3. Correlation Coefficient Matrix

	X2	X1	X3	X4	X5
X2	1.000000	0.905890	0.718280	0.045182	0.862867
X1	0.905890	1.000000	0.909601	0.379412	0.980837
X3	0.718280	0.909601	1.000000	0.566913	0.914905
X4	0.045182	0.379412	0.566913	1.000000	0.515225
X5	0.862867	0.980837	0.914905	0.515225	1.000000

From the correlation coefficient matrix, it can be seen that the correlation coefficients between the explanatory variables are high, which confirms the results of the economic significance test, and there does exist some multicollinearity.

- Modified Multicollinearity

Stepwise regression method was used to screen and eliminate the variables that caused multicollinearity. The specific steps are as follows: first use the explained variables Y to perform a simple regression on each of the considered explanatory variables, then based on the corresponding regression equation that contributes the most to the explained variables X_1, and then gradually introduce the remaining explanatory variables [2]. F-tests are performed after each explanatory variable is introduced, and t-tests are performed on the selected explanatory variables one by one. When the original explanatory variables are no longer significant due to the introduction of later explanatory variables, they are deleted to make sure that only significant variables are included in the regression equation before each new variable is introduced. This is an iterative process until neither significant explanatory variables are selected into the regression equation, nor are insignificant explanatory variables removed from the regression equation [4].

After stepwise regression, the variables removed were railway, highway mileage (X_3) and passenger traffic (X_4). Therefore, the revised explanatory variables are the value added of transportation, warehousing, and postal services (X_1), the number of employed persons on railways and highways (X_2), and the volume of freight transported (X_5).

Table.4. Stepwise Regression Results

	Coefficient	Std. Error	t-Statistic	Prob.
C	-120603.8	13379.84	-9.013847	0.0000
X1	17.26949	1.574459	10.96852	0.0000
X2	0.012363	0.004809	2.570681	0.0233
X5	0.049155	0.010192	4.822686	0.0003
R-squared	0.998756	Mean dependent var		358368.5
Adjusted R-squared	0.998468	S.D. dependent var		220498.8
S.E. of regression	8629.304	Akaike info criterion		21.16604
Sum squared resid	9.68E+08	Schwarz criterion		21.36209
Log likelihood	-175.9113	Hannan-Quinn criter.		21.18553
F-statistic	3477.914	Durbin-Watson stat		0.686184
Prob (F-statistic)	0.000000			

(2) Heteroscedasticity Test

Heteroscedasticity [7] means that the degree of dispersion of the observed values of the explanatory variables changes with the changes of the explanatory variables. The test is selected as the "White test".

Table.5. White Test Results

F-statistic	1.291662	Prob. F (9, 7)		0.3762
Obs*R-squared	10.61072	Prob. Chi-Square (9)		0.3033
Scaled explained SS	1.690910	Prob. Chi-Square (9)		0.9955
	Coefficient	Std. Error	t-Statistic	Prob.
C	1.50E+09	1.07E+09	1.398793	0.2046
X1	262033.4	163987.3	1.597888	0.1541
X1^2	13.24242	16.83283	0.786702	0.4573
X1*X2	-0.038221	0.034944	-1.093775	0.3103
X1*X5	-0.192553	0.243518	-0.790713	0.4551
X2	-706.0296	777.2922	-0.908320	0.3939
X2^2	0.000123	0.000150	0.816632	0.4410
X2*X5	0.000130	0.000186	0.697776	0.5078
X5	-1587.921	770.0767	-2.062030	0.0781
X5^2	0.000743	0.000883	0.842156	0.4275
R-squared	0.624160	Mean dependent var		56943743
Adjusted R-squared	0.140938	S.D. dependent var		43333002
S.E. of regression	40163462	Akaike info criterion		38.14398
Sum squared resid	1.13E+16	Schwarz criterion		38.63411
Log likelihood	-314.2238	Hannan-Quinn criter.		38.19270
F-statistic	1.291662	Durbin-Watson stat		1.311502
Prob (F-statistic)	0.376157			

According to the test results, $\chi^2(9) = 16.919 > n$, $R^2(9) = 16.919 > n$ we accept the null hypothesis, indicating that there is no heteroscedasticity in the random error in the model.

(3) Autocorrelation Inspection and Correction

- Autocorrelation Test

The Breusch-Godfrey test, or LM test, was used for model autocorrelation test [6]. It can be seen that the rejection of the null hypothesis indicates that the model has autocorrelation.

Table.6. LM Test Results

F-statistic	7.478823	Prob. F (1, 12)		0.0181
Obs*R-squared	6.527088	Prob. Chi-Square (1)		0.0106
	Coefficient	Std. Error	t-Statistic	Prob.
C	4037.331	11029.78	0.366039	0.7207
X1	0.838768	1.322302	0.634325	0.5378
X2	-0.000932	0.003944	-0.236356	0.8171
X5	-0.005246	0.008545	-0.613929	0.5507
RESID (-1)	0.657257	0.240336	2.734744	0.0181
R-squared	0.383946	Mean dependent var		-3.49E-11
Adjusted R-squared	0.178595	S.D. dependent var		7778.350
S.E. of regression	7049.627	Akaike info criterion		20.79927
Sum squared resid	5.96E+08	Schwarz criterion		21.04433
Log likelihood	-171.7938	Hannan-Quinn criter.		20.82363
F-statistic	1.869706	Durbin-Watson stat		1.341696
Prob (F-statistic)	0.180596			

- Modified Autocorrelation

The Cochranne-Orcutt iterative method [5] was used to make generalized differential regression for model autocorrelation correction.

Table.7. Cochrane-Octor Iteration Test Results

	Coefficient	Std. Error	t-Statistic	Prob.
C	-117029.1	17006.31	-6.881513	0.0000
X1	18.66527	1.510029	12.36086	0.0000
X2	0.007194	0.004614	1.559064	0.1473
X5	0.045505	0.011246	4.046407	0.0019
AR (1)	0.639277	0.200197	3.193234	0.0086
R-squared	0.999366	Mean dependent var		374499.0
Adjusted R-squared	0.999135	S.D. dependent var		217124.1
S.E. of regression	6386.372	Akaike info criterion		20.61203
Sum squared resid	4.49E+08	Schwarz criterion		20.85346
Log likelihood	-159.8962	Hannan-Quinn criter.		20.62439
F-statistic	4331.747	Durbin-Watson stat		1.652249
Prob (F-statistic)	0.000000			
Inverted AR Roots		.64		

The revised model is: $\hat{Y} = -117029.1 + 18.66527 X_1 + 0.007194 X_2 + 0.045505 X_5$ LM test on the revised model can be obtained:

Table.8. LM Test Results

F-statistic	0.971968	Prob. F (1, 10)		0.3474
Obs*R-squared	1.417384	Prob. Chi-Square (1)		0.2338
	Coefficient	Std. Error	t-Statistic	Prob.
C	8586.879	19125.95	0.448965	0.6630
X1	-0.095490	1.515052	-0.063027	0.9510
X2	0.002953	0.005506	0.536376	0.6034
X5	-0.005380	0.012513	-0.429951	0.6763
AR (1)	-0.269680	0.339128	-0.795216	0.4450
RESID (-1)	0.552720	0.560633	0.985884	0.3474
R-squared	0.088586	Mean dependent var		5.87E-07
Adjusted R-squared	-0.367120	S.D. dependent var		5468.962
S.E. of regression	6394.525	Akaike info criterion		20.64427
Sum squared resid	4.09E+08	Schwarz criterion		20.93399
Log likelihood	-159.1541	Hannan-Quinn criter.		20.65910
F-statistic	0.194394	Durbin-Watson stat		2.028447
Prob (F-statistic)	0.957706			

It can be seen at this time, $\chi^2(1) = 3.84146 > T$, $R^2 = 1.417384$, so accept the null hypothesis, indicating that the model no longer exists autocorrelation.

5. Model Summary and Application

This paper uses econometric methods to test and modify the original model for multiple collinearity, heteroscedasticity, and autocorrelation, excluding the two original variables of railway, highway mileage (X_3) and passenger traffic (X_4), and the revised explanatory variables. Add value for transportation, warehousing and postal services (X_1), employment of railways and highways (X_2), freight transportation (X_5) and establish a new model to draw conclusions:

The value added of transportation, warehousing and postal services, the employment of railways and highways, and the volume of freight transportation all have a positive effect on GDP, and the value added of transportation, warehousing and postal services, and the volume of freight transportation have a more significant impact on GDP. Significantly. For every 100 million yuan in added value of transportation, warehousing, and postal services, GDP will increase by 1,866,272.7 million yuan; for each additional person employed by railways and highways, GDP will increase by 90,719,400 million yuan; for every 10,000 tons of freight transportation, GDP will increase by 0.045505 billion yuan, in line with expected economic significance.

References

[1] Nilabhra Banerjee, Alec Morton, Kerem Akartunalı. Passenger demand forecasting in scheduled transportation [J]. European Journal of Operational Research, 2019.

[2] Ning Zhang, Peng Zhou, Chih-Chun Kung. Total-factor carbon emission performance of the Chinese transportation industry: A bootstrapped non-radial Malmquist index analysis [J]. Renewable and Sustainable Energy Reviews, 2015, 41.

[3] LIU Zhan-yu, SUN Xia-ling. Study on Carbon Emission Driving Factors and Low-carbon Policy of Transportation Industry in Henan Province [J]. International Journal of Plant Engineering and Management, 2018, 23 (01): 32-37.

[4] Giovani Magbiti Monzambe, Khumbulani Mpofu, Ilesanmi Afolabi Daniyan. Statistical Analysis of Determinant Factors and Framework Development for the Optimal and Sustainable Design of Municipal Solid Waste Management Systems in the Context of Industry 4.0 [J]. Procedia CIRP, 2019, 84.

[5] T.-W. Lee. Lagrangian transport equations and an iterative solution method for turbulent jet flows [J]. Physica D: Nonlinear Phenomena, 2020, 403.

[6] Fang Chuanglin, Sun Siao, Jia Shaofeng, Li Yuanyuan. Groundwater Level Analysis Using Regional Kendall Test for Trend with Spatial Autocorrelation [J]. Ground water, 2019, 57 (2).

[7] Science - Systems Science; Study Data from Southeast University Provide New Insights into Systems Science (Test for Heteroscedasticity In Partially Linear Regression Models) [J]. Science Letter, 2019.