Impact of Infrastructure Construction, Industrial Structure and Economic Development on Regional Highway Freight Volume

Shuang Yang [1,#], Yifan Zhang [1,#,*], Pengyu Dai [2,#]

[1]Department of logistics science and engineering, Anhui University, Anhui, 230601, China

[2]College of Humanities and Management, Hebei Agricultural University, Hebei, 061100, China

*Corresponding author: vj162774@163.com

#These authors contributed equally

Abstract: As an indispensable part of the urban economy, urban freight transport plays an essential role in promoting the development of the national economy. It is necessary to clarify how a series of socio-economic developments, such as infrastructure development and economic development, affect regional highway freight volumes. Based on the panel data of each city in China from 2005 to 2019, this study selects 14 indicators closely related to urban economy and highway freight volume and empirically analyzes the relationship between these 14 indicators and regional highway freight volume using panel regression and generalized additive model (GAM) regression. The results show that the infrastructure construction with various types of vehicles has a non-linear influence on highway freight volume. The proportion of secondary and tertiary industries shows that P values are all zero in the corresponding cointegration test and causality test, indicating that there is a cointegration relationship between them and highway freight volume. Both are one-way Granger reasons for urban freight volume.

Keywords: highway freight volume, GAM regression, fixed effect model, regional economic development

1. Introduction

As the central part of the transportation system, Highway transportation plays a vital role in social and economic development. Highway freight volume is an important indicator to measure competitiveness of highway transportation in the whole transportation system. It is also the basis index to formulate plans for highway infrastructure construction and set investment scale. Therefore, it is of great practical significance to study the influencing factors of highway freight volume.

Much research has been devoted to analyzing the main determinants of highway freight volume. Studies have ranged from focusing on economic factors [1] to social factors [2]. Other studies have compared determinants across cities to highlight potential differences and similarities [3]. These studies include factors related to society's industrial structure, consumer power, and transportation infrastructure. However, these studies mainly focus on forecasting highway freight volume, while these studies pay less attention to the indicators affecting highway freight volume.

The present study focuses on the indicators affecting highway freight volume by analyzing highway freight volume across 2005-2019 panel data at the provincial level in China. At the same time, this study selects 14 indicators related to highway freight volume via a thorough review of previous research. Although panel regression and GAM regression are compared in this study, the conclusion is drawn from regression results. The overall aim is to investigate the correlation between these indicators and highway freight volume.

Compared with the existing studies, this paper contributes as follows. First, this study takes those indicators that were rarely involved in previous studies but significantly impacted highway freight volume. Second, GAM regression is more explanatory than panel regression and can better deal with highly non-linear and non-monotonic relationships between the response and the explanatory variables. Third, GAM regression does not consider the correlation, collinearity, and cointegration among the response variables, making the regression results more reliable.

2. Literature review

Most studies have investigated the determinants of highway freight volume by focusing on the influences of GDP growth rate, total retail sales of consumer goods, and road area at the end of the year.A study took the highway freight volume and GDP from 1990 to 2014 as sample data, analyzed the function between highway freight volume and economic development using the regression fitting analysis method, and discussed the changing law of highway freight volume in China with GDP as the driving factor. The analysis shows that the response of highway freight volume to GDP changes is elastic in a particular range, and there is an apparent linear relationship between them [4].Road area at the end of the year belongs to the highway construction in China’s transportation infrastructure construction, an essential part of China’s logistics system. Dai et al. analyzed the relationship between China’s transportation construction and the development scale of the logistics industry[5].

Research on the relationship between highway freight volume and its influencing factors is usually based on the VAR model, the primary analysis model. For example, Gen et al. analyzed the dynamic relationship between highway freight volume, GDP, and transportation structure based on the VAR model [6]. Lu et al. focused on the relationship between Chongqing’s transportation demand and the economy [7]. Zhou et al. focused on comprehensive transportation mode to study the changing relationship between China’s economic development, industrial structure, and transportation structure development [8]. It is reliable to use the neural network method to establish a highway freight volume prediction model [9]. And Yang et al. used BP neural network to predict port throughput [10]. Each model has its applicability and can solve practical problems in different scenarios. Hence, the response model should be selected according to the characteristics of the data. The studies will discuss some rarely studied indicators that greatly influence highway freight volume. And the research logic of this study is as follows. First, this study divides these research indicators into three categories and then uses the extensive database to conclude an essential correlation between the three types of indicators. Second, this study selects the correct model to analyze the relationship between these indicators and highway freight volume. Third, this study determines the hands that were rarely involved in previous studies for specific analysis from these three types of indicators.

The present study addresses the influence of these indicators on regional highway freight volume and takes other main determinants identified by the literature review into account.

3. Methodology

3.1 Independent variables

To analyze the influence of various indicators on regional highway freight volume, an extensive database was collected consisting of 2005-2019 panel data by city level in China.As for the explained variable of highway freight volume, the variation trend of highway freight volume of significant cities in China is obtained by classifying the different data of the city panels according to the level of urban development. Through the analysis of previous literature, this study divides these indicators into three types they belong to infrastructure construction, industrial structure, and economic development. And the classification results are shown in Table.1.

Table.1. Classification of 14 indicators

Classification criteria	Indicator name and abbreviated form	
Infrastructure construction	**A1:** Road area at the end of the year (Road area)	**A2:** Area of the built-up area of jurisdiction (Built-up area)
	A3: Total bus passenger traffic (Total passenger)	**A4:** Number of public electric vehicles (Number of electric vehicles)
	A5: Number of taxi operating vehicles (Number of taxis)	
Industrial structure	**B1:** Proportion of secondary industry (Secondary industry)	**B2:** Proportion of tertiary industry (Tertiary industry)
Economic development	**C1:** Household population (Total population)	**C2:** Gross urban product (GUP)
	C3: Per capita gross urban product (per capita GUP)	**C4:** Total retail sales of consumer goods (Total retail)
	C5: Number of industrial enterprises (Industrial enterprises)	**C6:** Current assets of industrial enterprises (Current assets)
	C7: Fixed assets of industrial enterprises (Fixed assets)	

The selection of these indicators ensured that many factors could be taken into account. However, some cities havemissing data and outliers. Hence, this study deletes towns with many missing data, such as Xinjiang province. In addition, some outliers are processed by SPSS software. After the above data processing, a dataset with 63675 data volumes is obtained, and this dataset concludes with cities from the eastern region, western region, and central region. The indicators included with their summary statistics can be viewed in Table 2. The large span in city areas and data years ensured the reliability of regression results.

Table.2. The summary statistics of 14 indicators

	Mean	P50	Sd	Min	Max
HFV	9508	6531	9755	335	107064
Road area	1700	907	2334	38	24473
Built-up area	126	73	173	6	1515
Total passenger	20338	7170	43771	32	525606
Number of electric vehicles	1388	546	2895	11	38728
Number of taxis	3070	1463	5966	51	71517
Secondary industry	48	49	12	8	90
Tertiary industry	45	44	11	9	84
Total population	146	94	190	15	2479
GUP	116721089	3930900	27129975	67932	3.816×10^{8}
Per capita GUP	52104	43267	36036	2662	467749
Total retail	4974526	1697354	11082240	21853	1.585×10^{8}
Industrial enterprises	600	211	1269	6	18474
Current assets	8297889	2560717	20293227	11005	2.857×10^{8}
Fixed assets	5325136	2305812	9431556	10527	91918073

A strong relationship between infrastructure construction and highway freight volume was observed in all studies in the previous literature. However, former literature mainly focused on road construction and the city's acreage, and few studies focused on the traffic conditions, such as the number of buses. At the same time, the number of transportation plays an essential part in the traffic congestion. Hence, the number of different types of transport is taken into account in this study. However, it is not easy to collect data on various modes of transportation. Hence, the number of taxi-operated vehicles, the number of public electric vehicles, and the total passenger volume of buses represent the traffic conditions.

Freight volume is closely linked to a country's industrial design. Generally speaking, the development of the tertiary industry will not bring about a large amount of freight demand. Still, there will be a certain amount of new logistics demand, such as express delivery. The development of primary and secondary industries, which are closely related to the growth of freight volume, will bring a large amount of freight demand due to the characteristics of industrial development. However, the primary industry accounts for a small proportion of the three departments. [11]. Therefore, this study only considers the impact of the secondary industry and tertiary sectors on freight volume.

Economic development has a direct impact on freight volume. Previous studies generally considered the gross urban product, per capita gross urban product, and total retail sales of consumer goods. While other indicators affect the development of the economy, such as the population size and the number of industrial enterprises. According to previous studies, population growth has a particular impact on economic development while regional differences stil exist. Hence, the family size is taken into account in this study. In addition, previous studies did not consider the effect of industrial enterprises. However, industrial enterprises are the source of highway freight volume. Hence, this study studies the number of industrial enterprises, current assets of industrial enterprises, and fixed assets of industrial enterprises.

3.2 Generalized additive model

The generalized additive model (GAM) extends the additive model and can explain the non-linear relationship between response variables and impact factors. The only underlying assumption is that the functions are additive and that the components are smooth. The basic form of GAM is,

$$\mathrm{g(u\_i\ ) = \beta\_0 + s(x\_1\ ) + \cdots + s(x\_n\ ) + \varphi} \qquad (1)$$

In the expression, $g(u\_i\ )$ represents the join function, β_0 represents intercept, $s(x\_n)$ is a smooth function representing the complex relationship between highway freight volume and each variable, and φ defines the random residual term. When gam fits variables, it is necessary to set a smoother type for each variable to fit the local smoothing of response variables and independent variables. The choice of smoothing parameter values will strongly affect the smoothness of curves in the results, and even produce variable response trends in different states, which will lead to different understanding and interpretation of biological meanings. For example, *s ()* designates the smoother as spline smoothing. In this study, spline smoothing is chosen because of the different data nonlinearity. Although GAM, as a type of nonparametric smooth regression, is often used as an exploratory tool, R still provides a statistical method for the significance of independent variables, variance explanation rate, etc., which can be directly viewed through additive function. According to the essential function, this study establishes the GAM model based on the actual problem, and the expression is as follows,

$$\begin{gathered}\mathrm{g}(HFV) = \mathrm{s(road\ area) + s(built - up\ area) + s(total\ passenger)} \\ \mathrm{+s(number\ of\ electric\ vehicles) + s(number\ of\ taxis) + s(secondary\ industry)} \\ \mathrm{+s(tertiary\ industry) + s(total\ population) + s(GUP) + s(per\ capita\ GUP)} \\ \mathrm{+s(total\ retail) + s(industrial\ enterprises) + s(current\ assets) + s(fixed\ assets) + \varphi} \qquad (2)\end{gathered}$$

The strength of GAM is the ability to deal with highly non-linear and non-monotonic relationships between the response and the set of explanatory variables. GAM is sometimes referred to as data-rather than model-driven. This is because the data determine the nature of the relationship between the response and explanatory variable sets rather than assuming some form of parametric relationship.

4. Analyses and results

4.1 The result of panel regression

The study needs to conduct a unit root test on explanatory variables to check whether the data is stable. The stationarity requirement panel regression can only be carried out when the variables meet. For explanatory variables that do not pass the unit root test, the cointegration test was conducted in this study to test whether there was a particular cointegration relationship between explanatory variables and highway freight volume. The primary cointegration test also included Kao, Pedroni, and Westerlund. Different degrees of cointegration relationships can be seen from the number of passed tests. And the detailed results can be viewed in Table 3.

From Table 4, the road area at the end of the year's regression coefficient value is 1.929, indicating that the actual road area at the end of the year would have a significant positive impact on the highway freight volume. The total population shows a significant level of 0.05, and the regression coefficient value is 7.119, which indicates that the total population would have a significant positive impact on highway freight volume. As far as the industrial enterprises are concerned, it shows a significant level of 0.05. The regression coefficient value is 0.768, indicating that the number of industrial enterprises will significantly positively impact highway freight volume. As far as the total retail is concerned, the regression coefficient value is 0.000283, indicating that the total retail will significantly positively impact highway freight volume. The number of electric vehicles shows a significant level of 0.01. The regression coefficient value is 0.894, which indicates that the number of public electric vehicles will have a significant positive impact on highway freight volume.

Table.3. The result of the unit root test and the cointegration test

VARIABLES	Whether can participate in a panel regression	Unit root test ADF	LLC	IPS	Whether there is cointegration	Co-integration test Kao	Pedroni	Westerlund
Road area	Y	√	√	√	—	—	—	—
Built-up area	N	×	√	√	Y/+++	√	√	√
Total passenger	N	×	√	×	—	—	—	—
Number of electric vehicles	Y	√	√	√	Y/+++	√	√	√
Number of taxis	N	×	√	×	+++	√	√	√
Secondary industry	N	×	√	×	+++	√	√	√
Tertiary industry	N	×	√	×	Y/+++	√	√	√
Total population	Y	√	√	√	—	—	—	—
GUP	N	×	√	×	Y/+++	√	√	√
per capital GUP	N	×	√	×	Y/+++	√	√	√
Total retail	Y	√	√	√	—	—	—	—
Industrial enterprises	Y	√	√	√	—	—	—	—
Current assets	N	×	×	√	Y/+	×	×	√
Fixed assets	N	×	×	×	Y/+++	√	√	√

Table.4. The result of the fixed effect model

VARIABLES	FE
Road area	1.929***
Number of taxis	0.520***
Total population	7.119**
Total retail	0.000135***
Industrial enterprises	0.768***
Constant	3,337***
Observations	4,245
Number of cities	283
R-squared	0.244
Hausman	68.93

4.2 The result of GAM regression

According to the results, the adjusted decision coefficient R^2 is 0.624, and the variance explanation rate *IRV* is 63.2%. It indicates that the GAM model has a 63.2% ability to explain the overall change in highway freight volume.

Table.5. The result of GAM regression

Variable	edf	P	Variable	edf	P
S (Road area)	8.016	$< 2\times10^{-16}$ ***	S (Built-up area)	6.179	0.01120 *
S (Total passenger)	7.138	$< 2\times10^{-16}$***	S (Number of vehicles)	6.791	3.16×10-6 ***
S (Number of taxis)	7.800	$< 2\times10^{-16}$***	S(Secondary industry)	1.002	0.00570 **
S (Fixed assets)	1.016	0.05833.	S (Tertiary industry)	1.698	0.02468 *
S (Total population)	7.690	$< 2\times10^{-16}$ ***	S (GUP)	2.254	$< 2\times10^{-16}$ ***
S (per capital GUP)	5.996	8.11×10^{-7} ***	S (Total retail)	5.877	0.00892 **
S (Industrial enterprises)	5.110	$< 2\times10^{-16}$ ***	S (Current assets)	2.983	7.58×10-6 ***

Table 5 illustrates the relative importance between the 14 indicators and highway freight volume based on GAM regression. F factor with an enormous value indicates that this indicator has a more significant influence. From the results of GAM regression, the infrastructure construction and economic development indicators have a more substantial discount. And the maximum statistic is that the total population reached 33.091. And the degree of freedom can indicate the correlation between the indicators and regional highway freight volume. When the degree of freedom is close to 1, the linear relationship is represented. When the degree of freedom is greater than 1, it means that there is a curve relationship between them.

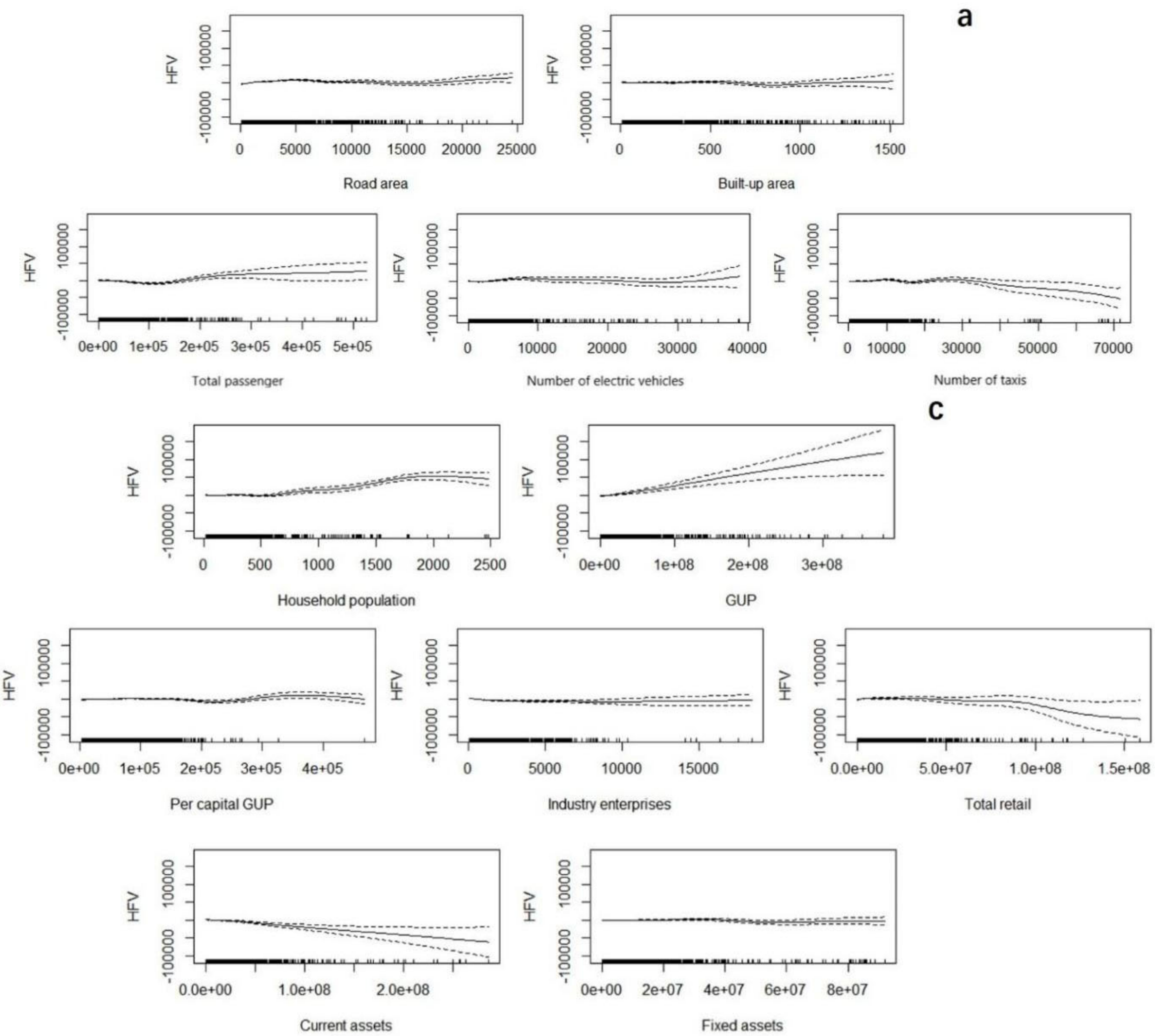

Figure 1. The influence trends of 14 indicators on highway freight volume from different levels

Fig.1 shows the variation trend of response variables after GAM model fitting. From this figure, the influence trends of various indicators on highway freight volume can be seen. And there is a linear relationship between the secondary industry, tertiary industry, current assets of industrial enterprises, and highway freight volume.

Compared with the panel and GAM regression results, this study finds that the influence trend of each indicator on highway freight volume is roughly the same in the two regressions. However, GAM regression is higher than panel regression in explaining each variable, reaching 63.2%. There is no unit root and cointegration test like panel regression, so the analysis process is much simpler. Based on Fig.1 and Table 5, this study selects three indicators that have different effects on highway freight volume, and these three indicators are clearly representative in both panel regression and GAM regression. In contrast, in the previous literature, few studies have been done.

4.3 Factor analysis

The first indicator is the usage of various types of vehicles from the infrastructure construction. From Fig.1, the result of GAM regression, we can see the influence of the number of different cars on highway freight volume. There is an apparent non-linear relationship between them, while with the increase of the number of different vehicles, the highway freight volume also increases to various degrees. Chen et al. selected the highway freight volume and other indicators from 1990 to 2012 as the main explanatory variables affecting the number of civil cars in China. They analyzed that the number of police vehicles in China is mainly affected by various taxes and highway freight volume. There is a significant positive correlation between the number of civil cars and freight volume [12]. Yang et al. studied the urban traffic efficiency in China, selecting urban GDP, population density, and other indicators as input indicators, passenger volume, and freight volume as output indicators. And using the spatial lag model to analyze the influencing factors of urban ecological traffic efficiency, it was found that the number of taxis was negatively correlated with urban traffic efficiency and negatively associated with highway freight volume [13]. However, this study has not found such an apparent linear relationship, but the development trend is similar to previous studies.

The second indicator is the proportion of secondary and tertiary industries from the industrial structure. The research on the influence of industrial structure on highway freight volume can be divided into the whole, the part, and the individual cases according to different research objects.

First, there is a long-term cointegration relationship between China's highway freight volume and the proportion of the secondary and the tertiary industry in the whole object. And the output value of the primary industry and the ratio of the primary industry have a continuously positive impact on freight volume. There is a one-way Granger causality between the proportion of the primary industry and highway freight volume. It is speculated that the secondary and tertiary industries will have a certain degree of negative impact on freight volume [14].

Furthermore, for some local objects, such as some cluster urban agglomerations with economic and transportation links, there is a one-way Granger causality between highway freight volume and the primary and secondary industries and a two-way Granger causality between highway freight volume and the tertiary industry [15]. The industrial structure of some cities also has different influences on the logistics volume. For example, the secondary industry in Dali has a high correlation with logistics demand, and the effect of the secondary industry on logistics demand is increasing. Similarly, Shanghai and Shenzhen, the first-tier cities, have some similarities. Shanghai's freight volume correlates with the secondary industry.

In contrast, the freight volume per unit GDP is positively correlated with the primary industry and negatively correlated with the secondary and tertiary industries. The freight volume of Shenzhen is cointegration with the secondary and tertiary industries. The proportion of Shenzhen's secondary and tertiary industries has a one-way Granger causality with freight volume, which is also the Granger cause of freight volume [16]. For cities with different development levels, the secondary industry is most related to urban logistics volume and urban road freight volume. The industrial structure has a significant influence on urban road freight volume.

Table.6. Cointegration and Granger causality tests

Types of tests	Statistics	HFV Secondary industry	HFV Tertiary industry
Kao test for cointegration	Modified Dickey-Fuller t	8.7175	8.5432
	Dickey-Fuller t	6.6260	6.5133
	Augmented Dickey-Fuller t	5.0395	4.8212
Granger causality test	Lag-order	1	1
	W-bar	1.8207	1.8332
	Z-bar	9.7630	9.8993
	Z-bar title	4.8736	4.9669

The final indicator is the gross urban product from the economic development. Previously, the analysis of highway freight volume and the economic level was based on provincial panel data. At the same time, it was found that there was a cointegration relationship between highway freight volume and GDP in different provinces, and the driving effect was different in other regions [17]. Overall, China's current domestic cargo turnover also showed a significant positive driving impact on the national economy. Still, the development of the national economy did not considerably impact the growth of cargo turnover [18]. However, the GDP has a continuous negative impact on highway freight volume [19]. Locally, there is a long-term equilibrium relationship between highway freight volume and economic indicators. For example, there is a one-way Granger causality between highway freight volume and highway mileage, number of industrial enterprises, and GDP in Shanghai. And from the existing literature research on urban freight transport and economic growth, most of them focus on the relationship between the increase in passenger traffic and urban economic growth caused by infrastructure construction such as high-speed rail. And the related infrastructure construction has a significant spatial spillover effect on China's urban economy [20].

The existing research lacks research on urban freight transport and urban economic growth, and the urban logistics volume index selection is not scientific. Therefore, it is necessary to explore the relationship between highway freight volume and urban economic growth in cities and conduct a classification analysis for cities with different development levels. Therefore, two indicators, gross urban product and highway freight volume are selected, and it is finally found that there is a particular rule between them. Fig.2 shows that the sample cities are grouped according to different development levels and added fitted lines.

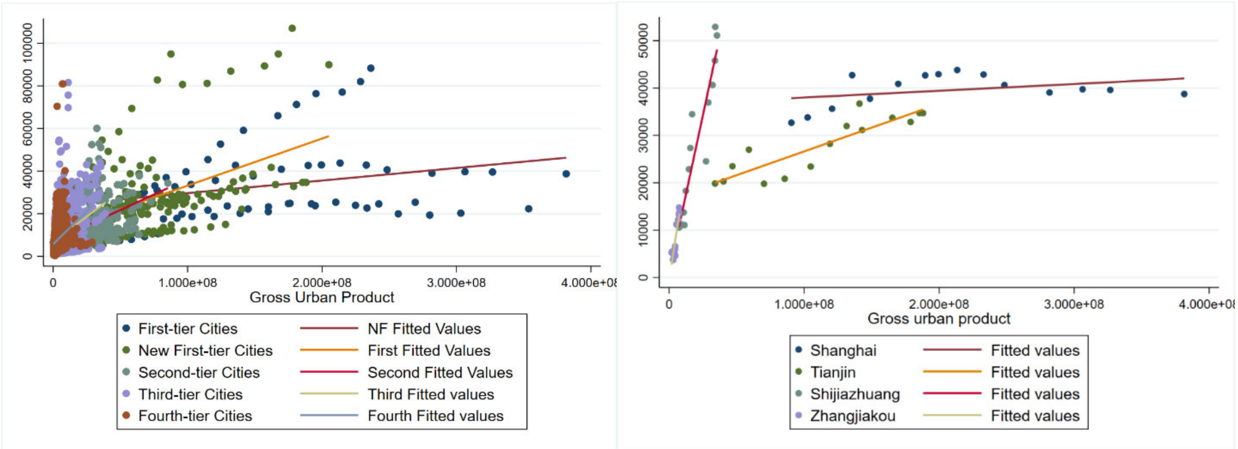

Figure 2. Highway freight volume and GUP in cities of different development levels and highway freight volume and GUP of selected cities

Fig.2 shows that from the First-tier Cities to the Fourth-tier Cities, the urban development level decreases while the slope of the fitting line between highway freight volume and the gross urban product increases. From this, it can be concluded that the promotion effect of the gross urban product on highway freight volume is more evident in cities with low development levels and vice versa.

5. Conclusions

This study analyzed the determinants of highway freight volume based on 2005-2019 panel data by provincial level in China with a specific focus on infrastructure construction, industrial structure, and economic development. The database collected almost all indicators related to all aspects of highway freight volume, and these indicators were identified by a thorough literature review of previous studies. The panel regression and GAM regression results both confirmed that the usage of various types of vehicles from the infrastructure construction has a non-linear influence on highway freight volume. Comparing the panel regression with GAM regression, it can be found that the GAM model has a 63.2% ability to explain the overall change in highway freight volume, and the relationship between each indicator and highway freight volume can be displayed visually by image. Meanwhile, GAM regression is more illustrative than panel regression. It can better handle the highly non-linear and non-monotonic relationships between the response and the explanatory variables. GAM regression does not consider the correlation, collinearity, and cointegration among the response variables, making the regression results more reliable. The deficiency of this research lies in that only three types of indicators are selected from the research variables for analysis. In addition, the explanation degree of the GAM model is not too high, and it can be improved.

References

[1] Lin Hang Fei, Luo Yulong. Cointegration analysis of the relationship between road freight volume and economic growth in Shanghai [J]. Journal of Tongji University, 2008,36 (10): 1378-1383.

[2] Chen Wei. Research on urban freight volume and its influencing factors [J]. Science and Education Guide-Electronic Edition (mid-year), 2015(7):109-109,183.

[3] Zhao Yongchao, Sun Youwang. The relationship between freight volume per unit GDP and urban industrial structure [J]. Transportation Science and Technology and Economy, 2009,11 (3): 121-122.

[4] Yao Jing, Li Shuang. Quantitative analysis of population and economic development [J]. The people of Northwest China, 2001 (1): 45-48.

[5] Lu Yi, Peng Wei, Wang Lizhi. Cointegration analysis of road freight volume, fuel price, and national economy [J]. Journal of transportation systems engineering and information technology, 2010,10 (1): 28-32.

[6] Geng Bo. Research on the dynamic relationship between China's freight volume, GDP, and transportation structure [J]. Practice and understanding of mathematics, 2014,44(16):8-12.

[7] Qi Xingmin, Xu Hai, Duan Chen, et al. Forecast of logistics freight volume in Xiangyang based on grey relational analysis and genetic neural network [J]. Logistics Technology, 2021,40 (7): 67-72,131.

[8] Dong Wang. Highway freight volume forecasting method based on neural network algorithm [J]. Journal of Beihua University (Natural Science Edition), 2014 (3): 417-420.

[9] Yang Sifan, Wei Meng, Ding Rong. Port cargo throughput forecast based on BP neural network with multi-factor selection and error correction [J]. Value Engineering, 2020,39(18):68-71.

[10] Yuan Yunxiao, Wang Baohai. Empirical analysis of the impact of freight volume on regional economic development based on the panel data of 31 provinces and cities in China [J]. China Market, 2018 (7): 14-17, 42.

[11] Cui Shudan, Li Xingye, Xiong Jing, Hua Maoming. Test the interaction between China's economic growth and freight turnover [J]. Statistics and Decision, 2008(08):75-77.

[12] Xiong Hao, Sun Youwang. Cointegration analysis of the relationship between freight volume and economic growth in China [J]. Systems Engineering, 2010,28(09):46-50.

[13] Gao Wenxiang, Lu Hong. Research on Logistics Demand of Dali Prefecture Based on Economic and Industrial Structure [J]. Railway transport and economy, 2015,37(07):11-15.

[14] Lin Hang Fei, Luo Yulong. Cointegration analysis of the relationship between road freight volume and economic growth in Shanghai [J]. Journal of Tongji University (Natural Science Edition), 2008(10):1378-1383.

[15] Dai Ying, Deng Yuchen, Song Han. Research on the Impact of China's Transportation Infrastructure Construction on the Development of Logistics Industry [J]. Railway transport and economy, 2014,36 (1): 71-76.

[16] Huang Juan, Liu Shuangxian. Freight volume forecast based on combination forecast [J]. Science and Technology and Innovation, 2021 (16): 94-96.