Research on the Relationship between Highway Trip Demand and Social Economy in the New Normal Economic stage of China Based on Panel Data Analysis

Zhuoqun Sun¹,², Yi Zhang¹,², Wenbo Ji³, Shengli Li³, Chao Wang¹,² and Zhirui Ye¹,², *

¹School of Transportation, Southeast University, Nanjing 210096, China.
²Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, Nanjing 210096, China.
³Shandong Provincial Communications Planning and Design Institute, Jinan, China.

*Corresponding author: yezhirui@seu.edu.cn

Abstract. As China entered the new normal of economy, the economic growth rate and industrial structure are undergoing huge changes, which has also profoundly affected the traffic volume of expressways. Therefore, analyzing the relationship between socio-economic and transportation demand is crucial for transportation planning and demand management. This study analyzed the interaction of socio-economic and trip demand from 2008 to 2017, which was a crucial period for Shandong’s industrial transformation and economic development. To achieve this goal, principal component analysis and panel data analysis methods were applied to our issue.

1. Introduction

China has entered a new normal stage of economy, and the economic growth rate and industrial structure are undergoing drastic changes. At the same time, the relationship between socio-economic and highway traffic is getting closer. In this context, we need to further study the impact between them.

2. Literature review

The forecast of traffic demand is the most important and basic step in the four-stage method. Traffic demand is a derivative demand of social and economic activities. Using multiple social-economic indicators to establish an economic forecasting model is an accurate and reliable approach to predict traffic demand. Prevedouros & Halkias (2016) utilizes socioeconomic indicators such as GDP and fuel quantity, and considers the time lag term to regress traffic demand. Hui & Zhong (2012) modeled freight traffic demand based on hierarchical regression. The development of machine learning and deep learning has also made the regression method into the field of traffic demand forecasting.

The traffic demand and economic data is obtained from continuous observation of each traffic community in a continuous period of time, which can be regarded as panel data in econometrics (Haughton & Haughton, 2011). At present, panel data model and dynamic panel data model have been applied in traffic demand prediction. González & Marrero (2012) used dynamic panel data model to quantitatively analyze the influence of road network growth and other factors on the induced increase
of road traffic in major Spanish regions from 1998 to 2006. Graham, Crotte, & Anderson (2009) estimated the influence of subway fare, income and service quality on traffic demand in 22 cities based on the dynamic panel method. Su (2010) studied the influence of urban spatial characteristics and other factors on traffic demand based on the panel data of 85 cities over 20 years by using the panel analysis method. Gomez, Vassallo, & Herraiz (2016) established a dynamic panel data model to explain the relationship between light vehicle traffic demand and key socio-economic variables over time.

Because transportation activities come from social and economic activities, transportation demand is a derivative demand of social economy. The complexity and diversity of social and economic activities result in complex factors which affecting transportation demand. In general, the development of transportation demand depends on factors such as industrial layout, population size and urban spatial layout. Socioeconomic needs are the essential for transportation demand. Many researches have used methods in their studies, including regression analysis, structural equation modeling, Granger causality test, etc. The method of regression analysis is the most common method. Matas & Raymond (2003) utilizes the elasticity coefficient of GDP and transportation demand to show that transportation demand is highly correlated with GDP. Analysis of socio-economic activities and traffic demand requires multiple causes and results, or factors which cannot be directly observed. Xing & Jing (2016) proposed structural equation modeling to explore the intrinsic link between economic measures, personal characteristics and other influencing factors, and the extent to the performance with different elements on traffic demand management (TDM). Choo & Mokhtarian (2007) based on the use of structural equation modeling to explore the causal relationship between transportation demand, communication, land use, economic activity and sociodemographic.

In the new-normal economy stage in China, highway plays an important role in social development, and economy situation is showing a tighter, more massive impact on highway traffic generation. It is feasible and necessary to re-analyze the relationship between highway traffic demand and socio-economic factors using panel data methods.

3. Methods

Panel data has n entities observations at equal or more than two periods. Sometimes panel data is also called longitudinal data as it adds a temporal dimension to cross-sectional data. Panel data is repeated observations of individuals at different time.

It can be denoted as follow

\[ (X_{it}, Y_{it}), i = 1, \ldots, n \text{ and } t = 1, \ldots, T \]  

where the index \( i \) refers to the entity while \( t \) refers to the time period.

If \( t \) is defined as a constant, \( (X_{i}, Y_{it}) \) is random variable in cross section; if \( i \) is defined as a constant, \( (X_{it}, Y_{it}) \) is time series on the longitudinal section.

Advantages of building models with panel data:

- There is a large number of observations, so that the sampling accuracy of the estimator can be improved.
- From fixed effect model, we can get consistent estimator or even the effective estimator of parameter.
- Researchers can analyze more dynamic information using panel data model than single-section data model.

3.1. Fixed Effects Panel model

The fixed effects regression model takes the form:

\[ Y_{it} = a_{i} + \beta_{1}X_{1, it} + \cdots + \beta_{k}X_{k, it} + u_{it} \quad i = 1, \ldots, n \text{ and } t = 1, \ldots, T \]
The $\alpha_i$ are entity-specific intercepts that capture heterogeneities across entities, and the variation is related to $X_{it}$. $X_{it}$ denotes a $K \times 1$ vector of the regression variable; $\beta_k$ is a $K \times 1$ vector of regression coefficients, $u_{it}$ is defined as the error value. Model a is named as entity fixed effects model. The strong assumption is $E(u_{it}|X_{it}) = 0$, $i = 1,2,...,N$. $\alpha_i$ is unobservable, and can describe the differences between models established by different individuals.

3.2. Random Effects Panel model
Consider the panel data model

$$y_{it} = \alpha_i + X_{it} \beta + \varepsilon_{it} \quad i = 1,2,...,N; t = 1,2,...,T \quad (3)$$

In this situation, $\alpha_i$ is a random variable and its distribution is independent of $X_{it}$. $X_{it}$ is a $K \times 1$ vector of regression variable. $\beta$ represents a $K \times 1$ vector of regression coefficients, $\varepsilon_{it}$ denotes the error value. This model is called individual random effects model. The assumptions is that $\alpha_i \sim itd(\alpha, \sigma^2_\alpha)$ and $\varepsilon_{it} \sim itd(0, \sigma^2_\varepsilon)$ are independent and identically distributed, and there is no limit to what kind of distribution.

4. Results

4.1. Data Collection
The data in this study mainly consists two parts. The first part is the socio-economic indicator data, which is obtained from the Statistical Yearbook of the Statistical Bureau of Shandong Province from 2008 to 2017; the other part is the highway trip data. This part of the data is collected by the transportation department of Shandong Province through the highway toll station.

4.2. Principle Component Analysis
In order to improve the accuracy of the model, several socio-economic factors were considered in the study. Therefore, it is necessary to carry out principal component analysis on socio-economic factors to eliminate the collinearity between independent variables. As a result, three principal components were extracted by the principal component analysis. The principal components carried 86.693% of the information contained in the original data, and each common factor contained more than 10% of the information.

Table 1 shows the rotated component matrix and the component score coefficient matrix in the process of principal component analysis. The rotated component matrix represents the correlation between each principal component and the original variables, and analyze which socio-economic aspects the principal component can represent. Principal component 1 can be considered that the principal component represents the degree of regional economic development. The principal component 2 can represent the degree of the Industrial structure optimization because it is obviously negatively correlated with the proportion of the secondary industry and positively correlated with the proportion of third industry. The principle component 3 can be considered to represent the residents’ living standards. Each principal component can be calculated as follows:

$$Z_1 = 0.956X_{GDP} - 0.410X_{\text{percentile of first industry}} - 0.937X_{\text{civilian vehicles}}$$
$$Z_2 = 0.032X_{GDP} + 0.663X_{\text{percentile of first industry}} + 0.116X_{\text{civilian vehicles}}$$
$$Z_3 = 0.208X_{GDP} - 0.406X_{\text{percentile of first industry}} - 0.051X_{\text{civilian vehicles}} \quad (4)$$
Table 1. Principal component analysis coefficient of social economic indicators variables

|                          | Rotational component coefficient matrix | Component score coefficient matrix |
|--------------------------|----------------------------------------|----------------------------------|
|                          | 1            | 2            | 3            | 1            | 2            | 3            |
| GDP                      | 0.879        | 0.404        | 0.148        | 0.956        | 0.032        | 0.208        |
| Percentile of first industry | -0.343      | -0.085       | -0.804       | -0.41        | 0.663        | -0.406       |
| Percentile of second industry | -0.214      | -0.876       | 0.339        | -0.657       | -0.176       | 0.682        |
| Percentile of third industry | 0.354        | 0.859        | 0.043        | 0.799        | -0.132       | -0.457       |
| Industrial Production    | 0.837        | 0.133        | 0.277        | 0.775        | -0.038       | 0.44         |
| Completion value of fixed assets investment | 0.8 | 0.493 | 0.151 | 0.947 | -0.015 | 0.098 |
| Absolute consumption of all residents | 0.256 | 0.658 | 0.637 | 0.664 | -0.665 | -0.141 |
| Retail sales of social consumer goods | 0.754 | 0.605 | -0.008 | 0.96 | 0.093 | -0.07 |
| CPI                      | 0.095        | 0.801        | 0.345        | 0.59         | -0.476       | -0.441       |
| Population               | 0.62         | 0.207        | -0.701       | 0.541        | 0.787        | -0.084       |
| Total value of imports and exports | 0.828 | 0.02 | 0.211 | 0.693 | 0.046 | 0.498 |
| Highway mileage           | 0.899        | 0.229        | -0.214       | 0.828        | 0.414        | 0.222        |
| Number of civilian vehicles | 0.752        | 0.572        | -0.025       | 0.937        | 0.116        | -0.051       |

4.3. Panel Data Analysis

The choice of fixed and random effects in the panel data model is fundamental to building a spatial panel model. In the study, the pooled model, the fixed effect panel data model and the random effect panel data model of passenger and freight trip were established. The effectiveness of the effect was tested by F test, and two types of effect were selected by Hausman test.

Table 2. Passenger trip generation panel data model

| Variable                  | Pooled Coefficient | Pooled Prob. | Fixed Effect Coefficient | Fixed Effect Prob. | Random Effect Coefficient | Random Effect Prob. |
|---------------------------|--------------------|--------------|--------------------------|--------------------|--------------------------|---------------------|
| Constant                  | 22311.40           | 0.00         | 22311.40                 | 0.00               | 22311.40                 | 0.00                |
| Principle factor 1        | 20588.28           | 0.00         | 41233.45                 | 0.00               | 27438.54                 | 0.00                |
| Principle factor 2        | 14333.57           | 0.00         | 15320.21                 | 0.00               | 13097.08                 | 0.00                |
| Principle factor 3        | 2453.25            | 0.04         | -11473.67                | 0.00               | -710.42                  | 0.75                |
| R-squared                 | 0.74               |              | 0.90                     |                    | 0.70                     |                    |
| Prob(F-statistic)         | 0.00               |              | 0.00                     |                    | 0.00                     |                    |

Hausman Test: 0.00

Table 3. Passenger trip attraction panel data model

| Variable                  | Pooled Coefficient | Pooled Prob. | Fixed Effect Coefficient | Fixed Effect Prob. | Random Effect Coefficient | Random Effect Prob. |
|---------------------------|--------------------|--------------|--------------------------|--------------------|--------------------------|---------------------|
| Constant                  | 22340.68           | 0.00         | 22340.68                 | 0.00               | 22340.68                 | 0.00                |
| Principle factor 1        | 20657.45           | 0.00         | 41051.07                 | 0.00               | 27494.99                 | 0.00                |
| Principle factor 2        | 14325.52           | 0.00         | 15258.61                 | 0.00               | 13066.23                 | 0.00                |
| Principle factor 3        | 2447.99            | 0.04         | -11348.81                | 0.00               | -743.99                  | 0.74                |
| R-squared                 | 0.74               |              | 0.90                     |                    | 0.70                     |                    |
| Prob(F-statistic)         | 0.00               |              | 0.00                     |                    | 0.00                     |                    |

Hausman Test: 0.00
Table 4. Freight trip generation panel data model

| Variable                | Pooled       | Fixed Effect        | Random Effect       |
|-------------------------|--------------|---------------------|---------------------|
|                         | Coefficient  | Prob.               | Coefficient         | Prob.               | Coefficient | Prob.               |
| Constant                | 39018.33     | 0.00                | 39018.33            | 0.00                | 39018.33    | 0.00                |
| Principle factor 1      | 26184.60     | 0.00                | 8430.01             | 0.13                | 19982.66    | 0.00                |
| Principle factor 2      | 11014.50     | 0.00                | 7460.72             | 0.00                | 8507.09     | 0.00                |
| Principle factor 3      | -4573.52     | 0.01                | 98.75               | 0.99                | -5533.11    | 0.09                |
| R-squared               | 0.64         | 0.85                | 0.31                |                      |             |                     |
| Prob(F-statistic)       | 0.00         | 0.00                | 0.00                |                      |             |                     |
| Hausman Test            |              |                     |                     |                      |             | 0.02                |

Table 5. Freight trip attraction panel data model

| Variable                | Pooled       | Fixed Effect        | Random Effect       |
|-------------------------|--------------|---------------------|---------------------|
|                         | Coefficient  | Prob.               | Coefficient         | Prob.               | Coefficient | Prob.               |
| Constant                | 38813.89     | 0.00                | 38813.89            | 0.00                | 38813.89    | 0.00                |
| Principle factor 1      | 26200.58     | 0.00                | 7344.50             | 0.18                | 19459.15    | 0.00                |
| Principle factor 2      | 10771.33     | 0.00                | 7130.61             | 0.00                | 8230.44     | 0.00                |
| Principle factor 3      | -4326.58     | 0.01                | 730.71              | 0.90                | -5253.34    | 0.10                |
| R-squared               | 0.64         | 0.85                | 0.59                |                      |             |                     |
| Prob(F-statistic)       | 0.00         | 0.00                | 0.00                |                      |             |                     |
| Hausman Test            |              |                     |                     |                      |             | 0.01                |

The results in the tables indicated that both the fixed effect and the random effect in the panel data model of passenger and freight trip are effective. On the other hand, the results of Hausman test indicates that the above panel fixed effects model is more applicable. From the coefficients of the model, it can be found that all three socio-economic principal components have a positive correlation with passenger traffic demand, and the effects of principal component 1 and principal component 2 are significantly greater than those of principal component 3. As for the freight transportation demand, principal component 2 is a significant influencing factor, and the effects of principal component 1 and principal component 3 are not significant.

5. Conclusion
The purpose of this article is to study the relationship between socio-economic factors and transport demand under the new normal stage of economy. The principal component analysis method is used to extract the valid information of the social and economic factors, and then the panel data analysis method is used to model and analyze the spatial relationship between the social economic principle components and the passenger and freight trip demand. The results show that the relationship between passenger trip demand, freight trip demand and the socio-economic under the new normal economic stage are becoming closer, and passenger transport demand is more susceptible to socioeconomic factors.

References
[1] Prevedouros, P and Halkias, B. Expressway Traffic Demand Forecasts in the Volatile Economic Environment of Greece, (Transportation Research Procedia, 2016), pp: 607-619.
[2] Hui, L. W. and Zhong, K. S. Hierarchical Regression Model for Analyzing Truck Freight Demand, (Applied Mechanics & Materials. 2012), pp: 2752-2756.
[3] Haughton, D and Haughton, J. “Chapter 9 Panel Data” in Living Standards Analytics. (Statistics
for Social & Behavioral Sciences, 2011), pp: 175-187.

[4] González, R. M. and Marrero, G. A., *Induced road traffic in Spanish regions: A dynamic panel data model*, (Transportation Research Part A: Policy and Practice, 2012), pp: 435-445.

[5] Graham, D. J., Crotte, A. and Anderson, R. J., *A dynamic panel analysis of urban metro demand* (Transportation Research Part E: Logistics and Transportation Review, 2009), pp: 787-794.

[6] Su, Q., *Travel demand in the US urban areas: A system dynamic panel data approach*, (Transportation Research Part A: Policy and Practice, 2010), pp: 110-117.

[7] Gomez, J., Vassallo, J. M., and Herraiz, I. *Explaining light vehicle demand evolution in interurban toll roads: a dynamic panel data analysis in Spain*, (Transportation, 2016), pp:677-703.

[8] Matas, A. and Raymond, J. *Demand Elasticity on Tolled Motorways*, (Journal of Transportation and Statistics, 2003), pp: 91-108.

[9] Xing, C., “Research on the effect of Traffic Demand Management of Beijing by Structural Equation Modeling” Master, thesis, Beijing Jiaotong University, 2016.

[10] Choo, S. Mokhtarian, P. L. *Telecommunications and travel demand and supply: Aggregate structural equation models for the US*, (Transportation Research Part A: Policy and Practice, 2007), pp: 4-18.

[11] JianGuo Qi, Hong Wang and Xuzhe Peng, et al. *The Connotation and Formation Mechanism of Chinese New Normal Stage of Economy*, (Economic Review Journal, 2015). (in Chinese).