



## ASSESSING EFFECTS OF BUS SERVICE QUALITY ON PASSENGERS' TAXI-HIRING BEHAVIOR

Hong-Wei WANG<sup>1</sup>, Zhong-Ren PENG<sup>2,3\*</sup>, Qing-Chang LU<sup>4</sup>,  
Daniel (Jian) SUN<sup>5</sup>, Cong BAI<sup>6</sup>

<sup>1,2,4,5,6</sup>*Center for ITS and UAV Applications, State Key Laboratory of Ocean Engineering,  
Shanghai Jiao Tong University, China*

<sup>3</sup>*Dept of Urban and Regional Planning, University of Florida, United States*

Received 30 November 2014; revised 9 July 2015, 2 November 2015; accepted 28 December 2015;  
published online 16 January 2017

**Abstract.** Due to low quality of bus service in a congested road network, some bus-waiting travelers would take taxis instead in order to save time or get to their destinations on time. However, the correlation between bus service quality and passengers' taxi-hiring behavior is essentially unknown. This paper aims to assess the effects of bus service quality on taxi-hiring behavior based on historical data from the Global Position Systems (GPS) equipped buses and taxis in the city of Shenzhen, China. The taxi-hiring behavior is captured by analyzing the taxi-data, such as the origins of passenger pick-up, destinations of passengers drop-off, and taxi paths from the taxi movement data. The quality of bus service is assessed based on the bus location information. Parametric, semiparametric and nonparametric models are developed to explore the effects of bus service quality on taxi-hiring behavior. The results indicate that bus speed, headway and stoppage time are the core factors affecting passengers' taxi-hiring behavior. Availability of metro, time of the day and bus route directions are the secondary important factors. This study shows that when buses run with relatively low and stable speed, taxi-hiring behavior is sensitive to the slight change of bus speed. More passengers would like to take taxis when bus speed starts to decline, or speed or stoppage time of buses tends to be irregular. However, the effects of bus headway on taxi-hiring behavior are more complicated. A specific turning point (coefficient of variability of mean headway  $\approx 0.7$ ) in the relationship between taxi-hiring behavior and bus headway is shown in this paper. Based on data mining and model development, this research presents details on attributes of bus service that drive passengers to switch to taxis and how changes in those attributes encourage modal shift from buses to taxis.

**Keywords:** quality of bus service, taxi-hiring behavior, data mining, model developing, geographic information system (GIS).

### Introduction

Transit systems are always congested in peak hours in East Asia mega-cities, such as Tokyo, Seoul, Hong Kong, Beijing, Shanghai, and Shenzhen. Low running speed, unpunctual arrival and full load are quite common, leading to passengers switch to taxis to save time. This phenomenon is related to two factors: quality of service and behavioral intention. Quality of service is the transit passengers' assessment of the standard of the service delivered (TRB 2013). Behavioral intention is transit passengers' judgment about the likelihood of continuing to use public transportation or willingness to recommend it to others (Rojo *et al.* 2012). Quality of service could affect perceived value, satisfaction and involvement, which could finally influence behavioral intentions of transit passengers (Lai, Chen 2011). This result indicates that

quality of service is the core factor affecting behavioral intentions.

Specifically, some attributes of service quality could significantly affect passengers' preferences to different transit modes. The quality of the information available to passengers about aspects of their journeys (times, frequencies, delays, etc.) is the most important attribute to affect users' willingness to pay for improving quality of service (Dell'Olio *et al.* 2011b). This means when bus reliability decreases under congestion, passengers start to feel difficult to get regular schedules or other information of bus operation, the willingness to pay for better service will go up. Moreover, travel time is an essential factor to determine users' preferences when faced with the introduction of two completely different public transport systems (Dell'Olio *et al.* 2012). Waiting time

\*Corresponding author. E-mail: [zrpeng@sjtu.edu.cn](mailto:zrpeng@sjtu.edu.cn)



and comfort are also shown to be the variables that public transport users most valued (Dell’Olio *et al.* 2012). If the comfort degree drops consistently and waiting time tends to be longer, some passengers would switch to taxis. Especially in Shenzhen, China, this phenomenon is quite common due to the congested bus system. Millions of passengers live in suburb and work in the downtown, generating huge traffic volume about 10 million trips per day (Xiao, Xu 2012; Lei 2009). During peak time periods, passengers have to wait 20 to 40 min due to bus overloading, and some of them even cannot catch their buses until passing 4 or 5 full-loaded buses (Xie *et al.* 2013). At Meilinguan hub station, a large bus hub, there are thousands of people waiting for buses in the morning peak from 7:00 to 10:00 (Li *et al.* 2009). Some buses even cannot close doors because of the overloaded passengers (Gu *et al.* 2010). The congestion of transit system in peak time results in the long, irregular and unbearable waiting time at stations, pushing some passengers to switch to taxis. However, exactly how attributes of bus service affect taxi-hiring behavior is still unclear.

Taxi-hiring behavior has been proven to have some cross-elasticity with other transportation modes. Passengers will choose alternative modes or services if buses arrive late, as they have many other options like taxis, alternative buses with different destinations along the same route, public motorcycles, the underground rail system (known as the MRT), tuk-tuks and vans (Hanaoka, Qadir 2009). The main factors affecting passengers’ preferences and bus-holding strategy are income, waiting time, service reliability and frequency (Redman *et al.* 2013). Wealthy people tend to be less sensitive to price and more sensitive to quality of service than poor people (Litman 2013). For example, the tendency to take a taxi will be higher as passengers’ monthly household income increases, while the MRT is more popular for those of lower monthly household income groups (Hanaoka, Qadir 2009). Moreover, higher value travels, such as business and commute travels, tend to be less price sensitive (Litman 2013). This means that different passengers who are price sensitive and waiting-time sensitive will have different preferences and limits of tolerance when buses are late.

According to previous research, bus headway was proven to have an essential influence on waiting time. Osuna and Newell (1972) conducted theoretical research and developed a model for the expected waiting time  $W$ , which was a function of the mean headway  $\mu$  and variations in the headway  $s^2$ :  $W = \mu \cdot \left( 1 + \frac{s^2}{\mu^2} \right)$ . Several empirical studies demonstrated this theoretical function was reasonable and found similar linear relationship between headway and waiting time from field observations. Salek and Machemehl (1999) developed a model, which was  $W = 2.0 + 0.3 \cdot \mu$  based on extensive empirical data. Fan and Machemehl (2002) used regression models to exam the effects of bus headway on waiting time, and the result was  $W = 2.0 + 0.3 \cdot \mu$ . The above research showed that headway was a main factor affecting waiting time, which was a core public transport vari-

able influencing quality of bus service and taxi-hiring behavior. Therefore, headway is paid more attention to in this study.

However, previous research had mainly focused on passengers’ taxi-hiring behavior from the viewpoint of characteristics of passengers and just had general discussions about the elasticity of taxis and other transportation modes. The studies of elasticity between buses and taxis in commute travel were limited. However, at some bus stations in metropolis, metro is far away to transfer, tuk-tuks and public motorcycles are forbidden, taxi is almost the only alternative mode when buses are late. Thus it is important to understand the details on attributes of bus operation that drive passengers to switch to taxis and how changes in attributes of bus service encourage modal shift from buses to taxis. The study on this issue would help bus companies and traffic administration departments have a better understanding of the severity of interplay between buses and taxis. To this end, this study analyzes the quantitative interplay between bus service quality and passengers’ taxi-hiring behavior, and identifies the key factors attracting passengers to hail taxis from waiting buses.

The remainder of the paper is structured as follows: first, the methodology, data description and specific workflows of data process are presented in Section 1. Next, in Section 2, parametric, semiparametric and non-parametric model are developed and compared. Then, in Section 3, model results and key factors contributing to passengers’ switch from buses to taxis are discussed. Finally, conclusions and recommendations for future work are proposed in the last section.

## 1. Methodology and data process

### 1.1. Methodology

To analyze the correlation between buses and taxis, two key points should be paid attention to: one is the definition of the taxis affected by bus service quality; the other is the methodology to recognize these taxis. Thus, there are three basic assumptions: First, if people have planned to take taxis, they would not go to bus stations to take taxis. Second, if passengers waiting on bus stops want to switch to taxis, the positions where they take taxis are always near the bus stations. Third, if people take buses to go to work, the workplace is usually not far from bus stations. Based on these three assumptions, if a taxi is hired around a bus station and drops off passengers not far from another station of the same bus route, this hail-taxi behavior is considered to be resulted from low bus service quality. That means if the OD (origin/destination) of taxis meets the above conditions, these taxis are regarded as the research objects. In this research, the total taxi and bus data in Shenzhen urban area are enormous. Therefore, the taxi and bus data in the Shennan Avenue need to be extracted into small data set.

A geo-processing workflow is created to capture these taxis from the small data set, based on ArcGIS 10.2 platform and the network data of Shenzhen. First, the buffer areas around bus stations, with reasonable distance to take taxis, are drawn in ArcMap. The pick-

up behavior of taxis is captured and recorded (see Section 1.3.1). Second, the largest reachable region from a bus station to nearby workplaces is defined, and all these regions along the bus routes are drawn in *ArcMap* (see Section 1.3.2). Third, the drop-off behavior of these taxis is captured, and the number of these taxis is recorded in the above-mentioned largest reachable regions (see Section 1.3.2).

The descriptive factors of bus service quality are calculated based on the bus data of Route 223 in Shennan Avenue. The processed data of taxis and buses are matched based on date and time. Availability of metro, AM or PM, weekday or weekend and bus directions are also considered. The entire matching data are listed in a table. Then parametric, semiparametric and nonparametric models are developed to analyze the correlation between bus service quality and passengers' taxi-hiring behavior around stations. Moreover, the relationships between passengers' taxi-hiring behavior and each factor of buses are explored.

## 1.2. Data description and extraction

This paper selected the Global Position Systems (GPS) data of taxis and buses in Shenzhen for 4 weeks including both weekdays and weekends from 12 October 2013 to 8 November 2013. Shennan Avenue is one of the busiest east-west main arterials in Shenzhen and always congested during peak time. A 21-kilometre section of Shennan Avenue is chosen as a case study segment. Route 223 has better GPS data quality and heavier passenger load, compared with other bus routes along the Shennan Avenue. Thus, Route 223 is selected to represent the condition of bus running on this avenue. In either direction, there are 38 stations, and the mean distance

between stations is 767 m. Considering the distribution of stations and whether the bus stops are around metro stations, 16 stations of Route 223 including 8 stations in eastbound direction and 8 stations in westbound are selected, as shown in Table 1.

*SQL Server 2012* platform is selected to extract taxi and bus data, according to the location coordinates, date, route number and other features. The GPS data of each day are divided into two parts based on morning peak from 7:30 to 9:30 and evening peak from 17:30 to 19:30. The taxi and bus data in off-peak time are not included. Therefore, both 21-day bus and taxi data are selected as 42 data files for each bus station. Totally, 16 stations generate 642 files in two directions. Moreover, the missing data and unreadable codes are deleted during the data extraction.

## 1.3. Taxi OD data collection

### 1.3.1. Capture pick-up behavior of taxis

Stations of Route 223 in both directions were drawn on the Shenzhen road network in *ArcMap* software. As demand increases, accessibility to public transport service should be increased and transfer distance ought to be dropped in congested network (Alonso *et al.* 2011). According to a previous study (Sun *et al.* 2011), the distance to transfer are 0.4 km (general distance), 0.6 km (acceptable distance), 0.8 km (maximal distance). Thus, 200 m is selected as a radius to draw station-centered buffer areas, considering the location distribution of taxi-hiring centering on stations. It is assumed that the bus service quality in one direction would basically affect the hail-taxi behavior in the same direction. Thus, the street centerline is used to clip the buffer areas.

Table 1. Information of selected bus stations

Station ID	Station name	Direction	Metro around	Distance to starting station [km]
5920	Investment Tower	eastbound	yes	11.2
7290	Taichong	eastbound	yes	0.9
5824	Zhuzilin	eastbound	yes	7.0
6078	Diwang Mansion	eastbound	yes	17.0
15865	Shennan Xiangmi	eastbound	no	8.9
5831	Shenhang Mansion	eastbound	no	7.8
6007	Tianmian	eastbound	no	13.8
6500	People Bridge	eastbound	no	17.5
6569	Guangshen Hotel	westbound	yes	2.0
6006	Shanghai Hotel	westbound	yes	6.2
6070	Diwang Building	westbound	yes	3.5
5856	China Merchant Bank Building	westbound	yes	12.5
6583	Jingpeng Building	westbound	no	1.5
6592	Metallurgical Mansion	westbound	no	1.0
20713	Shenhang Mansion	westbound	no	12.8
5910	SAR Newspaper	westbound	no	10.2

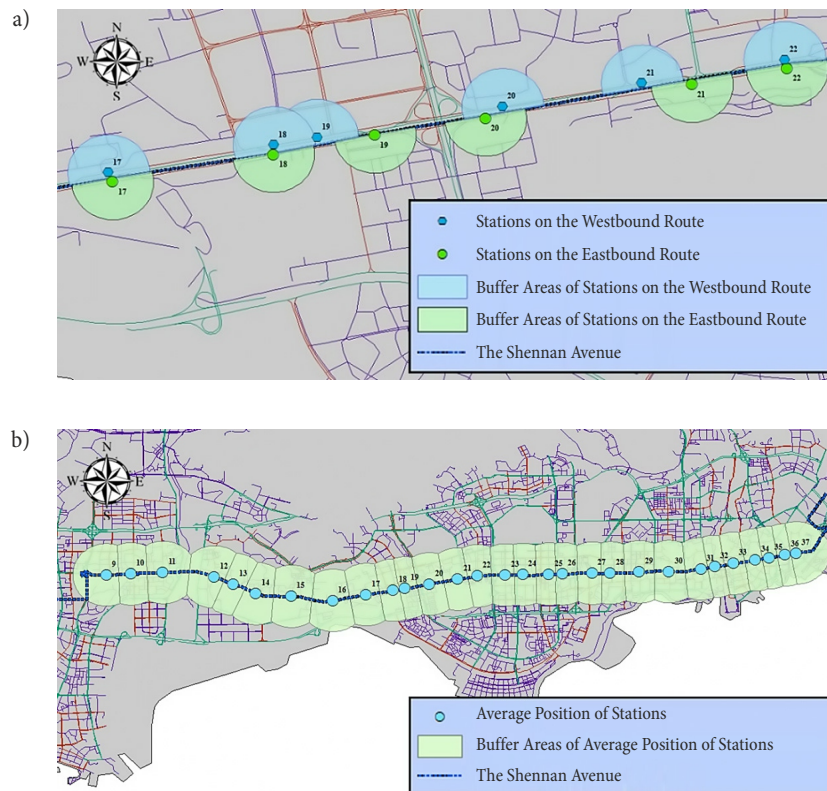


Figure 1. Buffer areas to capture: a – pick-up behavior of taxis; b – drop-off behavior of taxis

The half buffer areas covering stations in one direction are left and the other half are deleted, as shown in Figure 1a. The half buffer areas are marked with station ID. A geo-processing module is built to clip the taxi data based on the half buffer areas and assign the bus station ID field to taxi data. Finally, the taxi data are processed and output as *Excel* files.

In the taxi data, a field named *EMPTY\_LOAD* records the taxi load and empty state, where 1 represents load (with customers) and 0 represents empty (no customers). A procedure is applied to capture the moment when 0 changes to 1, which indicates that the taxi started to carry passengers. The number of hired taxis and the total number of taxis around the bus stations in peak time are counted. These two numbers are used to identify target bus stations for analysis. High value of the two numbers means that there are enough samples for research and this kind of stations could be included in the analysis.

### 1.3.2. Capture drop-off behavior of taxis

The crossings of the street centerline and the line connecting two opposite bus stations are selected as mean position of stations. 1 km is regarded as a suitable radius to draw station–workplace buffer areas, considering the maximal distance (0.8 km) and computational redundancy (Sun *et al.* 2011). Those buffer areas center on mean position of stations, as shown in Figure 1b. The taxi data are clipped by 1 km buffer areas and assigned corresponding bus station IDs field.

Drop-off behavior of taxis is captured with the following steps: First, the vehicle IDs of the hired taxis around bus stations are applied to filter the taxi data clipped by 1 km buffer areas. Second, the stations behind the hail-taxi station along Bus Route 223 in one direction are selected. The processed taxi data in Step 1 are filtered according to these selected stations. These two steps find all records of the trips of the taxis hired around the bus stations. After hired, *EMPTY\_LOAD* field of taxi keeps value of 1. Therefore, third, a procedure is applied to the data set of Step 2 to capture the moment when 1 changes to 0 at the first time, which indicates that the passengers get off the taxis and the taxis return to empty state. This step assists to find the destination of the taxi path.

Therefore, the taxi, with origin around a bus station and destination in a station–workplace buffer area, is considered to be hired due to low bus service quality. So far, the OD of the taxis is identified and the number of the taxis is counted. This number is regarded as the amount of taxis affected by bus service quality.

### 1.4. Bus data process

All the distances from starting station to selected station are measured in *Google Map*. The mean speed between starting station and the selected station for each bus is regarded as bus speed. For each selected station, the difference of station-arriving time between two adjacent buses of Route 223 is defined as headway. The difference

between station-arriving time and station-departure time of one bus is regarded as stoppage time. Thus, with SQL server 2012 platform, the mean and Coefficient of Variation (CV) of bus speed, headway and stoppage time could be calculated.

### 1.5. Data matching between taxi and bus calculated results

Variables describing taxi-hiring behavior and quality of transit service are matched based on time and bus station

ID. These variables contain TAXI, AVG\_SPEED, CV\_HEADWAY, METRO and etc., as described in Table 2 and Figure 2. In this paper, the number of taxis affected by bus service quality (TAXI) is selected as dependent variable. The independent variables include bus service attributes such as the mean and CV of the bus speed, headway and stoppage time, as well as route direction and availability of metro. A number of temporal variables are also used, such as peak time periods (morning peak from 7:30 to 9:30 and evening peak from 17:30 to 19:30).

Table 2. Descriptions of selected variables

<i>Definitions of the variables used in the regression model</i>					
Variable	Description				
TAXI	Number of taxis impacted by bus service quality				
AVG_SPEED	Mean bus speed [km/h]				
CV_SPEED	CV of bus speed [km/h]				
AVG_HEADWAY	Mean bus headway [min]				
CV_HEADWAY	CV of bus headway [min]				
AVG_STOPPAGE	Mean stoppage time of bus [min]				
CV_STOPPAGE	CV of bus stoppage time [min]				
AM_PEAK	AM peak: 1 and PM peak: 0				
METRO	METRO: 1 and NO METRO: 0				
DIRECTION	Eastbound: 1 and Westbound: 0				
<i>Descriptive statistics of numerical variables</i>					
Period	Variable	Mean	Median	Max	Min
All periods	TAXI	13.94	11.98	50.69	4.41
	AVG_SPEED [km/h]	20.36	20.69	33.65	9.53
	CV_SPEED	0.11	0.09	0.49	0.00
	AVG_HEADWAY [min]	8.71	8.31	17.26	5.49
	CV_HEADWAY	0.66	0.64	1.34	0.22
	AVG_STOPPAGE [min]	0.63	0.60	1.49	0.25
	CV_STOPPAGE	0.40	0.37	1.32	0.04
Metro only	TAXI	13.56	11.93	38.66	4.56
	AVG_SPEED [km/h]	20.02	20.30	33.65	10.66
	CV_SPEED	0.12	0.09	0.37	0.00
	AVG_HEADWAY [min]	8.78	8.38	23.26	5.49
	CV_HEADWAY	0.66	0.65	1.27	0.22
	AVG_STOPPAGE [min]	0.65	0.60	1.49	0.25
	CV_STOPPAGE	0.41	0.36	0.96	0.09
No metro only	TAXI	14.31	12.25	50.69	4.41
	AVG_SPEED [km/h]	20.70	21.15	30.06	9.53
	CV_SPEED	0.11	0.08	0.49	0.00
	AVG_HEADWAY [min]	8.64	8.29	19.01	5.54
	CV_HEADWAY	0.66	0.63	1.34	0.23
	AVG_STOPPAGE [min]	0.61	0.59	1.10	0.33
	CV_STOPPAGE	0.39	0.37	1.32	0.04
Eastbound only	TAXI	11.94	11.13	31.12	4.41
	AVG_SPEED [km/h]	24.23	23.77	33.65	18.46
	CV_SPEED	0.10	0.07	0.36	0.00
	AVG_HEADWAY [min]	8.33	8.27	15.73	5.49
	CV_HEADWAY	0.64	0.61	1.34	0.35
	AVG_STOPPAGE [min]	0.62	0.61	1.26	0.25
	CV_STOPPAGE	0.36	0.31	1.01	0.04
Westbound only	TAXI	15.93	13.63	50.69	4.56
	AVG_SPEED [km/h]	16.49	16.16	23.85	9.53
	CV_SPEED	0.13	0.11	0.49	0.00
	AVG_HEADWAY [min]	9.09	8.45	23.26	6.09
	CV_HEADWAY	0.68	0.66	1.29	0.22
	AVG_STOPPAGE [min]	0.63	0.58	1.49	0.33
	CV_STOPPAGE	0.44	0.41	1.32	0.09

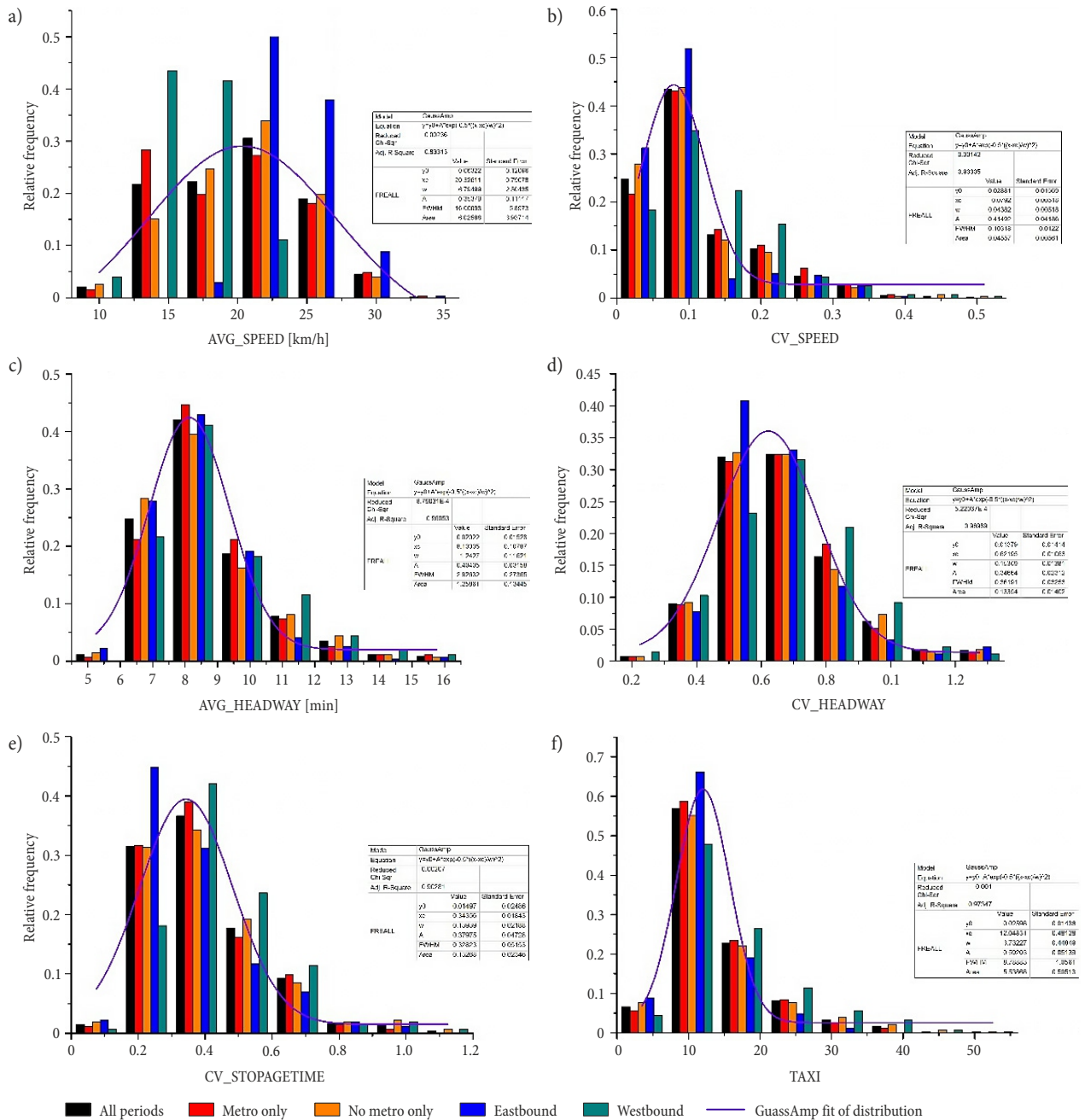


Figure 2. Descriptive statistics distribution of numerical variables: a – distribution of mean bus speed; b – distribution of coefficient of variability (CV) of bus speed; c – distribution of mean bus headway; d – distribution of CV of bus headway; e – distribution of CV of bus stoppage time; f – distribution of target taxis (the purple curve in each figure is fitting curve)

The GPS devices installed on vehicles record approximately 20 to 29 million data for taxis and 15 to 20 million data for buses per day in Shenzhen urban area. During the morning peak or evening peak each day, there are from 900 to 1050 thousand data for taxis and from 14 to 26 thousand data for Route 223 in the 21-kilometre section of Shennan Avenue. The matching records are generated from 16 stations in a continuous 21-day data set. This data set is calculated from totally about 19.95 million data of taxis and buses in the Shennan Avenue and extracted from 882 million data of all the taxis and buses in Shenzhen.

## 2. Statistical model development

### 2.1. An overview of parametric, nonparametric and semiparametric approaches

Multivariate regression techniques are often used to describe the dependence of a variable on explanatory variables, which include the parametric, nonparametric and semiparametric methods (Shively *et al.* 2010; Shively, Sager 1999). The parametric approach assumes that response function could be expressed parametrically and is often used to describe the dependence of a variable on explanatory variables.



However, a major disadvantage of the parametric approach is that a parametric function must be specified before fitting (Carslaw *et al.* 2007). Parametric model is limited when the presupposed model is inaccurate or the relationships between explanatory variables are hard to be represented as linear functions. Unlike parametric approaches, nonparametric approaches make no hypotheses about the parametric relationship between variables. The form of  $Y_i$  just depends on data itself. The additive nonparametric model can be described as (Hastie, Tibshirani 1990):

$$Y_i = A_i + \sum_{j=1}^n s_j(x_{ij}) + \varepsilon_i; \quad (1)$$

$$\varepsilon_i \sim N(0, \sigma^2),$$

where:  $Y_i$  is the concentration of the time series;  $s_j(x_{ij})$  is a smooth function of covariate  $x_j$ ;  $n$  is the number of covariates;  $\varepsilon_i$  is the residual.

The generality of the nonparametric model makes it very attractive when there is complicated nonlinearity in the multivariate case. Especially for transportation applications, there are many non-linear relationships among variables, sometimes the nonparametric approach is more effective (Li *et al.* 2009). However, there are also obvious shortcomings in nonparametric model. Nonparametric estimates are too flexible because the response functions totally depend on the observed data and often these fits are subject to high variance (Wan, Birch 2011).

The semiparametric approach is a combination of traditional parametric regression analysis and nonparametric regression methods. The semiparametric model can be given as follow (Ruppert *et al.* 2003):

$$Y_i = A_i + \sum_{i=1}^n \beta \cdot x_i + \sum_{j=1}^n s_j(x_{ij}) + \varepsilon_i; \quad (2)$$

$$\varepsilon_i \sim N(0, \sigma^2),$$

where:  $Y_i$  is the concentration of the time series;  $x_i$  is the  $i$ -th explanatory variables;  $s_j(x_{ij})$  is a smooth function of covariate  $x_j$ ;  $n$  is the number of covariates;  $\varepsilon_i$  is the residual, which is assumed to have constant variance.

Semiparametric approaches allow a mixture of parametric and nonparametric components. These characteristics lead that data suitable for parametric modeling are modeled that way and nonparametric components are used only where needed, which essentially combines the advantages of the parametric and nonparametric methods and avoid their disadvantages (Ruppert *et al.* 2003). The discussions above had implied that the semiparametric technique might be effectively applied to model the complex relationships between taxi-hiring behavior and bus service quality.

A key part in developing nonparametric and semiparametric models is how to find the smooth functions and how to control the degree of smoothness used. In this paper, we used a nonparametric approach proposed by Wood and Augustin (2002), which integrates model

selection and automatic smoothing parameter selection using penalized regression splines. Generalized Cross Validation (GCV) is applied to decide the choice of smoothing parameter. The semiparametric approach we used in this research was proposed by Ruppert *et al.* (2003). Smoothing parameter selection of semiparametric model are based on generalization of smoothing splines (French *et al.* 2001) and can be done via Restricted Maximum Likelihood (REML). Both approaches currently used depend on an unconditional Bayesian method, which has been shown to perform well through numerical simulations (Wood 2006) and offers a pragmatic solution to confidence interval estimation. All calculations were made by the R-programming language (Wood 2006). Nonparametric model was developed with package *mgcv* version 1.8-6 (Wood 2006) and semiparametric model was built via package *SemiPar* version 1.0-4.1 (Ruppert *et al.* 2003). More theoretical details of these methods are available in Wood and Augustin (2002), Wood (2003, 2006), Ruppert *et al.* (2003).

## 2.2. Model development

Statistical models were derived from parametric, nonparametric and semiparametric approaches. A key part to develop a semiparametric model is to define covariates suitable for parametric modeling and nonparametric modeling. However, a wide range of covariates in this semiparametric model result in that the combinations of parametric and nonparametric components are various. Thus, the process of semiparametric modeling needs to consider the performance of parametric and nonparametric approaches.

The performance of these three models could be assessed numerically with descriptive statistics. GPS data of taxis and buses in Shenzhen from 2 November to 8 November 2013 are applied to verify these models. Table 4 shows the details of model estimation compared with observations. Root Mean Squared Error (RMSE), Normalized Mean Bias (NMB), Mean Absolute Error (MAE), adjusted  $R^2$  and Theil's inequality coefficient  $T$  were employed to evaluate the model performance.  $T$  is a measure of a time series of estimated values compares to observed values, as expressed in Equation (3):

$$T = \frac{\sqrt{\frac{1}{n} \cdot \sum_{i=1}^n (P_i - O_i)^2}}{\sqrt{\frac{1}{n} \cdot \sum_{i=1}^n P_i^2 + \frac{1}{n} \cdot \sum_{i=1}^n O_i^2}}, \quad (3)$$

where:  $O_i$  represents the actual values;  $P_i$  represents predicted values;  $n$  is the total number of data.

$0 \leq T \leq 1$ ,  $T \rightarrow 0$  indicates that the model is good for estimation; otherwise, the model is not suitable for estimation.

At first, both parametric and nonparametric models are applied and compared, as shown in Equations (4)–(5) and Table 3. In the parametric model (Equation (4)), Table 3 shows that only bus speed, stoppage time, time period, metro and direction have linear

relationship with taxi-hiring behavior. However, bus headway is excluded from the parametric model due to the low statistical significance, which means there is no statistically linear relationship between taxi-hiring behavior and bus headway. This result is inconsistent with the common sense and findings in the literatures. In the nonparametric model, Table 3 indicates that bus speed and stoppage time fit for nonparametric modeling, but their significance decreases. This means that bus speed and stoppage time is more suitable for parametric modeling. Mean bus headway and irregularity of bus headway are significant in the nonparametric model. This result demonstrates that nonparametric model has a better performance in bus headway, comparing with parametric model. However, several significant Boolean variables in the parametric model are not included in the nonparametric model, such as metro and direction. This means that Boolean variables is not very suitable for nonparametric modeling. Figure 3 shows the individual smoothed model components for the nonparametric model, as defined in Equation (5).

$$\text{TAXI} = A_1 + \beta_1 \cdot \text{AVG\_SPEED} + \beta_2 \cdot \text{CV\_SPEED} + \beta_3 \cdot \text{AVG\_HEADWAY} + \beta_4 \cdot \text{CV\_HEADWAY} + \beta_5 \cdot \text{AVG\_STOPPAGE} + \beta_6 \cdot \text{CV\_STOPPAGE} + \beta_7 \cdot \text{AM\_PEAK} + \beta_8 \cdot \text{METRO} + \beta_9 \cdot \text{DIRECTION}; \quad (4)$$

$$\text{TAXI} = A_2 + s_1(\text{AVG\_SPEED}) + s_2(\text{CV\_SPEED}) + s_3(\text{AVG\_HEADWAY}) + s_4(\text{CV\_HEADWAY}) + s_5(\text{AVG\_STOPPAGE}) + s_6(\text{CV\_STOPPAGE}) + s_7(\text{AM\_PEAK}) + s_8(\text{METRO}) + s_9(\text{DIRECTION}); \quad (5)$$

$$\text{TAXI} = A_3 + \beta_1 \cdot \text{AVG\_SPEED} + \beta_2 \cdot \text{CV\_SPEED} + s_1(\text{AVG\_HEADWAY}) + s_2(\text{CV\_HEADWAY}) + s_3(\text{AVG\_STOPPAGE}) + \beta_3 \cdot \text{CV\_STOPPAGE} + \beta_4 \cdot \text{AM\_PEAK} + \beta_5 \cdot \text{METRO} + \beta_6 \cdot \text{DIRECTION}. \quad (6)$$

For the second step, semiparametric model was developed based on the performance of the parametric approach and the nonparametric approach. The covariates that are significant in parametric model are kept in the parametric components, while the others remained in nonparametric components, as shown in Equation (6). According to Table 3, almost all variables which are significant in parametric model or nonparametric model are included in semiparametric model. Significances of bus speed, stoppage time, time period and metro are similar in parametric and semiparametric models. Moreover, mean bus headway and irregularity of bus headway have better performance of confidence level in the semiparametric model, which means that bus headway is more suitable for semiparametric modeling.

Models are compared further based on descriptive statistics. In Table 4, the results indicate that semiparametric model has the highest adjusted  $R^2$  (0.671), comparing with parametric model (0.426) and nonparametric model (0.638). We found that nonparametric and semiparametric models had the similar adjusted  $R^2$ , but parametric model performed worse. This means that the effects of bus service quality on taxi-hiring behavior are closer to a kind of nonparametric relation. The RMSE, NMB, MAE and  $T$  show the same conclusion that the performance of nonparametric model and semiparametric model is similar, while fitting result of parametric model is the worst. In terms of model comparison, parametric model shows that it is a little restrictive when the relationships between explanatory variables are too complex to be represented as linear functions (Carslaw *et al.* 2007). Nonparametric model shows its superiority to fit complicated relationship between covariates, but it has some limitations when variables are Boolean or the relationship is linear. Semiparametric model contains the most variables and is able to explain the influence of bus headway on taxi-hiring behavior better.

All these three models were analyzed for a variety of statistical diagnostics. Residual autocorrelation can influence model performance and leads to standard er-

Table 3. Results of the parametric, nonparametric and semiparametric models

Variable	Parametric model			Nonparametric model		Semiparametric model	
	Estimated coefficients	Standardized coefficients	Sig.	Estimated coefficients	Sig.	Estimated coefficients	Sig.
Intercept	18.695		***	12.723	***	20.366	***
AVG_SPEED	-0.627	-0.418	***		**	-0.443	***
CV_SPEED	23.969	0.266	***		**	17.407	***
AVG_HEADWAY	-0.072	-0.020			**		***
CV_HEADWAY	-1.577	-0.040			*		**
AVG_STOPPAGE	1.027	0.025					
CV_STOPPAGE	15.274	0.372	***		**	12.833	***
AM_PEAK	0.537	0.038	*	1.391	**	1.473	**
METRO	-1.320	-0.093	**	-1.318		-1.751	*
DIRECTION	3.214	0.226	*	1.544		1.647	
Adjusted $R^2$	0.426			0.638		0.671	

Notes: Sig.: '\*\*\*' 0.001, '\*\*' 0.01, '\*' 0.05, '°' 0.1, '°' 1.



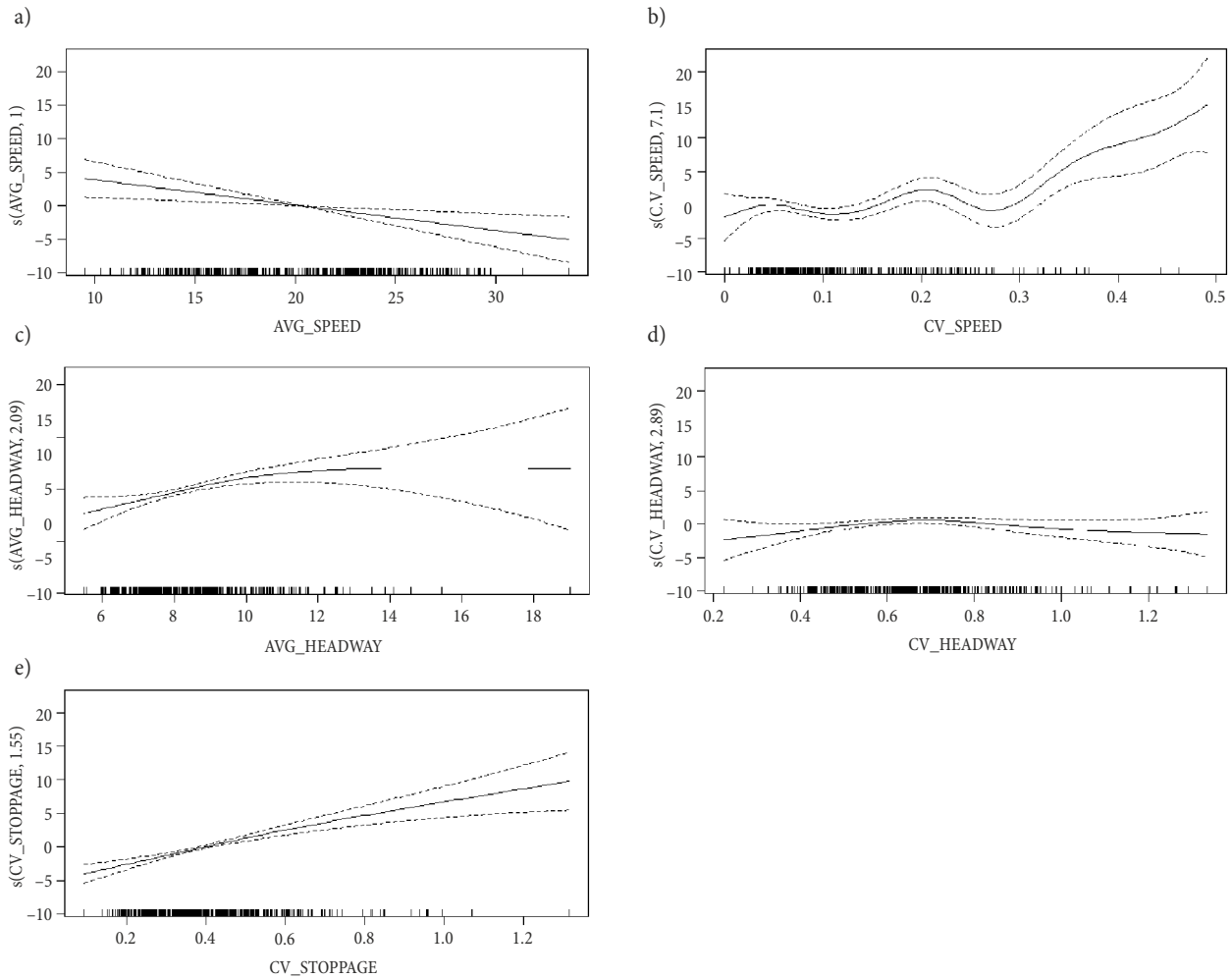


Figure 3. Fitted components of the nonparametric model: a – smooth function of mean bus speed; b – smooth function of coefficient of variability (CV) of bus speed; c – smooth function of mean bus headway; d – smooth function of CV of bus headway; e – smooth function of CV of bus stoppage time (the dashed lines are the estimated 95% confidence intervals; the vertical lines adjacent to the lower  $x$ -axis show the presence of data in (a)–(e))

Table 4. Descriptive statistics of parametric, nonparametric and semiparametric models for estimation

Statistics measure	Parametric model	Nonparametric model	Semiparametric model
Adjusted $R^2$	0.426	0.638	0.671
RMSE	6.472	5.871	5.805
NMB	-3.07%	-1.28%	-0.90%
MAE	4.395	4.302	4.276
$T$	0.115	0.093	0.088
Residual autocorrelation $\phi$	0.27	0.24	0.23

ror underestimated and model over-fitting. As shown in Table 4, residual autocorrelations of these three models varied between 0.23 and 0.27, which were acceptable. The residual distributions in all cases were close to normal, uniformly distributed and symmetrical. Furthermore, a Generalized Additive Mixed Model (GAMM) was used to model autocorrelation explicitly (Wood 2003). In GAMMs, the unknown coefficients or functions in the linear predictor are regarded as random variables; in order to make correlated data modelled

(Wood 2006). Thus, models shown in Eqs (4)–(6) were verified in GAMMs, and autocorrelations were modelled explicitly via function available in the *mgcv* package. The result indicated that the residual autocorrelation was not significant and well described by a lag-1 autoregressive process in all cases. This means that residual autocorrelation of all three models led to very minor influence on both the estimated trend and the uncertainty intervals. Thus, residual autocorrelations of these three models could be almost ignored.

### 3. Results and discussion

#### 3.1. Bus speed, headway and stoppage time

The mean bus speed is the most intuitively important factor to describe traffic congestion levels on the route. The CV of the bus speed describes the reliability of bus operation. The mean bus headway has been proven to have positive linear relationship with waiting time at bus stations (Fan, Machemehl 2002). The CV of the bus headway reflects the uniformity of bus arriving time and its high value indicates bus irregularity like bus bunch-

ing and big gaps between buses. The mean stoppage time of bus represents the time people could use to get on and off. Higher stoppage time allows people to have more chances to aboard and be less hurried. Moreover, the CV of the bus stoppage time reflects the variations of time for passengers to get on buses.

AVG\_SPEED and CV\_SPEED are included in parametric model, nonparametric model and parametric component of semiparametric model. Figure 3a shows that taxi-hiring behavior has decreased gradually with mean bus speed increasing, and Figure 4a presents

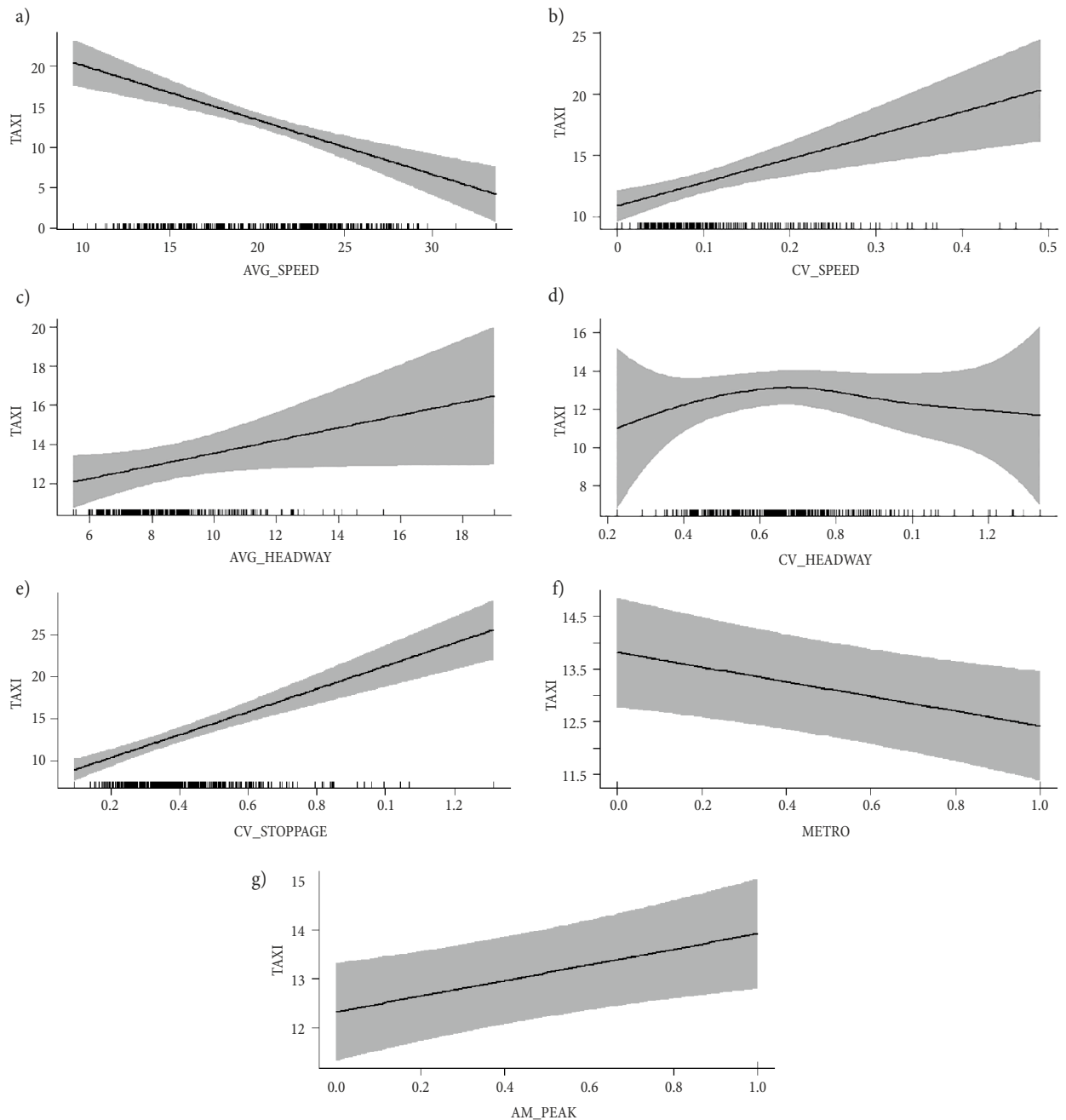


Figure 4. Fitted components of the semiparametric model: a – smooth function of mean bus speed; b – smooth function of coefficient of variability (CV) of bus speed; c – smooth function of mean bus headway; d – smooth function of CV of bus headway; e – smooth function of CV of bus stoppage time; f – smooth function of metro; g – smooth function of morning peak (the shaded area are the estimated 95% confidence intervals; the vertical lines adjacent to the lower x-axis show the presence of data in (a)–(g))

a more rapid drop in the semiparametric model. The unstandardized coefficient of AVG\_SPEED is  $-0.627$  in the parametric model, and  $-0.443$  in the semiparametric model. This means that TAXI will decrease by  $0.443$ – $0.671$  when all other variables are kept the same, for each km/h added to AVG\_SPEED. An examination of the uncertainty intervals of Figures 3a and 4a also suggest that the decline in TAXI has been statistically significant at the 95% confidence level. By contrast, Figure 4b indicates that taxi-hiring behavior has grown continuously with irregularity of bus speed increasing. Figure 3b also demonstrates that taxi-hiring behavior experiences a fluctuant rise when irregularity of bus speed goes up. In summary, the results show that in all the time periods, mean bus speed has negative influence on taxi-hailing but irregularity of bus speed has a positive effect. According to Table 3, in the parametric model, the standardized coefficient of AVG\_SPEED equals to  $-0.418$ , which is the highest among all the negative variables in absolute value. Meanwhile, the standardized coefficient of CV\_SPEED is  $0.266$  and ranks the second among all the positive variables. All of these analyses indicate the bus speed is an essential factor influencing taxi-hailing behavior. Moreover, the descriptive statistics of bus speed could also provide some reference. As shown in Table 2 and Figure 2a, AVG\_SPEED is  $20.36$  km/h among different peak time observations. The distribution (black bars) of AVG\_SPEED follows the similarly uniform distribution from  $14$  to  $26$  km/h. However, according to Table 2 and Figure 2b, the mean CV\_SPEED is  $0.11$  and about 70% of them are under  $0.1$  in each observed peak time. The analysis results of AVG\_SPEED and CV\_SPEED show that when buses run with relatively low and stable speed, taxi-hiring behavior is sensitive to the slight change of bus speed. More passengers would like to hire taxis when bus speed starts to decline or to be irregular.

AVG\_HEADWAY and CV\_HEADWAY are not included in the parametric model due to the lack of statistical significance, but they are significant covariates in both the nonparametric model and nonparametric component of the semiparametric model. This means that the effects of bus headway on taxi-hiring behavior are more complex and not just a simple linear relationship.

Figure 3c shows that taxi-hiring behavior experiences a continuous rise at first and reach the peak when mean headway increases to about 12 min. Then taxi-hiring behavior starts to keep stable, when all other covariates are kept the same. However, Figure 4c indicates a slightly different result that taxi-hiring behavior has a nearly linear increase with the increase of mean bus headway. The difference that taxi-hiring behavior shows different growth trends in Figures 3c and 4c when mean bus headway is longer than 12 min may be resulted from several reasons. First, smooth functions used in Figures 3c and 4c are estimated by different penalized spline smoothing. Penalized spline smoothers take on a number of forms (Ruppert *et al.* 2003). In this paper, the choice of smoothing parameter in nonparametric model is based on penalized regression splines and made

through GCV (Wood, Augustin 2002); Smoothing parameter selection of semiparametric model are based on generalization of smoothing splines (French *et al.* 2001) and can be done via REML. Second, Nonparametric methods are flexible in that the resulting estimates of mean response completely depend on the observed data itself (Wan, Birch 2011). Third, observed data of mean bus headway mainly distributed from 6–12 min, specifically, about 90% of them are distributed from 6.8 to 9.7 min, as shown in Figure 2c and the  $x$  axis of Figures 3c and 4c. Thus, when mean bus headway is shorter than 12 min, Figures 3c and 4c are mainly estimated based on observed data and have similar growth trends. When mean bus headway is longer than 12 min, observed points are fewer and different smooth functions lead to different fitting results. This means more people would like to hire taxis with mean bus headway closing to 12 min. However, when mean bus headway is over 12 min, the willing of people hiring taxis is still not clear enough according to Figures 3c and 4c.

Figure 3d shows that taxi-hiring behavior rises gradually with CV of bus headway increasing. When CV of bus headway grows to around 0.7, taxi-hiring behavior reaches the highest level. Then Figure 3d sees a slow and consistent drop. In Figure 4d, a similar change trends could be observed and taxi-hiring behavior reaches its peak at about 13 when CV of bus headway equals to 0.7. There is some indication that passengers could be divided into two groups: people sensitive to waiting time and people sensitive to price (Dell'Olio *et al.* 2011a). Both groups would like to wait for buses when headway just starts to be unstable. However, passengers sensitive to waiting time have a limit of tolerability for irregularity of bus headway, because irregular headway could make waiting time longer (Fan, Machemehl 2002), in this research, the limit is 70%. When irregularity of bus headway closes to this limit, more people would like to hire taxis; when irregularity of bus headway is over this limit, the percentage of people sensitive to waiting time in all passengers would decrease. People sensitive to price would comprise more proportion and less people would like to hire taxis although irregularity of bus headway becomes worse. The descriptive statistics of CV of bus headway could also provide us more details. According to Table 2 and Figure 2d, mean of CV of bus headway is 0.66 (close to 0.7) and the distribution (black bars) of CV of bus headway indicates that 70% of them are distributed between 0.5 and 0.7. This result indicates that bus headway drastically fluctuates in each observation and taxi-hiring behavior could almost reach its peak just when CV of bus headway shows normal performance in congested transit systems.

AVG\_STOPPAGE is excluded from all the models due to lack of significance. However, CV\_STOPPAGE is significant at the 99% confidence level in parametric model, nonparametric model and parametric component of semiparametric model. Figure 3e shows that taxi-hiring behavior grows substantially with CV of stoppage time increasing, and Figure 4e indicates a similar growth trends. The unstandardized coefficient of this variable is  $15.274$  in parametric model and  $12.833$  in paramet-

ric component of semiparametric model. This means that TAXI will increase about 12–15 with irregularity of bus stoppage time doubled, when all other variables are kept the same. Moreover, the standardized coefficient of CV\_STOPPAGE in parametric model is the highest (0.372) among all the variables, which indicates that this variable has essential influence on taxi-hiring behavior. Based on descriptive statistics in Table 2 and Figure 2e, the mean CV\_STOPPAGE in each observation is 0.40, and about 75% of them are distributed between 0.2 and 0.5 in each observed peak time. This result means that more passengers would like to hire taxis when stoppage time of bus starts to show more irregularity in peak time.

Service elasticity refers to how changes in transit service mileage, service-hours, frequency, and service quality (such as comfort) affect transit ridership (Litman 2013). Increased travel speed and reduced delay for a particular mode tends to attract travel from other modes on a corridor (Dell’Olio *et al.* 2011a, 2011b). In Table 3 and Figure 4a, 4b, for each km/h added to bus speed, taxi ridership shifted from transit would drop 0.443; increasing irregularity of bus speed by 10% was associated with a 1.741 increase in taxi ridership. The results mean that bus speed has a strong transit service elasticity and could affect the amount of induced travels in different modes. Furthermore, commute trips for each rider always have constant distances, thus travel time has a proportional relationship with transit speed, indicating that travel time and speed could be both regarded as indices to evaluate service elasticity in commute trips (Redman *et al.* 2013). Some transport economists supported the constant travel time budget hypothesis, which means the amount of time people devote to commute travel tends to remain constant (typically averaging 70–90 min daily), and found that the travel time between different modes significantly affected mode choice (Hanaoka, Qadir 2009), which evidenced the results of this paper. In Table 3 and Figure 4d, 4e, increasing irregularity of bus stoppage time by 10% was associated with a 1.283 increase in taxi ridership. Irregularity of bus headway produced a higher marginal utility for taxi riders attracted from transit before CV of bus headway equaled to 0.7. After that, an increased marginal disutility for irregularity of bus headway and taxi ridership could be observed, suggesting that fewer passengers had the willingness to switch to taxis from transit when waiting time became longer. This results indicated that the commute travelers were more time sensitive and had a limited tolerability for irregularity of service frequency (called a headway elasticity, Litman 2013), but the non-commute riders were more price sensitive. This means that the higher value travels, such as commute or business trips, tend to be less elastic than lower value travels, such as shopping or recreational trips (Rose, Hensher 2014).

In crowded transit system, buses with available room will spend more stoppage time to let passengers aboard. However, full-load buses are more common and just stop for a short while or even not stop. This leads to the large fluctuations of stoppage time and causes many passengers to wait for the next buses. Most of them need to wait for two or more buses; some of them even have

to wait for the 5th bus in order to get on (Xie *et al.* 2013). In this situation, the passengers’ waiting time mainly depends on the bus headway (Fan, Machemehl 2002) and the number of buses they have to wait before being able to get on. Moreover, lower bus speed and higher speed fluctuation mean more serious traffic congestion and higher passenger load. This makes waiting time become longer and once it starts to reach the limit of tolerance of passengers who are sensitive to time, some of them in a hurry to go to work or to get to meetings start to switch to taxis. This is consistent with the observations and local news reports (Xiao, Xu 2012; Xie *et al.* 2013; Li *et al.* 2009; Gu *et al.* 2010) about transit system in Shenzhen.

### 3.2. Metro

Metro is another factor affecting the hail-taxi behavior around bus stations owing to the cross-elasticity between them. Various research found metro and taxi elasticities often differed. In major cities, the cross-elasticity between them tends to be relatively low, typically in the  $-0.18$  range due to users’ different incomes and price sensitivity (Litman 2013). In this paper, Metro is included in the parametric model and parametric component of semiparametric model but removed from nonparametric model due to lack of significance. This means that the relationship between taxi-hiring behavior and metro is more suitable for parametric modeling. Figure 4f shows that taxi-hiring behavior decreases considerably with a metro station nearby. Specifically, according to unstandardized coefficient of METRO in parametric model and semiparametric model, taxi-hailing behavior with metro stations around will be 1.320–1.751 fewer than that far away from metro stations. However, the standardized coefficient of METRO is only  $-0.093$  in the parametric model, which is the lowest among all the variables left. This means that metro has negative effects on taxi-hiring behavior, but not very noticeable. Descriptive statistics in Table 2 also provide us similar results. Mean taxi-hiring behaviors in bus stations with and without metro around are 13.56 and 14.31 respectively, only small difference.

### 3.3. Direction

Direction is considered in the models, and most conditions of these two directions are similar but congestion degree between them is a little different. It is only included in parametric model and excluded from nonparametric model and semiparametric model because of insignificance. The result of parametric model shows that if the hail-taxi position is located on the westbound roadside, TAXI will be 3.214 more than the location on the eastbound. Descriptive statistics in Table 2 could provide more details. The mean TAXI values in eastbound and westbound are 11.94 and 15.93 respectively. Buses running in eastbound direction also have higher speed, lower headway and smaller stoppage time compared with westbound direction. Moreover, all of these three variables in eastbound direction have less irregularity. However, the standardized coefficient of DIRECTION is only 0.226, which is the lowest among the positive variables. This demonstrates a minimum effect.

### 3.4. Time period

AM\_PEAK is included in all three models. Figure 4g shows that taxi-hiring behavior decreases considerably when they happen in the morning peak. Specifically, according to unstandardized coefficients of AM\_PEAK in parametric model and semiparametric model, taxi-hailing behavior in morning peak will have a 0.537–1.473 increase comparing with that in evening peak. This means that people hurrying to work place in the morning have more willingness to hire taxis, comparing with people going home in the evening.

### Concluding remarks

This study aimed to contribute to a better understanding of relationship between the quality of bus service and taxi-hiring behavior in commute travels. As such, relevant research was sought to answer the following two questions: which attributes of bus operating drive passengers to switch to taxis. How changes in attributes of bus service encourage modal shift from buses to taxis. With GIS and database technology, passengers' taxi-hiring behavior was captured by OD of hired taxis via geo-processing methods, while the attributes of bus service were mined from huge data set. Both bus and taxi GPS data were matched and analyzed together with GIS.

Parametric, nonparametric and semiparametric models were developed and all applied to analyze the relationships between taxi-hiring behavior and bus service quality. The results showed that the semiparametric model had the highest adjusted  $R^2$  (0.671), comparing with that in the parametric model (0.426) and that in the nonparametric model (0.638). The comparison of RMSE, NMB, MAE and Theil's inequality coefficient  $T$  in different models also indicated that semiparametric model had the best estimation. Moreover, the analysis results in each model were basically consistent and could be verified mutually, which made conclusions more accurate. In sum, semiparametric model is very attractive when there is a complicated mixture of linearity and nonlinearity in the multivariate cases. Such as the case in this research, the effects of bus service quality on taxi-hiring behavior were complex, and a wide range of covariates need to be estimated. Nonparametric model shows its superiority when variables are non-Boolean and the relationships between explanatory variables are totally nonlinear. In this case, nonparametric approaches could reduce calculation amount a lot but still obtain results as good as nonparametric models. Parametric approaches or multiple linear regression models make it a little restrictive when covariates are derived from different research objects or the correlation between them is hard to be represented as linear functions. There is not much recommendation for parametric models when correlations among covariates are complex.

The research indicated that bus speed, headway and stoppage time were the core parameters affecting passengers' taxi-hiring behavior. Availability of metro, time of the day and bus route directions were the next impor-

tant factors. This study found that when buses run with relatively low and stable speed (mean speed  $\approx 20$  km/h and CV of mean speed  $\approx 0.1$ ), taxi-hiring behavior was sensitive to the slight change of bus speed. More passengers would like to hire taxis when bus speed started to decline, and speed or stoppage time of buses tend to be irregular. However, the effects of bus headway on taxi-hiring behavior were more complicated. A specific turning point (CV of mean headway  $\approx 0.7$ ) in the relationship between taxi-hiring behavior and bus headway was shown in this paper. When bus headway is close to the turning point, more passengers would like to hire taxis; with bus headway over this turning point, taxi-hiring behavior tends to be less. The result also showed that taxi drivers would get 1–3 more guests when driving in the morning peak, away from metro or in a more congested direction on a corridor.

This study is helpful in transportation policy making and beneficial to the future study about the cross-elasticity of different transit modes. Although bus speed, headway and stoppage time are not the causes of waiting time increase, they can show the transit system's congestion levels. Moreover, different degrees of congestion in crowded transit system affect the number of the passengers switching to taxis. Thus, bus speed, headway and stoppage time can be regarded as indices to evaluate the behavioral intention (transit passengers' judgment about the likelihood of continuing to take buses), and the elasticity of bus service quality and other transit modes. For traffic management departments, the frequency of this kind of forced transfer behavior between buses and taxis can be set as a new standard to evaluate the bus service quality. The government should also pay more attention to the locations of taxi stops near the bus stations without metro around or the bus stations in a more congested direction, because the taxi-hiring behavior is more frequent near such stations. When traffic management departments plan to set up new taxi stops, avoiding disturbing operation of buses and convenience for passengers hiring taxis near bus stations are key factors, which ought to be emphasized. Moreover, transit systems are congested in most mega-cities with high population density in East Asia, especially in China and Japan. Therefore, the research results of Shenzhen could be applied to these cities with congested traffic.

This work shows the specific relationship between bus service quality and taxi behavior near bus stations via data mining and model developing with parametric, semiparametric and nonparametric approaches. However, the methods to process the taxi data still have several limitations. First, only the destinations around bus stations along Shennan Avenue were captured. Because Shennan Avenue is one of the most congested road in Shenzhen, the phenomenon that passengers switch to taxis is common and more observations could be found on this road. However, the results are without comparison with other roads. Second, other bus routes along Shennan Avenue except for Route 223 were not taken into account. Although Route 223 are busier and has

better GPS data quality, the lack of other routes will still affect the definition of statistics results. Third, the observation interval (2 hours in peak time) is a little big. Because even in Shennan Avenue, the target taxi-hiring behavior is only several dozens at one station during peak time. If the observation interval (such as 15 or 30 min) is shorter, the number of target taxis is too small to capture. But buses are too many in 2-hour observation interval, this makes it hard to find which buses directly affect the target taxi-hiring behavior. In future studies, the comprehensive interplay between buses and taxis under network with more destinations should be focused. Moreover, the GPS data with longer observation periods (e.g. 3 months) and more advanced methods of data processing should be addressed in the future.

### Acknowledgements

The authors would like to thank Wenqiao Sun and Weikao Wu for critically reading and painstakingly editing the manuscript, and for their helpful comments and improvement of the text.

This research is funded by the National Natural Science Foundation of China (No 51408356). Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors.

### References

- Alonso, B.; Moura, J. L.; Dell'Olio, L.; Ibeas, Á. 2011. Bus stop location under different levels of network congestion and elastic demand, *Transport* 26(2): 141–148. <http://doi.org/10.3846/16484142.2011.584960>
- Carslaw, D. C.; Beevers, S. D.; Tate, J. E. 2007. Modelling and assessing trends in traffic-related emissions using a generalised additive modelling approach, *Atmospheric Environment* 41(26): 5289–5299. <http://doi.org/10.1016/j.atmosenv.2007.02.032>
- Dell'Olio, L.; Ibeas, Á.; Dominguez, A.; Gonzalez, F. 2012. Passenger preference analysis: light rail transit or bus versus car, *Transport* 27(3): 276–285. <http://doi.org/10.3846/16484142.2012.719839>
- Dell'Olio, L.; Ibeas, Á.; Cecin, P. 2011a. The quality of service desired by public transport users, *Transport Policy* 18(1): 217–227. <http://doi.org/10.1016/j.tranpol.2010.08.005>
- Dell'Olio, L.; Ibeas, Á.; Cecin, P.; Dell'Olio, F. 2011b. Willingness to pay for improving service quality in a multimodal area, *Transportation Research Part C: Emerging Technologies* 19(6): 1060–1070. <http://doi.org/10.1016/j.trc.2011.06.004>
- Fan, W.; Machemehl, R. B. 2002. Characterizing bus transit passenger waiting times, in *2nd Material Specialty Conference of the Canadian Society for Civil Engineering*, 5–8 June 2002, Montreal, Quebec, Canada, 1–10.
- French, J. L.; Kammann, E. E.; Wand, M. P. 2001. [Semiparametric nonlinear mixed-effects models and their applications]: comment, *Journal of the American Statistical Association* 96(456): 1285–1288. <http://doi.org/10.1198/016214501753381913>
- Gu, S. C.; Luo, G. Z.; Zhang, Y. B. 2010. The way home is crowded and sad, *Shenzhen Special Zone Daily* 8. Available from Internet: [http://sztqb.sznews.com/html/2010-08/02/content\\_1178035.htm](http://sztqb.sznews.com/html/2010-08/02/content_1178035.htm) (in Chinese).
- Hanaoka, S.; Qadir, F. M. 2009. Passengers' perceptions and effects of bus-holding strategy using automatic vehicle location technology, *Journal of Advanced Transportation* 43(3): 301–319. <http://doi.org/10.1002/atr.5670430304>
- Hastie, T. J.; Tibshirani, R. J. 1990. *Generalized Additive Models*. Chapman and Hall/CRC. 352 p.
- Lai, W.-T.; Chen, C.-F. 2011. Behavioral intentions of public transit passengers: the roles of service quality, perceived value, satisfaction and involvement, *Transport Policy* 18(2): 318–325. <http://doi.org/10.1016/j.tranpol.2010.09.003>
- Lei, D. 2009. Buses, cannot move!, *Shenzhen Evening News* 3. Available from Internet: [http://wb.sznews.com/html/2009-03/23/content\\_558803.htm](http://wb.sznews.com/html/2009-03/23/content_558803.htm) (in Chinese).
- Li, B.; Zheng, J. Y.; Hu, J. W. 2009. Ten thousand people wait buses at Meilinguan station, *Shenzhen Economic Daily* 12. Available from Internet: [http://szsb.sznews.com/html/2009-12/15/content\\_890742.htm](http://szsb.sznews.com/html/2009-12/15/content_890742.htm) (in Chinese).
- Litman, T. 2013. *Understanding Transport Demands and Elasticities How Prices and Other Factors Affect Travel Behavior*. Victoria Transport Policy Institute, Canada. 76 p. Available from Internet: <http://www.vtpi.org/elasticities.pdf>
- Osuna, E. E.; Newell, G. F. 1972. Control strategies for an idealized public transportation system, *Transportation Science* 6(1): 57–72. <http://doi.org/10.1287/trsc.6.1.52>
- Redman, L.; Friman, M.; Gärling, T.; Hartig, T. 2013. Quality attributes of public transport that attract car users: a research review, *Transport Policy* 25: 119–127. <http://doi.org/10.1016/j.tranpol.2012.11.005>
- Rojó, M.; Gonzalo-Orden, H.; Dell'Olio, L.; Ibeas, Á. 2012. Relationship between service quality and demand for inter-urban buses, *Transportation Research Part A: Policy and Practice* 46(10): 1716–1729. <http://doi.org/10.1016/j.tra.2012.07.006>
- Rose, J. M.; Hensher, D. A. 2014. Demand for taxi services: new elasticity evidence, *Transportation* 41(4): 717–743. <http://doi.org/10.1007/s11116-013-9482-5>
- Ruppert, D.; Wand, M.; Carroll, R. J. 2003. *Semiparametric Regression*. Cambridge University Press. 404 p.
- Salek, M.-D.; Machemehl, R. B. 1999. *Characterizing Bus Transit Passenger Wait Times*. Report No SWUTC/99/167211-1. Southwest Region University Transportation Center, Texas Transportation Institute, The Texas A&M University System. 230 p. Available from Internet: <http://static.tti.tamu.edu/swutc.tamu.edu/publications/technical-reports/167211-1.pdf>
- Shively, T. S.; Kockelman, K.; Damien, P. 2010. A Bayesian semi-parametric model to estimate relationships between crash counts and roadway characteristics, *Transportation Research Part B: Methodological* 44(5): 699–715. <http://doi.org/10.1016/j.trb.2009.12.019>
- Shively, T. S.; Sager, T. W. 1999. Semiparametric regression approach to adjusting for meteorological variables in air pollution trends, *Environmental Science & Technology* 33(21): 3873–3880. <http://doi.org/10.1021/es990286b>
- Sun, D.; Peng, Z.-R.; Shan, X.; Chen, W.; Zeng, X. 2011. Development of web-based transit trip-planning system based on service-oriented architecture, *Transportation Research Record: Journal of the Transportation Research Board* 2217: 87–94. <http://doi.org/10.3141/2217-11>
- TRB. 2013. *Transit Capacity and Quality of Service Manual*. TRB's Transit Cooperative Highway Research Program (TCRP) Report 165. Transportation Research Board (TRB), Washington, DC. Available from Internet: <http://www.trb.org/Main/Blurbs/169437.aspx>

- Wan, W.; Birch, J. B. 2011. A semiparametric technique for the multi-response optimization problem, *Quality and Reliability Engineering International* 27(1): 47–59.  
<http://doi.org/10.1002/qre.1106>
- Wood, S. N. 2006. *Generalized Additive Models: an Introduction with R*. Chapman and Hall/CRC. 410 p.
- Wood, S. N. 2003. Thin plate regression splines, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 65(1): 95–114. <http://doi.org/10.1111/1467-9868.00374>
- Wood, S. N.; Augustin, N. H. 2002. GAMs with integrated model selection using penalized regression splines and applications to environmental modelling, *Ecological Modelling* 157(2–3): 157–177.  
[http://doi.org/10.1016/S0304-3800\(02\)00193-X](http://doi.org/10.1016/S0304-3800(02)00193-X)
- Xiao, H.; Xu, J. H. 2012. More than ten million people take buses every day in Shenzhen, *Shenzhen Economic Daily* 10. Available from Internet: [http://szsb.sznews.com/html/2012-10/29/content\\_2254119.htm](http://szsb.sznews.com/html/2012-10/29/content_2254119.htm) (in Chinese).
- Xie, L. W.; Chen, J. J.; Luo, J. J. 2013. Waiting 1 hour to get on the buses, the reporters of the crystal news experienced the worst buses, *The Crystal News* 11. Available from Internet: [http://jb.sznews.com/html/2013-11/05/content\\_2673778.htm](http://jb.sznews.com/html/2013-11/05/content_2673778.htm) (in Chinese).