

## Research Article

# Investigating Travel Flow Differences between Peak Hours with Spatial Model with Endogenous Weight Matrix Using Automatic Vehicle Identification Data

Yiwei Zhou <sup>1,2</sup>, Zhaocheng He <sup>3</sup>, Jin-Yong Chen <sup>4</sup>, Linglin Ni,<sup>5</sup> and Jieshuang Dong<sup>1</sup>

<sup>1</sup>Business School, University of Shanghai for Science and Technology, 516 Jungong Rd, Shanghai 200093, China

<sup>2</sup>Center for Supernetworks Research, University of Shanghai for Science and Technology, 334 Jungong Rd 200093, Shanghai, China

<sup>3</sup>Guangdong Provincial Key Laboratory of Intelligent Transportation System, School of Intelligent Systems Engineering, Sun Yat-sen University, Guangzhou 510275, China

<sup>4</sup>School of Automotive and Transportation Engineering, Guangdong Polytechnic Normal University, Guangzhou 510665, China

<sup>5</sup>Beijing Wuzi University Logistics School, Beijing 101149, China

Correspondence should be addressed to Jin-Yong Chen; [chenjinyong@gpnu.edu.cn](mailto:chenjinyong@gpnu.edu.cn)

Received 11 August 2022; Revised 23 November 2022; Accepted 30 November 2022; Published 28 December 2022

Academic Editor: Peter J. Jin

Copyright © 2022 Yiwei Zhou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The rapid urbanization has brought great challenges to the transportation network. However, travel flow at peak hours is not always the same. It is important to investigate how travel flow differs between peak hours to capture travel flow patterns and influential factors to facilitate traffic management and urban planning. This paper establishes a spatial model with endogenous weight matrix (SARBP-EWM) to investigate the travel flow differences between morning and evening peaks on both weekday and weekend based on automatic vehicle identification (AVI) data and point of interest (POI) data in Xuancheng, China. The results confirm strong spatial effects and endogeneity issue. Besides, facility variables such as number of offices and number of clinics reveal strong negative impacts on travel flow differences on both weekday and weekend, while the number of middle school shows significantly positive relation with travel flow differences. In addition, the endogenous weight matrix on both weekday and weekend is successfully estimated and compared. It is found that TAZ pairs tend to be clustered with lower spatial weights on weekday, while they are more randomly distributed with higher spatial weights at weekend. Based on the results above, the policies proposed from Xuancheng 14<sup>th</sup> Five-Year Plan are evaluated and discussed. The above empirical analysis quantifies impacts from key factors on urban travel flow differences between peak hours and provides important references for urban planning and policy making.

## 1. Introduction

The rapid urbanization process in China has brought much challenges to urban transportation network, especially the excessive travel flow at peak hours. It not only adds travel time cost but also lowers network efficiency. However, travel flow at peak hours is not always the same. People may commute to work in the morning peak while they go for entertainment in the evening peak. The travel flow patterns at peak hours can be totally different. Therefore, it is worthwhile to study the origin-destination (OD) travel flow

patterns at peak hours and discover key influential factor, which will provide important insights for the traffic management and urban planning.

Many studies have been conducted on the travel flow analysis in peak hours. However, few studies focus on the travel flow differences between peak hours. The travel flow pattern could be different during morning peak and evening peak. It is important to capture such different travel patterns and find out influential factors. Jia et al. [1] pointed out that A.M. peak urban traffic could be different from P.M. peak. The P.M. peak was usually regarded as the mirror image of

A.M. peak traffic. Fosgerau and Fukuda [2] confirmed that heteroskedasticity phenomenon traffic conditions in the peak periods are more variable than those in the regular periods. Ni et al. [3] noted that the travel flow features cannot be captured by merely integrating the A.M. and P.M. travel flows. Urban travel flow has different patterns in morning peak and evening peak hours due to different travel purposes such as commuting or entertainment. For example, Hu et al. [4] used the taxi GPS trajectory data to analyze the coupling relationship between regional taxi demand and social development by building the coupling coordination degree model (CCDM). Results showed that in the morning rush, taxi orders flow from residential to office; the reverse applies during the evening rush.

Despite these studies, few studies have considered spatial correlations in travel flow analysis. Spatial correlations exist among travelers' mode choice, daily traffic, and peak hour travel and have been confirmed by previous studies [5–7]. Ni et al. [3] analyzed urban travel flow by developing a spatial autocorrelation model based on mobile phone data in Hangzhou, China. Results concluded that the model ignoring the spatial autocorrelation tends to underestimate the impacts of influence factors on travel flows. Chu et al. [8] developed a multiscale convolutional long short-term memory network (MultiConvLSTM) deep learning model, which considered both temporal and spatial correlations to predict the future travel demand and OD flows. However, few studies have considered endogeneity in travel flow analysis. Endogeneity exists among traffic participants and has been studied by many researchers [9–11]. However, traditional spatial models assume an exogenous weight matrix, which may lead to biased estimation or even false conclusions. Therefore, it is critical to consider spatial correlations and endogeneity in urban travel flow analysis to better capture travel flow patterns and quantify impacts of key factors.

To the best knowledge of the authors, few existing studies have systematically investigated OD travel flow differences between peak hours while considering spatial correlations, not to mention endogeneity due to its complexity. This paper addresses the endogeneity issue by using a spatial binary probit model with endogenous weight matrix (SARBP-EWM) to identify such endogeneity and quantify the impacts of key factors based on automatic vehicle identification (AVI) data collected from Xuancheng, China.

## 2. Literature Review

**2.1. Urban Travel Flow Analysis.** Many studies have been conducted on urban travel flow analysis. Various land use variables and facility variables were found to have an influence on urban travel flow. Tsai et al. [12] investigated the relationship between public transport demand and land use characteristics in the Sydney Greater Metropolitan Area using a geographically weighted regression (GWR) approach. Results revealed that impacts of land use characteristics on public transport demand vary spatially. Zhou and Wang [13] developed structural equation modeling (SEM) to investigate relation between online shopping and

shopping trips and other influential factors. They found that residents' location in urban center is strongly associated with their propensity to shop online, which leads to increase of shopping trips. Wu et al. [14] proposed a novel spatiotemporal random effects (STRE) model to predict urban travel flow using data from loop detectors. Zheng and Liu [15] used connected vehicle (CV) trajectory and signal status data to estimate traffic volumes at signalized intersections. Li et al. [16] built a deep feature fusion model to predict space-mean-speed using heterogeneous data. Samara et al. [17] proposed a novel approach for estimating vehicle travel time distribution using copula-based discrete convolution.

Although numerous studies have been conducted on urban travel flow analysis, few studies have focused on travel flow at peak hours. Kumar and Vanajakshi [18] performed short-term traffic flow prediction by establishing a seasonal autoregressive integrated moving average (ARIMA) model with limited input data. Results indicated that travel flows in morning peak were larger than evening peak in three consecutive days. Yang and Qian [19] acknowledged correlations of travel time between morning peak and afternoon peak. Morning peak travel time was used when predicting afternoon travel time. Shen et al. [20] investigated car travel demand in Hangzhou by establishing a geographically and temporally weighted regression (GTWR) model. The results confirmed peak phenomena and found that the influence of built environment and household properties on car travel demand varies with space and time. In general, few studies have been conducted on travel flow at or between peak hours. Further studies are needed to investigate travel flow patterns at peak hours.

**2.2. Spatial Effects in Travel Flow Analysis.** Spatial effects have major impacts on urban travel flow. Many studies have been conducted to consider spatial effects in travel flow analysis. Bhat and Zhao [21] performed spatial analysis of activity stop generation by establishing a multilevel mixed logit model to address the spatial heterogeneity across TAZs. LeSage and Pace [22] proposed spatial weight structure including origin dependence, destination dependence, and OD dependence in the standard spatial autoregressive models for analyzing flow patterns. LeSage and Thomas-Agnan [23] further provided expressions for calculating partial derivatives of the above model to quantify spatial dependence between OD flows. The model was then applied to analyze commuting flows of 60 regions in Toulouse, France. Kerkman et al. [24] analyzed spatial dependence among public transit passenger flows in an urban region in the Netherlands by developing five distinct spatial interaction models (SIMs). Results indicated that spatial autocorrelation effects cannot be neglected in travel flow analysis. Zheng and Geroliminis [25] formulated an optimization framework of equitable congestion pricing schemes for multimodal networks with heterogeneous population. The results justified the need for a value-of-time (VOT)-based pricing among groups with different behaviors and cost savings.

Despite numerous studies on spatial effects, few studies have considered endogeneity in travel flow analysis. Endogeneity occurs when one's travel behavior and the whole network affect each other, which will lead to biased estimation or even false conclusions. Mokhtarian and Cao [6] reviewed major methods on addressing the residential self-selection issue on travel behavior. Schatzmann et al. [26] applied a spatial autoregressive model to study OD public transportation commuting flows between municipalities in Switzerland. They used an instrumental variable approach to account for endogeneity and showed that income differences are underestimated in the gravity and spatial models if assumed exogenous. Guevara et al. [27] proposed the multiple indicator solution (MIS) method in a stated preference (SP) experiment to correct for endogeneity due to omitted crowding in public transport choice. Results suggest that endogeneity issue may arise if indicators are only weakly correlated with the omitted attribute. Guerrero et al. [28] proposed a control function updated (CFU) method to correct endogeneity issue in transport modeling. The results indicated that the new CFU approach showed statistically significant improvements over the classical approach in all scenarios tested.

**2.3. AVI Data in Travel Flow Analysis.** With the rapid development of telecommunications technology, new types of data with spatial and temporal information such as call detail record (CDR) data, location-based social media (LBSM) data, and taxi GPS data emerge and are receiving more attention in recent years [29–32]. Besides these data sources, the AVI data are considered an emerging data source in recent years. The AVI provides rich information including detector number, license plate, vehicle passing time, and time stamp, which could be used to reconstruct the vehicle trajectory and obtain OD flow information. Ahmed and Abdel-Aty [33] used AVI data for real-time crash prediction research. The results showed that the likelihood of a crash is statistically related to speed data obtained from AVI segments. Zhan et al. [34] proposed a queue length estimation model using license plate (LPR) recognition data, which provided an efficient queue length estimation at the lane level in real time. Zheng et al. [35] used data from automated number plate recognition (ANPR) cameras to study the travel time reliability. Fekih et al. [36] proposed a framework to extract dynamic trip flows and travel demand patterns from cellular signaling data to estimate aggregate trip by time of day.

As for the travel flow studies using AVI data, Chen et al. [37] proposed a copula-based approach to model arterial travel time distribution (TTD), which was examined with AVI data and next generation simulation (NGSIM) trajectory data. The results demonstrate the advantage of the proposed copula-based approach. Zhao et al. [38] collected 24 h AVI data from Wuhan, China, to investigate weekly travel patterns of private vehicles and identified four types of commuters. The results revealed six variations of the travel demand on weekdays and weekends. Huang et al. [39] proposed a semisupervised deep learning based model that

appropriately combines both AVI and smartphone trajectory data during training. The model can provide OD estimation and prediction services on larger spatial areas beyond the limited spatial coverage of AVI data. Cao et al. [40] proposed a new method to recover day-to-day dynamic OD flows using both connected vehicle (CV) trajectories and AVI observations. The results indicated that the proposed method requires very few AVI detectors and CV trajectories to achieve competitive estimation performance against two benchmark models.

### 3. Data Description

Two major data sources are used in this paper. First, the OD flow data of morning and evening peaks are extracted from AVI data gathered from Xuancheng City, Anhui Province, China. The AVI system automatically identifies license plates when vehicles leave stop line and stores relative message as AVI data. Each AVI record mainly includes information such as detector number, license plate, vehicle passing time, and time stamp. Besides, the vehicle trajectory can be reconstructed based on the above information. In addition, the OD flow data, which contain license plate number, departure time, arrival time, and OD pair, are further generated from trajectory data. Based on process above, we finally extracted the OD flow data between 7:30 A.M. and 10:30 A.M. in morning peak and between 5:30 P.M. and 8:30 P.M. in the evening peak in four consecutive weeks from September 2 to 29, 2019. The original OD pairs recorded in the AVI database are stored in the form of roads. To facilitate this study, each OD road was further identified and matched to the corresponding TAZ. Finally, we take the average traffic volume of 20 workdays and 8 weekends, respectively, to obtain the average travel flow of each TAZ.

Second, the key urban facilities such as offices, supermarkets, schools, hospitals, hotels, sports centers, and bus stations have major influence on the residents' mobility. In this paper, the points of interest (POIs) of these different key facilities were obtained on the AutoNavi Map [41] by using web crawler technology. The whole Xuancheng area is divided into 32 TAZs. The above travel flow data and facility data are integrated for each TAZ.

**3.1. Dependent Variable.** The dependent variable needs to represent the travel flow differences between morning and evening peaks. First, total travel flow for 32 TAZs is calculated. As stated above, the average OD travel flows between TAZ  $i$  and TAZ  $j$  are derived from the AVI database. The total outgoing travel flow from TAZ  $i$  is calculated by taking sum of outgoing travel flows  $q_{ij}^{\text{out}}$  with origin TAZ  $i$ . Similarly, the total incoming travel flow to TAZ  $i$  is calculated by taking sum of incoming travel flows  $q_{ji}^{\text{in}}$  with destination TAZ  $i$ . Therefore, the total travel flow to and from TAZ  $i$  is calculated as follows:

$$Q_i = \sum_j q_{ij}^{\text{out}} + \sum_j q_{ji}^{\text{in}}. \quad (1)$$

The mere absolute travel flow difference is sensitive to various factors. For example, travel flow in TAZs with larger areas may vary to a large extent, while travel flow in TAZs with smaller areas may only change in a small degree. Therefore, the relative travel flow fluctuation is derived to measure the relative travel flow differences between peak hours, where the difference between morning and evening peak travel flows is divided by evening peak travel flow to obtain the travel flow fluctuations. The travel flow fluctuation is calculated as follows:

$$F_i = \frac{Q_i^m - Q_i^e}{Q_i^e}, \quad (2)$$

where  $F_i$  represents the travel flow fluctuation between morning peak and evening peak for TAZ  $i$ ,  $Q_i^m$  and  $Q_i^e$  represent the travel flow of TAZ  $i$  in the morning peak and evening peak, respectively, and  $Q_i^m$  and  $Q_i^e$  are calculated using (1) by adding the inbound traffic and outbound traffic of TAZ  $i$  in the morning peak and evening peak, respectively.

Since the travel patterns may be different on weekday and weekend, the average weekday (20 workdays) travel flow fluctuation and average weekend (8 weekends) travel flow fluctuation are calculated, respectively, in the paper. The travel flow fluctuation distributions for 32 TAZs on weekday and weekend are illustrated in Figures 1 and 2. In addition, the cumulative distributions of average weekday and weekend travel flow fluctuations are calculated and presented in Figure 3 for further analysis.

Generally speaking, TAZs in suburban areas have larger fluctuations, which indicates that there are significant travel flow differences between morning and evening peaks, while downtown areas have relatively smaller fluctuations. For TAZs that have negative fluctuations, they are mainly located in city central areas.

The cumulative distributions in Figure 3 show that both weekday and weekend fluctuations vary greatly from  $-20\%$  below to  $60\%$  above in general. It indicates that travel flow fluctuations between morning peak and evening peak among TAZs vary to a large extent. According to Figure 3, since less than  $10\%$  fluctuations are negative, most travel flow fluctuations are positive, suggesting that morning peak travel flow is dominating evening peak travel flow in most TAZs. Besides, weekday fluctuations are slightly larger than weekend fluctuations in general for given cumulative percentage.

To determine whether there is significant difference between morning and evening peak travel flows, a threshold needs to be determined. Since the travel flow at weekday and weekend has different fluctuation levels and may have different travel patterns, it is not appropriate to set a unified fluctuation level for them. Fluctuation threshold needs to be determined separately for weekday and weekend to achieve balanced fluctuation quantiles and eliminate random errors. As shown in Figure 3, weekday fluctuations vary from  $-40\%$  to  $60\%$ . To maintain balanced distributions of fluctuations, a  $30\%$  fluctuation threshold is set for weekday, which corresponds to roughly  $50\%$ . It means that travel flow fluctuations with negative fluctuations and fluctuations less than  $30\%$  are

considered insignificant, taking up about  $50\%$ , while fluctuations larger than  $30\%$  are considered significant difference between morning and evening peaks, taking up equally about  $50\%$ . Similarly, a  $20\%$  fluctuation threshold is set for weekend fluctuations, which corresponds to about  $50\%$  to achieve even distributions for fluctuations smaller and larger than the threshold.

Based on the thresholds above, a binary dependent variable is defined to determine whether morning peak travel flow is significantly greater than evening peak travel flow. The binary dependent variable is defined as follows:

$$Y_i^* = \begin{cases} 1, & \text{if } F_i > \alpha, \\ 0, & \text{if } F_i \leq \alpha, \end{cases} \quad (3)$$

where  $Y_i^*$  is the binary dependent variable indicating whether morning peak travel flow is significantly greater than evening peak travel flow for TAZ  $i$ .  $\alpha$  is the fluctuation threshold. As discussed above, it is set as  $30\%$  for weekday fluctuations and  $20\%$  for weekend fluctuations. The dependent variable takes a value of 1 if the absolute value of travel flow fluctuation is larger than the threshold and takes a value of 0 otherwise.

Dependent variables for both weekday and weekend travel flow differences are generated using equation (3). The spatial distributions of dependent variable for both weekday and weekend are illustrated in Figures 4 and 5. Percentage of dependent variable taking values of 1 and 0 for both weekday and weekend is calculated, respectively, and summarized in Table 1.

According to Figures 4 and 5, most TAZs reveal significant travel flow differences in suburban TAZs on both weekday and weekend. For travel flow differences on weekday, 16 out of 32 TAZs take value 1 according to Table 1, taking up  $50.0\%$  of all 32 TAZs. These TAZs with significant travel flow differences are mostly located in suburban areas. Only a few TAZs with significant travel flow differences are located in downtown areas. Similar travel patterns are also discovered in weekend travels as shown in Figure 5, where 19 TAZs show significant travel flow differences in suburban areas, taking up  $59.4\%$  of all 32 TAZs. The rest of the 13 TAZs with insignificant travel flow differences are all located in downtown areas, taking up  $40.6\%$  of all 32 TAZs. For weekend travels, TAZ 2, TAZ 9, and TAZ 14 change from being insignificant on weekday to being significant, while the rest remain the same on both weekday and weekend.

Spatial distributions above show that travel flows have greater differences in suburban areas than central areas across weekday and weekend. Such phenomena have been found and investigated by previous studies [16, 18, 20, 42]. It is possible that central areas have more regular travel patterns due to commuting travels, while suburban areas have more random travel patterns. For Xuancheng, this phenomenon is uniform across many TAZs. This paper reveals the mechanism of such phenomenon and quantifies the influence of potential factors.

To examine the spatial effect of travel flow differences among TAZs, Moran's  $I$  of dependent variable is calculated at  $0.025$  with  $p$  value  $0.027$ . It indicates that there are significant

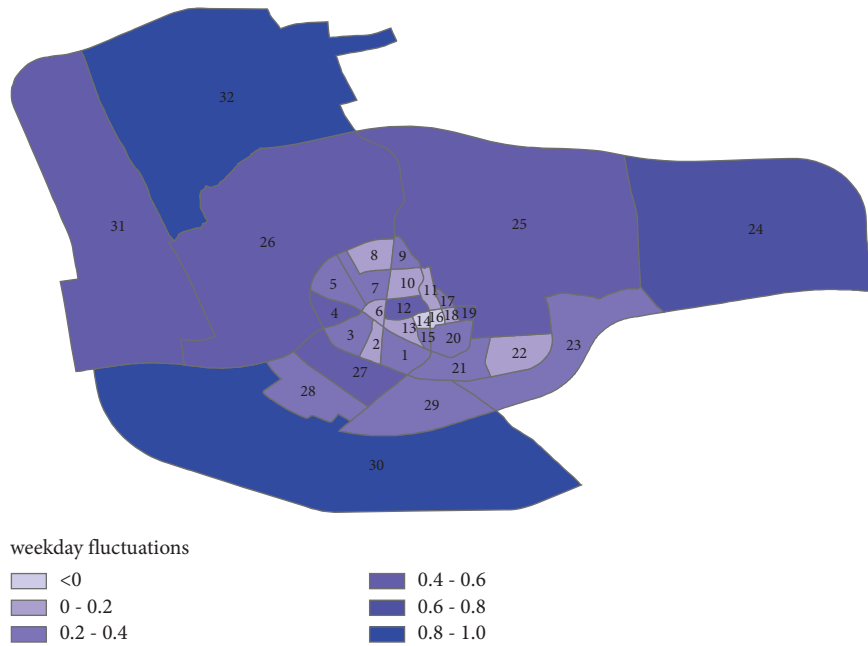


FIGURE 1: Average weekday travel flow fluctuations.

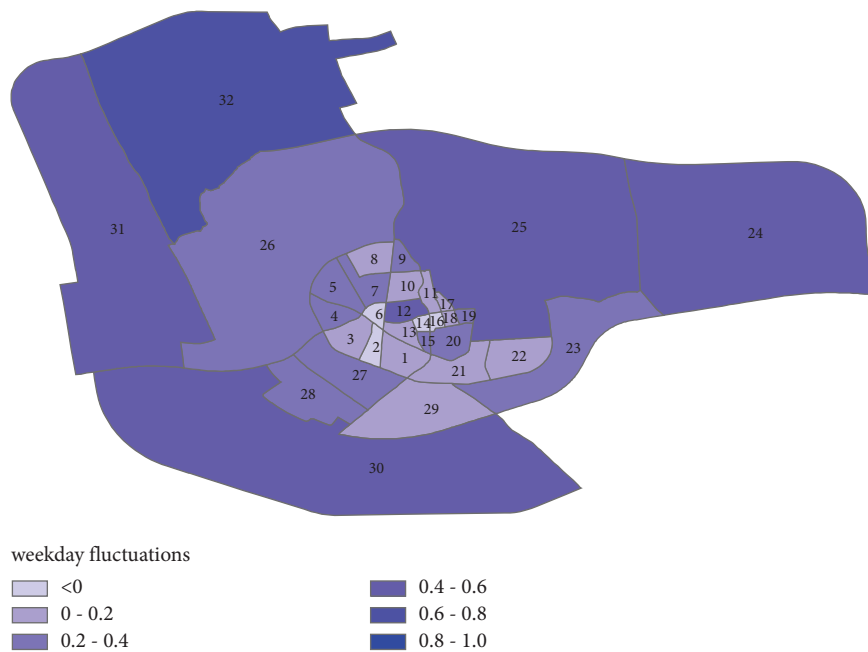


FIGURE 2: Average weekend travel flow fluctuations.

positive spatial correlations among dependent variables. Also, TAZs with significant travel flow differences tend to cluster. Therefore, the spatial correlations should be taken into account to explore the influential factors of travel flow differences.

**3.2. Explanatory Variables.** Explanatory variables based on current dataset include public facilities such as office, supermarket, middle school, clinic, inn, and sports

center, which can have major impacts on travel flow. To avoid multicollinearity issue, we choose only one variable for each type of facility. Also, only variables used in the final model are presented. Such model could be extended by incorporating new types of variables which provide meaningful results. For these facilities, this paper uses point of interest (POI) data from online map service to obtain detailed number of facilities for each TAZ [41]. Distribution of these facilities is shown in Figure 6.

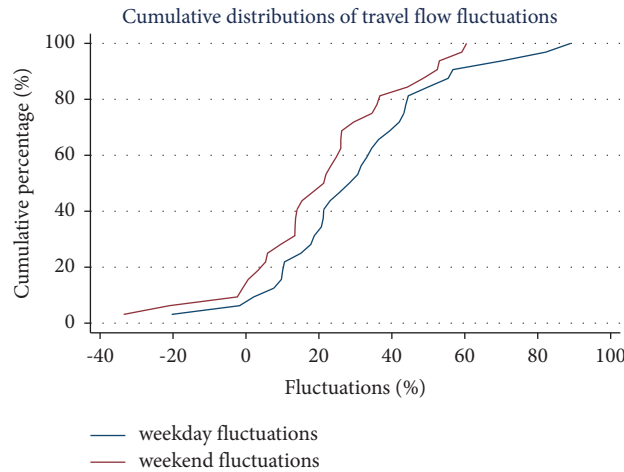


FIGURE 3: Cumulative distribution of average weekday and weekend travel flow fluctuations.

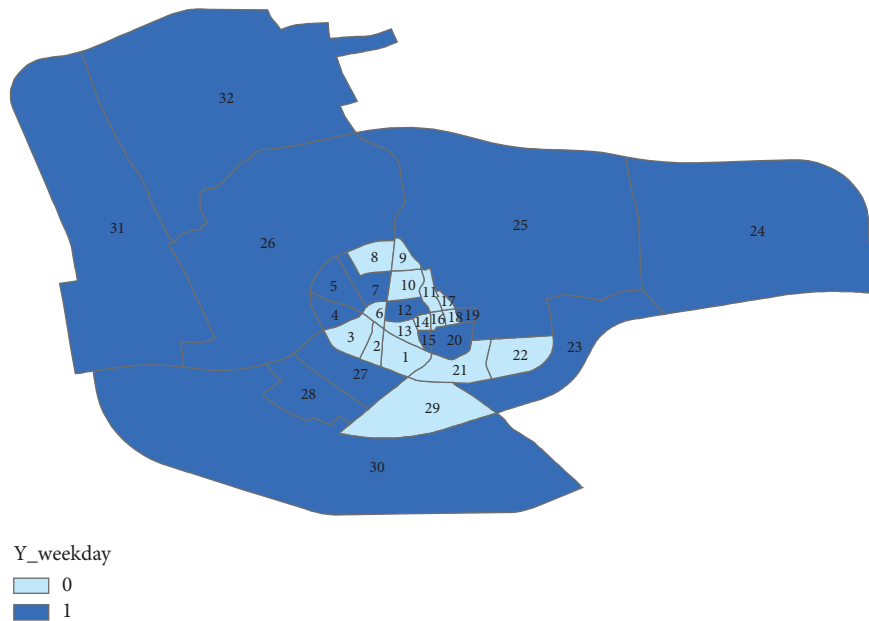


FIGURE 4: Spatial distribution of average weekday travel flow differences.

Offices are the major sources of travel flow as people need to go to work during morning peak and go back home in the evening peak. The commuting travel flow induced by office buildings takes large portion of total travel demand and may cause various issues like traffic congestion [43]. The spatial distribution of offices is illustrated in Figure 6(a) where offices are found to cluster densely in city center areas while there are fewer offices in suburban areas. This is contrary to the spatial distributions of travel flow differences in Figures 4 and 5 where city center areas have no significant travel flow differences and suburban areas have major travel flow differences. For example, TAZs in city center area like TAZs 1, 20, and 27 have higher number of offices but correspond to insignificant travel flow differences.

Supermarkets are major sources of shopping activities, attracting numerous people to travel to buy life necessities every day. Supermarkets are very popular in China as they

are located in almost every community and play major roles in providing daily services for local residents. Researchers have studied the impacts of supermarket on travel flows [44, 45]. Therefore, the number of supermarkets in TAZs is considered in this paper. According to Figure 6(b), TAZs in outskirts areas have higher number of supermarkets, which are consistent with travel flow difference patterns shown in Figure 4. For example, TAZs in outskirts areas like TAZ 25 and TAZ 26 have relatively higher number of supermarkets which correspond to significant travel flow differences according to Figure 4, while TAZ 2, TAZ 3, and TAZ 6 have smaller number of supermarkets and correspond to insignificant travel flow differences.

Schools are considered another major source of travel flow. Students go to school in the morning and go back home in the evening. Students themselves and accompanying

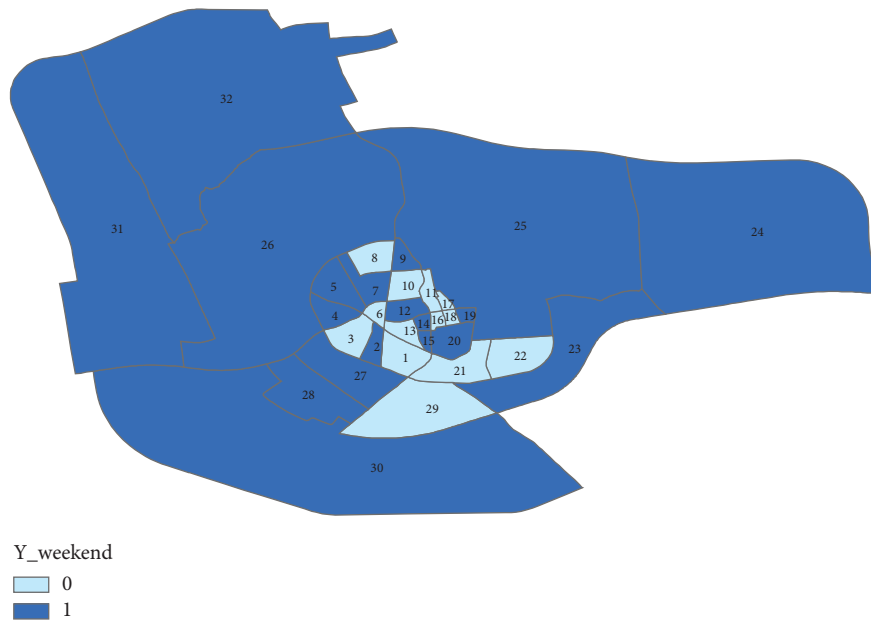


FIGURE 5: Spatial distribution of average weekend travel flow differences.

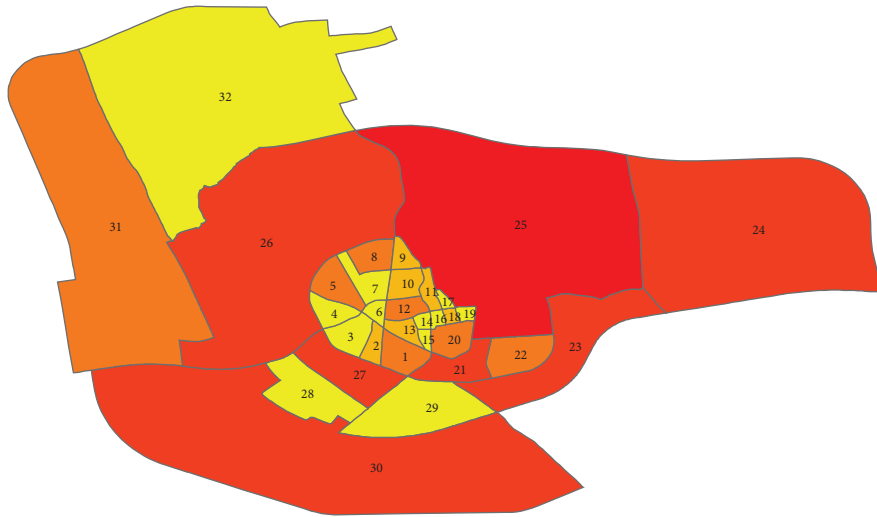
TABLE 1: Summary of dependent variable.

Dependent variable	Threshold (%)	Value 1		Value 0	
		Number	Percentage	Number	Percentage
Weekday fluctuations (Y_weekday)	30	16	50.0	16	50.0
Weekend fluctuations (Y_weekend)	20	19	59.4	13	40.6



(a)

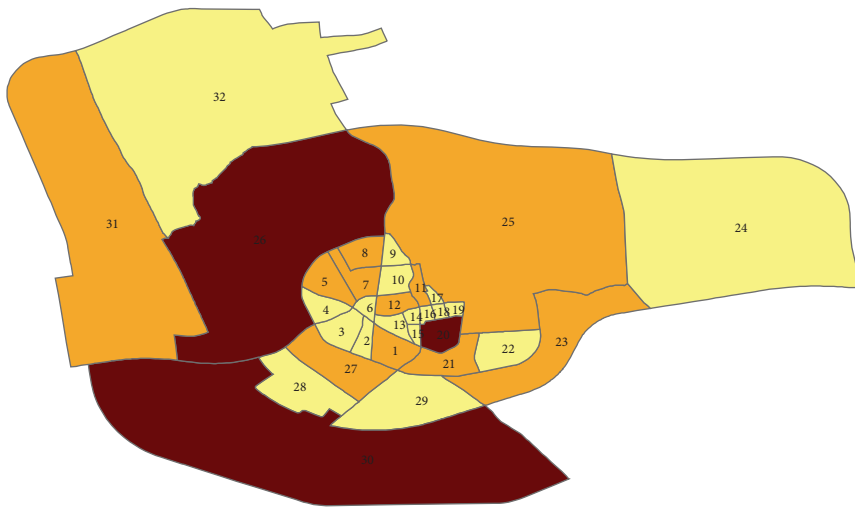
FIGURE 6: Continued.



number of supermarket

0 - 9	31 - 44
10 - 16	45 - 84
17 - 30	

(b)



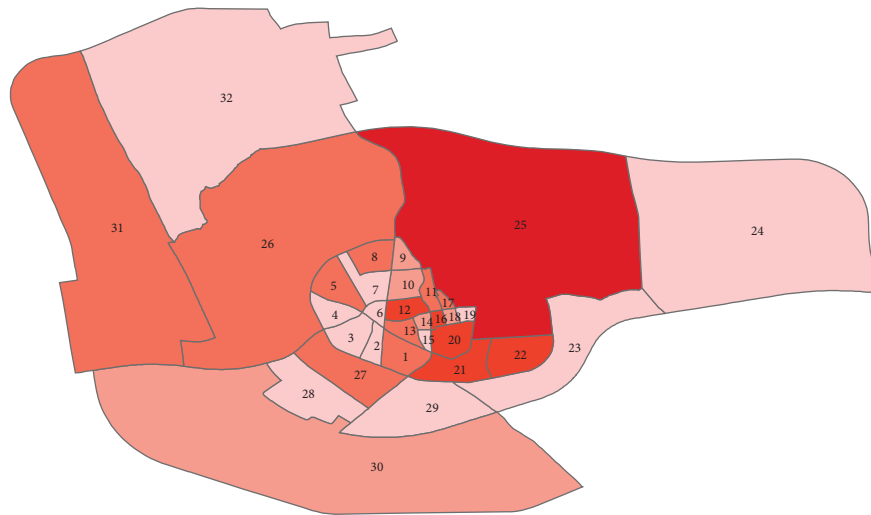
number of middle school

0
1
2

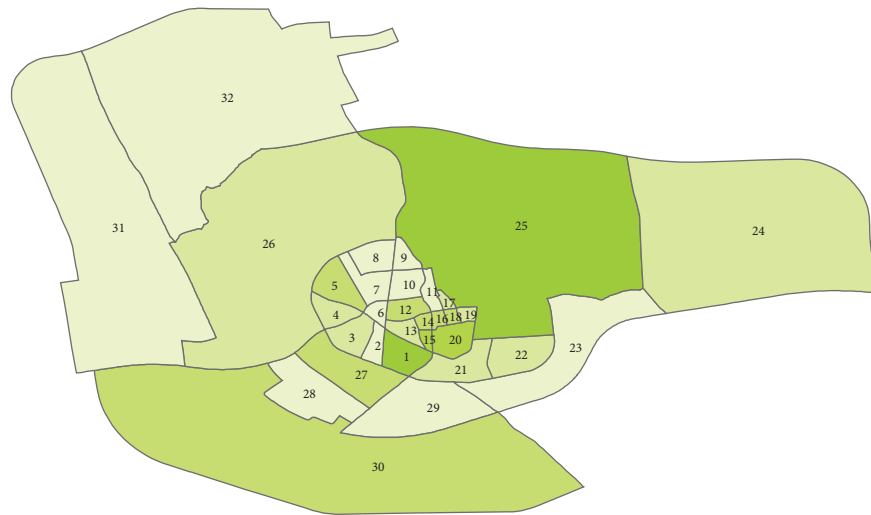
(c)

FIGURE 6: Continued.





(d)



(e)

FIGURE 6: Continued.



FIGURE 6: Spatial distributions of facilities in Xuancheng, China. (a) Office distributions. (b) Supermarket distributions. (c) Middle school distributions. (d) Clinic distributions. (e) Inn distributions. (f) Sports center distributions.

parents generate large travel demand. Therefore, schools cannot be neglected in this study. Ikeda et al. [46] discovered that schools can generate active school travels under certain built environment conditions. The spatial distribution of middle schools is presented in Figure 6(c) where suburban TAZs have higher number of middle schools than urban TAZs. Such spatial distributions are consistent with travel flow difference distributions with outskirts areas being significant.

For medical facilities, clinics are considered another source of travel flow. Clinics exist in almost every community in China now, which provide necessary medical services for local residents. They together with high-grade hospitals have medical systems to provide medical services from all levels. Cheng et al. [47] analyzed the spatial correlation of residents' accessibility to these medical facilities and confirmed spatial imbalance among these hospitals. Similarly, the spatial distribution of clinics is shown in Figure 6(d) where suburban TAZs, especially TAZs in east areas like TAZ 20, TAZ 21, TAZ 22, and TAZ 25, have higher number of clinics than urban TAZs.

Besides, recreation facilities like inns and sports centers also generate substantial travel demand. People visiting Xuancheng for business or tourism would choose to stay in local inns. Such recreation facilities can generate travel flow, which had been studied by previous research. Schirpke et al. [48] mapped recreation flows in the Alpine Space area and found significant spatial pattern differences between mountain areas and lowlands. Similarly, the spatial distributions of inns are illustrated in Figure 6(e) where suburban TAZs have relatively higher number of inns than central areas. In addition, sports centers show similar patterns according to Figure 6(f) where suburban TAZs in east and

south areas have higher number of sports centers. People travel to sports centers in leisure time for relaxation, which generates potential travel demand.

**3.3. Endogenous Variable.** By considering endogeneity, an endogenous variable needs to be defined to represent the interdependency structure among TAZs. This paper chooses population density as the endogenous variable. Population density is the necessary condition for regional development and social and economic activities, which represents the urbanization level of TAZs. The population density differences between TAZs can indicate the general urbanization level differences. The endogeneity of population density has also been studied by previous research. Zhao and Kaestner [49] addressed the possible endogeneity of population density by using a two-step instrumental variable approach to investigate the effects of urban sprawl on obesity. Therefore, population density is chosen as endogenous variable in this paper. The population density distribution is illustrated in Figure 7.

**3.4. Indicator Variables for Endogenous Variable.** The endogenous variable could be influenced by various indicator variables. Road density reflects regional characteristics and has potential influence on the endogenous variable. Therefore, road length is chosen to be the indicator variable for the endogenous variable. The spatial distribution of road length is illustrated in Figure 8.

According to Figure 8, it is clear that urban TAZs have higher road density in city center. For TAZs in urban areas, they have relatively lower road density. This is understandable as urban areas tend to have more connected roads

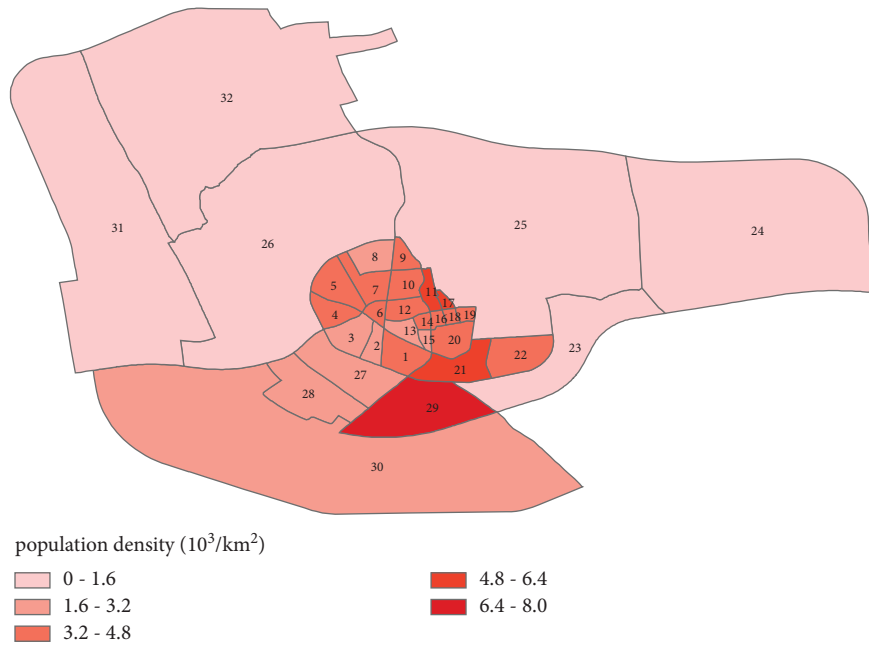


FIGURE 7: Spatial distributions of population density in Xuancheng, China.

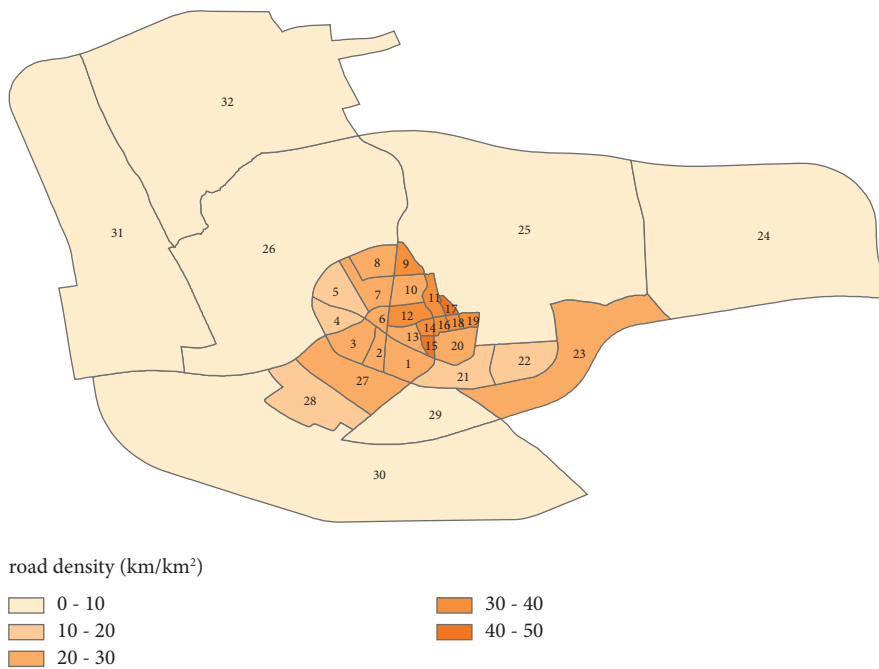


FIGURE 8: Spatial distributions of road density in Xuancheng, China.

for travel and economic activities, while suburban TAZs have much larger areas and fewer roads, which lead to lower road density.

Another important explanatory variable for endogenous variable is transit accessibility. Transit accessibility ensures necessary passenger travels and goods exchange, thus maintaining normal operation of the city. It has potential influence on the endogenous variable. In minor cities like Xuancheng, the major transit mode is bus. The bus stop

density represents the availability of bus transit in TAZs. Therefore, the bus stops density is chosen as explanatory variable for endogenous variable. The spatial distributions of bus stop density are illustrated in Figure 9.

As seen from Figure 9, TAZs in city center areas have higher bus stop density, while suburban TAZs have relatively lower bus stop density. In practice, city center areas have more frequent transit travel demand, which leads to higher bus stop density in downtown areas, while suburban areas



FIGURE 9: Spatial distributions of bus stop density in Xuancheng, China.

have fewer bus stops and larger areas, thus having lower bus stop density.

**3.5. Travel Time.** The travel impedance between TAZ pairs is also defined to indicate spatial correlations between TAZ pairs. By considering comprehensive travel impedance between TAZ pairs, this paper chooses average travel time in the A.M. to indicate travel impedance between TAZ pairs. Travel time is derived by taking average of each vehicle's travel time between TAZ pairs, which is obtained based on trajectory from AVI database.

The average travel time between TAZ  $i$  and TAZ  $j$  is calculated as follows:

$$T_{ij} = \frac{\sum_k^{N_{ij}} t_{ij}^k}{N_{ij}}, \quad (4)$$

where  $T_{ij}$  is the average travel time between TAZ  $i$  and TAZ  $j$ ;  $t_{ij}^k$  is the travel time for vehicle  $k$  between TAZ  $i$  and TAZ  $j$ ; and  $N_{ij}$  is the number of vehicles traveling from origin  $i$  to destination  $j$ .

The statistics of the above explanatory variables are summarized in Table 2. There is no multicollinearity between explanatory variables, as the maximum variance inflation factor (VIF) is 4.44. The travel time has 1,024 observations since it is defined for each TAZ pair. All other explanatory variables are defined at the zonal level and have 32 observations.

## 4. Methodology

**4.1. Model Establishment.** A spatial autoregressive binary probit model with endogenous weight matrix (SARBP-EWM) is established in this paper to investigate the travel flow differences between peak hours. The SARBP-EWM

model was first proposed by Zhou et al. [50] to account for the endogeneity problem in spatial econometrics. By applying the SARBP-EWM model, the time subscript  $t$  in the original model is set to 1 since this paper only focuses on average daily travel flow. Besides, the geographic fixed effect  $M$  and state fixed effect  $N$  are held as constants. Therefore, the SARBP-EWM model is reduced as follows:

$$\mathbf{Y}^* = \rho \mathbf{W}(\mathbf{Z}) \mathbf{Y}^* + \mathbf{X} \boldsymbol{\beta} + \mathbf{E}, \quad (5)$$

where dependent variable  $\mathbf{Y}^* = (y_1^*, y_2^*, \dots, y_n^*)'$  is a  $n \times 1$  vector of binary variable indicating the travel flow differences ( $Y_{\text{weekday}}$  and  $Y_{\text{weekend}}$ ) between morning peak and evening peak in TAZs.  $y_i^* = 1$  if travel flow fluctuation is larger than threshold for TAZ  $i$  according to (3);  $y_i^* = 0$  otherwise.

For other variables,  $\mathbf{X}$  is  $n \times k_1$  matrix of explanatory variables.  $\mathbf{W}$  is an  $n \times n$  spatial weight matrix that represents the relative weights between all TAZ pairs. The weights are defined by the endogenous variable  $\mathbf{Z}$ .  $\boldsymbol{\beta}$  is  $k_1 \times 1$  the vector of coefficients for corresponding explanatory variables. As stated above, the explanatory variables include number of offices (office), number of supermarkets (supermarket), number of middle schools (mid\_school), number of clinics (clinic), number of inns (inn), and number of sports centers (sports).  $\mathbf{E} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)'$  is an  $n \times 1$  error term vector of (5).

**4.2. Definition of Weight Matrix.** The weight element  $w_{ij}$  stands for the element on the  $i^{\text{th}}$  row and  $j^{\text{th}}$  column of weight matrix  $\mathbf{W}$ , which indicates the relative weight between TAZ  $i$  and  $j$ . As noted in (5),  $\mathbf{W}$  is defined as the function of the endogenous variable  $\mathbf{Z}$  so that

$$w_{ij} = F(\mathbf{z}_i, \mathbf{z}_j), \quad i, j \in n, \quad (6)$$

TABLE 2: Descriptive statistics of variables.

Variable	Name	No. of observations	Description	Mean	Std	Min	Max
Dependent variable	Y_weekday	32	Binary variable indicating travel flow difference between morning and evening peaks on weekday (1 if flow fluctuation is larger than threshold; 0 otherwise)	0.500	0.508	0	1
	Y_weekend	32	Binary variable indicating travel flow difference between morning and evening peaks on weekend (1 if flow fluctuation is larger than threshold; 0 otherwise)	0.594	0.499	0	1
Explanatory variables	Office	32	Number of offices in TAZs	1.531	2.155	0	7
	Supermarket	32	Number of supermarkets in TAZs	19.69	17.33	0	84
	mid_school	32	Number of middle schools in TAZs	0.531	0.671	0	2
	Clinic	32	Number of clinics in TAZs	1.938	2.341	0	10
	Inn	32	Number of inns in TAZs	2.625	4.263	0	19
	Sports	32	Number of sports centers in TAZs	0.656	1.181	0	6
Endogenous variable	pop_den	32	Population density in TAZs (10 <sup>3</sup> /km <sup>2</sup> )	3.159	1.634	0.266	7.035
Explanatory variables for endogenous variable	road_density	32	Road density in TAZs (km/km <sup>2</sup> )	22.98	11.82	3.382	47.05
	bus_den	32	Number of bus stop density in TAZs (/km <sup>2</sup> )	7.623	5.983	0	23.45
Travel time	t_time	1,024	Travel time between centroids of origin and destination TAZs (m)	17.27	4.903	6.761	36.06

where  $F(\bullet)$  is the function that can take various forms such as a generalized Euclidean distance or a gravity model that considers key socioeconomic factors.  $z_i = (z_{i1}, z_{i2}, \dots, z_{ip})$  is a  $1 \times p$  vector of TAZ  $i$ 's demographic or economic characteristics. Common ways to define weight element include using geographic distances, social network [51], length of road [52], or various economic quantities among variables

[53]. For example, Lee and Yu [54] defined  $w_{ij}$  by combining the demographic and economic distance with geographical distance.

In this paper, we define the weight element by replacing the geographical distance with travel time. So, the weight element  $w_{ij}$  in this paper is defined as follows:

$$w_{ij} = T_{ij}^{-\gamma_0} |z_{i1} - z_{j1}|^{-\gamma_1} |z_{i2} - z_{j2}|^{-\gamma_2} \dots |z_{ip} - z_{jp}|^{-\gamma_p}, \quad i, j \in n. \quad (7)$$

Since only population density (pop\_den) is chosen as the endogenous variable,  $p = 1$ . The weight element  $w_{ij}$  is only defined by travel time  $T_{ij}$  and differences of population density  $z_i - z_j$  between TAZ  $i$  and  $j$ . And  $\gamma = (\gamma_0, \gamma_1, \gamma_2, \dots, \gamma_p)$  are  $(p+1) \times 1$  estimable parameters which correspond to travel time and population density differences, respectively.  $T_{ij}$  is the average travel time between TAZ  $i$  and TAZ  $j$  in (4).

**4.3. Definition of Endogenous Variable.** The SARBP-EWM model has a distinct feature to incorporate an endogenous variable and its influential factors through entry equation as follows:

$$\mathbf{Z} = \tilde{\mathbf{X}}\tilde{\boldsymbol{\beta}} + \Delta. \quad (8)$$

The endogenous variable  $\mathbf{Z} = (z'_1, z'_2, \dots, z'_n)'$  is an  $n \times p$  matrix consisting of  $p$  endogenous variables representing the interdependency structure among TAZs. Also,  $z_i = (z_{i1}, z_{i2}, \dots, z_{ip})$  is a  $1 \times p$  vector indicating TAZ  $i$ 's interdependency structure. The selection of variable  $\mathbf{Z}$  needs

to have major influence on the interdependency structure among TAZs while having correlation with dependent variable  $\mathbf{Y}^*$ . As stated above, population density (pop\_den) is chosen as the endogenous variable, so  $p = 1$ .

The endogenous variable  $\mathbf{Z}$  in (8) is influenced by its explanatory variable  $\tilde{\mathbf{X}}$  which is an  $n \times k_2$  matrix consisting of  $k_2$  explanatory variables.  $\tilde{\boldsymbol{\beta}}$  is  $k_2 \times 1$  the vector of parameters which corresponds to  $\tilde{\mathbf{X}}$ . For this paper, the explanatory variables include road length (road\_length) and bus stop density (bus\_den).

$\Delta = (\delta'_1, \delta'_2, \dots, \delta'_n)'$  is an  $n \times p$  error term vector of (8). The endogeneity occurs when the error terms  $\mathbf{E}$  and  $\Delta$  are correlated. Assuming  $\varepsilon_{it}$  and  $\delta_{it}$  meet i.i.d conditions and follow joint normal distribution across all  $i$ 's and  $j$ 's with mean 0 and variance-covariance matrix  $V$ , the jointly normal distribution can be written as follows:

$$(\varepsilon_i, \delta_i) \sim N_{p+1} \left( 0, \begin{pmatrix} \sigma_\varepsilon^2 & \sigma'_{\varepsilon\delta} \\ \sigma_{\varepsilon\delta} & V_\delta \end{pmatrix} \right) = N_{p+1}(0, V), \quad (9)$$

or

$$\varepsilon_i | \delta_i \sim N(\sigma'_{\varepsilon\delta} V_{\delta}^{-1} \delta_i, \sigma_{\varepsilon}^2 - \sigma'_{\varepsilon\delta} V_{\delta}^{-1} \sigma_{\varepsilon\delta}). \quad (10)$$

**4.4. Model Estimation.** The model estimation uses a Bayesian Markov Chain Monte Carlo (MCMC) method to estimate parameters. The MCMC method decomposes a set of parameters of a complex model into a sequence of sublayers, addressing one parameter each time [55]. It updates one parameter each time and uses it for the next sampling process.

Next, we substitute (10) into equation (5) and to derive the likelihood function which can be written as

$$(\mathbf{I} - \rho \mathbf{W}) \mathbf{Y}^* = \mathbf{X} \boldsymbol{\beta} + (\mathbf{Z} - \tilde{\mathbf{X}} \tilde{\boldsymbol{\beta}}) \boldsymbol{\eta} + \boldsymbol{\xi}, \quad (11)$$

where  $\boldsymbol{\eta} = \mathbf{V}_{\delta}^{-1} \sigma_{\varepsilon\delta}$  is a  $p \times 1$  vector. Error term  $\boldsymbol{\xi} \sim N(\mathbf{0}, \sigma_{\xi}^2 \mathbf{I}_n)$  is independent of  $\mathbf{Y}^*$  and follows the normal distribution with mean  $\mathbf{0}$  and variance  $\sigma_{\xi}^2 \mathbf{I}$  where  $\sigma_{\xi}^2 = \sigma_{\varepsilon}^2 - \sigma'_{\varepsilon\delta} V_{\delta}^{-1} \sigma_{\varepsilon\delta}$ . To write all parameters for both independent variable and endogenous variable together as  $\boldsymbol{\beta}^* = (\boldsymbol{\beta}, \tilde{\boldsymbol{\beta}})$ , the conditional likelihood function can be generally written as

$$f(\mathbf{Y}^* | \mathbf{Z}, \rho, \boldsymbol{\gamma}, \boldsymbol{\beta}^*, \mathbf{V}) \propto \left( (\sigma_{\xi}^2)^{-n/2} \times |\mathbf{S}_n(\rho, \boldsymbol{\gamma})| \times \exp\left(-\frac{\mathbf{H}_n' \mathbf{H}_n}{2\sigma_{\xi}^2}\right) \right), \quad (12)$$

where  $\mathbf{S}_n(\rho, \boldsymbol{\gamma}) = \mathbf{I}_n - \rho \mathbf{W}(\boldsymbol{\gamma})$  and  $\mathbf{H}_n = (\mathbf{I} - \rho \mathbf{W}) \mathbf{Y}^* - \mathbf{X} \boldsymbol{\beta} - (\mathbf{Z} - \tilde{\mathbf{X}} \tilde{\boldsymbol{\beta}}) \boldsymbol{\eta}$ .

Based on the likelihood function above, the posterior distributions for  $\rho, \boldsymbol{\gamma}, \boldsymbol{\beta}^*, \mathbf{V}$  are derived and presented in Table 3.

More detailed derivation process of these parameters' posterior distribution functions can be found in Zhou's work [56].

The conditional posterior of the  $i^{\text{th}}$  element of dependent variable  $y_i^*$  actually follows a truncated normal distribution below:

$$f(y_i^* | \mathbf{Z}, \mathbf{Y}^*, y_{-i}^*, \rho, \boldsymbol{\gamma}, \boldsymbol{\beta}^*, \mathbf{V}) \sim \begin{cases} \mathbf{N}_i(\mathbf{S}_n^{-1} \mathbf{H}_n, \sigma_{\xi}^2 (\mathbf{S}_n' \mathbf{S}_n)^{-1}), \text{ left truncated at } 0, \text{ if } y_i = 1, \\ \mathbf{N}_i(\mathbf{S}_n^{-1} \mathbf{H}_n, \sigma_{\xi}^2 (\mathbf{S}_n' \mathbf{S}_n)^{-1}), \text{ right truncated at } 0, \text{ if } y_i = 0, \end{cases} \quad (13)$$

where  $\mathbf{N}_i(\mathbf{S}_n^{-1} \mathbf{H}_n, \sigma_{\xi}^2 (\mathbf{S}_n' \mathbf{S}_n)^{-1})$  refers to the  $i^{\text{th}}$  element of the multivariate normal distribution  $\mathbf{N}(\mathbf{S}_n^{-1} \mathbf{H}_n, \sigma_{\xi}^2 (\mathbf{S}_n' \mathbf{S}_n)^{-1})$ .

## 5. Results

To estimate parameter values, the SARBP-EWM model was run using MATLAB for 10,000 iterations.  $\boldsymbol{\beta}^*$  and  $\mathbf{V}$  are sampled using the MCMC method.  $\rho$  and  $\boldsymbol{\gamma}$  are sampled using the MH method since they do not follow standard distributions. The first 6000 iterations are set as "burnt-in" to allow parameters to gradually converge to true values. Therefore, the last 4000 iterations are used to obtain coefficient estimates. In addition to the SARBP-EWM model, a binary probit model is also run to make comparison. The posterior distributions of all parameters of SARBP-EWM model on weekday are illustrated in Figure 10. The estimation results for both weekday and weekend and binary probit model are summarized in Table 4.

The model is well estimated in general with most parameters following their expected posterior distributions in Table 3. According to Figure 10, posterior distributions for facility variables  $\boldsymbol{\beta}$  and transit accessibility variables  $\tilde{\boldsymbol{\beta}}$  suggest that they follow normal distributions, which meets with their expected posterior distributions. The elements in covariance matrix  $\mathbf{V}$  reveal inverse Wishart distributions, which are consistent with their derived distributions in Table 3.

However, spatial autocorrelation coefficient  $\rho$  reveals two major peaks with mean value around 0.10 according to Figure 10(a). Also, indicator variable coefficients  $\gamma_0, \gamma_1$  in Figures 10(b) and 10(c) show that travel time (travel\_time) coefficient  $\gamma_0$  has mean value around 0.08 while population density (pop\_den) coefficient  $\gamma_1$  has one major peak and few local minimums and maximums, which leads to mean value around 0.27. The distributions of  $\rho, \boldsymbol{\gamma}$  follow non-standard distributions as their derived posterior distributions in Table 3 do not follow standard distributions. The distribution of coefficients  $\tilde{\boldsymbol{\beta}}$  and  $\tilde{\boldsymbol{\beta}}$  is illustrated in Figures 10(d)–10(m), which follow standard multivariate normal distributions. Also, elements of covariance matrix  $\mathbf{V}$  are presented in Figures 10(n) and 10(o). The impacts of variables are presented in Table 4 and are discussed in detail as follows.

**5.1. Spatial Effects.** The spatial autocorrelation coefficient is estimated to be significantly positive in both weekday and weekend using the SARBP-EWM model, which suggests that there is strong spatial autocorrelation among TAZs regarding travel flow differences between morning peak and evening peak. The significant travel flow differences in one TAZ tend to spread to surrounding TAZs, and such spatial spillover effect cannot be neglected. Such phenomenon is

TABLE 3: Parameters' posterior distributions.

Variable	Prior distribution	Posterior distribution	Sampling method
$\rho, \gamma$	$\rho \sim U(-1, 1)$ $\rho_0 = 0.1$ $\gamma \sim \mathbf{N}_{p+1}(\gamma_0, \mathbf{R}_0)$ $\gamma_0: (p+1) \times 1$ vector of 0 $\mathbf{R}_0: \mathbf{I}_{p+1} \times 10^{12}$	$f(\mathbf{Y}^*   \mathbf{Z}, \rho, \gamma, \beta, \mathbf{V}) \propto$ $ \mathbf{S}_n(\rho, \gamma)  \times \exp(-\mathbf{H}_n' \mathbf{H}_n / 2\sigma_\xi^2)$	Metropolis–Hastings (MH)
$\beta^*$	$\beta^* \sim \mathbf{N}_{k_1+pk_2}(\beta_0, \mathbf{B}_0)$ $\beta_0: (k_1 + pk_2) \times 1$ vector of 0 $\mathbf{B}_0: \mathbf{I}_{k_1+pk_2} \times 10^6$	$\mathbf{N}_{k_1+pk_2}(\mathbf{T}_\beta, \mathbf{A}_\beta^{-1})$ where $\mathbf{T}_\beta = \mathbf{A}_\beta^{-1} (\mathbf{B}_0^{-1} \beta_0 + \sum_{i=1}^n \mathbf{x}_{i \beta}' \mathbf{V}^{-1} \mathbf{y}_{i \beta}^*)$ $\mathbf{A}_\beta^{-1} = (\mathbf{B}_0^{-1} + \sum_{i=1}^n \mathbf{x}_{i \beta}' \mathbf{V}^{-1} \mathbf{x}_{i \beta}^*)^{-1}$	Multivariate normal distribution
$\mathbf{V}$	$\mathbf{V} \sim \mathbf{W}_{p+1}^{-1}(\Psi, \nu)$ $\Psi: \mathbf{I}_{p+1}$ $\nu: 10$	$\mathbf{W}_{p+1}^{-1}(\Psi + \sum_{i=1}^n h_i h_i', \nu + n)$ where $h_i$ is the $i^{\text{th}}$ element of $\mathbf{H}_n$	Inverse-Wishart distribution

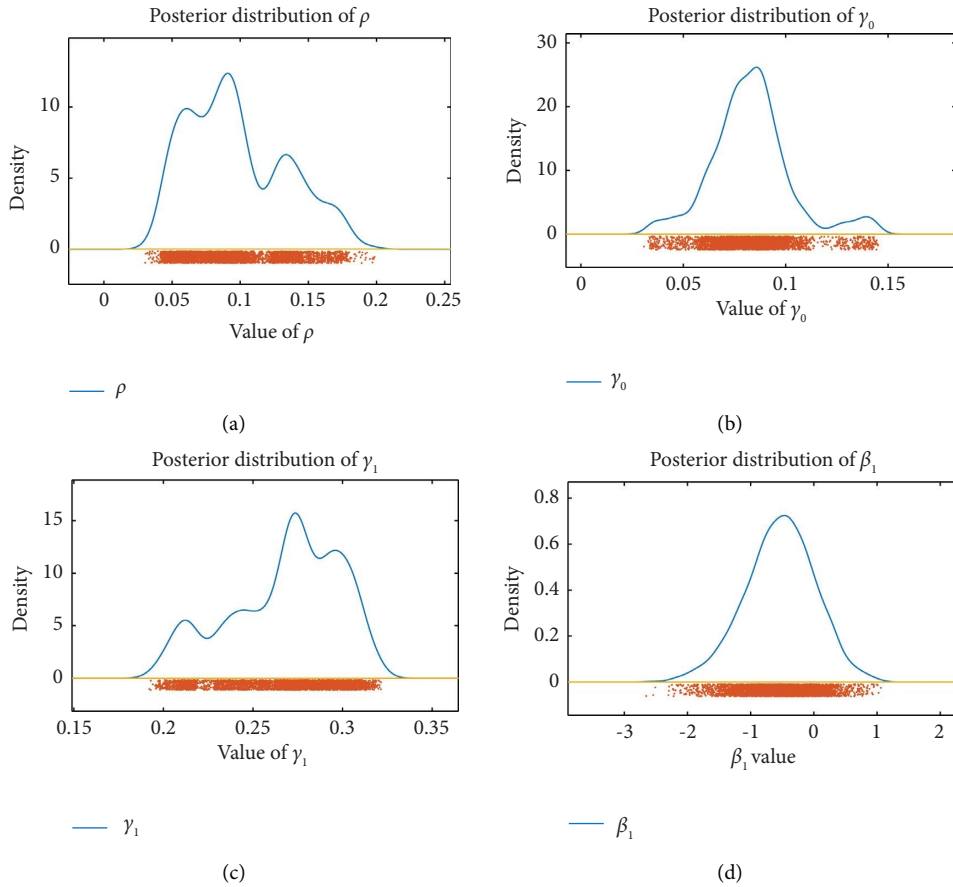


FIGURE 10: Continued.

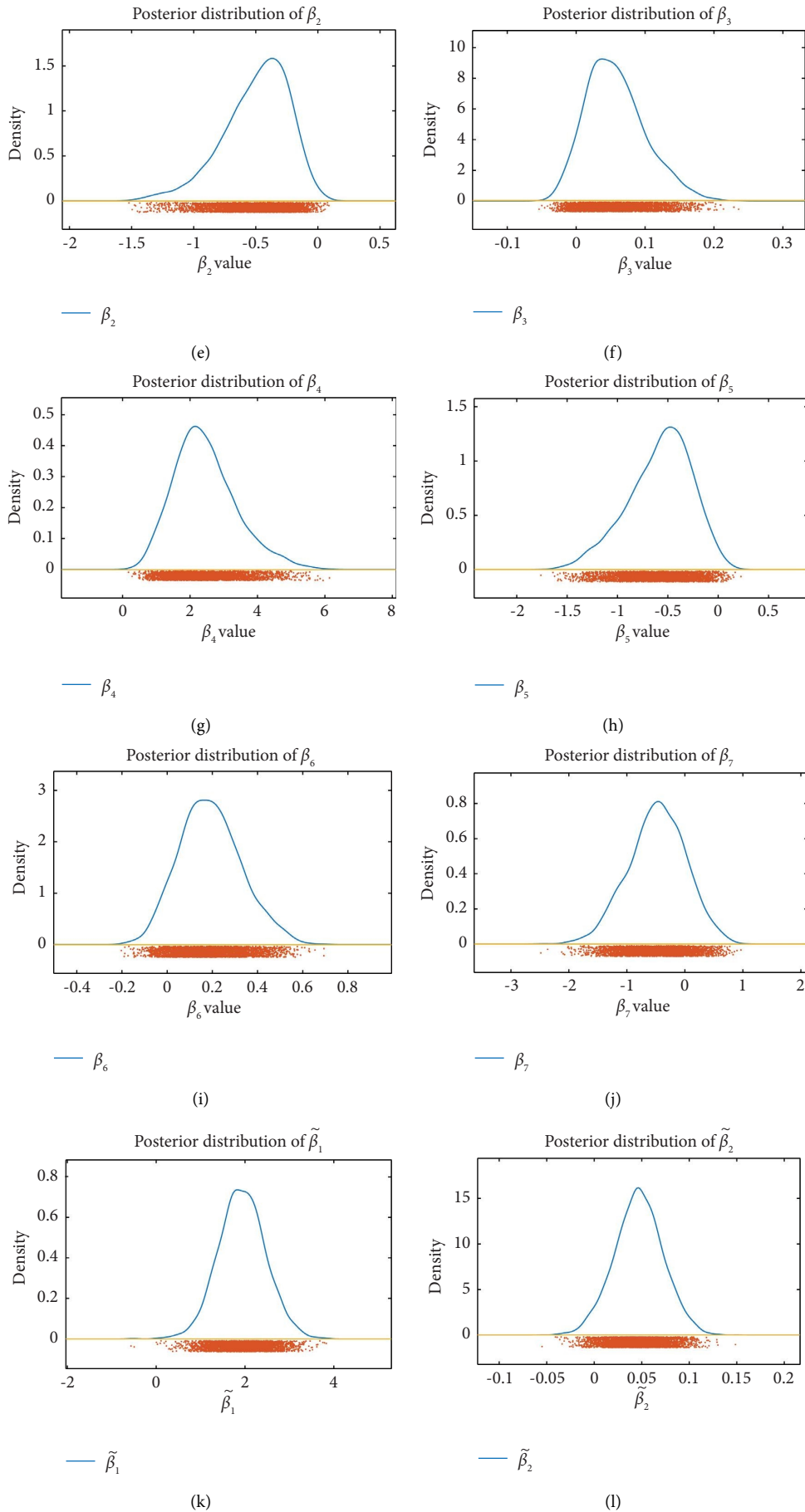


FIGURE 10: Continued.



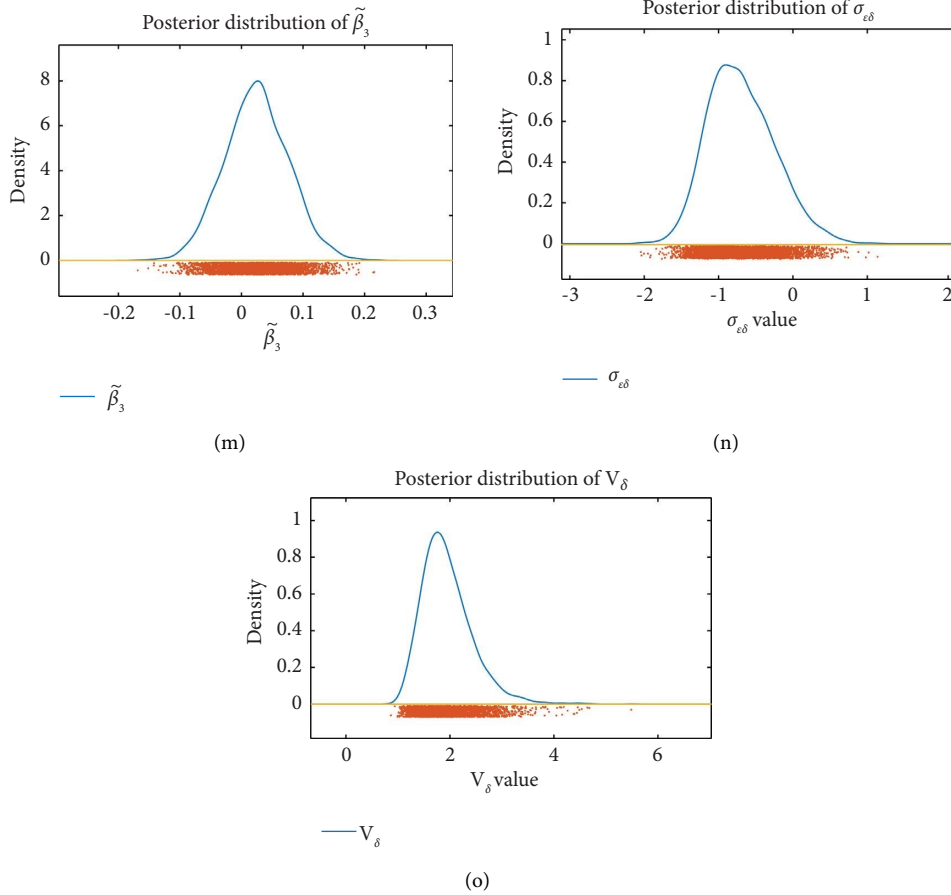


FIGURE 10: Posterior distributions of parameters. (a) Posterior distribution of  $\rho$ . (b) Posterior distribution of  $\gamma_0$ . (c) Posterior distribution of  $\gamma_1$ . (d) Posterior distribution of  $\beta_1$ . (e) Posterior distribution of  $\beta_2$ . (f) Posterior distribution of  $\beta_3$ . (g) Posterior distribution of  $\beta_4$ . (h) Posterior distribution of  $\beta_5$ . (i) Posterior distribution of  $\beta_6$ . (j) Posterior distribution of  $\beta_7$ . (k) Posterior distribution of  $\tilde{\beta}_1$ . (l) Posterior distribution of  $\tilde{\beta}_2$ . (m) Posterior distribution of  $\tilde{\beta}_3$ . (n) Posterior distribution of  $\sigma_{\epsilon\delta}$ . (o) Posterior distribution of  $V_\delta$ .

TABLE 4: Estimation results of the SARBP-EWM model and binary probit model for weekday and weekend.

Explanatory variables	SARBP-EWM (weekday)		SARBP-EWM (weekend)		Binary probit model (weekday)		Binary probit model (weekend)		
	Coef.	t value	Coef.	t value	Coef.	t value	Coef.	t value	
Spatial autocorrelation coefficient	$\rho$	0.10	2.67	0.09	2.98**	NA	NA	NA	NA
Coefficient for endogenous variable ( $\gamma$ )	travel_time	0.08	4.25**	0.01	0.36	NA	NA	NA	NA
	pop_den	0.27	8.77**	0.35	7.62**	NA	NA	NA	NA
Coefficients for facility variables ( $\beta$ )	Constant	-0.55	-0.98	0.03	0.05	-0.37	-0.74	0.16	0.34
	Office	-0.50	-1.89*	-0.39	-2.05**	-0.42	-1.72*	-0.36	-1.70**
	Supermarket	0.06	1.31	0.05	1.41	0.05	1.49	0.05	1.47
	mid_school	2.44	2.63**	1.42	2.26**	1.88	2.38**	1.20	1.81*
	Clinic	-0.59	-1.85*	-0.44	-1.92*	-0.51	-2.15**	-0.50	-2.29**
	Inn	0.19	1.38	0.12	1.14	0.12	1.28	0.13	1.05
Coefficients for transportation variables ( $\tilde{\beta}$ )	Sports	-0.49	-0.99	-0.17	-0.44	-0.65	-1.57	-0.33	-0.74
	Constant	1.92	3.60**	1.81	3.50**	NA	NA	NA	NA
	road_density	0.05	1.81**	0.05	2.05**	NA	NA	NA	NA
Covariance matrix ( $\mathbf{V}$ )	bus_den	0.02	0.46	0.02	0.53	NA	NA	NA	NA
	$\sigma_\epsilon^2$	1.25 (fixed)	NA	1.25 (fixed)	NA	NA	NA	NA	NA
	$\sigma_{\epsilon\delta}$	-0.70	-1.58	-0.98	-2.92**	NA	NA	NA	NA
	$V_\delta$	1.94	3.96**	1.93	4.08**	NA	NA	NA	NA

\*\*0.05 level; \*0.1 level; NA: not applicable.

also confirmed in previous research studies where the change of travel flow differences in a TAZ will have positive impact on the travel flow differences of other TAZs [57].

**5.2. Indicator Variables and Endogeneity.** The SARBP-EWM model contains indicator variables which define the endogenous weight matrix, which is jointly defined by travel time (*travel\_time*) and population density (*pop\_den*) according to (7). Population density is estimated to be significantly positive in both weekday and weekend, indicating that it is a good indicator variable for defining the endogenous weight matrix. In fact, population density is an important indicator of urbanization level of an area. The population density difference measures the urbanization level difference between OD pairs, which defines the relative weight between OD pairs. Therefore, population density is selected to represent the interdependency structure between OD pairs.

The endogeneity among TAZs is measured by the term  $\sigma_{\epsilon\delta}$  in covariance matrix  $\mathbf{V}$ , which represents the errors and their covariance between (5) and (8). The covariance  $\sigma_{\epsilon\delta}$  is estimated to be insignificant with mean value of  $-0.70$  on weekday and being significant with mean value of  $-0.98$  at weekend, indicating that there is significant endogeneity among TAZs at weekend. Such endogeneity is neglected in traditional spatial models, which may lead to biased estimation and misleading conclusions. It suggests that there is negative convolution between travel flow differences and population density. With the increase of population density, travel flow differences tend to decrease as there is balanced inflow and outflow travel flow with the regional development. However, when the TAZ is at initial development stage with low population density, there could be great travel flow differences between morning peak and evening peak.

For the error term  $V_{\delta}$  in (8), it shows positive and significant correlation with endogenous variable population density in both weekday and weekend, indicating that the unobserved variable in (8) is strongly correlated with endogenous variable. This may be because current dataset does not provide other variable information such as car ownership or per capital income, which may help explain the endogenous variable. However, it can be understood that this is by far the best result we can get based on current dataset.

**5.3. Endogenous Weight Matrix.** The SARBP-EWM model has the distinct feature to allow an endogenous weight matrix, which indicates the relative weights of influence between TAZ pairs. For this paper, such influence means the travel flow differences and the surrounding affected TAZs, similar to the “peak spreading” phenomenon. To further illustrate and compare the relative weights among TAZ pairs, weight elements of all TAZ pairs on weekday and weekend are categorized and presented in Figures 11 and 12, respectively.

According to Figure 11, most TAZs have relatively low spatial weights in weekday as indicated by 987 light red blocks with spatial weights  $0.001$ – $0.060$ , taking up 96.4% of all TAZ pairs, while TAZ pairs with high spatial weights

consist of 23 TAZ pairs in  $0.041$ – $0.060$  range (red color) and 14 TAZ pairs in  $0.080$  and above range (dark red color), together taking up 3.6% of all TAZ pairs. Figure 11 suggests that TAZs with high spatial weights tend to cluster, which means that they tend to spread their travel differences to surrounding TAZs, similar to the “peak spreading” phenomenon. Intuitively, such flow imbalance comes from surround TAZs and cannot be instantly diminished. Weight element distribution in weekend shows similar patterns. Besides, the clustered blocks are mostly identical to those in weekday, indicating that these blocks easily reveal significant travel differences, whether being weekday or weekend. To be specific, 984 TAZs have relatively low spatial weights, taking up 96.1% of all TAZ pairs. Also, TAZ pairs with high spatial weights consist of 24 TAZ pairs in  $0.041$ – $0.060$  range (red color) and 16 TAZ pairs in  $0.080$  and above range (dark red color), together taking up 3.9% of all TAZ pairs. TAZ pairs with high spatial weights also tend to cluster in weekend. Previous research studies have confirmed such finding where the change of travel flow differences in an TAZ will have positive impact on the travel flow differences of other TAZs [22, 57, 58].

To further illustrate the endogenous weight matrix, spatial distributions of weight elements for typical TAZs are presented in Figure 13. The results indicate that TAZ 7 has high spatial weights with TAZ 1, TAZ 4, TAZ 5, TAZ 14, and TAZ 22 in weekday. It suggests that urban TAZ has strong impacts on travel flow differences of surrounding and relevant TAZs. Similarly, TAZ 8 reveals high spatial weights with TAZ 2 and TAZ 3 in the city south direction and TAZ 13 and TAZ 15 in city southeast direction in weekend, indicating that urban TAZ has strong impacts on travel flow differences of surrounding and relevant TAZs. Both surrounding and relevant TAZs with high spatial weights tend to cluster in both weekday and weekend.

**5.4. Effects of Explanatory Variables.** Facility variables like the number of offices, supermarkets, middle schools, clinics, inns, and gymnasiums in TAZs may have major impacts on the travel flow differences between morning peak and evening peak. Estimation results using the SARBP-EWM model show that facility variables have significant impacts on travel flow differences on both weekday and weekend in general. Impacts on travel flow fluctuations are consistent on both weekend and weekday. The specific impact of each facility variable is described as follows:

- (a) The number of offices (*office*) in TAZs is estimated to be significantly negative with dependent variable on both weekday and weekend. It suggests that a higher number of offices in origin TAZs are associated with lower travel flow differences between peak hours. However, in previous studies, offices are usually thought as major sources of travel demand [39, 43, 59]. However, since this paper focuses on the travel flow differences, offices tend to generate large but relatively equal commuting travel demand as people need to go to work in the morning peak and leave office in evening peak. The equally large

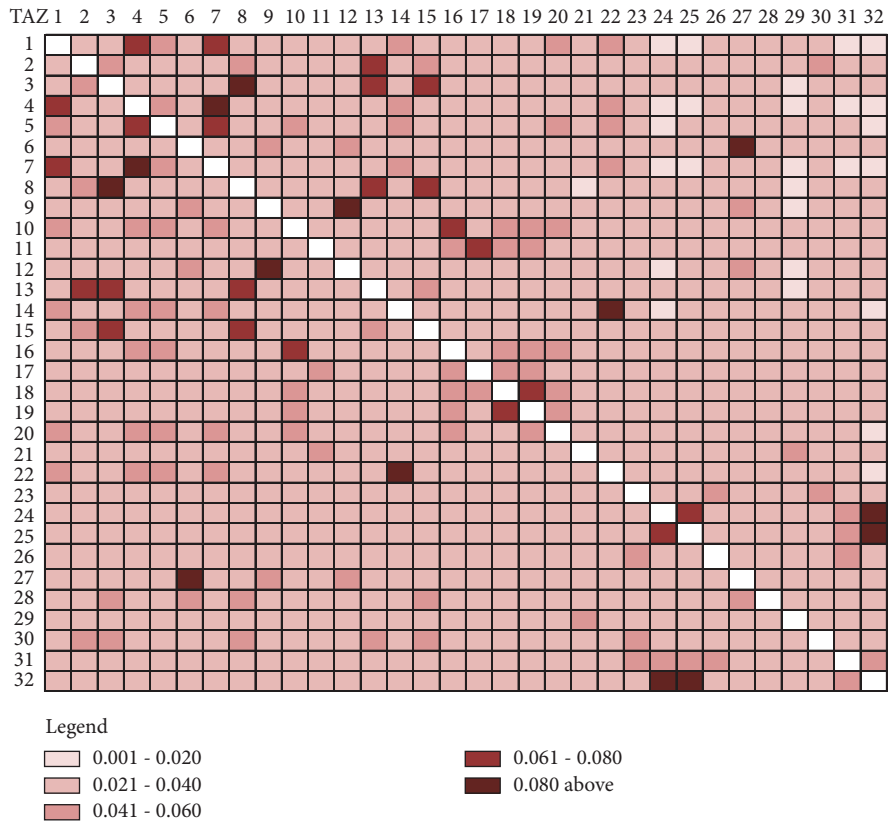


FIGURE 11: Weight element categorization in weekday.

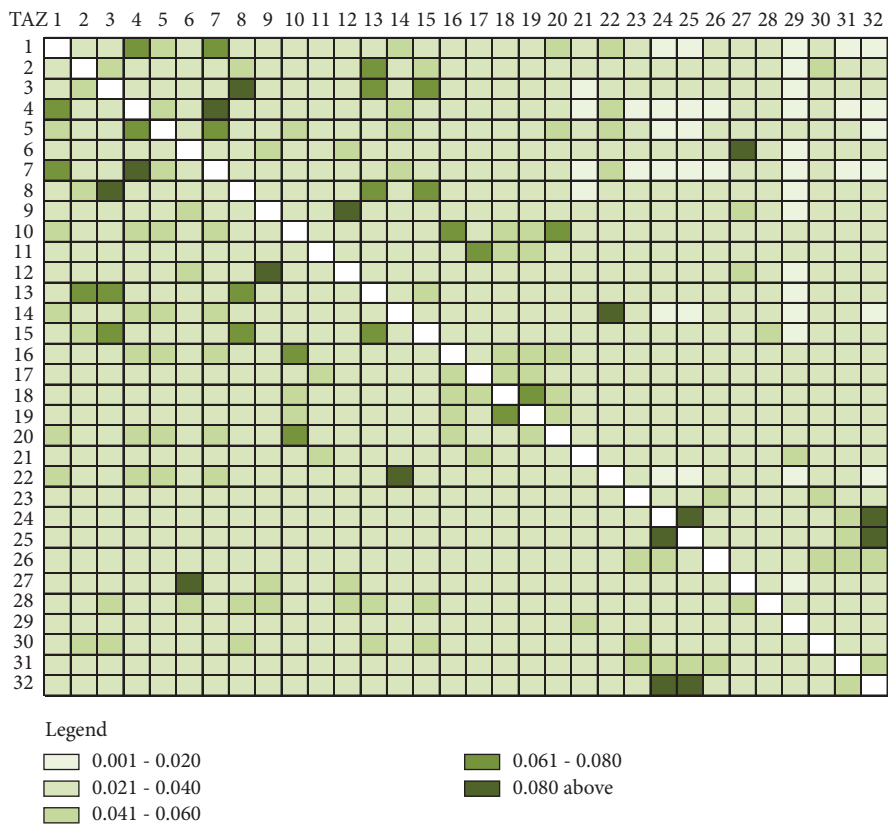


FIGURE 12: Weight element categorization in weekend.

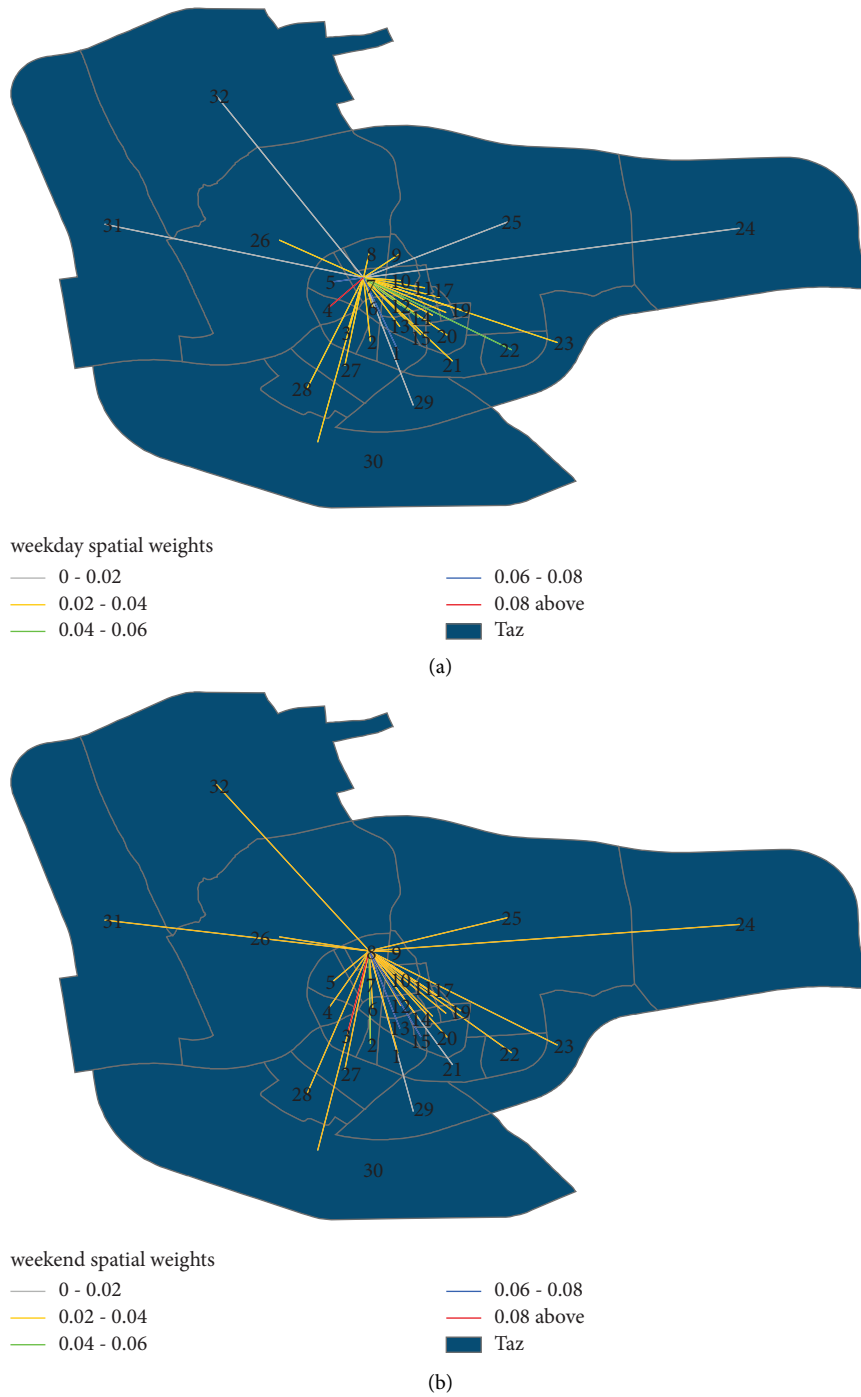


FIGURE 13: Spatial distributions of weight elements for typical TAZs in weekday and weekend. (a) Spatial distribution of weight elements for TAZ 7 in weekday. (b) Spatial distribution of weight elements for TAZ 8 in weekend.

commuting travel flow actually does not contribute much to the travel flow differences between peak hours. Besides, in urbanized TAZs with more office buildings, the commuting travel flow takes large portion of total travel flow, thus stabilizing the fluctuation level of travel flow between peak hours. Such result is consistent with result in the binary probit model.

(b) The number of supermarkets (supermarket) reveals insignificant relation with travel flow differences on both weekday and weekend. In fact, supermarkets in China are usually located at urbanized area which serve local residents with snacks, food, and life necessities. Local residents tend to buy food and drinks in the morning after getting up and pick some life necessities after coming back in the evening.

Therefore, supermarket may generate equal travel flows in the morning and in the evening, thus contributing little to the travel differences between morning peak and evening peak.

- (c) The number of middle schools (*mid\_school*) in TAZs shows significant and positive relation with dependent variable in both SARBP-EWM model and binary probit model across weekday and weekend, which indicates that the change of number of middle schools in TAZs has major impact on the travel flow differences between morning peak and evening peak. Intuitively, schools are major travel flow sources as students need to go to school in the morning and leave school in the evening during weekdays [46]. However, the number of schools also has major impacts on travel flow differences at weekend because students take “weekend” classes at school, which is common in less urbanized cities like Xuancheng as students choose to take “weekend” classes to enhance their studies to get better exam grades. Therefore, the number of schools in TAZs has major impacts on the travel flow differences on both weekday and weekends.
- (d) The number of clinics (*clinic*) shows significant and negative relation with dependent variable in both weekday and weekend. It means that a higher number of clinics are associated with lower travel flow differences between peak hours. In fact, hospital’s impact on travel demand varies with time and space according to previous studies [20, 47]. Results indicate that clinics in Xuancheng have negative impacts on travel flow differences. It may be because people are busy working or relaxing during daytime while they are more flexible to visit clinics during evening peak hours.
- (e) The number of inns (*inn*) reveals insignificant relation with dependent variable in both weekday and weekend. It indicates that the number of inns has little influence on travel flow differences between peak hours. It is possible that for small cities like Xuancheng in China, inns are not the major travel flow sources that would not affect travel flow differences.
- (f) The number of sports centers (*sports*) shows insignificant relation with dependent variable in both weekday and weekend. It suggests that the number of sports center has no major impact on travel flow differences between peak hours. In practice, sports centers are not major sources of travel flows. People go to sports centers occasionally for sports activities. Therefore, sports centers have little influence on the travel flow differences.

*5.5. Effects of Indicator Variables for Endogenous Variable.* The endogenous variable population density in (8) is also affected by explanatory variables like road density and bus stop density. The road density (*road\_density*) in TAZs

reveals significant and positive correlations with population density in both weekday and weekend using the SARBP-EWM model, suggesting that higher road density is associated with higher population density. Usually, areas with higher urbanization level tend to have both higher population density and higher road density, as urbanized areas would attract more residents and have more roads to accommodate frequent travel demands.

Another important factor that may affect population density is the transit accessibility. The results show that bus stop density (*bus\_den*) is insignificantly associated with population density. In practice, bus stop density in small cities like Xuancheng may not have major impacts on population density.

In general, the SARBP-EWM model successfully identified significant spatial effects and endogeneity. It reveals the hidden influential factors which are not discovered by the traditional binary probit model and quantifies their impacts on travel flow differences in both weekday and weekend. Facility variables like number of offices, middle schools, and clinics have major impacts on travel flow differences. Besides, road density shows significant relation with endogenous variable population density.

## 6. Discussion of Policy Implications

Xuancheng has released the “Xuancheng 14th Five-Year Plan for National Economic and Social Development and the Long-Range Objectives through the Year 2035 (X145Plan)” on May, 2021 [60], which sets detailed targets for city’s future development. X145Plan proposes to promote industrial development, enhance public service infrastructures, and build more advanced transportation systems. These policies have potential impacts on travel flow differences based on results above. Therefore, it is necessary to analyze how these policies would affect the travel flow differences in different TAZs.

According to X145Plan, Xuancheng plans to promote industrial platform construction from several aspects. First, the X145Plan proposes to enhance Xuancheng economic and technical development zone (TAZ 28) by focusing on renewable energy, equipment manufacturing, food and drug, electronic information, etc. Second, the X145Plan aims to develop Xuancheng Modern Service Industrial Park (TAZ 24) by promoting digital economy, logistics, agricultural products, etc. [60]. These industrial promotion policies would lead to the emergence of many office buildings, which would decrease local travel flow differences based on results in Table 4. Therefore, industrial promotion policies should be advocated as they enhance economy while decreasing travel flow differences. However, decreasing travel flow differences do not mean decreasing absolute traffic volume. City planners and policy makers still need to be cautious not to add too much traffic to the city.

In addition, Xuancheng aims to build multilevel consumption platforms to facilitate consumption goods circulation according to X145Plan. The X145Plan proposed to build new featured streets including Huchengfang (TAZ 14) and Doufuxiang (TAZ 16), enhance business circles

TABLE 5: Impacts of X145Plan’s major policies on typical TAZs.

Directly affected TAZs and impacts on travel flow differences from major policies					Indirectly affected TAZs from directly affected TAZs on weekday (red blocks)	Indirectly affected TAZs from directly affected TAZs at weekend (red blocks)
Building offices	Building consumption platforms	Building schools	Building medical facilities	Building sports facilities		
NA	TAZ 1 (?)	NA	NA	NA	TAZ 4(?) TAZ 5(?) TAZ 7(?) TAZ 14(?) TAZ 20(?) TAZ 22(?)	TAZ 4(?) TAZ 5(?) TAZ 7(?) TAZ 14(?) TAZ 20(?) TAZ 22(?)
NA	TAZ 2 (?)	NA	NA	NA	TAZ 3(?) TAZ 8(?) TAZ 13(?) TAZ 15(?) TAZ 30(?)	TAZ 3(?) TAZ 8(?) TAZ 13(?) TAZ 15(?) TAZ 30(?)
NA	NA	TAZ 11 (+)	NA	NA	TAZ 16(+) TAZ 17(+) TAZ 18(+) TAZ 19(+)	TAZ 16(+) TAZ 17(+) TAZ 18(+) TAZ 19(+)
NA	NA	NA	TAZ 12 (-)	NA	TAZ 6(-) TAZ 9(-) TAZ 27(-)	TAZ 6(-) TAZ 9(-) TAZ 27(-)
NA	TAZ 14 (?)	NA	NA	NA	TAZ 1(?) TAZ 4(?) TAZ 5(?) TAZ 7(?) TAZ 22(?)	TAZ 1(?) TAZ 4(?) TAZ 5(?) TAZ 7(?) TAZ 22(?)
NA	TAZ 16 (?)	NA	NA	NA	TAZ 4(?) TAZ 5(?) TAZ 7(?) TAZ 10(?) TAZ 18(?) TAZ 19(?) TAZ 20(?)	TAZ 4(?) TAZ 5(?) TAZ 10(?) TAZ 18(?) TAZ 19(?) TAZ 20(?)
NA	NA	TAZ 20 (+)	NA	NA	TAZ 1(+) TAZ 4(+) TAZ 5(+) TAZ 7(+) TAZ 10(+) TAZ 16(+) TAZ 19(+)	TAZ 1(+) TAZ 4(+) TAZ 5(+) TAZ 7(+) TAZ 10(+) TAZ 16(+) TAZ 19(+)
TAZ 24 (-)	TAZ 24 (?)	NA	NA	NA	TAZ 25(?) TAZ 31(?) TAZ 32(?)	TAZ 25(?) TAZ 31(?) TAZ 32(?)
NA	TAZ 27 (?)	NA	NA	NA	TAZ 6(?) TAZ 9(?) TAZ 12(?)	TAZ 6(?) TAZ 9(?) TAZ 12(?)
TAZ 28 (-)	NA	NA	NA	NA	TAZ 3(-) TAZ 6(-) TAZ 8(-) TAZ 15(-) TAZ 27(-)	TAZ 3(-) TAZ 6(-) TAZ 8(-) TAZ 15(-) TAZ 27(-)
NA	NA	NA	TAZ 30 (-)	TAZ 30 (?)	TAZ 2(?) TAZ 3(?) TAZ 8(?) TAZ 13(?) TAZ 15(?) TAZ 23(?)	TAZ 2(?) TAZ 3(?) TAZ 8(?) TAZ 13(?) TAZ 15(?) TAZ 23(?)

TABLE 5: Continued.

Directly affected TAZs and impacts on travel flow differences from major policies					Indirectly affected TAZs from directly affected TAZs on weekday (red blocks)	Indirectly affected TAZs from directly affected TAZs at weekend (red blocks)
Building offices	Building consumption platforms	Building schools	Building medical facilities	Building sports facilities		
NA	NA	TAZ 31 (+)	NA	NA	TAZ 23(+)	TAZ 23(+)
					TAZ 24(+)	TAZ 24(+)
					TAZ 25(+)	TAZ 25(+)
					TAZ 26(+)	TAZ 26(+)
					TAZ 32(+)	TAZ 32(+)

“+” means increased travel flow differences; “-” means decreased travel flow differences; “NA” means not available; “?” means ambiguous impacts on travel flow differences.

including Xinglong International Plaza (TAZ 1) and Wanda Plaza (TAZ 2), and promote consumption goods retail through platforms such as Maisha shopping mall (TAZ 27) and Xuancheng agricultural product wholesale market (TAZ 24) [60]. These new or enhanced consumption platforms such as business circles and shopping malls would have insignificant impact on the travel flow differences based on estimation results in Table 4. However, this does not mean city planners and decision makers can build shopping malls at will. They still need to be cautious when planning building new shopping centers to avoid adding too much travel flow to the transportation network.

Regarding public service infrastructures like educational, medical, and sports facilities, X145Plan proposed to strengthen public service infrastructures by building or extending educational, medical, and sports facilities including University for the elderly (TAZ 20), Caijinhu Middle School (TAZ 31), No. 3 Middle School extension (TAZ 11), People’s Hospital Extension Phase II (TAZ 12), City Center for Disease Control (TAZ 30), City Sports Center Stadium (TAZ 30), and so on [60]. Middle schools have positive impacts on travel flow differences according to results above. However, hospitals have negative impacts on travel flow differences. Also, sports facilities have little impact on travel flow differences. Therefore, the total impacts of these public service infrastructures are ambiguous. For city planners and decision makers, they need to carefully evaluate possible impacts on travel flow differences under different policies.

Therefore, to evaluate impacts on travel flow differences of major policies of X145Plan, the typical directly and indirectly affected TAZs based on estimation results are summarized in Table 5.

The directly affected TAZs are obtained from X145Plan, and impacts are derived based on estimation results. In addition to these directly affected TAZs, some TAZs would be indirectly affected spatially according to the estimated endogenous weight matrix. For example, TAZ 11 is directly affected by building No. 3 Middle School extension and would spatially affect travel flow differences on nearby TAZ 16, TAZ 17, TAZ 18, and TAZ 19 on both weekday and weekend. However, for some areas such as TAZ 24, building offices and consumption platform would cause negative and ambiguous impacts separately. Therefore, the total impacts on spatially affected TAZs are ambiguous. The above results

provide detailed policy implications based on X145Plan, which would facilitate urban planning and policy making.

For transportation-related policies, X145Plan proposes to build more advanced transportation network by “connecting each county with roads” [60]. Therefore, more roads would be built in TAZs and between TAZs. According to estimation results, an increase in road density is associated with higher population density, which indicates that TAZs with higher road density tend to have higher population density. In fact, for minor cities like Xuancheng, TAZs with higher road density are mainly urban TAZs with more developed facilities and services, which tend to attract more residents. This provides important policy implications for urban planners to build more roads to help boost economy and attract population. This caters to the traditional Chinese saying that “Building the road is the first step to become rich.”

## 7. Conclusions

This paper investigates the travel flow differences between morning and evening peaks based on AVI data in Xuancheng, China. A spatial model with endogenous weight matrix is established to investigate influential factors considering the endogeneity issue. The results confirmed strong spatial effects and endogeneity among TAZs. As for influential factors, number of offices and number of clinics are found to have negative relation with the travel flow differences on both weekday and weekend, while the number of middle schools shows strong positive relation with dependent variable. In addition, the spatial weight matrices for both weekday and weekend are estimated and compared. Spatial weights in weekday tend to cluster with lower weights while they are randomly distributed with higher weights in weekend. Based on the results, policy recommendations are evaluated and proposed.

The main contributions of this paper are summarized as follows. (a) This paper utilizes AVI data to investigate the travel flow differences between peak hours on both weekday and weekend. The results confirm strong spatial correlations among TAZ pairs on both weekday and weekend. (b) Endogeneity among TAZs is considered and quantified. This paper is among the few studies considering endogeneity by applying a spatial model with an endogenous weight matrix.

The endogenous weight matrix is successfully estimated. (c) This paper quantifies impacts of key factors on travel flow differences. The results suggest that facility variables such as number of offices, supermarkets, middle schools, and clinics have major influence on the travel flow differences. (d) This paper provides major policy implications based on X145Plan. Policies on enhancing industrial development, building more schools, and improving medical services are evaluated, and affected TAZs are identified.

Future work mainly includes two aspects. First, the travel flow fluctuation threshold can be determined using more systematic approaches. This paper chooses fluctuation threshold based on its approximate cumulative distribution. It can be further elaborated to be able to judge significant travel flow differences while tolerating random variations. Second, more efficient algorithms could be developed to obtain more accurate OD flow based on AVI data. This paper reconstructs vehicle trajectories based on AVI data from intersections, which still leave much uncertainty on vehicle trajectories. More efficient trajectory reconstruction techniques could be adopted to recover precise vehicle trajectories to provide more accurate data for analysis.

## Data Availability

The data are not publicly available due to privacy or ethical restrictions but are available from the corresponding author on reasonable request.

## Disclosure

This study was performed at the Sun Yat-sen University (SYSU) Research Center of ITS in the context of the collaboration with the Joint Research and Development Laboratory of Smart Policing in Xuancheng Public Security.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Authors' Contributions

Jin-Yong Chen and Yiwei Zhou were responsible for study conception and design. Zhaocheng He, Jin-Yong Chen, and Yiwei Zhou collected the data. Yiwei Zhou, Jieshuang Dong, and Linglin Ni analyzed and interpreted the results. Yiwei Zhou was responsible for draft manuscript preparation. All authors reviewed the results and approved the final version of the manuscript.

## Acknowledgments

This paper was financially supported by the Ministry of Education of China Humanities and Social Sciences Youth Fund Project (22YJC790189) and Shanghai University Young Teachers Cultivation and Support Project.

## References

- [1] S. Jia, H. Peng, and S. Liu, "Urban traffic state estimation considering resident travel characteristics and road network capacity," *Journal of Transportation Systems Engineering and Information Technology*, vol. 11, no. 5, pp. 81–85, 2011.
- [2] M. Fosgerau and D. Fukuda, "Valuing travel time variability: characteristics of the travel time distribution on an urban road," *Transportation Research Part C: Emerging Technologies*, vol. 24, pp. 83–101, 2012.
- [3] L. Ni, X. C. Wang, and X. M. Chen, "A spatial econometric model for travel flow analysis and real-world applications with massive mobile phone data," *Transportation Research Part C: Emerging Technologies*, vol. 86, pp. 510–526, 2018.
- [4] B. Hu, S. Zhang, Y. Ding, M. Zhang, X. Dong, and H. Sun, "Research on the coupling degree of regional taxi demand and social development from the perspective of job–housing travels," *Physica A: Statistical Mechanics and Its Applications*, vol. 564, Article ID 125493, 2021.
- [5] S. L. Handy and K. J. Clifton, "Local shopping as a strategy for reducing automobile travel," *Transportation*, vol. 28, no. 4, pp. 317–346, 2001.
- [6] P. L. Mokhtarian and X. Cao, "Examining the impacts of residential self-selection on travel behavior: a focus on methodologies," *Transportation Research Part B: Methodological*, vol. 42, no. 3, pp. 204–228, 2008.
- [7] L. Daqing, J. Yinan, K. Rui, and S. Havlin, "Spatial correlation analysis of cascading failures: congestions and blackouts," *Scientific Reports*, vol. 4, no. 1, p. 5381, 2014.
- [8] K. F. Chu, A. Y. S. Lam, and V. O. K. Li, "Deep multi-scale convolutional LSTM network for travel demand and origin-destination predictions," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 8, pp. 3219–3232, 2020.
- [9] J. Bolland and D. Ashmore, "Traffic peak spreading in congested urban environments," in *Proceedings of the from 25th Australasian Transport Research Forum*, Canberra, Australia, May 2002.
- [10] W. Cheng, G. S. Gill, T. Sakrani, D. Ralls, and X. Jia, "Modeling the endogeneity of lane-mean speeds and lane-speed deviations using a Bayesian structural equations approach with spatial correlation," *Transportation Research Part A: Policy and Practice*, vol. 116, pp. 220–231, 2018.
- [11] F. Zhou, C. Li, Z. Huang, R. Xu, and W. Fan, "Fare incentive strategies for managing peak-hour congestion in urban rail transit networks," *Transportmetrica: Transportation Science*, vol. 18, pp. 166–187, 2020.
- [12] C. H. Tsai, C. Mulley, and G. T. Clifton, "Examining local interaction between public transport demand and land use characteristics using geographically weighted regression," in *Transportation Research Board 92nd Annual Meeting*, pp. 13–0317, TRB, Washington, DC, USA, 2013.
- [13] Y. Zhou and X. Wang, "Explore the relationship between online shopping and shopping trips: an analysis with the 2009 NHTS data," *Transportation Research Part A*, vol. 70, pp. 1–9, 2014.
- [14] Y. J. Wu, F. Chen, C. T. Lu, and S. Yang, "Urban traffic flow prediction using a spatio-temporal random effects model," *Journal of Intelligent Transportation Systems*, vol. 20, no. 3, pp. 282–293, 2016.
- [15] J. Zheng and H. X. Liu, "Estimating traffic volumes for signalized intersections using connected vehicle data," *Transportation Research Part C: Emerging Technologies*, vol. 79, pp. 347–362, 2017.
- [16] L. Li, X. Qu, J. Zhang, Y. Wang, and B. Ran, "Traffic speed prediction for intelligent transportation system based on a deep feature fusion model," *Journal of Intelligent Transportation Systems*, vol. 23, no. 6, pp. 605–616, 2019.



- [17] A. Samara, F. Rempe, and S. Gottlich, "A novel approach for vehicle travel time distribution: copula-based dependent discrete convolution," *Transportation Letters*, vol. 14, no. 7, pp. 740–751, 2021.
- [18] S. V. Kumar and L. Vanajakshi, "Short-term traffic flow prediction using seasonal ARIMA model with limited input data," *European Transport Research Review*, vol. 7, no. 3, p. 21, 2015.
- [19] S. Yang and S. Qian, "Understanding and predicting travel time with spatio-temporal features of network traffic flow, weather and incidents," *IEEE Intelligent Transportation Systems Magazine*, vol. 11, no. 3, pp. 12–28, 2019.
- [20] X. Shen, Y. Zhou, S. Jin, and D. Wang, "Spatiotemporal influence of land use and household properties on automobile travel demand," *Transportation Research Part D*, vol. 84, Article ID 102359, 2020.
- [21] C. R. Bhat and H. Zhao, "The spatial analysis of activity stop generation," *Transportation Research Part B: Methodological*, vol. 36, no. 6, pp. 557–575, 2002.
- [22] J. P. LeSage and R. K. Pace, "Spatial econometric modeling of origin-destination flows," *Journal of Regional Science*, vol. 48, no. 5, pp. 941–967, 2008.
- [23] J. P. LeSage and C. Thomas-Agnan, "Interpreting spatial econometric origin-destination flow models," *Journal of Regional Science*, vol. 55, no. 2, pp. 188–208, 2015.
- [24] K. Kerkman, K. Martens, and H. Meurs, "A multilevel spatial interaction model of transit flows incorporating spatial and network autocorrelation," *Journal of Transport Geography*, vol. 60, pp. 155–166, 2017.
- [25] N. Zheng and N. Geroliminis, "Area-based equitable pricing strategies for multimodal urban networks with heterogeneous users," *Transportation Research Part A: Policy and Practice*, vol. 136, pp. 357–374, 2020.
- [26] T. Schatzmann, G. Sarlas, and K. W. Axhausen, "Spatial modelling of origin-destination commuting flows in Switzerland," in *Proceedings of the Transportation Research Board 98th Annual Meeting*, Washington, DC, USA, May 2019.
- [27] C. A. Guevara, A. Tirachini, R. Hurtubia, and T. Dekker, "Correcting for endogeneity due to omitted crowding in public transport choice using the Multiple Indicator Solution (MIS) method," *Transportation Research Part A: Policy and Practice*, vol. 137, pp. 472–484, 2020.
- [28] T. E. Guerrero, C. A. Guevara, E. Cherchi, and J. D. D. Ortúzar, "Forecasting with strategic transport models corrected for endogeneity," *Transportmetrica: Transportation Science*, vol. 18, no. 3, pp. 708–735, 2021.
- [29] L. Alexander, S. Jiang, M. Murga, and M. C. González, "Origin-destination trips by purpose and time of day inferred from mobile phone data," *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 240–250, 2015.
- [30] R. J. Lee, I. N. Sener, and J. A. Mullins, "An evaluation of emerging data collection technologies for travel demand modeling: from research to practice," *Transportation Letters*, vol. 8, no. 4, pp. 181–193, 2016.
- [31] F. Luo, G. Cao, K. Mulligan, and X. Li, "Explore spatio-temporal and demographic characteristics of human mobility via Twitter: A case study of Chicago," *Applied Geography*, vol. 70, pp. 11–25, 2016.
- [32] F. Zhao, L. Fu, M. Zhong et al., "Development and validation of improved impedance functions for roads with mixed traffic using taxi GPS trajectory data and simulation," *Journal of Advanced Transportation*, vol. 2020, Article ID 7523423, 12 pages, 2020.
- [33] M. M. Ahmed and M. A. Abdel-Aty, "The viability of using automatic vehicle identification data for real-time crash prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 2, pp. 459–468, 2011.
- [34] X. Zhan, R. Li, and S. V. Ukkusuri, "Lane-based real-time queue length estimation using license plate recognition data," *Transportation Research Part C: Emerging Technologies*, vol. 57, pp. 85–102, 2015.
- [35] F. Zheng, J. Li, H. Van Zuylen, X. Liu, and H. Yang, "Urban travel time reliability at different traffic conditions," *Journal of Intelligent Transportation Systems*, vol. 22, no. 2, pp. 106–120, 2018.
- [36] M. Fekih, L. Bonnetain, A. Furno et al., "Potential of cellular signaling data for time-of-day estimation and spatial classification of travel demand: a large-scale comparative study with travel survey and land use data," *Transportation Letters*, vol. 14, no. 7, pp. 787–805, 2021.
- [37] M. Chen, G. Yu, P. Chen, and Y. Wang, "A copula-based approach for estimating the travel time reliability of urban arterial," *Transportation Research Part C: Emerging Technologies*, vol. 82, pp. 1–23, 2017.
- [38] Y. Zhao, X. Zhu, W. Guo, B. She, H. Yue, and M. Li, "Exploring the weekly travel patterns of private vehicles using automatic vehicle identification data: a case study of wuhan, China," *Sustainability*, vol. 11, no. 21, p. 6152, 2019.
- [39] T. Huang, Y. Ma, Z. T. Qin et al., "Origin-destination flow prediction with vehicle trajectory data and semi-supervised recurrent neural network," in *Proceedings of the 2019 IEEE International Conference on Big Data (Big Data)*, pp. 1450–1459, IEEE, Los Angeles, CA, USA, December 2019.
- [40] Y. Cao, K. Tang, J. Sun, and Y. Ji, "Day-to-day dynamic origin-destination flow estimation using connected vehicle trajectories and automatic vehicle identification data," *Transportation Research Part C: Emerging Technologies*, vol. 129, Article ID 103241, 2021.
- [41] A. N. Map, "Route navigation," 2021, <https://www.gaode.com/>.
- [42] Q. Sun, Y. Sun, L. Sun et al., "Research on traffic congestion characteristics of city business circles based on TPI data: the case of Qingdao, China," *Physica A: Statistical Mechanics and Its Applications*, vol. 534, Article ID 122214, 2019.
- [43] X. Liu, L. Gong, Y. Gong, and Y. Liu, "Revealing travel patterns and city structure with taxi trip data," *Journal of Transport Geography*, vol. 43, pp. 78–90, 2015.
- [44] J. T. Zamora and M. F. Alexis, "Comparison of trip attraction between a mall and a school in the City of Dagupan, Philippines," in *Proceedings of the 12th Eastern Asia Society for Transportation Studies Conference 2017*, pp. 85–96, Ho Chi Minh City, Vietnam, November 2017.
- [45] S. Meena, G. R. Patil, and A. Mondal, "Understanding mode choice decisions for shopping mall trips in metro cities of developing countries," *Transportation Research Part F*, vol. 46, pp. 133–146, 2019.
- [46] E. Ikeda, T. Stewart, N. Garrett et al., "Built environment associates of active school travel in New Zealand children and youth: a systematic meta-analysis using individual participant data," *Journal of Transport & Health*, vol. 9, pp. 117–131, 2018.
- [47] G. Cheng, X. Zeng, L. Duan et al., "Spatial difference analysis for accessibility to high level hospitals based on travel time in Shenzhen, China," *Habitat International*, vol. 53, pp. 485–494, 2016.
- [48] U. Schirpke, C. Meisch, T. Marsoner, and U. Tappeiner, "Revealing spatial and temporal patterns of outdoor

- recreation in the European Alps and their surroundings,” *Ecosystem Services*, vol. 31, pp. 336–350, 2018.
- [49] Z. Zhao and R. Kaestner, “Effects of urban sprawl on obesity,” *Journal of Health Economics*, vol. 29, no. 6, pp. 779–787, 2010.
- [50] Y. Zhou, X. Wang, and J. Holguín-Veras, “Discrete choice with spatial correlation: A spatial autoregressive binary probit model with endogenous weight matrix (SARBP-EWM),” *Transportation Research Part B*, vol. 94, pp. 440–455, 2016.
- [51] R. T. Leenders, “The specification of weight structures in network autocorrelation models of social influence,” SOM Research Reports, No. 02B09. University of Groningen, Groningen, Netherlands, 2002.
- [52] Y. Kamarianakis and P. Prastacos, “Space-time modeling of traffic flow,” *Computers & Geosciences*, vol. 31, no. 2, pp. 119–133, 2005.
- [53] C. Lam and P. C. Souza, “Estimation and selection of spatial weight matrix in a spatial lag model,” *Journal of Business & Economic Statistics*, vol. 38, no. 3, pp. 693–710, 2020.
- [54] L.-F. Lee and J. Yu, “QML estimation of spatial dynamic panel data models with time varying spatial weights matrices,” *Spatial Economic Analysis*, vol. 7, no. 1, pp. 31–74, 2012.
- [55] J. P. LeSage and R. K. Pace, *Introduction to Spatial Econometrics*, Chapman & Hall/CRC, Taylor & Francis Group, Boca Raton, FL, USA, 2009.
- [56] Y. Zhou, “Discrete choice modeling with interdependencies: a spatial binary probit model with endogenous weight matrix,” Retrieved from ProQuest Dissertations & Theses database, Doctoral dissertation, Rensselaer Polytechnic Institute, Troy, NY, USA, 2015.
- [57] M. S. Middela and G. Ramadurai, “Incorporating spatial interactions in zero-inflated negative binomial models for freight trip generation,” *Transportation*, vol. 12, pp. 1–22, 2020.
- [58] C. Thomas-Agnan and J. P. LeSage, “Spatial econometric OD-flow models,” in *Handbook of Regional Science*, M. Fischer and P. Nijkamp, Eds., Springer, Berlin, Germany, 2014.
- [59] J. Tang, F. Gao, F. Liu, W. Zhang, and Y. Qi, “Understanding spatio-temporal characteristics of urban travel demand based on the combination of GWR and GLM,” *Sustainability*, vol. 11, no. 19, pp. 5525–5619, 2019.
- [60] Xuancheng Municipal Government, “14th five-year plan for national economic and social development and the long-range Objectives through the year 2035 (X145Plan),” 2021, <http://www.xuancheng.gov.cn/OpennessContent/show/2137029.html>.