한국ITS학회논문지 J. Korea Inst. Intell. Transp. Syst.

공간 상관성을 고려한 서울시 택시통행의 영향요인 분석

Identifying Key Factors to Affect Taxi Travel Considering Spatial Dependence: A Case Study for Seoul

이 향 숙* · 김 지 윤** · 추 상 호*** · 장 진 영**** · 최 성 택*****

* 주저자 : 인천대학교 동북아물류대학원 부교수
** 공저자 : 홍익대학교 도시계획과 박사수료
*** 교신저자 : 홍익대학교 건설도시공학부 교수
**** 공저자 : 홍익대학교 도시계획과 연구교수
***** 공저자 : 조지아공과대학교 토목·환경공학과 박사 후 연구원

Hyangsook Lee* · Ji yoon Kim** · Sangho Choo*** · Jin young Jang**** · Sung taek Choi*****

* Graduate School of Logistics, Incheon National University
** Dept. of Urban Planning, Hongik University
*** Dept. of Urban Design & Planning, Hongik University
**** Dept. of Urban Planning, Hongik University
***** School of Civil and Environmental Eng.. Georgia Institute of Technology, Atlanta, USA
† Corresponding author : Sangho Choo, shchoo@hongik.ac.kr

Vol.18 No.5(2019)
October, 2019
pp.64~78

pISSN 1738-0774
eISSN 2384-1729
https://doi.org/10.12815/kits.2019.18.5.64

Received 6 September 2019
Revised 24 September 2019
Accepted 10 October 2019

요 약

본 논문은 공간 상관성을 고려하여 서울의 택시통행에 영향을 미치는 요인을 분석한 것으로 택시의 GPS 자료를 이용하였다. 먼저 이들 자료를 이용하여 택시통행의 평균 통행시간, 평균 통행거리, 시공간적 분포 등의 통행특성을 분석하였다. 분석결과, 상당수의 택시통행이 평일 오전 첨두(8-9시)와 심야(0-1시) 시간대에 집중하는 것으로 나타났으며, 평균 통행거리와 통행시간은 각각 5.9km와 13분으로 나타났다. 이는 택시가 주로 단거리 통행수단이나 심야에 대중교통의 대체수단으로 활용되고 있음을 보여주고 있다. 다음으로 Moran's I 검정을 통해 교통존 기반의 택시통행들이 공간적으로 상관성이 있음을 규명하였다. 따라서 이를 고려한 택시통행에 대한 공간회귀모형(공간시차모형과 공간오차모형)을 추정하였으며, 인구사회 변수(가구수, 고령자수, 여성인구비, 자동차대수 등), 교통서비스 변수(지하철역수, 버스 정류장수 등), 토지이용 변수(인구밀도, 고용밀도, 주거면적 등) 등을 독립변수로 고려하였다. 모형분석 결과 이들 변수들이 통계적으로 유의하게 택시통행에 영향을 미치는 것으로 나타났다.

핵심어 : 택시, GPS, Moran's I 검증, 공간 상관성, 공간회귀모형

ABSTRACT

This paper explores key factors affecting taxi travel using global positioning system(GPS) data in Seoul, Korea, considering spatial dependence. We first analyzed the travel characteristics of taxis such as average travel time, average travel distance, and spatial distribution of taxi trips according to the time of the day and the day of the week. As a result, it is found that the most taxi trips were generated during the morning peak time (8 a.m. to 9 a.m.) and after the midnight (until 1 a.m.) on weekdays. The average travel distance and travel time for taxi trips were 5.9 km and 13 minutes, respectively. This implies that taxis are mainly used for short-distance travel and as an alternative to public transit after midnight in a large city. In addition, we identified that taxi trips were spatially correlated at the traffic analysis zone(TAZ) level through the Moran's I test. Thus, spatial regression models (spatial-lagged and spatial-error models) for taxi trips were developed, accounting for socio-demographics (such as the number of households, the number of elderly people, female ratio to the total population, and the number of vehicles), transportation services (such as the number of subway stations and bus stops), and land-use characteristics (such as population density, employment density, and residential areas) as explanatory variables. The model results indicate that these variables are significantly associated with taxi trips.

Key words : Taxi, Global Positioning System(GPS) data, Moran's I test, Spatial dependence, Spatial regression models

Ⅰ. Introduction

Taxi has been traditionally evaluated as a useful quasi-public transit mode in metropolitan cities, occasionally providing a higher level of service than other public transit modes(Kattan et al., 2010). Evidently, the mode share of taxis accounts for a major portion of public transit in densely populated cities; that is, taxis play a non-negligible role in the public transit system. In the case of New York City, for example, the number of for-hire vehicle registrations, including taxis, green cabs, black cars and private cab companies, has rapidly increased, reaching 72,000 in 2015(NYC Department of Transportation, 2016).

Seoul is ranked as one of the world's most populated cities, having more than ten million population with well-connected public transit system. In 2013, (1) bus and subway, so-called as major public transit, accounted for 65.6% of the entire transportation, while taxi constituted only 6.9%; (2) among 70,000 registered taxis in Seoul almost 49,000 taxis were driver-owned vehicles (the rest of them were owned by companies), and (3) the number of taxis per one thousand people in Seoul was 6.7, which is the highest number than those in other major developed cities such as Tokyo (6.4), Beijing (3.4), London (3.3), New York (1.6) and Paris (1.3). Meanwhile, the taxi industry has been in a difficult situation due to emerging competitors such as ride-hailing service (e.g., Uber, Lyft) and high-quality public transit services, resulting in decreased demand for the original taxi services, together with a skeptical perspective that an oversupply of cabs may lead to increases of zero-occupied trips, consequently causing urban problems of traffic congestion and energy consumption(Qian and Ukkusuri, 2015).

Recently, taxis have been equipped with Global Positioning System(GPS) devices, recording vehicle's travel information including travel time, locations where they pick up and drop off passengers, travel speed, and travel distance. The GPS data is very useful when simply observing travel characteristics of taxis (e.g., route choice) or investigating spatial and temporal characteristics of the passenger ridership in more depth. Sizable studies have focused on taxi data extracted from GPS devices. There are three case studies, for example, using GPS data in New York. Zhan et al.(2015) estimated hourly link travel times using taxi origin/destination(O/D) trip data in New York City which is extracted from GPS devices. Hochmair(2016) found that taxi trips were positively associated with the number of subway stations, but negatively related to the number of bus stops, using taxi GPS data in New York. Wang and Ross(2017) explored the spatial relationship between the distribution of taxi trips(O/D) and locations of transit stations using taxi GPS data in New York, finding a complementary relationship between taxis and transit. Similarly, Veloso et al.(2011) analyzed taxi-GPS traces in Lisbon, Portugal to visualize the spatiotemporal variation of taxi services. Employing an exploratory analysis, the authors demonstrated relationships between pick-up and drop-off locations, concluding that taxi service locations were closely related to public transportation terminals. Zheng et al.(2011) addressed a few problems in existing urban planning approaches using the taxi GPS trajectories in Beijing, China. They analyzed the distribution of region pairs with salient traffic problems, stratified by the type of day(work days or non-work days), presenting the linking structure and correlation among the regions. Using empirical taxi GPS trace data in San Francisco, Hoque et al.(2012) conducted a comprehensive analysis, covering instantaneous velocity profile, spatiotemporal distribution, connectivity of vehicle communications, clustering, hot spot, trip duration, empty cruise interval, and so on, focusing on mobility of urban taxi caps. Nicholas(2012) conducted descriptive, spatial and statistical analyses with taxi trip data in Arlington County to investigate travel characteristics of taxis (e.g., speed, distance and frequency of taxi trips, occupancy, and fare distribution), as well as spatial movement patterns such as taxi trip routes, spatial distribution, and O/D pairs. Shi et al.(2008) proposed a GPS and Geographical Information System(GIS) integrated system for urban traffic flow analysis. They had chosen Urban GIS-Transport data to construct GIS maps in Shanghai, China, and then collected real-time GPS data in order to implement the location amendment. The empirical experiments synthesizing such data proved the efficiency of the new system. Tang et al.(2015) also constructed a taxi trip O/D matrix using taxi GPS data in Harbin, China to develop the trip distribution model based on the entropy maximizing method.

Some researchers have explored a connection between travel characteristics of the taxi mode and social/land-use traits. Qi et al.(2011), for instance, measured the relationships between the get-on/off characteristics of taxi users and social functions of city regions using taxi GPS panel data in Hangzhou, China. The authors found that the volume of taxi passengers' riding and alighting in a region partly explained the social activity dynamics in such areas. Another study by Liu et al.(2012) analyzed temporal variations of both pick-ups and drop-offs and their association with land use characteristics using the taxi GPS data in Shanghai. They categorized the study area into six traffic sub-areas based on an indicator showing how much the numbers of pick-ups and drop-offs are balanced and corresponding distinctive temporal patterns, land use types, and land use intensity. Kim(2015) explored key factors influencing taxi travel in Seoul using the taxi GPS data in Seoul. He developed regression models for taxi departure trips on a weekday, incorporating socio-economic traits, land use characteristics, and public transit use variables. The model results showed that these variables significantly affected taxi travel. Yang et al.(2018) explored the correlation among taxi demand, land use pattern, and accessibility to other modes, using GPS data in Washington, DC. The authors found that the demand for taxis was significantly related to urban form characteristics (e.g., residential density and employment density). In classifying the city of Beijing, China, using point-of-interest data and taxi GPS trajectory data, Wang et al.(2018) took spatial semantics and interactions into account, thereby dividing the city into three types of urban functional regions: commercial, residential, and industrial districts. However, most previous studies have adopted descriptive approaches focusing on taxi movement characteristics, and the limited number of studies attempted to find what kind of specific traits affect taxi travel.

Given the background above, this paper identifies key factors affecting taxi travel using GPS data in Seoul, contemplating spatial correlation which has not been considered in the literature. We first explore taxi travel characteristics (average travel time, average travel distance, and spatial distribution of taxi trips) and conduct spatial dependence tests to examine whether taxi trips are spatially correlated at the traffic analysis zone(TAZ) level. We then develop two types of spatial regression models (spatial lagged and spatial error models) for taxi trips, incorporating socio-demographic traits, transportation services and land use characteristics as explanatory variables.

The remainder of the present study is as follows. Section 2 is concerned with GPS data profile of taxis in Seoul. Section 3, afterwards, presents spatial-temporal characteristics of taxi travel. Section 4 then moves on to the estimation results of the spatial regression models, providing key findings and implications. Lastly, Section 5 summarizes the research including a discussion about future research.

Ⅱ. Data Description

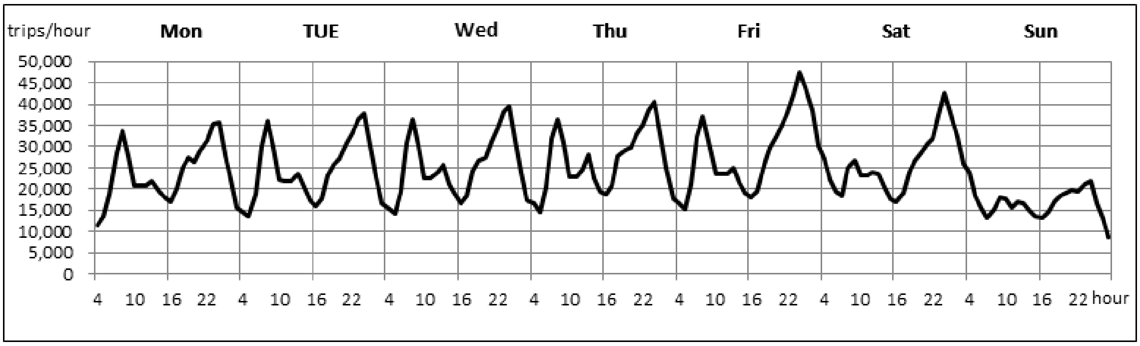
1. Data Collection

In 2013, the Seoul government imposed a new law stating that all company-owned taxis have to install GPS devices (a part of digital tachograph equipment). GPS data from the devices includes departure(pick-up) and arrival(drop-off) time, travel distance, fare, and GPS coordinates of origin and destination for each taxi (passenger) trip. These data can be used to analyze travel characteristics of taxis including taxi driving patterns, and temporal and spatial distribution. Given that background above, this study used GPS data provided by Korea Smart Card Corporation－a credit card payment system provider collecting and managing GPS data in Seoul(The Seoul Government, 2013).

Due to data availability, we collected weekly GPS data, particularly location information for departure and arrival trips, ranging from March 17 to March 23, 2014, which was extracted from company-owned taxis, due to a practical limitation (installing a GPS device was not mandatory for driver-owned taxis). It might be controversial, but we assure that our data is reliable since nearly a half of the taxis were still included, which is ample size to investigate taxi travel characteristics. The initial number of total taxi trips (or transactions) was about 4.88 millions. In the data cleaning process, we excluded 2.4% of the total trips from the original dataset due to missing values or errors (e.g., time, code, logical or payment errors). Consequently, the final dataset includes 4.76 million taxi trips. The daily time period covers from 4 a.m. to 3:59 a.m. on the following day (a company-owned taxi is shared with two drivers, and the night shift generally occurs around 4 a.m.).

2. General Characteristics

We identified the trip characteristics of taxis in Seoul; it includes trip distribution, travel distance, and travel time with respect to the time of the day and the day of the week (weekday, Saturday, Sunday). The average number of daily taxi trips was 710 thousand trips (34.8 trips per taxi) on weekdays, 719 thousand trips (35.2 trips per taxi) on Saturday, and 487 thousand trips (26.4 trips per taxi) on Sunday. Fridays show the largest average number of daily taxi trips during weekdays, with 777 thousand trips (37.6 trips per taxi) on such days. In analyzing the taxi departure trip distribution by hour, a repetitive up and down cycle was observed (see Fig. 1), presenting two distinct peaks regularly (the morning peak (8 to 9 a.m.) and midnight peak (midnight to 1 a.m.)). This is perhaps because (1) many travelers chose taxis in the morning to reach their office on time, and (2) most public transit services do not run from midnight to early morning. We can confirm that night peaks are more remarkable than morning peaks (especially, exceeding 45 thousand trips/hour were made on Friday). On the contrary, weekend taxi travel showed a different pattern as expected. On Saturday the amount of trips increased after midnight (0 to 1 a.m.), showing that many people traveled using taxis at night, while it showed relatively low numbers in the morning largely due to absence of commute trips. The number of trips on Sunday did not even meet the volume recorded in the daytime on weekdays and Saturday. The data do not show any peak patterns during the day (it showed a small peak in the evening period, but it is not still distinct from other days).



<Fig. 1> Temporal distribution of taxi departure trips by the day of the week

The average taxi travel distance was 5.9 km on weekdays and Saturday, and 5.8 km on Sunday, implying that taxis are relatively used for short distance trips. The trips less than 5 km and 10 km accounted for 60.3% and 82.6%, respectively. The travel distance pattern by the time of the day seems to be similar across seven days. The average travel distance by the time of the day did not demonstrate a significant variation during the daytime (ranging from 4 to 5 km) and tended to be longer from night to dawn (the average trip distance was 7.2 km between 9 p.m. and 6 a.m., showing a remarkable difference from the daytime pattern). During the midnight peak (0 to 1 a.m.), the average travel distance was the longest (8.9 km, on average). We speculate that other alternative modes were not available at this period (i.e., transfer from/to a different mode is limited), thereby it increased the average taxi travel distance.

The average travel time on weekdays, Saturday, and Sunday were 13.0, 12.7, and 11.0 minutes, respectively. There was no clear difference between weekdays and Saturday, but the average trip duration on Sunday was relatively short perhaps because of lower traffic congestion. The portion of the travel times less than 10 minutes accounted for almost 72% of total trips on weekdays and Saturday, and 83% on Sunday. In addition, morning and night peak periods demonstrate slightly longer travel times on weekdays, implying that traffic congestion negatively affects travel time regardless of how long riders travel.

Ⅲ. Spatial-temporal Characteristics of Taxi Travel

In this section, we analyze spatial distribution of taxi travel and the number of departure and arrival trips by the day of the week (weekdays, Saturday, Sunday, and a specific time period) at the administrative district level (referred to as the TAZ level in Seoul).

1. Spatial Distribution

On weekdays, the areas with the largest number of taxi departures were Gangnam (about 12,900-17,100 trips/day), Yeouido (about 12,100-13,800 trips/day) and Jongno (about 6,600-11,700 trips/day), where a lot of office buildings are densely located (see Fig. 2). We can easily find that Gangnam and Yeouido are ranked on the top, but districts ranked 3rd seem somewhat different. Jongno in the departure TAZ category was replaced with Heohyeon or Chungdam in the arrival TAZ category (many traditional markets and shopping centers are located in the latter two districts) as well as a minor shift on Saturday. On Sunday, we observed a considerable increase in both trips at Seogyo which is a well-known place for its popularity of shopping and entertainment. It indirectly implies that areas where major public institutions, companies, and recreational and commercial facilities are located produced and attracted a lot of taxi trips. Not surprisingly, as the number of arrival trips increases, the number of departure trips also increases, which is an easily expected outcome.

The top-ranked areas tend to be geographically located in the center of Seoul. <Fig. 2> depicts the departure and arrival taxi trips at the TAZ level in Seoul and suburbs. There is no significant difference in the spatial distribution of taxi trips between weekdays and Saturday, except for darker areas on the weekend (a darker TAZ indicates more trips). On Sunday, we observed somewhat different spatially distributed patterns－a clear drop in taxi trips.

2. Spatial-temporal Distribution

We further analyze spatial distribution of taxi trips with respect to four major time periods (7 to 9 a.m., 2 to 4 p.m., 6 to 8 p.m. and 11 p.m. to 1 a.m.). All cases demonstrate that both departure and arrival taxi trips were largely concentrated in the central business areas of Seoul including Gangnam, Yeouido, and Jongno, together with Sangam, where TV media centers and major stadiums are located, Gasan, where a digital industry complex facility is placed, and Hangangro, which has a large-scale electronic shopping mall. The first time period (7 to 9 a.m.) showed that people actively used taxis as expected (recall the morning peak-hour is 8-9 a.m.). From 2 to 4 p.m. (1st non-peak hour), both movements within Seoul and inter-regional trips (i.e., travel across a boundary) were relatively small compared to those made in other time periods. During the afternoon peak hour (6-8 p.m.), movements tended to increase in Seoul as well as inter-region travel from Seoul to adjacent new satellite cities such as Goyang, Bucheon, and Guangmyung. From 11 p.m. to 1 a.m., taxi tips were likely to partly increase; especially, there was a considerable increase in arrival trips (most of them seem to return to homes located in suburban areas).

On Saturday, we find similar spatial distributions, but the overall number of taxi trips was smaller than weekdays. Interestingly, there were huge taxi departures and arrivals from 11 p.m. to 1 a.m. We speculate that most of taxi users engaged in social/recreational activities till late at night in Seoul, then they chose taxis to return home (as mentioned above other modes are usually unavailable after midnight). Findings that departures were heavily concentrated in Gangnam, Jongno, Yeouido and Chungdam, where recreational and commercial facilities are densely located, and the arrival areas were mostly residential ones in Seoul or its suburbs support our speculation. There was a considerable decrease in the total number of taxi trips on Sunday compared to weekdays and Saturday as expected. No remarkable changes or peaks were observed during the period between 11 p.m. to 1 a.m.; we assume that this is partly because many people did not move or returned home early to prepare for the upcoming week.

<Fig. 2> Spatial distribution of departure/arrival trips

Ⅳ. Spatial Regression Models for Taxi Trips

1. Model Specification

We developed the spatial regression model to account for spatial dependency which may lead to a biased estimation result (we conjectured that TAZs are not spatially independent, demonstrating that association exists in

Section 3), thereby determining what factors affect the amount of taxi trips, stratified by the day of the week (weekday, Saturday and Sunday). This approach can be distinguished from existing studies mostly adopting ordinary regression approaches to identify determinants (e.g., Yang et al., 2018; Sahaller, 2005; Yang and Gonales, 2014). As proven above, taxi trips seem to be spatially correlated, perhaps due to land use patterns. In estimating the spatial regression model, we first conducted spatial cluster analysis (referred to as the Moran's I statistics) to double check whether the travel distribution is spatially correlated, then classified subgroups based on their spatial correlation. Based on that classification, we estimated two different regression models－a spatially lagged (SL) model and a spatial error (SE) model－using GeoDa 0.9, which is a free software used to conduct spatial data analysis and modeling. The former one assumes that the neighboring values of the dependent variables affect the value of the dependent variable itself, and the latter one assumes that the spatial association issue can raise in error terms of linear models (Ward and Gleditsch, 2007).

As Anselin(1988) described, the spatial lagged model can be written as:

$$y = \rho Wy + X\beta + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I),$$

where y is a vector of a dependent variable, ρ is a spatial autoregressive coefficient, W is a matrix of spatial weights, X is a matrix of explanatory variables, β is a vector of coefficients, ε is a vector of error terms which is normally distributed with zero mean and constant variance σ^2, and I is an identity matrix.

The spatial error model can be expressed as:

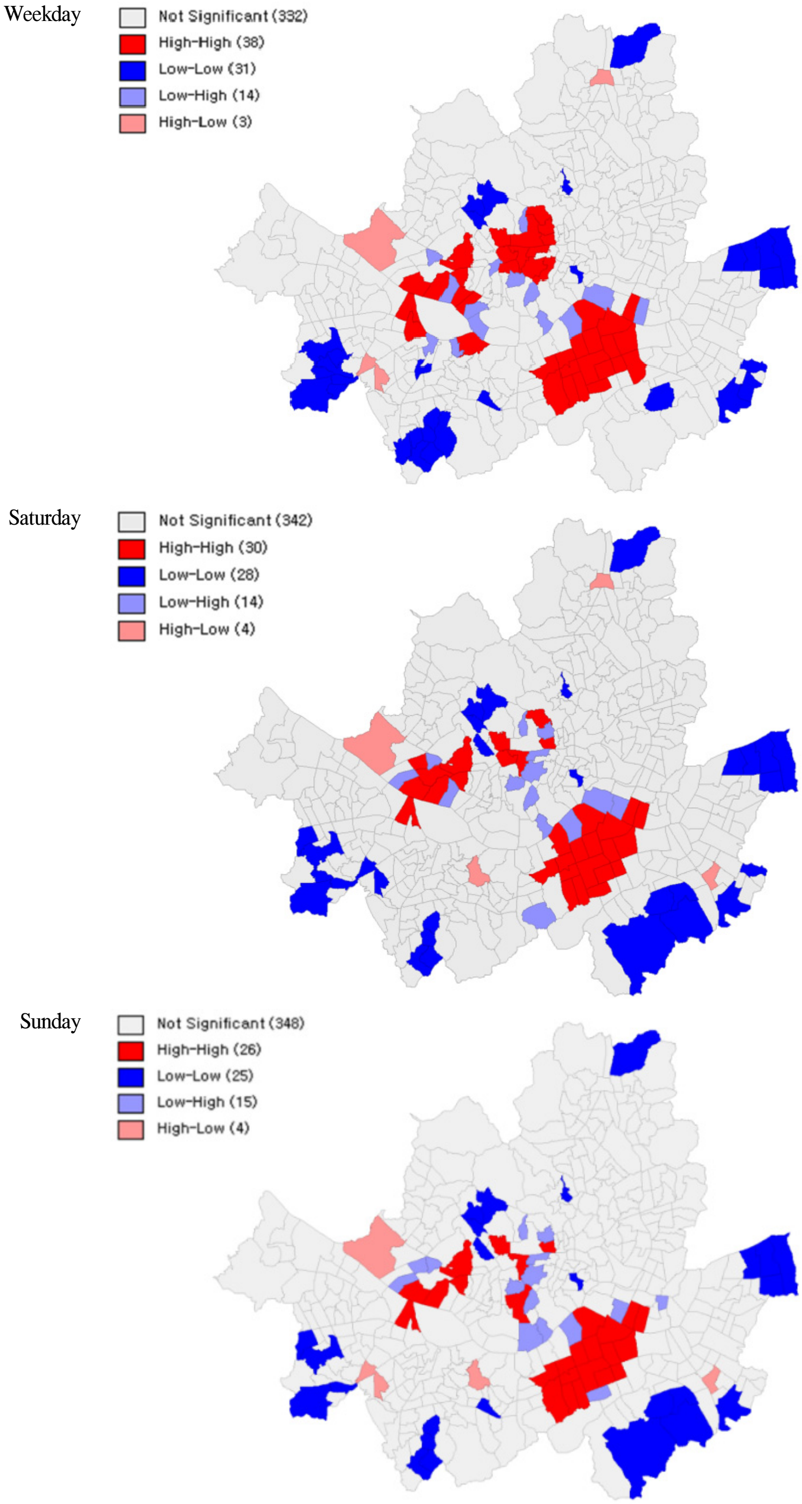
$$y = X\beta + \lambda W\xi + \epsilon,$$

where λ indexes a vector of a coefficient of the spatially lagged autoregressive error, and ξ is a vector of a spatial component of an error term. To develop final models, maximum likelihood estimation method is conventionally used.

We incorporated socio-demographic traits, transportation attributes, and land use characteristics as explanatory variables in the spatial regression models for taxi departure trips. Explanatory variables which are suspicious of multicollinearity (the variation inflation factor (VIF $\geq$ 5) were excluded in the final models.

2. Moran's I Test of Taxi Departure Trips

We conducted Moran's I test for the regression models for taxi departure trips to examine spatial autocorrelation; it showed that there was strong spatial autocorrelation of the residuals (see Tables 1 and 2). Thus, we concluded that the spatial regression model is appropriate for the analysis of taxi departure trips. <Fig. 3> visualizes the result of the local Moran's I test for the taxi trips by the day of the week by clustering them into five groups. Lots of primary and secondary central business areas fell into the High-High (HH) cluster (38 TAZs out of 418 TAZs) on weekdays, which are painted with red on maps (hot spots). On the contrary, 31 TAZs belonged to the Low-Low (LL) cluster (cold spots) on weekdays. Most of them are located at outskirts of Seoul.



<Fig. 3> Local Moran's I cluster map for taxi departure trips

3. Model Results

We present model estimation results for whole day taxi trips, together with the ordinary regression model as a benchmark model in <Table 1>. All significant parameters of ρ and λ demonstrate that taxi travel patterns in Seoul are spatially correlated. The goodness-of-fit measurements (R^2s) range 0.56-0.73, showing respectable model fits which is better than those of the reference model. This is an easily expected outcome in terms of the motivation our research. Interestingly, we confirm that estimated coefficients in the reference models tend to be overestimated, which hints that our approach can correct the biased estimation result.

Turning to the interpretation of individual coefficients, some socio-demographic traits are significantly associated with volume of taxi ridership. The number of households has a positive impact on taxi trips across all nine models, meaning that the more households live in a TAZ, the more taxi trips are generated. We can view this variable as a useful indicator representing the demand of taxi trips, which is align with a conventional wisdom of the trip generation model. We also find that the portion of female has a positive relationship with taxi trips. The concerns about safety while traveling at night is presumably involved in this relationship. The number of elderly people is negatively related to taxi trips (albeit it was not significant in the weekend SE model), which makes sense because elderly people are generally less active compared to younger groups due to physical limitations. The number of companies, which leads to business-related taxi trips, has positive impacts on taxi trips.

Regarding transportation attributes, there is a strong evidence of relationship between subways and taxis, which is similar to previous studies (e.g., Hochmair, 2016). The number of subway stations is positively related to taxi trips whereas the number of bus stops is not significant across the models. Based on those two opposite outcomes, we presume that (1) taxi trips are largely connected with subways, not with buses, and (2) taxi users may prefer to use subways as their primary mode (in this case the taxi mode could be an access/egress mode when traveling). The number of vehicles has a non-negligible positive impact on taxi ridership, but it is significant only in the weekday models.

With respect to land use characteristics, population density appears across nine models with negative signs. In general, that variable is expected to have a positive relationship (particularly in Seoul), but the outcome shows an opposite association. Employment density and residential area ratio seem to positively affect taxi trips while sales facility floor area ratio and commercial area ratio appear not to be important in explaining taxi departure trips. A few variables show inconsistent impacts. Office facility floor area ratio, for example, has an only significant influence on taxi trips on weekday, and building floor area entropy－which is indicative of the level of mixed land use－is statistically significant in weekend models, but not significant in other models.

The estimation results of spatial regression models by peak type are presented in <Table 2>. Impacts of explanatory variables are similar to those in the weekday models in <Table 1>. The socio-demographic traits (number of households, gender, and population density), business characteristics (the number of companies and employment density) and accessibility to subway (number of subway stations) have consistent impacts across three time periods. In comparison of magnitudes of the explanatory variables, the off-peak time model tends to show smaller impact on taxi ridership than others (morning and afternoon peak time models), and the number of elderly people is likely to be less significant in the morning peak models (elderly people’s inclination to avoid traffic congestion might be involved in this relationship).

<Table 1> Spatial regression model results by the day of the week

Variable	Weekday			Saturday			Sunday		
	OLS[a]	SL[b]	SE[c]	OLS	SL	SE	OLS	SL	SE
No. of Households	0.117*	0.106*	0.0938*	0.105*	0.0882*	0.0629*	0.0675*	0.0595*	0.0444*
Female ratio (%)	81.521*	71.787*	65.919*	71.808*	62.225*	55.119*	38.675*	34.780*	31.097*
No. of students	0.0030	0.0047	0.0090	0.00240	0.00428	0.0103	-0.000338	0.000549	0.004253
No. of elderly people	-0.253*	-0.240*	-0.235*	-0.208*	-0.183*	-0.141	-0.126*	-0.115*	-0.093
No. of companies	0.335*	0.360*	0.343*	0.390*	0.403*	0.381*	0.223*	0.229*	0.218*
No. of subway stations	164.602*	153.641*	139.916*	157.061*	149.611*	144.395*	105.078*	100.755*	97.044*
No. of bus stops	-0.741	0.540	1.211	0.417	1.974	3.587	0.931	1.649	2.694
No. of vehicles	0.0966*	0.0911*	0.1129*	0.0334	0.0309	0.0404	0.0131	0.0129	0.0195
Population density(person/ha)	-1.761*	-1.448*	-1.039*	-1.900*	-1.557*	-1.191*	-1.262*	-1.092*	-0.841*
Employment density(person/ha)	3.488*	2.463*	2.642*	1.901*	1.086	1.481*	1.300*	0.922*	1.099*
Residential area ratio(%)	11.208*	9.806*	8.228*	15.118*	13.519*	12.967*	9.797*	8.994*	8.583*
Commercial area ratio(%)	5.495	4.161	10.340	1.786	0.987	9.185	2.482	1.996	6.693*
Office facility floor area ratio(%)	43.306*	38.248*	35.754*	15.623	12.053	10.249	3.112	1.562	0.591
Sales facility floor area ratio(%)	-9.001	-8.638	-8.921	-2.684	-2.016	-3.472	-1.086	-0.700	-1.551
Building floor area entropy[d]	602.631	409.300	759.251	1311.193*	1111.708*	1376.018*	986.521*	903.696*	1090.455*
ρ		0.212*			0.231*			0.184*	
λ			0.429*			0.406*			0.375*
R^2	0.690	0.709	0.725	0.556	0.580	0.598	0.548	0.563	0.583
No. of cases (TAZs)	418	418	418	418	418	418	418	418	418
Likelihood-ratio test		23.8749*	35.792*		18.735*	27.995*		11.099*	22.850*
Moran's I	6.6183*			11.929*			5.488*		
LM-Lag	30.773*			22.445*			12.970*		
Robust LM-Lag	4.725*			1.896			0.0556		
LM-error	38.379*			26.822*			21.5002*		
Robust LM-error	12.330*			6.274*			8.585*		

Notes: * means statistically significance at α=0.1.
[a] Ordinary regression model
[b] Spatially lagged model
[c] Spatial error model
[d] It was calculated by Shannon's entropy using the following formula

$$SE = -\sum_{i=1}^{n} \frac{P_i \log(P_i)}{\log(n)},$$

where is number of building groups and is each group's floor area ratio.

<Table 2> Spatial regression model results by the time of the weekday

Variable	Morning peak (7~9 AM)			Off-peak (2~4 PM)			Afternoon peak (6~8 PM)		
	OLS	SL	SE	OLS	SL	SE	OLS	SL	SE
No. of households	0.00787*	0.00613*	0.00558*	0.00648*	0.00590*	0.00539*	0.00977*	0.00949*	0.00909*
Female ratio (%)	6.484*	4.453*	4.058*	5.733*	4.942*	4.394*	5.745*	5.460*	5.310*
No. of students	-0.00088	-0.00040	-0.00007	0.000599	0.000670	0.000879	0.000658	0.000675	0.000797
No. of elderly people	-0.013	-0.011	-0.0118*	-0.0111*	-0.0108*	-0.0110*	-0.0192*	-0.0193*	-0.0208*
No. of companies	0.01612*	0.0214*	0.01815*	0.0185*	0.0201*	0.0189*	0.0224*	0.0232*	0.0222*
No. of subway stations	11.567*	8.691*	4.536	10.494*	9.674*	9.129*	16.560*	16.064*	15.128*
No. of bus stops	-0.095	0.139	0.338	-0.0128	0.07138	0.1085	0.0620	0.1101	0.186
No. of vehicles	0.0152*	0.0133*	0.0162*	0.00585*	0.00567*	0.00712*	0.00808*	0.00808*	0.0101*
Population density(person/ha)	-0.165*	-0.1132*	-0.0803*	-0.1331*	-0.1102*	-0.0846*	-0.167*	-0.154*	-0.122*
Employment density(person/ha)	0.313*	0.13732*	0.1678*	0.2793*	0.2072*	0.215*	0.362*	0.323*	0.319*
Residential area ratio(%)	0.993*	0.758*	0.579*	0.637*	0.551*	0.463*	0.684*	0.640*	0.529*
Commercial area ratio(%)	0.1161	0.1774	0.755	0.785*	0.653*	0.991*	0.995*	0.931*	1.223*
Office facility floor area ratio(%)	5.919*	4.505*	4.007*	2.511*	2.165*	2.025*	2.480*	2.353*	2.413*
Sales facility floor area ratio(%)	-0.599	-0.681	-0.654	-0.482	-0.458	-0.467	-0.926	-0.913	-0.898
Building floor area entropy[d]	13.226	-12.855	18.854	21.893	12.896	37.668	25.150	18.770	36.222
ρ		0.403*			0.213*			0.0977*	
λ			0.643*			0.383*			0.285*
R^2	0.620	0.705	0.737	0.657	0.677	0.687	0.651	0.655	0.667
No. of samples (TAZs)	418	418	418	418	418	418	418	418	418
Likelihood-ratio test		92.300*	114.756*		21.380*	25.887*		4.110	13.835
Moran's I	11.929*			5.488*			4.190*		
LM-Lag	122.302*			27.206*			5.095*		
Robust LM-Lag	22.061*			6.228*			0.116		
LM-error	129.803*			25.890*			14.579*		
Robust LM-error	29.562*			4.912*			9.599*		

Notes: * means statistically significance at α=0.1.

[a] Ordinary regression model

[b] Spatially lagged model

[c] Spatial error model

[d] It was calculated by Shannon's entropy using the following formula

$$SE = -\sum_{i=1}^{n} \frac{P_i \log(P_i)}{\log(n)},$$

where is number of building groups and is each group's floor area ratio.

In sum, we confirm that (1) taking spatial correlation into account is essential to analyze taxi ridership (technically departure trips of the taxi mode) based on the outcome that our spatial regression models provide better model fits than conventional regression models, and (2) socio-demographic traits, transportation attributes, and land use characteristics affect (partly at least) taxi departure trips in various ways.

Ⅴ. Conclusions

This paper explores temporal and spatial travel characteristics of taxis (particularly focusing on taxi departure trips) to identify determinants using taxi GPS data in Seoul collected from digital tachograph. The GPS device provides exclusive information including boarding and alighting time of passengers, distance, and location. To correct unexpected bias caused by spatial dependence, we adopted the spatial regression approach. As a result, we confirm that our proposed models are useful to identify what factors affect behavioral characteristics of the taxi mode (particularly, taxi departure trips); specifically, socio-demographic traits, transportation attributes, and land use characteristics appear to be important in determining taxi ridership. Key findings and implications can be summarized as follows.

First, we conducted descriptive analysis to investigate taxi travel characteristics focusing on the number of trips, travel distance, and travel time by the day of the week and type of peak time. The number of average daily trip per taxi was 34.8 trips on weekdays, 35.2 trips on Saturday, and 26.4 trips on Sunday. Friday showed the most frequent trips per taxi (37.6) while the smallest number of average trip per taxi was made on Sunday. With respect to peak time, the largest taxi trips were produced during the morning peak time (8-9 a.m.) and after midnight time period (0-1 a.m.) on weekdays. The average travel distance and time were 5.9 km and 13 minutes, respectively.

Secondly, we visualized the spatial distribution of departure and arrival taxi trips on maps. On weekdays and Saturday, Gangnam, Yeouido and Jongno showed sizable departure and arrival taxi trips. In the case of Seogyo (a specific TAZ in Seoul), there was a considerable increase in the number of trips from/to that TAZ on Sunday. As expected, as the number of arrival trips increases, the number of departure trips is likely to increase, and no noticeable difference at a specific time period between weekdays and Saturday was observed, but we found a clear drop in taxi trips on Sunday.

Finally, two spatial regression models－the spatially lagged and spatial error models－were estimated to identify key factors affecting taxi departure trips (i.e., taxi ridership). We found some intriguing relationships, which mostly align with earlier studies. First of all, our spatial regression models exhibited a better model fit than ordinary regression models in terms of the goodness-of-fit and interpretability. The number of households is positively associated with taxi trips, and females are likely to travel more using taxis. Interestingly, the taxi mode has a positive relationship with the subway mode, which might be an empirical evidence implying that taxis can complement the limitations of public transit (technically those of subways); that is, people could choose taxis as an access/egress mode when they use subways as their primary mode. Regarding land use characteristics, *population density* negatively influences taxi trips during the weekend while *employment density* positively affects taxi trips during weekdays. Both *residential area* and *office facility floor area ratio* are positively related to taxi trips across three types of the day (weekday, Saturday, and Sunday). The mixed land use factor is also positively related to taxi ridership on the weekend. Given that results above, we suggest that accounting for spatial dependence, then employing the spatial regression approach is essential to analyze spatially distributed travel patterns.

This research may be likely to be of interest to scholars, public-sector planners and policy-makers, researchers and other transportation-related professionals because we provide a better understanding of taxi travel characteristics. However, we still call for additional data collection and studies to overcome two limitations: (1) the omission of information related to passenger characteristics and (2) exclusion of driver-owned taxi data. If these issues can be handle with in the near future, further empirical studies will provide a comprehensive understanding of future taxi travel demand.

ACKNOWLEDGEMENTS

This work was supported by Incheon National University Research Grant in 2015. The earlier version of this paper was presented at the annual meeting of Western Regional Science Association in 2016.

REFERENCES

Anselin L.(1988), *Spatial Econometrics: Methods and Models*, Dordrecht, the Netherlands: Kluwer.

Hochmair H.(2016), "Spatiotemporal pattern analysis of taxi trips in New York City," *Transportation Research Record, Journal of Transportation Research Record*, no. 2542, pp.45-56.

Hoque M. A., Hong X. and Dixon B.(2012), "Analysis of mobility patterns for urban taxi cabs," *International Conference on Computing, Networking and Communications, Mobile Computing and Vehicle Communications Symposium*, pp.756-760.

Kattan L., Barros A. and Wirasinghe S. C.(2010), "Analysis of work trips made by taxi in Canadian cities," *Journal of Advanced Transportation*, vol. 44, no. 1, pp.11–18.

Kim J.(2015), *A study on key factors to affect taxi travel using DTG data in Seoul*, Master Thesis, Hongik University.

Liu Y., Wang F., Xiao Y. and Gao S.(2012), "Urban land uses and traffic 'source-sink areas': Evidence from GPS-enabled taxi data in Shanghai," *Landscape and Urban Planning*, vol. 106, pp.73-87.

Nicholas J. S.(2012), *Examination of taxi travel patterns in Arlington County*, Master Thesis, Virginia Tech.

NYC Department of Transportation(2016), *New York City Mobility Report*.

Qi G., Li X., Li S., Pan G. and Wang Z.(2011), "Measuring social functions of city regions from large-scale taxi behaviors," *Pervasive Computing and Communications Workshops*, pp.384-388.

Qian X. and Ukkusuri S. V.(2015),"Spatial variation of the urban taxi ridership using GPS data," *Applied Geography*, vol. 59, pp.31-42.

Schaller B.(2005), "A regression model of the number of taxicabs in U.S. cities," *Journal of Public Transportation*, vol. 8, no. 5, pp.63-78.

Shi W., Kong Q. and Liu Y.(2008), "A GPS/GIS integrated system for urban traffic flow analysis," *Proceedings of the 11th International IEEE Conference on Intelligent Transportation Systems*, pp.844-849.

Tang J., Liu F., Wang Y. and Wang H.(2015), "Uncovering urban human mobility from large scale taxi GPS data," *Physical A: Statistical Mechanics and its Applications*, vol. 438, pp.140-153.

The Seoul Government(2013), *Research on operation cost analysis and improvement of fare structure of taxi*, The Seoul Government.

Veloso M., Phithakkitnukoon S. and Bento C.(2011), "Urban mobility study using taxi traces," *TDMA '11 Proceedings of the 2011 International Workshop on Trajectory Data Mining and Analysis*, pp.23-30.

Wang F. and Ross C. L.(2017), "New potential for multimodal connection: Exploring the relationship between taxi and transit in New York City (NYC)," *Transportation*, pp.1-22. https://doi.org/10.1007/s11116-017-9787-x.

Wang Y., Gu Y., Dou M. and Qiao M.(2018), "Using spatial semantics and interactions to identify urban functional regions," *ISPRS International Journal of Geo-Information*.

Ward M. D. and Gleditsch K. S.(2007), *An Introduction to Spatial Regression Models in the Social Sciences*, Available at http://dces.wisc.edu/wp-content/uploads/sites/30/2013/08/W4_W7_WardGleditsch.pdf.

Yang C. and Gonzales E.(2014), "Modeling taxi trip demand by time of day in New York City," *Transportation Research Record, Journal of Transportation Research Record*, no. 2429, pp.110-120.

Yang Z., Franz M. L., Zhu S., Mahmoudi J., Nasri A. and Zhang L.(2018), "Analysis of Washington, DC taxi demand using GPS and land-use data," *Journal of Transport Geography*, vol. 66, pp.35-44.

Zhan X., Hasan S., Ukkusuri S. V. and Kamga C.(2015), "Urban link travel time estimation using large-scale taxi data with partial information," *Transportation Research Part C*, vol. 33, pp.37-49.

Zheng Y., Liu Y., Yuan J. and Xie X.(2011), "Urban computing with taxicabs," *UbiComp '11a Proceedings of the 13th International Conference on Ubiquitous Computing*, pp.89-98.