This is an author-produced, peer-reviewed version of this article.
The final publication is available at www.sciencedirect.com. Copyright restrictions may apply
https://doi.org/10.1016/j.cities.2022.103637.

Understanding Taxi Ridership with Spatial Spillover Effects and Temporal Heterogeneity

Pengyu Zhu; Jie Huang*; Jiaoe Wang; Yu Liu; Jiarong Li; Mingshu Wang; Wei Qiang

Abstract

In urban transportation systems, taxis are regarded as flexible, convenient, and time-saving. Taxi demand is affected by various built-environment factors and by the time of the day. Although many studies have investigated correlations between taxi demand and the built environment, the direct and spillover effects of built environment factors on taxi demand have not been examined at a fine spatial scale. To address this gap in the literature, this paper employs spatial econometric models using GPS-tracked taxi trips, mobile signaling data, and points of interest (POIs) to study taxi demand in Beijing at a 1-kilometer square grid resolution. The results show that, in the morning and evening peak hours, road network density has the strongest (positive) direct and indirect impact on taxi ridership. A relationship is also found between public transportation and taxi ridership: bus coverage has positive direct effects and insignificant indirect effects on taxi pick-ups and drop-offs, while subway coverage has negative indirect effects, suggesting that it may absorb taxi demand from surrounding grids. Results also indicate that various built-environment factors affect taxi demand differently at morning and evening peak times. This study reveals the complex nature of taxi ridership and has important implications for policymakers, transport planners, and other stakeholders in megacities around the world.

1. Introduction

Taxis play a vital role in urban transportation because of the flexibility of the door-to-door service they offer and their 24/7 availability. A good understanding of the temporal and spatial distribution of taxi trips as well as the factors that have substantial influences on taxi demand can help governments and policymakers design a well-connected multi-modal transportation system. A multimodal transportation system can address the increasing congestion and emission problems in metropolitan areas and better satisfy passengers' needs for transit (Çetin & Yasin Eryigit, 2011; Schaller, 2005).

Many studies have examined the relationship between the built environment and travel behavior in terms of distance, duration, and mode choice (Cervero 2002; Dong and Zhu 2015; Ewing 2015;

This is an author-produced, peer-reviewed version of this article.
The final publication is available at www.sciencedirect.com. Copyright restrictions may apply
https://doi.org/10.1016/j.cities.2022.103637.

Wang, 2001; Antipova et al., 2011; Ewing and Cervero, 2010; Ewing et al., 2015; Zhou et al., 2019; Zhu, 2012, 2013; Zhu et al. 2013, 2017, 2018, 2020). These studies have generally agreed that built environment factors, including residential and employment density, land use diversity, distance to public transit, and destination accessibility may all significantly influence travel behavior. Recently, researchers have begun to explore the impacts of socio-demographic and built environment factors on taxi demand and related travel behavior. For example, McNally (2008) proposed that population, employment, and other socio-demographic factors can all affect the number of taxi passengers. Qian and Ukkusuri (2015) found that a lower median income level is associated with a smaller number of taxi trips in particular places in New York City. Most recently, Yu and Peng (2019) suggested that built environment factors also significantly impact the demand for ride-sourcing services. Following the theoretical framework that has been generalized by this research stream, this study investigates the relationships between taxi ridership, the urban built environment, and neighborhood socioeconomic factors. Since many cities lack disaggregated data at a fine geographical level, the *first* contribution of this paper is to offer an exemplary framework for researchers to incorporate and utilize various types of big data, including taxi origin and destination data (O-Ds), mobile signaling data, points of interest (POIs), and other web data.

Although some studies have examined the built environment and taxi ridership, few have considered the spatial autocorrelations associated with both taxi ridership and built environment factors. Moreover, few studies have identified the spatial spillover effects of the various built environment factors that influence taxi demand. To fill this gap, this study uses spatial econometric models that take into consideration the spatial autoregressive process. By introducing the spatial weight matrix, we can more accurately estimate the direct (local) and indirect (spillover) effects of the explanatory variables on the outcome variable (Anselin, 1988). Therefore, the *second* contribution of this research is to provide a more comprehensive understanding of how different built environmental characteristics and socioeconomic variables influence taxi ridership via their local and spillover effects.

Furthermore, the spatial distributions of taxi ridership exhibit different patterns during morning versus evening peak hours (Liu, Wang, Xiao, & Gao, 2012; Zhu, Huang, Guibas, & Zhang, 2013). It is reasonable to speculate that there is some level of temporal heterogeneity in the relationships we want to test. Hence, the *third* contribution of this paper is to explore how built environment factors affect taxi O-Ds differently at different times (i.e., morning vs. evening).

Applying an innovative approach that combines spatial-temporal big data analytics with traditional spatial economic models, this study provides a comprehensive picture of how various built environment and neighborhood socioeconomic factors influence taxi ridership, both in local neighborhoods via direct effects and in nearby neighborhoods via spillover effects. We find that these built environment factors have different direct and indirect impacts on taxi ridership and that these effects vary by time (i.e., morning vs. evening peak). For example, road network density may not only directly increase local taxi demand but also have spillover effects inducing more taxi ridership in neighboring areas. Similarly, different types of public transportation have different

This is an author-produced, peer-reviewed version of this article.
The final publication is available at www.sciencedirect.com. Copyright restrictions may apply https://doi.org/10.1016/j.cities.2022.103637.

impacts on taxi ridership. Bus coverage has positive direct effects on local taxi ridership but insignificant spillover effects. However, subway coverage shows negative indirect effects on taxi pick-ups and drop-offs during both peaks, suggesting that it may absorb taxi demand from nearby cells. Additionally, the results show different relationships between various categories of points of interest (POI) and taxi ridership. During the morning peak, the number of POIs in public management and services increase local taxi demand, while those in residences and related facilities have negative indirect effects on taxi pick-ups in surrounding areas. Meanwhile, POIs in commercial and recreational services and POIs in manufacturing and offices both have positive effects on local taxi pick-ups during the evening peak. POIs in transportation services have a positive direct impact on taxi demand during both morning and evening peaks.

Based on these results, we suggest that transportation management agencies should pay close attention to the direct and spatial spillover effects of various built environment factors on taxi ridership. For example, because road network density not only affects local taxi ridership but also demand in adjacent areas, it is necessary for transportation planners to comprehensively consider the road layout in surrounding areas. Moreover, the interaction between different public transportation modes and taxi usage should be included in taxi demand modeling and multi-modal transportation planning. As ride-sourcing services such as Uber and Lyft become increasingly popular around the world, this research also has important planning implications for the improved integration of these services into the existing multi-modal transportation system.

2. Literature review

Taxi demand in cities is usually imbalanced, with temporal and/or spatial gaps between taxi services and demand. Imbalanced taxi demand results in empty-load vehicle running, traffic congestion, and air pollution. Hence, it is essential that urban transport planners have an in-depth understanding of taxi demand. Many studies have confirmed an imbalanced spatial distribution of taxi demand. For example, about 90 percent of taxi trips were found to take place in downtown areas (i.e., Manhattan) in New York City (Qian and Ukkusuri, 2015); taxi demand was found to differ between urban districts in Munich (Jager et al., 2016); and pick-ups and drop-offs of taxi trips were found to be imbalanced at a local spatial scale in Shanghai (Liu et al., 2012). With this in mind, researchers have employed Geographically Weighted Regression (GWR) models in taxi demand analysis to better explain the imbalance of taxi ridership (Qian and Ukkusuri, 2015; Li et al., 2019; Chen et al. 2021; Wang and Noland, 2021; Yuan et al. 2021). Spatial heterogeneity and the non-linear spatial patterns of taxi demand have been considered in their models. However, spatial spillover effects remain under-researched.

Spatial spillover effects refer to the interaction effects among nearby geographical units due to their spatial dependence (i.e., spatial autocorrelation). Some preliminary studies have taken spatial autocorrelation into consideration in analyzing the factors influencing taxi demand or ridership

This is an author-produced, peer-reviewed version of this article.
The final publication is available at www.sciencedirect.com. Copyright restrictions may apply
https://doi.org/10.1016/j.cities.2022.103637.

(see a summary in Supplementary File). For instance, two conference papers, Correa et al. (2017) and Pan et al. (2019) both used linear models, spatial error models, and spatial lag models to examine the spatial distribution of traditional taxis, e-hailing taxis, and/or Uber ridership in New York City. Lavieri et al. (2018) developed a spatial lag multivariate count model to explore the factors attracting ride-sourcing trips in Austin. Ni & Chen (2020) used K-means clustering and spatial lag model to explore the impact of built-environment features on the use of dockless bike sharing and taxis to serve as transfer modes for metro in Beijing. Similarly, Zhang et. al (2020) adopted mixed modeling structure of spatial lag and simultaneous equation models to investigate the influencing factors on traditional taxi and app-based taxi demand in New York City. However, a thorough review of the existing literature shows that *no* empirical research to date has considered: 1) the spatial autocorrelation in explanatory variables among geographical units (i.e., existing research only considers the spatial autocorrelation in outcome variable); 2) taking important further steps to actually calculate the spatial *spillover effects* and differentiate them from direct effects. Therefore, a key contribution of this paper is to fill these gaps and provide more in-depth understanding of not only the direct (local) effects but also the spatial spillover effects of various factors on taxi demand.

In addition to the spatial dimension, taxi demand imbalance also exists on a temporal scale. Taxi demand varies significantly over the course of the day (Liu et al., 2015) and over different days of the week (weekdays vs. weekends) (Zhao et al., 2016; Wang et al., 2020). Some studies have attempted to include temporal dynamics in taxi demand prediction (Phithakkitnukoon, 2010; Veloso, 2011). Moreira-Matias (2013) combined three different time-series models with real-time taxi trip data to predict short-term demand with demand uncertainty. In transport geography, some studies have utilized the spatiotemporal characteristics of taxi trips to forecast passenger demand, such as Lee et al. (2008) and Yuan et al. (2011). More recently, machine learning algorithms have been used in taxi ridership prediction (Shao et al., 2015; Zhao et al., 2016; Zhou et al., 2019). In sum, taxi demand prediction models highlight the importance of the temporal heterogeneity in taxi ridership. In this regard, our paper also attempts to explain the temporal variations in taxi ridership in relation to various influencing factors; in particular, a comparison is made between morning and evening peak hours.

Lastly, empirical research has started to explore specifically how taxi ridership may be influenced by built environment factors and socio-demographic variables (Comito et al., 2015; Comito Qian and Ukkusuri, 2016; Wang and Mu, 2018). For instance, Yang and Gonzales (2014) found that taxi ridership in New York is significantly correlated with population density, employment density, and education levels. Liu et al. (2020) found a strong correlation between taxi ridership and land use mix, population density, and road junctions in Beijing. Some research has also linked taxi trips to urban functions (Liu et al., 2021; Keler et al., 2020; Gong et al, 2016; Zhou et al., 2015; Hu et al., 2021). Moreover, research on the relationships between taxis and other transportation modes, particularly public transit, has become increasingly important with the development of on-demand e-hailing platforms (Gonzales et al., 2014; Schaller, 2005; Ulak et al., 2020). Overall, most of these studies have utilized traditional census data to measure built environment at different

This is an author-produced, peer-reviewed version of this article.
The final publication is available at www.sciencedirect.com. Copyright restrictions may apply
https://doi.org/10.1016/j.cities.2022.103637.

geographical levels. Yet census data at a fine geographic scale are not readily available for many cities in developing countries. Our research adopts a novel approach that utilizes various big data and open-source data, such as cell phone data, POI data, web-crawled housing transaction data, to measure a variety of built environment and socio-demographic variables. This approach could provide a useful framework for researchers in developing countries to carry out their own research on related topics.

3. Data and methodology

3.1 Study area, variables and data source

3.1.1 Study area

Beijing, the study area, had a population of 21.7 million and an urban area of 16,410 km^2 in 2017. Ding and Zhao (2014) found that Beijing's spatial structure fits the monocentric city model in their study on land development, housing prices, and residential and employment distributions. In addition, Beijing has a six-ring road network, which has been used in some previous studies as the boundary to analyze urban transportation issues in Beijing (Kong et al., 2017; Yao, Wu, Zhu, Gao,& Liu, 2019). However, in this study, our analysis covers only the area within the fifth ring road (Fig. 1) because of the spatial limitations of the mobile signaling data used in our models. Among Beijing's sixteen municipal districts, two districts (Dongcheng and Xicheng) are entirely covered by our study area, and five (Haidian, Shijingshan, Chaoyang, Daxing, and Fengtai) are partly covered. In terms of spatial resolution, we acknowledge the modifiable area unit problem (MAUP), which means that the results will vary according to the scale of the research unit, resulting in statistical bias in spatial analysis (Openshaw & Taylor, 1981). With this in mind, we divide the study area into 1 km-by-1 km square cells, following the method used in many previous studies (Kong, Liu, Wang, Tong & Zhang, 2017; Liu, Wang, Xiao and Gao, 2012; Liu, Gong, Gong and Liu, 2015; Louail et al., 2014). This results in 683 grid cells where both taxi trip data and mobile signaling data are available.

This is an author-produced, peer-reviewed version of this article.
The final publication is available at www.sciencedirect.com. Copyright restrictions may apply
https://doi.org/10.1016/j.cities.2022.103637.

Fig. 1 The research region

3.1.2 Dependent variables

The dependent variables of this study are taxi trip origins and destinations, which are represented by the number of taxi trips originating from each cell in the morning peak hours, the number of taxi trips ending in each cell in the morning peak hours, the number of taxi trips originating from each cell in the evening peak hours, and the number of taxi trips ending in each cell in the evening peak hours (Table 1).

In this study, taxi trip records were extracted from GPS trajectory data generated by all taxis in Beijing from April 1st to 26th 2015, with an average of 17,984 taxis each day in our study area for this time frame. Every taxi trip record in this dataset contains the geographic locations (i.e., longitude and latitude) of a taxi every 30 seconds and the time and location of each pick-up and drop-off. With the successive trajectories of each taxi, each trip can be extracted and expressed in the following form: longitude and latitude of the pick-up location, longitude and latitude of the drop-off location, pick-up time, and drop-off time. An example trip is [(116.29253, 39.86538), (116.28003, 39.82736), 2015-4-15 15:36:31, 2015-4-15 15:49:39]. Because the dataset includes all taxis that operated in our study area during the study period, it can be considered to comprehensively reflect the real spatial and temporal distributions of taxi trips in Beijing. The average number of taxi trips starting or ending in the study area was approximately 340,000 per day, with a total of 26 days. All taxi trips were validated according to their travel distance and duration. That is, taxi trip records were excluded if the travel distance was less than 10 meters and the duration was less than 10 seconds. For each grid cell, we counted the number of taxi pick-ups and the number of drop-offs for each hour of each day. We then divided peak hours into the

This is an author-produced, peer-reviewed version of this article.
The final publication is available at www.sciencedirect.com. Copyright restrictions may apply
https://doi.org/10.1016/j.cities.2022.103637.

morning peak (7 AM – 9 AM) and evening peak (5 – 7 PM). The average morning (or evening) peak-hour taxi trip origins (or destinations) in a grid cell were calculated as the total number of taxi pick-ups (or drop-offs) in that cell during morning (or evening) peak hours for all weekdays from April 1st to 26th, 2015, divided by the number of weekdays. Finally, we related taxi trip O-Ds with socioeconomic variables, public transit coverage, level of land development intensity, and different categories of POIs, all of which are aggregated at the same grid level. These explanatory variables are discussed below in detail.

3.1.3 Independent variables

The independent variables in this study were mobile signaling, POIs, coverage of subway stations and bus stops, average housing price, and other built environment factors such as road network, percent of road area, number of buildings, and average number of stories of buildings (Table 1).

(1) Mobile signaling data

In China, fine-grained and accurate employment and residential population data are unavailable for many cities. In such a situation, mobile signaling data can be an alternative for estimating the number of workers and residents at specific locations (Ding, Niu,& Song, 2016; Louail et al., 2014). This study employs mobile phone data from China Unicom, offered by Smart Steps Co., Ltd[1]. It is worth noting that China Unicom is one of the three largest telecommunication corporations, and its market share in Beijing is 29.6%. The other two corporations are China Mobile and China Telecom.

Each mobile user was identified with an anonymous and unique ID from the original mobile signaling data. The time and duration of every user in a defined local service area were recorded according to the records of the base transceiver station the user's mobile phone had connected with. With the original information, the program offered by Smart Steps identifies the employment and residential cell for each user. The identification rules are as follows:

The work cell for a user is the cell in which the user stays most frequently between 9 AM and 5 PM during all weekdays within a month. The residential cell for a user is the cell in which the user stays most frequently between 9 PM and 5 AM (of the next day) within a month. The numbers of employees and residents were then aggregated by 1 km-by-1 km grid cells and recorded in the dataset, allowing the residential and employment density in the grid network to be calculated.

[1]Smart Steps is a data-sourcing company providing Mobile Signaling Data products for China Unicom (Smart Steps Digital Technology CO., LTD). In this paper, the authors obtained resident and employment distributions in 1km*1km grid cells from Smart Steps.

This is an author-produced, peer-reviewed version of this article.
The final publication is available at www.sciencedirect.com. Copyright restrictions may apply
https://doi.org/10.1016/j.cities.2022.103637.

(2) POI data

The original POI data, including 140,337 POIs located in the research area, fell within sixteen categories (finance and insurance services, hotels, living services, shopping, scenic spots, catering, sports and entertainment, companies and enterprises, business offices, residential buildings and facilities, government agencies and social organizations, hospitals and clinic services, living services, educational and social services, public facilities, and transport facilities) and were obtained from Baidu Map Services (http://map.baidu.com). Baidu is the largest web map service provider in China (Yao et al., 2017) with a large-scale user group. Since many original categories are further divided into multiple subcategories, overlapping often occurs between sub-categories (e.g., some restaurants not only belong to the Chinese restaurants subcategory under the category of catering services but also belong to the hotels category). Thus, reclassification is necessary. Referring to the classification system adopted in China's standard land use planning, the POI data were reclassified into five categories: commercial and recreational services, manufacturing and offices, residence and related facilities, public management and services, and transportation services.[2] The final classification is shown in Fig. 2. Then, the number of POIs in each new category was aggregated by grid cells.

[2] The number of POIs for transportation service are not included in our main models because of their overlap with bus and subway coverage, hence incurring potential collinearity issues.

This is an author-produced, peer-reviewed version of this article.
The final publication is available at www.sciencedirect.com. Copyright restrictions may apply
https://doi.org/10.1016/j.cities.2022.103637.

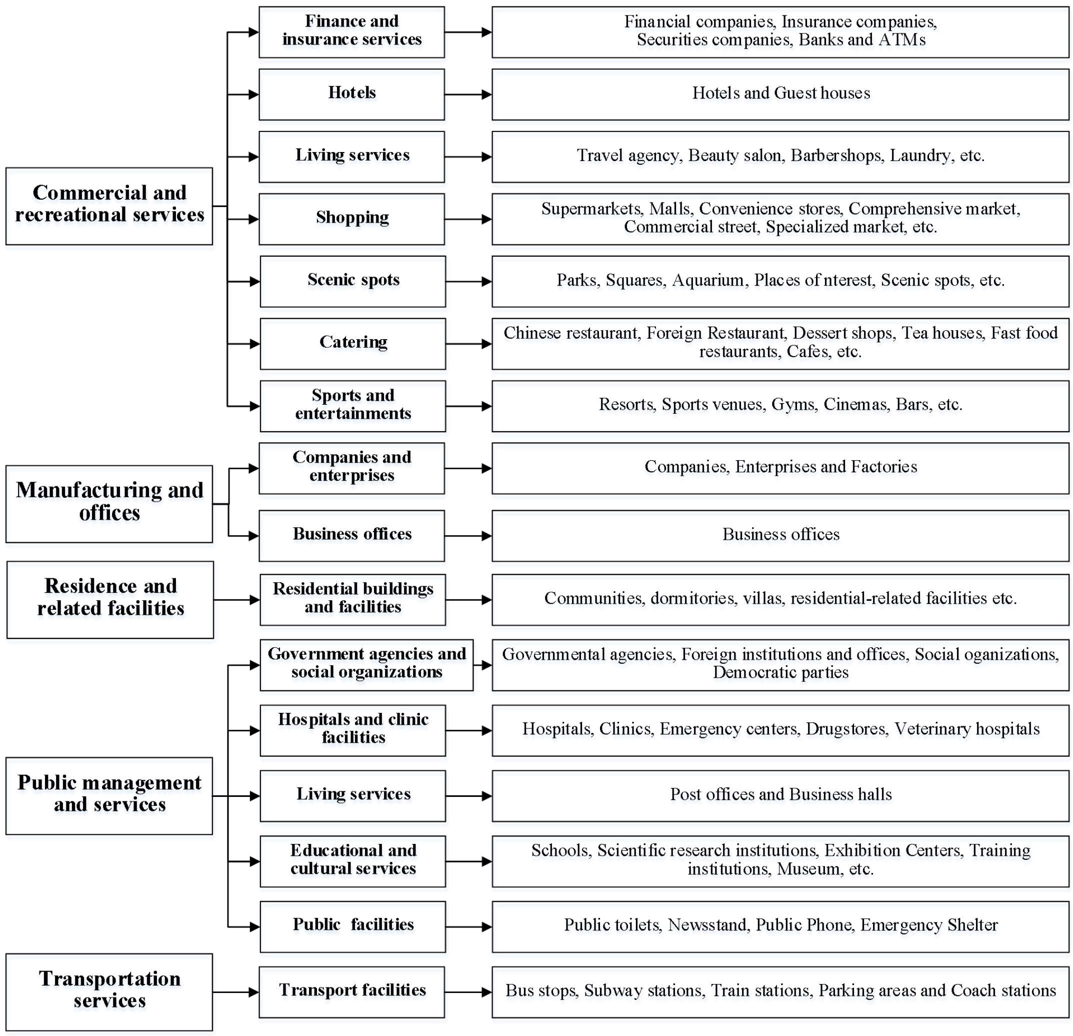

Fig. 2 The reclassification of different categories of POIs

(3) Coverage of subway stations and bus stops

The coordinates of bus stops and subway stations within the study area were obtained based on the POI data. The buffer area of each subway station or bus stop was generated using the coordinates as the center and 400m or 200m as the radius from the subway station or bus stop. The coverage of subway stations (or bus stops) in each cell was calculated as the ratio of the subway (or bus) buffer area to the total land area of the cell.

(4) Average housing price

The original locations and prices of housing units for sale were collected from a housing transaction platform (https://bj.lianjia.com) that is widely used in China. As there are no official

This is an author-produced, peer-reviewed version of this article.
The final publication is available at www.sciencedirect.com. Copyright restrictions may apply
https://doi.org/10.1016/j.cities.2022.103637.

data on the actual transaction prices of real estate in Beijing, this study adopts the housing prices published on this platform. The average housing price of each grid cell was estimated using the prices of all housing units for sale in the cell.

However, housing prices were missing for some cells within our study area because of the lack of housing transactions in those cells as reported by the platform. In previous research, spatial interpolation methods such as the kriging method and inverse distance weighted method are widely used to estimate land prices in unknown areas of cities (Chica-Olmo, 2007; Chica-Olmo, Cano Guervos, & Chica Olmo, 2013; Hu, Cheng, Wang, & Xie, 2012; Hu, Yang, Li, Zhang, & Xu, 2016; Martínez, Lorenzo, &Rubio, 2000; Zhang, Tan, & Tang, 2015). Moreover, the kriging method has been proven to have the advantage of achieving faster and better global predictions when there is limited sample data (Montero-Lorenzo, 2009). Therefore, in this study, we used areal kriging, a kriging-based disaggregation technique in the Geostatistical Analyst extension of ArcGIS10.5, to address the problem of missing data in some cells (Fig.3 (a)) by replacing missing values with interpolated values. This allowed for the collection of data over one set of polygons and predictions for a different set of polygons (Krivoruchko, Gribov, & Krause,2011). The final interpolation result is shown in Fig.3(b).

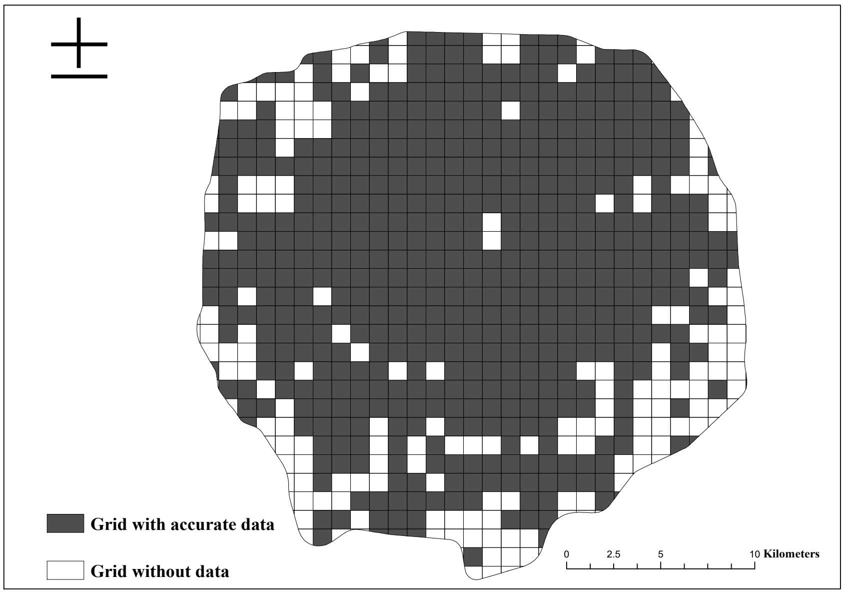

Fig. 3(a)

This is an author-produced, peer-reviewed version of this article.
The final publication is available at www.sciencedirect.com. Copyright restrictions may apply
https://doi.org/10.1016/j.cities.2022.103637.

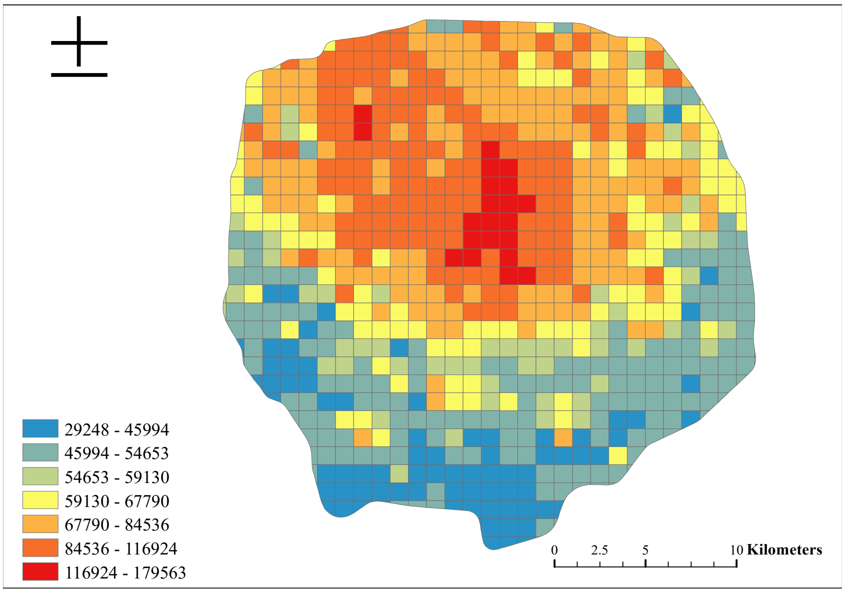

Fig. 3(b)

Fig. 3 The original average housing price data (a) and the predicted average housing prices after areal kriging interpolation (b)

(5) Other built environment factors

Road network data in Beijing were obtained from Open Street Map (www.openstreet.com), a collaborative project that includes a free editable map and geographic data, including street maps, that has been widely used as a data source for road networks (e.g., Wang et al., 2020). Then, the road area within each grid was calculated by multiplying the length of the road, the number of lanes, and the lane width (usually 3-4 m, depending on the class of the road) to obtain the total road area in the grid and the road area ratio of the grid. For example, the width of an expressway with six lanes is estimated to be 30 m. Note that we used the standard hierarchical road system in China to calculate road width.

The spatial data of buildings were obtained from Gaode Map Services (http://lbs.amap.com/), one of the main online map service providers in China. The total number of buildings and the average number of stories of buildings in each grid were aggregated using these data. Due to data limitations, data for all independent variables in this study were obtained for 2018, which were the closest data available to match the taxi trip records for our study period.

3.1.4 Descriptive statistics

The descriptive statistics for all independent and dependent variables are summarized in Table 1.

This is an author-produced, peer-reviewed version of this article.
The final publication is available at www.sciencedirect.com. Copyright restrictions may apply
https://doi.org/10.1016/j.cities.2022.103637.

The average number of taxi trips in the evening peak hours, including both origins and destinations, was larger than that in the morning peak hours, while the maximum number of taxi drop-offs in the evening peak hours was lower than that in the morning peak hours. This may be because people may have less diversified travel purposes and more concentrated and fixed destinations during morning peak hours than during evening peak hours. Additionally, in order to alleviate potential data heterogeneity issues and calculate elasticities, variables are transformed into logarithms in our models, except the ratio of land covered by 200m radius from bus stops in the grid cell, the ratio of land covered by 400m radius from subway stations in the grid cell, and the ratio of road area to total land area in the grid cell. We have tested for potential multicollinearity among explanatory variables using Variance Inflation Factor (VIF). As shown in Appendix 1, the maximum VIF value of all explanatory variables across all four models is 5.7, with a mean around 3.2, indicating no strong multicollinearity exists[3].

Table 1. Summary statistics (Obs. = 683)

Variables	**Mean**	**Std. Dev.**	**Min**	**Max**
Number of taxi trip originations in the morning peak	17.55	18.92	0.00	102.53
Number of taxi trip originations in the evening peak	20.02	22.67	0.00	124.14
Number of taxi trip destinations in the morning peak	15.78	19.57	0.03	226.53
Number of taxi trip destinations in the evening peak	19.80	21.34742	0.00	136.34
Residential density (per km^2)	5161.90	5040.97	16.00	35188.00
Employment density (per km^2)	6707.98	4732.45	14.00	22976.00
Average housing price per m^2 (CNY)	69620.46	21431.22	29248.00	179562.50
Ratio of land covered by 200m radius from bus stops in the grid cell	0.37	0.22	0.00	0.91
Ratio of land covered by 400m radius from subway stations in the grid cell	0.18	0.20	0.00	0.86
Ratio of road area to total land area in the grid cell	0.10	0.053	0.00	0.25
Total number of buildings in a grid cell	340.20	167.76	24.00	1791.00
Average number of stories of buildings in a grid cell	3.79	1.92	1.14	12.91
Number of POIs in commercial and recreational services	52.84	52.46	0.00	390.00
Number of POIs in manufacturing and offices	6.62	8.59	0.00	54.00
Number of POIs in residence and related facilities	18.94	15.68	0.00	87.00
Number of POIs in public management and services	65.33	58.32	0.00	291.00

[3] For VIF cutoff values commonly used by other research, please refer to Montgomery et al., 2012; Gareth et al., 2013; Chatterjee et al., 2015; Zhu, 2021.

This is an author-produced, peer-reviewed version of this article.
The final publication is available at www.sciencedirect.com. Copyright restrictions may apply
https://doi.org/10.1016/j.cities.2022.103637.

3.2 Methods

3.2.1 Test for spatial autocorrelation

Changes in number of taxi trips may be spatially dependent due to geographical proximity. The extent of spatial autocorrelation can be measured using Moran's I index. Following Moran (1948), the global Moran's I index is defined as follows:

$$I = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij} (x_i - \bar{x})(x_j - \bar{x})}{S^2 \sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij}} \tag{1}$$

where S^2 is the sample variance, n is the total number of grid cells, x_i (x_j) is the value of the attribute considered in area i (j), and W_{ij} represents the elements of the spatial weight matrix, which is a binary contiguity matrix in this study. Positive values (usually between 0 and 1) of I indicate a positive spatial autocorrelation in the analyzed variable; that is, high (low) values surround a high (low) value. In contrast, negative values (usually between -1 and 0) of I indicate a negative spatial autocorrelation, with a high (low) value surrounded by low (high) values. If Moran's I value is close to 0, spatial independence is suggested. Strong spatial autocorrelation means that a spatial model should be adopted in order to obtain unbiased estimates.

3.2.2 Spatial econometric model

This research focuses on taxi trip origins and destinations in the morning and evening peak hours. A series of spatial econometric models, such as the general nesting spatial (GNS) model, spatial Durbin model (SDM), spatial lag model (SLM), and spatial error model (SEM), can be used for our analyses. Following the specific-to-general rule to compare different spatial models, we selected SDM to estimate the impacts of built environment factors such as public transit coverage and land development intensity (e.g., the total number of buildings in a grid cell, the average number of stories of buildings in a grid cell), socio-economic variables, and different categories of POIs on taxi demand. The detailed model selection process is presented in section 4.

Overall, SDM includes both endogenous and exogenous interaction effects, which will help protect against omitted variable bias. In addition, LeSage and Pace (2009) pointed out that even if the true data generation progress is the SLM or SEM, the use of SDM will ensure unbiased estimates for the explanatory variable parameters. Thus, this study applied SDM to estimate the impacts of various factors on the spatiotemporal distributions of taxi pick-ups and drop-offs. We divided the analyses into four models – the taxi origin models and taxi destination models during morning peak and evening peak hours. With these SDM models, we can not only estimate the direct effects of the explanatory variables on taxi trip origins and destinations in a local grid cell,

This is an author-produced, peer-reviewed version of this article.
The final publication is available at www.sciencedirect.com. Copyright restrictions may apply
https://doi.org/10.1016/j.cities.2022.103637.

but also measure their impacts on taxi trip O-Ds in neighboring cells (i.e., the spatial spillover effects).

$$TAXI_i = \rho \sum_{j=1}^{N} W_{ij} \times TAXI_j + \beta X_i + \gamma \sum_{j=1}^{N} W_{ij} \times X_j + \mu + \varepsilon \quad (2)$$

$$\varepsilon \sim N(0, \sigma^2 I_n) \quad (3)$$

where ρ represents the endogenous effect, β represents the direct effect of explanatory variables, and γ represents its spatial spillover effect, μ is the intercept, and ε is the error term. In this study, W is specified as a row-normalized binary weight matrix. The off-diagonal elements w_{ij}=1 if units *i* and *j* share a common border, and zero otherwise. I_n, is an N-dimensional identity matrix. *TAXI* represents the number of taxi trip origins or destinations within the grid during the morning peak or evening peak. *X* is a collection of thirteen explanatory variables: residential density, employment density, average housing price, ratio of land covered by 200m radius from bus stops in the grid cell, ratio of land covered by 400m radius from subway stations in the grid cell, ratio of road area to total land area in the grid cell, total number of buildings in a grid cell, average number of stories of buildings in a grid cell, number of POIs in commercial and recreational services, number of POIs in manufacturing and office, number of POIs in residence and related facilities, number of POIs in public management and services, and number of POIs in transportation services. Note that in this study, residential density is only included in the morning peak pick-up model and evening peak drop-off model, while employment density is only included in the morning peak drop-off model and evening peak pick-up model because of the nature of commuting trips.

Additionally, because of the influence of feedback loops, the coefficient in spatial econometric models may be biased, leading to erroneous conclusions. LeSage and Pace (2009) proposed that the point estimations of spatial models may lead to the inaccurate interpretation caused by feedback loop effects, and the partial derivative represents a more valid basis for testing whether the spillover effects exist. Therefore, we further calculate the direct, indirect, and total effects using the spatial decomposition technique. According to the definition, the direct effect is the average extent to which the local outcome variable changes when a particular element of an explanatory variable in that unit itself changes. The indirect effect denotes the average impact of changing a particular element of an explanatory variable on the outcome variable of neighboring units or the average response of the outcome variable to the change in an explanatory variable from neighboring units. The total effect is defined as the sum of these two effects.

4. Results

4.1 Model selection

Several tests were performed to identify the most appropriate model. First, a spatial autocorrelation test was applied to test for spatial dependence of the dependent variables. Table 2 lists the Moran's I

This is an author-produced, peer-reviewed version of this article.
The final publication is available at www.sciencedirect.com. Copyright restrictions may apply
https://doi.org/10.1016/j.cities.2022.103637.

indexes of the taxi O-D points in the four different models. Moran's I index is an indicator of global spatial autocorrelation. As shown in the table, the Moran's I index of all four models exhibits statistically significant and large positive values, indicating that the distribution of taxi O-D points demonstrates high and positive spatial autocorrelation during the morning and evening peaks. Therefore, spatial econometric models that address these spatial effects should be adopted to accurately estimate the impact of different socioeconomic and built environmental factors on taxi trip origins and destinations.

The procedure for model selection follows a general-to-specific rule (Elhorst, 2014). Therefore, the SDM was constructed as a starting point, including the spatial lags of both the dependent and explanatory variables. The Lagrange multiplier (LM) (Burridge, 1980) tests and robust LM tests (Anselin, Bera, Florax, & Yoon, 1996) for spatial lag and spatial error of four different models were applied to test whether the SDM should be degraded to an SLM or SEM. Under the null hypothesis that there is no spatial autocorrelation, the LMlag and LMerr statistics test the spatial lag and spatial errors, respectively (Pelin, 2016). The results are illustrated in Table 2, indicating that both the hypothesis of no spatially lagged dependent variable and the hypothesis of no spatially autocorrelated error term must be rejected at the 1% significance level. Additionally, we calculated the AIC, BIC and Log-likelihood values of all SDM, SLM and SEM models (as shown in Table 3 and Appendix 2). All these indicators suggest SDM consistently outperforms other models. In sum, we can conclude that SDM is the most suitable model for this research.

Table 2. The Global Moran's I, LM Test and Robust LM Test of four different models

	Morning peak taxi pick-ups	Morning peak taxi drop-offs	Evening peak taxi pick-ups	Evening peak taxi drop-offs
Moran's I	0.702***	0.765***	0.741***	0.731***
LMerr	121.535***	93.620***	95.029***	86.268***
Robust LMerr	51.148 ***	28.124***	14.734***	25.225***
LMlag	107.132***	127.982***	171.889***	135.792***
Robust LMlag	36.745***	62.486***	91.595***	74.748***

Note: ***, **, and * indicate 1%, 5%, and 10% confidence levels, respectively.

4.2 Empirical results of SDM

Table 3 reports the regression results of all four SDM models. The coefficients of the spatial lag term (W×Y) are significant and positive in all four models, consistent with our model selection results. Table 4 decomposes the direct effects (local effects), indirect effects (spillover effects) and total effects of all explanatory variables on taxi origins and destinations during morning and evening peak hours.

This is an author-produced, peer-reviewed version of this article.
The final publication is available at www.sciencedirect.com. Copyright restrictions may apply
https://doi.org/10.1016/j.cities.2022.103637.

Table 3. Spatial regression results of four SDM models for taxi pick-ups and drop-offs during morning and evening peak hours

		Morning peak taxi pick-ups	Morning peak taxi drop-offs	Evening peak taxi pick-ups	Evening peak taxi drop-offs
Residential and employment densities	Employment density (log)		0.192*** (6.37)	0.161*** (5.19)	
	Residential density (log)	0.101*** (3.08)			0.078*** (2.87)
Housing price	Average housing price (log)	0.137 (1.28)	0.061 (0.6)	0.183* (1.76)	0.137 (1.54)
Public transportation	Ratio of land covered by 200m radius from bus stops in the grid cell	1.363*** (10.07)	0.834*** (6.60)	1.244*** (9.53)	0.956*** (8.55)
	Ratio of land covered by 400m radius from subway stations in the grid cell	0.098 (0.90)	0.285*** (2.81)	0.156 (1.48)	0.171* (1.90)
Road network density	Ratio of road area to total land area in the grid cell	4.208*** (0.90)	3.567*** (7.48)	4.574*** (9.31)	3.099*** (7.36)
Land development intensity	Total number of buildings in a grid cell (log)	0.014 (0.25)	0.05 (0.95)	0.009 (0.16)	0.038 (0.79)
	Average number of stories of buildings in a grid cell (log)	0.811*** (9.43)	0.412*** (5.20)	0.442*** (5.41)	0.582*** (8.20)
Public services	Number of POIs in commercial and recreational services (log)	0.096*** (2.80)	-0.03 (-0.95)	0.138*** (4.21)	0.208*** (7.40)
	Number of POIs in manufacturing and offices (log)	0.037 (1.37)	0.127*** (4.87)	0.184*** (6.81)	0.092*** (4.15)
	Number of POIs in residence and related facilities (log)	0.093** (2.26)	0.015 (0.40)	-0.002 (-0.06)	0.086** (2.51)
	Number of POIs in public management and services (log)	0.165*** (4.09)	0.117*** (3.10)	0.094** (2.42)	0.098*** (2.93)
	Constant	-3.526*** (-2.99)	-2.391 (-2.16)	-3.867*** (-3.39)	-3.16*** (-3.20)
Residential and employment densities	W×Employment density (log)		-0.019 (-0.25)	-0.049 (-0.64)	
	W×Residential density (log)	-0.017 (-0.24)			0.006 (0.10)
Housing price	W× Average housing price (log)	0.011 (0.17)	-0.013 (-0.2)	0.062 (0.95)	0.067 (1.24)
Public transportation	W×Ratio of land covered by 200m radius from bus stops	-0.893*** (-2.72)	-0.222 (-0.73)	-0.788** (-2.50)	-0.35 (-1.30)
	W×Ratio of land covered by 400m radius from subway stations	-0.481* (-1.77)	-0.546** (-2.17)	-0.652** (-2.51)	-0.636*** (-2.83)

This is an author-produced, peer-reviewed version of this article.
The final publication is available at www.sciencedirect.com. Copyright restrictions may apply
https://doi.org/10.1016/j.cities.2022.103637.

Road network density	W×Road area ratio	-0.942 (-0.88)	0.259 (0.26)	-1.085 (-1.05)	0.657 (0.74)
Land development intensity	W×Total number of buildings in a grid cell (log)	0.093 (0.78)	0.001 (0.00)	-0.073 (-0.64)	-0.088 (-0.88)
	W×Average number of stories of buildings in a grid cell (log)	-0.384** (-1.97)	-0.064 (-0.37)	-0.115 (-0.64)	-0.297* (-1.86)
Public services	W×Number of POIs in commercial and recreational services (log)	0.025 (0.31)	0.02 (0.27)	0.079 (1.02)	0.055 (0.82)
	W×Number of POIs in manufacturing and offices (log)	0.114* (1.83)	-0.026 (-0.41)	-0.035 (-0.53)	0.02 (0.39)
	Number of POIs in residence and related facilities (log)	-0.214** (-2.34)	-0.134 (-1.59)	-0.043 (-0.49)	-0.145** (-1.92)
	W×Number of POIs in public management and services (log)	-0.006 (-0.06)	0.08 (0.95)	0.04 (0.46)	-0.01 (0.00)
	W×Y	0.489*** (9.46)	0.478*** (8.9)	0.429*** (7.79)	0.415*** (7.50)
	AIC	922.15	862.03	823.47	653.30
	BIC	1035.31	975.19	936.63	766.46
	Log-likelihood	-436.07	-406.01	-386.74	-301.65
	Obs.	683	683	683	683

Z-values in parentheses. Note: ***, **, and * indicate 1%, 5%, and 10% confidence levels, respectively.

Table 4. Decomposition of the direct, indirect and total effects of explanatory variables on taxi pick-ups and drop-offs

	Morning peak taxi pick-ups	Morning peak taxi drop-offs	Evening peak taxi pick-ups	Evening peak taxi drop-offs
		Direct effects		
Employment density (log)		0.197*** (6.69)	0.163*** (5.35)	
Residential density (log)	0.103*** (3.21)			0.08*** (3.03)
Average housing price (log)	0.143 (1.32)	0.062 (0.61)	0.192** (1.84)	0.145 (1.62)
Ratio of land covered by 200m radius from bus stops in the grid cell	1.346*** (9.89)	0.848*** (6.69)	1.227*** (9.43)	0.959*** (8.61)
Ratio of land covered by 400m radius from subway stations in the grid cell	0.065 (0.57)	0.254** (2.43)	0.117 (1.1)	0.136 (1.48)
Ratio of road area to total land area in the grid cell	4.296*** (8.51)	3.717*** (7.89)	4.631*** (9.56)	3.221*** (7.76)
Total number of buildings in a grid cell (log)	0.022 (0.39)	0.052 (1.00)	0.004 (0.07)	0.033 (0.71)

This is an author-produced, peer-reviewed version of this article.
The final publication is available at www.sciencedirect.com. Copyright restrictions may apply
https://doi.org/10.1016/j.cities.2022.103637.

Average number of stories of buildings in a grid cell (log)	0.813*** 9.46	0.422*** (5.34)	0.447*** (5.5)	0.579*** (8.22)
Number of POIs in commercial and recreational services (log)	0.101*** (2.96)	-0.03 (-0.94)	0.147*** (4.51)	0.217*** (7.77)
Number of POIs in manufacturing and offices (log)	0.047* (1.76)	0.13*** (4.97)	0.187*** (6.96)	0.096*** (4.37)
Number of POIs in residence and related facilities (log)	0.08* (1.96)	0.006 (0.15)	-0.005 (-0.14)	0.079** (2.34)
Number of POIs in public management and services (log)	0.171*** (4.25)	0.127*** (3.40)	0.099** (2.58)	0.100*** (3.04)
Indirect effects				
Employment density (log)		0.124 (1.04)	0.031 (0.28)	
Residential density (log)	0.056 (0.47)			0.059 (0.68)
Average housing price (log)	0.136 (1.29)	0.027** (0.27)	0.220** (2.25)	0.190** (2.57)
Ratio of land covered by 200m radius from bus stops in the grid cell	-0.392 (-0.73)	0.301 (0.62)	-0.399 (-0.86)	0.071 (0.19)
Ratio of land covered by 400m radius from subway stations in the grid cell	-0.752 (-1.60)	-0.698 (-1.63)	-0.916** (-2.26)	-0.866** (-2.53)
Ratio of road area to total land area in the grid cell	1.944 (1.18)	3.349** (2.23)	1.37 (0.96)	2.979** (2.47)
Total number of buildings in a grid cell (log)	0.175 (0.89)	0.041 (0.23)	-0.109 (-0.66)	-0.111 (-0.78)
Average number of stories of buildings in a grid cell (log)	0.024 (0.08)	0.226 (0.80)	0.117 (0.44)	-0.085 (-0.37)
Number of POIs in commercial and recreational services (log)	0.124 (0.95)	0.009 (0.08)	0.215* (1.92)	0.218** (2.28)
Number of POIs in manufacturing and offices (log)	0.23** (2.21)	0.059 (0.57)	0.069 (0.7)	0.089 (1.18)
Number of POIs in residence and related facilities (log)	-0.293 (-1.94)	-0.215 (-1.58)	-0.069 (-0.53)	-0.168 (-1.52)
Number of POIs in public management and services (log)	0.131 (0.89)	0.232* (1.73)	0.125 (0.98)	0.062 (0.57)
Total effects				
Employment density (log)		0.322*** (2.65)	0.194* (1.7)	
Residential density (log)	0.159 (1.31)			0.139 (1.58)
Average housing price (log)	0.279 (1.62)	0.089 (0.54)	0.412*** (2.63)	0.334*** (2.69)
Ratio of land covered by 200m radius from bus stops in the grid cell	0.954* (1.69)	1.149** (2.25)	0.828* (1.7)	1.03** (2.53)
Ratio of land covered by 400m radius from subway stations in the grid cell	-0.687 (-1.35)	-0.444 (-0.96)	-0.799 (-1.82)	-0.73** (-1.98)
Ratio of road area to total land area in the grid cell	6.24*** (3.61)	7.066*** (4.5)	6.001*** (4.04)	6.20*** (4.95)
Total number of buildings in a grid cell (log)	0.197 (0.98)	0.093 (0.52)	-0.105 (-0.62)	-0.078 (-0.54)
Average number of stories of buildings in a grid cell (log)	0.836 (2.51)	0.648** (2.18)	0.564* (2)	0.494** (2.05)
Number of POIs in commercial and recreational services (log)	0.225 (1.63)	-0.021 (-0.17)	0.362*** (3.07)	0.435*** (4.35)

This is an author-produced, peer-reviewed version of this article.
The final publication is available at www.sciencedirect.com. Copyright restrictions may apply
https://doi.org/10.1016/j.cities.2022.103637.

Number of POIs in manufacturing and offices (log)	0.277** (2.55)	0.189* (1.74)	0.256** (2.49)	0.185** (2.36)
Number of POIs in residence and related facilities (log)	-0.213 (-1.35)	-0.209 (-1.48)	-0.074 (-0.55)	-0.09 (-0.78)
Number of POIs in public management and services (log)	0.302* (1.94)	0.359** (2.56)	0.225* (1.69)	0.162 (1.44)

Z-values in parentheses. Note: ***, **, and * indicate 1%, 5%, and 10% confidence levels, respectively.

This is an author-produced, peer-reviewed version of this article.
The final publication is available at www.sciencedirect.com. Copyright restrictions may apply
https://doi.org/10.1016/j.cities.2022.103637.

4.2.1 Residential and employment densities

As discussed in the methodology section, the morning peak pick-up and evening peak drop-off models include the residential density in a grid cell, while the morning peak drop-off and evening peak pick-up models include the employment density. First, in the morning peak taxi pick-up model, we find residential density has a significant and positive direct effect, indicating that a 1% increase in residential density is associated with a 0.10% increase in the number of taxi trip origins in the local grid. Meanwhile, no significant indirect effects are found. These results are similar to the findings of Zhang et al. (2018), who focused on influencing factors to determine the ridership distribution of taxi services in New York City. Commuters from home to workplaces or to transit stations are probably the main group of passengers that taxis pick up in the morning peak hours; thus, locations with a higher residential density usually have larger taxi demand.

In the morning peak taxi drop-off model, employment density shows a similar tendency, with significant and positive direct impacts. A 1% increase in employment density can directly increase morning taxi trip drop-offs in the local grid by 0.20%. Additionally, such an increase will lead to a 0.32% rise in morning taxi drop-offs for the entire research area, according to the estimated total effects. We did not find significant spillover effects, suggesting that employment density of a grid cell or block does not affect the number of taxi trip drop-offs in surrounding areas. This is reasonable because taxis provide door-to-door services and the drop-off locations of most taxi trips are close to the destinations that the passengers want to go (e.g., workplace, transit station).

In the evening peak taxi pick-up model, the direct effect of employment density is also significant and positive, indicating that a 1% increase in employment density can increase the number of taxi pick-ups in local grids by 0.16%. This confirms that higher job density incurs more taxi demand. Meanwhile, the insignificant spillover effects suggest again that this variable has a negligible impact on taxi demand in the surrounding areas.

Lastly, in the evening peak taxi drop-off model, residential density is also found to have a significant and positive direct effect on taxi drop-offs, indicating that a 1% increase in residential density can increase taxi drop-offs in the local grid by 0.08%. This finding is in line with Liu et al. (2012), who investigated the temporal variations of pick-ups and drop-offs and related them to different land-use features. They claimed that a typical residential area is a source area for taxi trips in the morning but a sink area in the evening.

Overall, it is interesting to note that both residential density and employment density only have statistically significant direct effects but no significant spillover effects in all models. Moreover, cross-model comparisons (based on Chow test) suggest that the elasticity of taxi ridership (as measured in pick-ups and drop-offs) with respect to employment density is slightly larger than the elasticity with respect to residential density. This might be because taxi trips in employment

This is an author-produced, peer-reviewed version of this article.
The final publication is available at www.sciencedirect.com. Copyright restrictions may apply
https://doi.org/10.1016/j.cities.2022.103637.

(sub)centers are more clustered than in residential areas.

4.2.2 Public transportation

Regarding public transportation factors, bus and subway coverage have different effects on taxi ridership. Bus coverage has significant and positive direct effects in all four models, but no significant indirect effects. Meanwhile, subway coverage shows positive direct effects on taxi drop-offs in the morning peak, as well as significant and negative indirect effects on taxi pick-ups and drop-offs in the evening peak. Our estimates indicate that, in the morning peak models, a 10 percentage point increase in the ratio of bus 200-meter or transit 400-meter catchment areas leads to a 8.85% (note that $100*(e^{0.0848}-1) = 8.850$) and a 2.57% increase in taxi drop-offs in the local grid, respectively, while they both have no significant spillover effects on surrounding grids. In the evening peak models, bus coverage has similar direct effects on taxi pick-ups and drop-offs compared to morning peak, with no statistically significant spillover effects; for subway coverage, only the spillover effects are significant -- a 10 percentage point increase in the ratio of subway 400-meter catchment area would lead to a 8.75% decrease in taxi pick-ups and a 8.30% decrease in drop-offs in adjacent grids.

A comparison of the results across all four models suggests that, while bus and subway both have some level of positive direct effect on taxi pick-ups or drop-offs during different times of the day, they show different patterns of indirect effects. Subway coverage has a significant and negative indirect effect, suggesting it may absorb taxi demand from surrounding areas, whereas bus coverage's indirect effect is statistically insignificant. Certainly, a possible explanation for the insignificant spillover effects of buses is that the effective service range of a bus stop is typically smaller than our unit of analysis (i.e., 1km*1km grids), such that the spillover effects cannot be identified at this spatial scale. Analyses at a finer scale may be able to detect indirect effects of bus coverage on taxi O-Ds.

4.2.3 Land development intensity

Land development intensity is measured by the total number of buildings and the average number of stories in a grid cell. While results show that the number of buildings has neither direct nor indirect effects on taxi O-Ds, the average number of stories is estimated to have significant and positive direct effects in all four models. Note that the indirect effects of the average number of stories are also insignificant. In the morning peak hours, a 1% growth in the average number of stories in a grid can increase taxi pick-ups in that grid by 0.81% and taxi drop-offs by 0.42%. In the evening peak hours, a 1% growth in the average number of stories can increase taxi pick-ups and drop-offs in the local grid by 0.45% and 0.58%, respectively. These results suggest that areas containing more high-rise buildings (e.g., residential, commercial and office) are likely to generate more taxi demand, while the number of buildings have little impact after all other variables have

This is an author-produced, peer-reviewed version of this article.
The final publication is available at www.sciencedirect.com. Copyright restrictions may apply
https://doi.org/10.1016/j.cities.2022.103637.

been controlled.

4.2.4 Housing prices

For average housing prices, we find significant positive direct and indirect effects on taxi O-Ds in the evening peak pick-up model. A 1% increase in housing prices increases evening taxi pick-ups both in the local grid and in surrounding grids by 0.19% and 0.22%, respectively. Interestingly, we also find that housing prices have little influence on taxi ridership in the morning peak, after controlling for other variables. Although people living in more expensive communities may be financially more capable of taking taxis, they are also more likely to drive their own cars in the morning peak hours, hence offsetting their need for taxis. Note that in the evening peak taxi drop-off model, this variable also shows no significant direct effects, which is consistent with the morning peak taxi pick-up model, because commuters leave home in the morning and return home in the evening.

4.2.5 Road network density

The road network density, as measured by the ratio of road area to total land area in the grid cell, is a good indicator of urban transport infrastructure quality. We find this variable has significant positive direct and indirect effects on taxi O-Ds in all models, suggesting that an improvement in road network density will increase taxi demand in both the local area and surrounding areas. This is likely because areas with a dense road network, such as business districts and urban centers, usually have high concentration of jobs and large travel demand. In addition, a higher exposure to available taxis on the road and less waiting time for passengers may also increase taxi ridership.

4.2.6 Points of Interest

In general, various POIs show different direct effects and indirect effects on taxi O-Ds. First, the number of POIs in public management and services have strong positive impacts on local taxi pick-ups and drop-offs in both peak times. For example, a 1% increase in the number of public management and service POIs will directly increase local taxi pick-ups by 0.17% and drop-offs by 0.13% in the morning, but has no significant spillover effects. Moreover, the number of POIs in manufacturing and offices and the number of POIs in commercial and recreational services both show significant and positive direct effects on taxi O-Ds, but the former has some positive spillover effects in the morning peak and the latter only has (positive) spillovers in the evening peak.

Note that we did not consider the POIs for transportation service in our main models because of their potential collinearity with bus and subway coverage. However, as a robustness check, we

This is an author-produced, peer-reviewed version of this article.
The final publication is available at www.sciencedirect.com. Copyright restrictions may apply
https://doi.org/10.1016/j.cities.2022.103637.

also tested adding the number of POIs for parking (i.e., a subset of transportation services that is less likely to be correlated with bus and subway coverage) into our models. The estimation results are shown in Appendix 3. Although the VIF of parking POIs is as high as 7.25, we find no major differences in model estimates compared to our original models, in terms of both the sign and the magnitude of most coefficient estimates.

5. Conclusions

This research provides some insight into the complex spatial and temporal patterns of taxi ridership via the application of a variety of big data and a comprehensive investigation of its relationships with built environment and neighborhood socioeconomic factors. Moreover, we adopt a spatial econometrics model to examine not only the direct effects but also the spillover effects of these factors on the spatial and temporal variations of taxi pick-ups and drop-offs. The findings have important implications for urban planners and policymakers in their efforts to improve the balance between taxi services and demand, reduce traffic congestion, and enhance the efficiency of the multi-modal transportation system. Based on multi-sourced big data, this study also provides a useful framework to generate various built environment variables that are not directly provided by government agencies in many countries, such as residential and employment densities, the number of POIs, public transit stations, and median housing price.

Our results show that road network density has the largest impact on taxi ridership. It increases both taxi pick-ups and drop-offs during morning and evening peak hours. More importantly, it increases taxi O-Ds not only in the local grid cell but also in surrounding cells. These findings suggest that simply increasing the road density in one small area but not the surrounding areas may cause traffic bottlenecks and result in traffic congestion in the whole area. Transportation planners should comprehensively consider the road layout in the entire area in order to reduce traffic congestion.

We also find that the two public transit modes have different effects on taxi ridership. Bus coverage has significant and positive direct effects on taxi O-Ds during both morning and evening peak hours but no spillover effects. Meanwhile, subway coverage has significant and positive direct effects in the morning peak taxi drop-off model, but its spillover effects are found to be significant and negative in two evening peak models. While the strong negative indirect effects of subway coverage indicate that the subway stations may absorb taxi demand in surrounding grids, the positive direct effects of both bus and subway coverage suggest that they may increase taxi demand in the local grid. Such findings are useful for designing and improving the multi-modal transportation system to better integrate bus, subway, and taxi usage, in order to enhance mobility, reduce traffic congestion, and promote environmental sustainability.

This is an author-produced, peer-reviewed version of this article.
The final publication is available at www.sciencedirect.com. Copyright restrictions may apply https://doi.org/10.1016/j.cities.2022.103637.

As for land development intensity, the results show that the direct effects of the average number of stories on taxi ridership are significant and positive in all four models, which is likely related to the large demand for taxis generated by people living or working in high-rise buildings where parking is often insufficient. Transportation planners need to pay special attention to the development of high-rise buildings and make sure the transportation impact analysis of these projects takes into account the additional taxi demand caused by such development.

Moreover, both residential and employment densities have significant and positive direct effects on taxi O-Ds in two peak periods, but no significant indirect effects are found. As many large cities around the world continue to face increasing residential density and employment density, planners need to improve their travel demand models to specifically incorporate the extra taxi trips generated by the higher densities, thereby improving the accuracy of those forecasting models.

Lastly, different categories of POIs also affect taxi ridership, among which the POIs in transportation services consistently have positive and direct effects in all four models. Other types of POIs, such as those in public management and services and those in manufacturing and offices, also show some level of positive direct effect on taxi pick-ups and drop-offs. These findings suggest that urban travel demand modeling should also take into account the number of various POIs because they may affect taxi ridership and hence influence traffic.

To some extent, ride-sourcing services such as DiDi in China and Uber are quite similar to taxi services, especially when many taxi services around the world have launched their own mobile apps in recent years to better assist dispatching vehicles to areas with high demand for taxis. With ride-sourcing services becoming increasingly popular around the world, this research also has important planning implications for better integrating these services into the existing multi-modal transportation system and improving overall transport efficiency.

References

Al-Ayyash, Z., Abou-Zeid, M., Kaysi, I., 2016. Modeling the demand for a shared-ride taxi service: An application to an organization-based context. *Transport Policy*, 48, 169–182.

Anselin, L. (1988). *Spatial econometrics: Methods and models*. Kluwer Academic, Boston.

Anselin, L., Bera, A. K., Florax, R., & Yoon, M. J. (1996). Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics*, 26, 77-104.

Antipova, A., Wang, F., & Wilmot, C. (2011). Urban land uses, socio-demographic attributes and commuting: A multilevel modeling approach. *Applied Geography*, 31, 1010–1018.

Beijing Municipal Bureau of Statistics. (2016). *Beijing Statistical Yearbook*. Beijing: China

This is an author-produced, peer-reviewed version of this article.
The final publication is available at www.sciencedirect.com. Copyright restrictions may apply
https://doi.org/10.1016/j.cities.2022.103637.

Statistics Press.

Burridge, P. (1980). On the cliff-ord test for spatial correlation. *Journal of Royal Statistical Society*, Series B, 42, 107-108.

Cervero, R. (2002). Built environments and mode choice: toward a normative framework. *Transportation Research Part D: Transport and Environment*, 7(4), 265–284.

Çetin, T., & Yasin Eryigit, K. (2011). Estimating the effects of entry regulation in the Istanbul taxicab market. *Transportation Research Part A: Policy and Practice*, 45(6), 476–484.

Chatterjee, S., & Hadi, A. S. (2015). Regression analysis by example. John Wiley & Sons.

Chen, C., Feng, T., Ding, C., Yu, B., & Yao, B. (2021). Examining the spatial-temporal relationship between urban built environment and taxi ridership: Results of a semi-parametric GWPR model. Journal of Transport Geography, 96, 103172.

Chica-Olmo, J. (2007). Prediction of housing location price by a multivariate spatial method: cokriging. *Journal of Real Estate Research*, 29(1), 95-114.

Chica-Olmo, J., Cano-Guervos, R., & Chica-Olmo, M. (2013). A Coregionalized Model to Predict Housing Prices. *Urban Geography*, 34(3), 395–412.

Ding, C., & Zhao, X. (2014). Land market, land development and urban spatial structure in Beijing. *Land Use Policy*, 40, 83-90.

Ding, L., Niu, X, & Song, X. (2016). Measuring the employment center system in Shanghai central city: A study using mobile phone signaling data. *Acta Geographica Sinica*, 71(3), 484-499.

Elhorst, J.P. (2014). *Spatial Econometrics: From Cross Sectional Data to Spatial Panels*. Heidelberg: Springer Press.

Ewing, R., & Cervero, R. (2010). Travel and the Built Environment. *Journal of the American Planning Association*, 76(3), 265–294.

Ewing, R., Tian, G., Goates, J.P., Zhang, M., Greenwald, M.J., Joyce, A., Kircher, J., Greene, W., (2015). Varying influences of the built environment on household travel in15 diverse regions of the United States. *Urban Stud*. 52, 2330–2348.

Gareth, J., Daniela, W., Trevor, H., & Robert, T. (2013). *An introduction to statistical learning: with applications in R*. Spinger.

This is an author-produced, peer-reviewed version of this article.
The final publication is available at www.sciencedirect.com. Copyright restrictions may apply
https://doi.org/10.1016/j.cities.2022.103637.

Gong, L., Liu, X., Wu, L., Liu, Y., 2016. Inferring trip purposes and uncovering travel patterns from taxi trajectory data. *Cartography and Geographic Information Science*, 43(2), 1–12.

Gonzales, E. J., Yang, C., Morgul, E. F., Ozbay, K., & Mineta National Transit Research Consortium. (2014). Modeling taxi demand with GPS data from taxis and transit (No. CA-MNTRC-14-1141). *Mineta National Transit Research Consortium*.

Hu, S., Cheng, Q., Wang, L., & Xie, S. (2012). Multifractal characterization of urban residential land price in space and time. *Applied Geography*, 34, 161–170.

Hu, S., Yang, S., Li, W., Zhang, C., & Xu, F. (2016). Spatially non-stationary relationships between urban residential land price and impact factors in Wuhan city, China. *Applied Geography*, 68, 48–56.

Hu, S., Gao, S., Wu, L., Xu, Y., Zhang, Z., Cui, H., & Gong, X. (2021). Urban function classification at road segment level using taxi trajectory data: A graph convolutional neural network approach. *Computers, Environment and Urban Systems*, 87, 101619.

Jager, B., Wittmann, M., Lienkamp, M. (2016). Analyzing and modeling a city's spatiotemporal taxi supply and demand: a case study for Munich. *Journal of Traffic and Logistics Engineering*. 4(2), 147–153.

Keler, A., Krisp, J. M., & Ding, L. (2020). Extracting commuter-specific destination hotspots from trip destination data–comparing the boro taxi service with Citi Bike in NYC. *Geo-Spatial Information Science*. https://doi.org/10.1080/ 10095020.2019.1621008.

Kim, K., 2018. Exploring the difference between ridership patterns of subway and taxi: Case study in Seoul. *Journal of Transport Geography*. 66, 213–223.

Kong, X., Liu, Y., Wang, Y., Tong, D., & Zhang, J. (2017). Investigating public facility characteristics from a spatial interaction perspective: A case study of Beijing hospitals using taxi data. *ISPRS International Journal of Geo-Information*, 6(2), 38.

Krivoruchko, K., Gribov, A., & Krause, E. (2011). Multivariate areal interpolation for continuous and count data. *Procedia Environmental Sciences*, 3, 14-19.

Lavieri, P. S., Dias, F. F., Juri, N. R., Kuhr, J., & Bhat, C. R. (2018). A model of ridesourcing demand generation and distribution. *Transportation Research Record*, 2672(46), 31-40.

LeSage, J. and Pace, R.K. (2009). *Introduction to Spatial Econometrics*. New York: CRC Press.

This is an author-produced, peer-reviewed version of this article.
The final publication is available at www.sciencedirect.com. Copyright restrictions may apply
https://doi.org/10.1016/j.cities.2022.103637.

Li, B., Cai, Z., Jiang, L., Su, S., & Huang, X. (2019). Exploring urban taxi ridership and local associated factors using GPS data and geographically weighted regression. *Cities*, 87, 68-86.

Liu, Q., Ding, C., & Chen, P. (2020). A panel analysis of the effect of the urban environment on the spatiotemporal pattern of taxi demand. *Travel Behaviour and Society*, 18, 29-36.

Liu, X., Gong, L., Gong, Y., & Liu, Y. (2015). Revealing travel patterns and city structure with taxi trip data. *Journal of Transport Geography*, 43, 78-90.

Liu, Y., Wang, F., Xiao, Y., & Gao, S. (2012). Urban land uses and traffic 'source-sink areas': Evidence from GPS-enabled taxi data in Shanghai. *Landscape and Urban Planning*, 106(1), 73-87.

Liu, Y., Singleton, A., Arribas-bel, D., & Chen, M. (2021). Identifying and understanding road-constrained areas of interest (AOIs) through spatiotemporal taxi GPS data: A case study in New York City. Computers, *Environment and Urban Systems*, 86, 101592.

Liu, Q., Ding, C., & Chen, P. (2020). A panel analysis of the effect of the urban environment on the spatiotemporal pattern of taxi demand. *Travel Behaviour and Society*, 18, 29-36.

Louail, T., Lenormand, M., Ros, O. G. C., Picornell, M., Herranz, R., Frias-Martinez, E., Ramasco, J.J. & Barthelemy, M. (2014). From mobile phone data to the spatial structure of cities. *Scientific reports*, 4, 5276.

Matías Gámez Martínez, José María Montero Lorenzo, & Noela García Rubio. (2000). Kriging methodology for regional economic analysis: estimating the housing price in albacete. *International Advances in Economic Research*, 6(3), 438-450.

McNally, M. G. (2008). *The four-step model*. UC Irvine: Center for Activity Systems Analysis. Retrieved from http://escholarship.org/uc/item/0r75311t.

Montero-Lorenzo, J.-M., Larraz-Iribas, B., & Páez, A. (2009). Estimating commercial property prices: an application of cokriging with housing prices as ancillary information. *Journal of Geographical Systems*, 11(4), 407–425.

Montgomery, D., Peck, E., & Vining, G. (2012). *Wiley series in probability and mathematical statistics* (2nd ed.). New York: John Wiley & Sons, 1992. Introduction to linear regression analysis.

Moran, P. A. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society: Series B (Methodological)*, 10(2), 243-251.

This is an author-produced, peer-reviewed version of this article.
The final publication is available at www.sciencedirect.com. Copyright restrictions may apply
https://doi.org/10.1016/j.cities.2022.103637.

Nam, D., Hyun, K., Kim, H., Ahn, K., & Jayakrishnan, R. (2016). Analysis of grid cell–based taxi ridership with large-scale GPS data. *Transportation Research Record*, 2544(1), 131-140.

Neoh, J.G., Chipulu, M., Marshall, A. (2017) What encourages people to carpool? an evaluation of factors with meta-analysis. *Transportation*. 44, 423–447.

Ni, Y., & Chen, J. (2020). Exploring the Effects of the Built Environment on Two Transfer Modes for Metros: Dockless Bike Sharing and Taxis. *Sustainability*, 12(5), 2034.

Openshaw, S. and Taylor, P. J. 1981. The modifiable areal unit problem. *In Quantitative Geography*. London: Routledge & Kegan Paul.

Pan, R., Zhang, S., Yang, H., Xie, K., & Wen, Y. (2019, October). Analysis of Spatial Equity in Taxi Services: A Case Study of New York City. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)* (pp. 2659-2664). IEEE.

Pelin Akçagün. (2016). Provincial growth in turkey: a spatial econometric analysis. *Applied Spatial Analysis and Policy*, 10(2), 1-29.

Qian, X., & Ukkusuri, S. V. (2015). Spatial variation of the urban taxi ridership using GPS data. *Applied Geography*, 59, 31–42.

Schaller, B. (2005). A regression model of the number of taxicabs in US cities. *Journal of Public Transportation*, 8(5), 63.

Ulak, M. B. , Yazici, A. , & Aljarrah, M. . (2020). Value of convenience for taxi trips in new york city. *Transportation Research Part A Policy and Practice*, 142, 85-100.

Wang, F., Antipova, A., & Porta, S. (2011). Street centrality and land use intensity in Baton Rouge, Louisiana. *Journal of Transport Geography*, 19, 285–293.

Wang, M., & Mu, L. (2018). Spatial disparities of Uber accessibility: An exploratory analysis in Atlanta, USA. *Computers Environment and Urban Systems, 67*, 169-175.

Wang, M., Chen, Z., Mu, L. & Zhang, X. (2020). Road network structure and ride-sharing accessibility: Evidence from a network science perspective. *Computers Environment and Urban Systems, 80*, 101430.

Willis, G., & Tranos, E. (2021). Using 'Big Data' to understand the impacts of Uber on taxis in New York City. *Travel Behaviour and Society*, 22, 94–107.

Wang, J., Huang, J., & Du, F. (2020). Estimating spatial patterns of commute mode preference

This is an author-produced, peer-reviewed version of this article.
The final publication is available at www.sciencedirect.com. Copyright restrictions may apply
https://doi.org/10.1016/j.cities.2022.103637.

in Beijing. Regional Studies, *Regional Science*, 7(1), 382-386.

Wang, S., & Noland, R. B. (2021). Variation in ride-hailing trips in Chengdu, China. Transportation Research Part D: Transport and Environment, 90, 102596.

Yang, C., & Gonzales, E. J. (2014). Modeling taxi trip demand by time of day in New York City. *Transportation Research Record*, 2429(1), 110-120.

Yao, X., Wu, L., Zhu, D., Gao, Y., & Liu, Y. (2019). Visualizing spatial interaction characteristics with direction-based pattern maps. *Journal of Visualization*, 1-15.

Yao, Y., Liu, X., Li, X., Zhang, J., Liang, Z., & Mai, K., et al. (2017). Mapping fine-scale population distributions at the building level by integrating multisource geospatial big data. *International Journal of Geographical Information Science*, 31(5-6), 1220-1244.

Yu, H., & Peng, Z. R. (2019). Exploring the spatial variation of ridesourcing demand and its relationship to built environment and socioeconomic factors with the geographically weighted poisson regression. *Journal of Transport Geography*. 75(2019), 147-163.

Yuan, C., Duan, Y., Mao, X., Ma, N. & Zhao, J. (2021). Impact of the mixed degree of urban functions on the taxi travel demand. PLOS ONE, 16 (3), s. e0247431. doi:10.1371

Zhang, W., Le, T. V., Ukkusuri, S. V., & Li, R. (2018). Influencing factors and heterogeneity in ridership of traditional and app-based taxi systems. *Transportation*.47:971–996.

Zhang, W., Le, T. V., Ukkusuri, S. V., & Li, R. (2020). Influencing factors and heterogeneity in ridership of traditional and app-based taxi systems. *Transportation*, 47(2), 971-996.

Zhang, Z., Tan, S., & Tang, W. (2015). A GIS-based spatial analysis of housing price and road density in proximity to urban lakes in Wuhan City, China. *Chinese Geographical Science,* 25(6), 775–790.

Zhou, Y., Fang, Z., Thill, J. C., Li, Q., & Li, Y. (2015). Functionally critical locations in an urban transportation network: Identification and space-time analysis using taxi trajectories. *Computers, Environment and Urban Systems*.52:34-47.

Zhou, X., Wang, M., & Li, D. (2019). Bike-sharing or taxi? Modeling the choices of travel mode in Chicago using machine learning. *Journal of Transport Geography*. 79:102479.

Zhu, B., Huang, Q., Guibas, L., & Zhang, L. (2013). Urban population migration pattern mining based on taxi trajectories. In *3rd international workshop on mobile sensing: the future, brought to you by big sensor data*, Philadelphia, USA.

This is an author-produced, peer-reviewed version of this article.
The final publication is available at www.sciencedirect.com. Copyright restrictions may apply
https://doi.org/10.1016/j.cities.2022.103637.

Zhu, P., Ho, S. N., Jiang, Y., & Tan, X. (2020). Built environment, commuting behaviour and job accessibility in a rail-based dense urban context. *Transportation Research Part D: Transport and Environment*, 87, 102438.

Zhu, P., Wang, L., Jiang, Y., & Zhou, J. (2018). Metropolitan size and the impacts of telecommuting on personal travel. *Transportation*, 45(2), 385-414.

Zhu, P., Zhao, S., Wang, L., & Al Yammahi, S. (2017). Residential segregation and commuting patterns of migrant workers in China. *Transportation Research Part D: Transport and Environment*, 52, 586-599.

Zhu, P. (2013). Telecommuting, household commute and location choice. *Urban Studies*, 50(12), 2441-2459.

Zhu, P., Dong, H., & Wu, C. (2013). Do Residents of Smart Growth Neighborhoods in Los Angeles, California, Travel "Smarter"? Transportation research record, 2397(1), 61-71.

Zhu, P. (2012). Are telecommuting and personal travel complements or substitutes? The Annals of Regional Science, 48(2), 619-639.

Appendix 1 Results of VIF test.

	Morning peak taxi pick-ups	**Morning peak taxi drop-offs**	**Evening peak taxi pick-ups**	**Evening peak taxi drop-offs**
Employment density (log)		4.09	4.09	
Residential density (log)	3.27			3.27
Average housing price (log)	1.53	1.59	1.59	1.53
Ratio of 200m radius from bus stops to grid cell	2.57	2.58	2.58	2.57
Ratio of 400m radius from subway stations to grid cell	1.44	1.44	1.44	1.44
Road area ratio	1.96	1.94	1.94	1.93
Total number of buildings in a grid cell (log)	2.58	2.54	2.54	2.58
Average number of stories of buildings in a grid cell (log)	3.04	3.00	3.00	3.04
Number of POIs in commercial and recreational services(log)	5.25	5.22	5.22	5.25
Number of POIs in manufacturing and offices (log)	1.93	2.70	2.93	2.36
Number of POIs in residence and related facilities (log)	5.43	5.25	5.25	5.43
Number of POIs in public management and services (log)	5.70	5.68	5.68	5.70
Mean VIF	3.19	3.28	3.28	3.19

Appendix 2 Results of SLM and SEM Models

		Morning peak taxi pick-ups		Morning peak taxi drop-offs		Evening peak taxi pick-ups		Evening peak taxi drop-offs	
		SLM	SEM	SLM	SEM	SLM	SEM	SLM	SEM
Residential and employment densities	Employment density (log)			0.18***	0.03***	0.15***	0.22***		
	Residential density (log)	0.08***	0.03***					0.02***	4.84***
Housing price	Average housing price (log)	0.10	0.1***	0.1***	0.06	0.15*	0.42***	0.07***	4.47***
Public transportation	Ratio of 200m radius from bus stops to grid cell	1.27***	0.14***	0.14***	0.81***	1.16***	1.23***	0.11***	8.02***
	Ratio of 400m radius from subway stations to grid cell	0.03	0.11*	0.11*	0.22**	0.06	0.27***	0.09	3.1***
Road network density	Road area ratio	4.49***	0.51***	0.51***	3.93***	4.68***	4.74***	0.39***	7.48***
Land development intensity	Total number of buildings in a grid cell (log)	0.07	0.06	0.06	0.06	-0.02	-0.03	0.04	-0.10
	Average number of stories of buildings in a grid cell (log)	0.76***	0.09***	0.09***	0.41***	0.42***	0.48***	0.07***	8.42***
Public services	Number of POIs in commercial and recreational services(log)	0.12***	0.03***	0.03***	-0.01	0.17***	0.12***	0.03***	6.42***
	Number of	0.04*	0.03***	0.03***	0.12***	0.17***	0.2***	0.02***	5.37***

POIs in manufacturing and offices (log)								
Number of POIs in residence and related facilities (log)	0.05	0.04***	0.04***	-0.01	-0.02	0.02	0.03*	2.88***
Number of POIs in public management and services (log)	0.16***	0.04***	0.04***	0.12***	0.08**	0.15***	0.03***	4.48***
Constant	-3.33***	1.20***	1.20***	-2.49***	-3.28***	-6.45***	0.79***	-5.66***
W×Y	0.26***		0.05***		0.28***		0.02***	
W×E		0.05***		0.06***		0.57***		9.57***
Obs.	683	683	683	683	683	683	683	683
AIC	942.837	941.77	816.49	891.47	866.09	905.46	661.37	719.57
BIC	1006.21	1005.14	879.86	954.84	929.46	929.46	724.74	782.94
Log-likelihood	-457.42	-456.88	-394.24	-431.74	-419.04	-438.73	-316.69	-345.78
R-square	0.8897	0.8817	0.8897	0.8686	0.9091	0.8966	0.9204	0.9073

Note: The superscripts ***, **, and * indicate that the coefficient is statistically significant at the 1%, 5%, and 10% level

Appendix 3 Robustness check: SDM estimation results (with POIs in parking services included)

		Morning peak taxi pick-ups	Morning peak taxi drop-offs	Evening peak taxi pick-ups	Evening peak taxi drop-offs
Residential and employment densities	Employment density (log)		0 .17*** (5.85)	0.14*** (4.63)	
	Residential density (log)	0.10** (3.18)			0.08** (3.01)
Housing price	Average housing price (log)	0.15 (1.47)	0.09 (0.98)	0.21* (2.13)	0.15 (1.72)
Public transportation	Ratio of land covered by 200m radius from bus stops in the grid cell	1.22*** (9.17)	0.68*** (5.56)	1.09*** (8.62)	0.82*** (7.51)
	Ratio of land covered by 400m radius from subway stations in the grid cell	0.12 (1.12)	0.31** (3.24)	0.19* (1.84)	0.19* (2.23)
Road network density	Ratio of road area to total land area in the grid cell	3.79*** (7.60)	3.14*** (6.84)	4.15*** (8.76)	2.68*** (6.59)
Land development intensity	Total number of buildings in a grid cell (log)	-0.01 (-0.15)	0.03 (0.59)	-0.01 (-0.23)	0.01 (0.32)
	Average number of stories of buildings in a grid cell (log)	0.68*** (7.92)	0.27*** (3.44)	0.30*** (3.73)	0.45*** (6.46)
Public services	Number of POIs in commercial and recreational services (log)	0.04 (1.26)	-0.09** (-2.79)	0.08* (2.55)	0.16*** (5.66)
	Number of POIs in manufacturing and offices (log)	-0.01 (-0.48)	0.08** (3.05)	0.14*** (5.10)	0.044* (1.97)
	Number of POIs in residence and related facilities (log)	0.07 (1.68)	-0.01 (-0.36)	-0.03 (-0.81)	0.06 (1.85)
	Number of POIs in public management and services (log)	0.11** (2.81)	0.06 (1.55)	0.03 (0.92)	0.05 (1.39)
	Number of POIs in parking services (log)	0.25*** (6.51)	0.29*** (8.06)	0.28*** (7.68)	0.25*** (7.86)
	Constant	-3.49** (-3.04)	-2.42* (-2.28)	-3.87*** (-3.52)	-3.06** (-3.22)
Residential and employment densities	W×Employment density (log)		-0.07 (-0.90)	-0.04 (-0.54)	
	W×Residential density (log)	-0.07 (-1.01)			-0.04 (-0.75)
Housing price	W× Average housing price (log)	-0.02 (-0.37)	-0.06 (-1.01)	0.01 (0.22)	0.03 (0.65)
Public transportation	W×Ratio of land covered by 200m radius from bus stops	-0.85*** (-2.67)	-0.17 (-0.58)	-0.74* (-2.44)	-0.31 (-1.19)

	W×Ratio of land covered by 400m radius from subway stations	-0.55**	-0.59*	-0.69**	-0.69***
		(-2.08)	(-2.46)	(-2.77)	(-3.21)
Road network density	W×Road area ratio	-0.53	0.86	-0.52	1.06
		(-0.50)	(0.90)	(-0.52)	(1.25)
Land development intensity	W×Total number of buildings in a grid cell (log)	0.21	0.11	0.03	0.02
		(1.69)	(0.96)	(0.27)	(0.19)
	W×Average number of stories of buildings in a grid cell (log)	-0.21*	0.11	0.06	-0.12
		(-1.08)	(0.64)	(0.33)	(-0.78)
Public services	W×Number of POIs in commercial and recreational services (log)	0.04	0.03	0.08	0.06
		(0.45)	(0.39)	(1.10)	(0.98)
	W×Number of POIs in manufacturing and offices (log)	0.12*	-0.01	-0.01	0.03
		(1.89)	(-0.10)	(-0.23)	(0.58)
	Number of POIs in residence and related facilities (log)	-0.22**	-0.15	-0.06	-0.15**
		(-2.43)	(-1.84)	(-0.70)	(-2.01)
	W×Number of POIs in public management and services (log)	0.04	0.13	0.09	0.05
		(0.41)	(1.52)	(1.04)	(0.66)
	W×Number of POIs in parking services (log)	-0.13	-0.16*	-0.16*	-0.15*
		(-1.37)	(-1.92)	(-1.95)	(-1.96)
	W×Y	0.50***	0.49***	0.47***	0.44***
		(9.38)	(9.38)	(8.84)	(8.08)
	Obs.	683	683	683	683
	R-square	0.89	0.90	0.91	0.93

Note: The superscripts ***, **, and * indicate that the coefficient is statistically significant at the 1%, 5%, and 10% level.