Hindawi
Journal of Advanced Transportation
Volume 2019, Article ID 3689389, 14 pages
https://doi.org/10.1155/2019/3689389

Research Article

Analyses of the Imbalance of Urban Taxis' High-Quality Customers Based on Didi Trajectory Data

Beibei Hu,[1] Xuanxuan Xia ,[1] Xinyi Shen,[2] and Xianlei Dong [1]

[1] *School of Business, Shandong Normal University, Jinan, 250358, China*
[2] *Beijing Academy of Educational Sciences, Beijing, 100031, China*

Correspondence should be addressed to Xuanxuan Xia; xiaxuanxuansdnu@163.com and Xianlei Dong; sddongxianlei@163.com

Received 17 February 2019; Revised 16 May 2019; Accepted 17 June 2019; Published 3 July 2019

Academic Editor: Oded Cats

The distribution of high-quality customers (hereafter HQC) for taxis (including ride-hailing cars) has a significant impact on drivers' revenue and taxis' operation efficiency. Based on the taxi global positioning system data, we construct an evaluation model of passengers to discuss the imbalance of consumption in taxi market. The profit margin for each taxi order is calculated, and then a grid-based method is used to distinguish the HQC and the regions with potential higher benefits. We analyze the HQC's distribution of taxis (including ride-hailing cars) in different areas and in different time periods. We find that the HQC are distributed mainly on the periphery of the main urban area, which indicates that traffic condition is even worse in the urban center because of factors such as congestion. The HQC are more concentrated on workdays and more scattered on nonworkdays, which implies that the public have different travel habits and demands on workdays and nonworkdays. The proportion of HQC in each administrative district or functional zone is not always positively correlated to either the proportion of total orders or the total HQC. This indicates that the distribution of HQC in each administrative district or functional zone is imbalanced. The proportion of orders and that of HQC are roughly the same in the temporal dimension, being higher in the morning and evening rush hours. Compared with the distribution of the HQC of ride-hailing, that of taxis is more imbalanced in the temporal dimension. Relevant departments should further coordinate taxi pricing, strengthen market control, and promote sustainable development in the taxi and ride-hailing markets.

1. Introduction

Taxis are an important supplement in urban public transportation. Compared with other traditional traffic modes, they are more comfortable, flexible, and convenient with full time [1]. In recent years, along with the development of computer science and information technology, online ride-hailing services (including taxis and some private cars) have emerged to meet consumers' travel needs. According to official figures, the number of online ride-hailing users in China has reached 336 million by the end of 2017 [2]. The scale of the taxi market in China is becoming unprecedented, but it is accompanied by a series of socioeconomic problems, such as the imbalance of income between taxi drivers [3] and the passenger themselves taking great risks [4]. With the rapid expansion of online ride-hailing services, many conflicts of interests between online ride-hailing vehicles and traditional taxis have appeared and are becoming increasingly acute [5]. These conflicts result in shrinking of the traditional taxi market [6] and decline in the income of taxi drivers [7]. Meanwhile, many violent incidents have broken out between traditional taxi drivers and online ride-hailing drivers [8]. The existence of the conflicts above has significantly negative impacts on order in the taxi market.

The traditional taxi market has typical oligopoly characteristics [9]. However, the entry of the ride-hailing services has broken the original balance in taxi market, thus increasing the requirements for taxi market regulation. The distribution of passengers in the taxi market is random, dynamic, and complex, and it is imbalanced in both temporal and spatial dimensions [10]. The distributional imbalance not only aggravates the management of the taxi market [11] but also affects its operating efficiency. The imbalance might result in the waste of related resources and may also cause a shortage of demand and a concomitant excess supply in the taxi market, that is, coexistence of difficulty in "taking a taxi" and "finding customers" [12]. Thus, understanding and then mastering the spatial-temporal distribution of taxi passengers, especially the HQC, can improve passengers' travel experience and promote the operational efficiency of the taxi market. It is of great importance to optimize resource allocation, promote resource utility effectivity, and improve the refined management in the taxi market.

In recent years, with the development of the global positioning system (GPS) and technology for analyzing big data, it is becoming possible to use GPS real-time positional data of taxis to study related problems. Previous research was mainly concerned with (a) using trajectory data to obtain traffic real-time status [13–15], passenger travel status [16, 17], and personnel real-time flow [18, 19] to either predict urban road congestion [20] or provide reference data for urban traffic planning; (b) exploring the spatial-temporal distribution characteristics related to traffic, such as the spatial-temporal distribution of traffic travel supply and demand [21–25], the spatial-temporal distribution of taxis and passengers [26], the spatial-temporal distribution of the stream of people and traffic flow [27–30], and occupational and residential traffic [31], by studying the statistical rules of various elements of transportation and providing either theoretical or technical support for traffic management; and (c) the reality of taking advantage of trajectory data to study the efficiency of taxi operation and drivers' service profit margins. Examples are exploring the balance between urban taxi supply and passenger travel to evaluate the operational efficiency based on the capacity utilization data and vacancy data of taxis [32, 33], studying the search behavior of taxis in vacancy and simulating the effects of supply and demand balance for taxis [34], establishing taxi and ride-hailing service models to quantify drivers' service profit margins to improve taxi drivers' revenue [5], and constructing a spatial-temporal trajectory model of the balanced distribution with load to optimize the information matching between passengers and drivers and improve the operational efficiency of taxis [35–37].

In summary, the previous studies based on taxi GPS trajectory data have focused mainly on issues such as the spatial-temporal distribution of passengers or vehicles, the efficiency of taxi operations, and the profit margins of drivers. Despite the important achievements made by the study of traffic positional data, there are relatively few studies exploring the imbalance of the taxi market from the perspective of passengers' profit margins. Based on trajectory data, this paper will construct a passenger evaluation model from the perspective of profit margins, proposing the selection conditions of HQC. Finally, this paper takes Beijing as an example to compare the HQC's spatial-temporal distribution of taxis and ride-hailing vehicles, which can explain the imbalance of the taxi market. That can provide the corresponding data and theoretical support for government regulations and relevant departments to make decisions. In other respects, it can also make an important contribution to the sustainable development of the urban taxi market.

2. Data Processing and Methodology

2.1. Data Processing. The Didi Service platform contains 565,582 orders for taxis and ride-hailing vehicles in Beijing, including 91,986 orders for traditional taxis and 473,596 orders for online ride-hailing services. Cleaning the data, such as deleting duplicate and inaccurate data, leaves 506,940 valid orders (74,686 taxi orders and 432,254 ride-hailing orders) remaining. The trajectory data of the orders each contain five fields: order ID (after desensitization treatment), time stamp of the trajectory point, the corresponding longitude coordinates, the corresponding latitude coordinates, and instantaneous velocity.

HQC are mainly determined by passengers' demand in different grids, which are featured by the distribution of OD trips at different locations and time. The distribution of OD trips contributes to the service profit margin of taxis in different regions and different time periods, which leads to an imbalance in the distribution of HQC. As is well-known, both the pick-up locations and the destinations are very important for trips. With the service profit margin calculated, the pick-up locations and the destinations are included in the model. However, when we distinguish the HQC points, only the origins of trips are focused because these origins reflect people's travel demand to taxis. Therefore, we analyze the distribution of HQC to understand the imbalance of the taxi market based on passengers' pick-up points. The process of data preprocessing is as follows:

(1) The travel time, travel distance, average instantaneous speed, and low-speed travel time (the travel time of the instantaneous speed less than 12km/h) are calculated for each order. Due to the fact that a low-speed fare will be charged if the speed is lower than 12 km/h, the boundary of low-speed is set as 12km/h. And the geographical location and the time of the passenger pick-up points are recorded.

(2) The latitude and longitude grids are divided and the total number of orders is counted in each grid. Specifically, the division standards take 39.4333°N, 115.4167°E as the origin and take the ground distance of 100m, 500m, and 1km as the side lengths, respectively. However, when the grid is 100m×100m, too many grids with only a single order occur (more than 90%), resulting in insignificantly statistical results for the distribution of HQC. We then divide the grid size as 500m×500m and 1km×1km, respectively, and reach some similar conclusions. Referring to previous partition methods of grid [36, 38–40], we take the grid size of 1km×1 km as an example to describe the results (please refer to the Appendix Figures 7 and 8 for the details of results as per taking the grid size as 500m×500m). Consequently, Beijing is divided into 31,150 latitude and longitude grids (the latitude span of each grid is 0.009° and the longitude span is 0.01174°). Based on this, we can further record the center latitude and longitude of each grid and count the total number of orders for each grid.

(3) The center latitude and longitude of each grid are matched to the detailed address, to further classify the grids into the administrative districts and functional zones (including education, traffic, leisure, commerce, medical treatment,

Table 1: Charge standard of Beijing taxi.

Category	Fare
Base rate (0–3 km)	13 yuan
Mileage fee	2.3 yuan/km
Low speed fee	Below 12 km/h: added 2 km's rental per 5 minutes during the morning and evening rush hours, added 1 km's rental during other times (excl. empty cruise fee)
Empty cruise fee	Over 15 km carrying passengers one way, added 50% of the basic unit price; round-trip passengers (the starting point and end point are within 2 km (incl. 2 km)) excluding empty cruise fee.
Night-time charge	23:00 (incl. 23:00) to 5:00 next day (excl. 5:00) operation: added 20% of the basic unit price.
carpooling charge	For the carpooling mileage, charged 60% of the payable

residence, government, culture, life service, catering, shopping, and hotel) in Beijing. Then, the number of orders in each administrative district and functional zone in Beijing is counted by hour. According to the "Guiding Opinions of the Beijing Municipal People's Government of the CPC Beijing Municipal Committee on Functional Position and Evaluation Indicators," Beijing is generally divided into four major functional areas, namely, the core districts of capital function (including Dongcheng and Xicheng), the urban function extended districts (including Haidian, Chaoyang, Fengtai, and Shijingshan), the new districts of urban development (including Tongzhou, Shunyi, Daxing, Changping, and Fangshan), and the ecological preservation development districts (including Mentougou, Huairou, Pinggu, Miyun, and Yanqing).

(4) Each order's profit margin (a.k.a. the net income per unit of time) is calculated. The taxi and ride-hailing fee is based on the "Taxis Price Standard," issued by the Beijing Development and Reform Commission on January 22, 2017, and the "User Guide" and "Driver's Guide," published by the Didi Chuxing, respectively. See Tables 1 and 2 for details.

2.2. Methodology. We first define what the HQC and HQC points are and then describe how to distinguish and quantify those HQC.

HQC Points. The HQC points are defined as the locations where the taxi drivers have a great probability to pick up the orders with higher service profit margins.

HQC. The HQC are defined as the orders in these HQC points, i.e., the orders with higher probability of obtaining higher service profit margins.

The efficiency of the taxi is related to the distribution of its orders (or its passengers). To quantify accurately the passengers' distribution of HQC of urban taxis, this paper conducts an evaluation model of passengers based on the service profit margins [5] of taxis. In other words, the drivers' service profit margin of orders is considered as an indicator to distinguish the HQC points. The factors influencing taxi drivers' service profit margins include pricing rules, capacity utilization rate, travel distance, speed, and travel time. Traditionally, the complete service process ($o \longrightarrow d$) of taxis (ride-hailing services) includes seeking passengers, waiting for passengers, and delivering passengers from the starting point to the destination. The profit calculation formula of taxis ($\overline{F}^{T}_{o\longrightarrow d}$) to complete a service process ($o \longrightarrow d$) is as follows:

$$\begin{aligned}\overline{F}^{T}_{o\longrightarrow d} &= \frac{\sum_{o\in O,d\in D} N^{v}_{od} F_{od}}{\sum_{o\in O,d\in D} N^{v}_{od}} \\ &= \frac{\sum_{o\in O,d\in D} N^{v}_{od}\left[F_a + p^{T}_{b}\left(S_{od}-s\right) + F_{low-speed} + F_{empty}\right]}{\sum_{o\in O,d\in D} N^{v}_{od}} \\ &= F_a + \overline{F}_{low-speed} + \overline{F}_{empty} + p^{T}_{b}\cdot\left(\overline{S}_{od}-s\right)\end{aligned} \quad (1)$$

Here, N^{v}_{od} denotes the number of vehicles carrying passengers from o to d per unit time. F_{od} means the taxi fare from area o to d. p^{T}_{b} is the basic unit price of mileage per taxi. $F_a, F_{low-speed}, F_{empty}$ are the base rate, the low-speed fee, and the empty cruise fee accumulated during one service, respectively. S_{od}, s represent the distance and the distance included in the base rate, respectively. And $\overline{F}_{low-speed}, \overline{F}_{empty}, \overline{S}_{od}$ refer to average low speed fee, average empty cruise fee, and average distance, respectively.

In the same way, the profit calculation formula of ride-hailing ($\overline{F}^{R}_{o\longrightarrow d}$) to complete a service process ($o \longrightarrow d$) is as follows:

$$\begin{aligned}\overline{F}^{R}_{o\longrightarrow d} &= \frac{\sum_{o\in O,d\in D} N^{v}_{od} F_{od}}{\sum_{o\in O,d\in D} N^{v}_{od}} \\ &= \frac{\sum_{o\in O,d\in D} N^{v}_{od}\left[p^{R}_{b}\cdot S_{od} + p_c\cdot h_{od} + p^{R}_{empty}\cdot\left(S_{od}-s'\right)\right]}{\sum_{o\in O,d\in D} N^{v}_{od}} \\ &= p^{R}_{b}\cdot\overline{S}_{od} + p_c\cdot\overline{h}_{od} + p^{R}_{empty}\cdot\left(\overline{S}_{od}-s'\right)\end{aligned} \quad (2)$$

Here, $p^{R}_{b}, p_c, p^{R}_{empty}$ represent the fare for unit mileage, the fee for unit service time, and the empty cruise fee per unit of ride-hailing, respectively. s', h_{od} indicate the distance over which an empty cruise fee is charged and the travel time from district o to district d, respectively.

Due to the different pricing standards of taxis and ride-hailing services under different service mileages, the profit calculation formula of taxis and ride-hailing will also be different. And, disregarding the difference in fuel cost between empty cruising and traveling with passengers, this paper discusses and calculates the profit margin based on the service mileage, which is the following three cases [5].

Table 2: Charge standard of Beijing ride-hailing.

	Normal type		Enjoyed type	
Mileage	1.25 yuan/km		1.6 yuan/km	
	workdays	non-workdays	workdays	non-workdays
Duration fee	21:00–06:00 0.8 yuan/min	21:00–06:00 0.8 yuan /min	21:00–06:00 0.8 yuan/min	21:00–06:00 0.8 yuan /min
	06:00–10:00 0.64 yuan /min	12:00–15:00 0.64 yuan /min	06:00–10:00 0.64 yuan /min	12:00–15:00 0.64 yuan /min
	17:00–21:00 0.64 yuan /min	others 0.48 yuan /min	17:00–21:00 0.64 yuan /min	others 0.48 yuan /min
	others 0.4 yuan /min		others 0.4 yuan /min	
Empty cruise fee	0.64 yuan/km (over 20 km)		0.68 yuan/km (over 20 km)	
Minimum charge	11 yuan		12.8 yuan	

(1) When the service mileage of taxis is $0 < S_{od} \leq s$ and the profit margin of ride-hailing is $F^R_{o\longrightarrow d} \leq F_{\min}$, the expected profit margin of taxis and ride-hailing can be described as $\overline{E}^T_{o\longrightarrow d}$ and $\overline{E}^R_{o\longrightarrow d}$, which are shown as follows:

$$\overline{E}^T_{o\longrightarrow d} = \frac{F_a + p_{low} \cdot t_{low}}{\overline{S}_{od}} \cdot \overline{V}^T_{od} \cdot \varepsilon^T_0 - E^T{}_w - E^T{}_{fuel} - \phi_M \tag{3}$$

$$\overline{E}^R_{o\longrightarrow d} = \frac{F_{\min} - F_M}{\overline{S}_{od}} \cdot \overline{V}^R_{od} \cdot \varepsilon^R_0 - E^R_w - E^R_{fuel} \tag{4}$$

Here, p_{low}, t_{low} represent the low speed fee per unit and the duration of low-speed driving in service, respectively. $\varepsilon^T_0, \varepsilon^R_0$ denote the capacity utilization rate of taxi and the capacity utilization rate of ride-hailing, respectively. $\phi_M, F_{\min}, F_M$ represent monthly rents per unit time, the lowest revenue for each ride, and the information fee collected for every transaction by the online ride-hailing platform, respectively. $E^T_w, E^R_w, E^T_{fuel}, E^R_{fuel}$ represent the waiting cost of taxi, the waiting cost of ride-hailing, the fuel cost of taxi, and the fuel cost of ride-hailing, respectively. $\overline{V}^T_{od}, \overline{V}^R_{od}$ denote the average speed of taxi and the average speed of ride-hailing. Among these,

$$E^T_w = \phi^w \cdot \varepsilon^T_1, E^T_{fuel} = \frac{L \cdot S_{od} \cdot p_{oil}}{100\overline{h}_{od}} \cdot \left(1 - \varepsilon^T_1\right) \tag{5}$$

$$E^R_w = \phi^w \cdot \varepsilon^R_1, E^R_{fuel} = \frac{L \cdot S_{od} \cdot p_{oil}}{100\overline{h}_{od}} \cdot \left(1 - \varepsilon^R_1\right) \tag{6}$$

Here, $\phi^w, \varepsilon^T_1, \varepsilon^R_1, L, p_{oil}$ represent waiting cost per unit time, the proportion of the time the taxi spends waiting, the proportion of the time the ride-hailing vehicle spends waiting, the fuel usage per 100 km, and the fuel price, respectively.

(2) When the service mileage of taxis is $s < S_{od} \leq s'$ and the profit margin of ride-hailing is $F^R_{o\longrightarrow d} > F_{\min}$, and the service mileage of ride-hailing is $0 < S_{od} \leq s'$, the expected profit margin of taxis and ride-hailing can be described as $\overline{E}^T_{o\longrightarrow d}$ and $\overline{E}^R_{o\longrightarrow d}$, which are shown as follows:

$$\overline{E}^T_{o\longrightarrow d} = \left(\frac{F_a - p^T_b \cdot s + p_{low} \cdot t_{low}}{\overline{S}_{od}} + p^T_b\right) \cdot \overline{V}^T_{od} \cdot \varepsilon^T_0 - E^T_w - E^T_{fuel} - \phi_M \tag{7}$$

$$\overline{E}^R_{o\longrightarrow d} = \left(p^R_b - \frac{F_M}{\overline{S}_{od}}\right) \cdot \overline{V}^R_{od} \cdot \varepsilon^R_0 + p_c \cdot \varepsilon^R_0 - E^R_w - E^R_{fuel} \tag{8}$$

(3) When the service mileage of taxis and ride-hailing is $S_{od} > s'$, the expected profit margin of taxis and ride-hailing can be described as $\overline{E}^T_{o\longrightarrow d}$ and $\overline{E}^R_{o\longrightarrow d}$, which are shown as follows:

$$\overline{E}^T_{o\longrightarrow d} = \left(\frac{F_a - p^T_b \cdot s + p_{low} \cdot t_{low} - p^T_{empty} \cdot s'}{\overline{S}_{od}} + p^T_b + p^T_{empty}\right) \cdot \overline{V}^T_{od} \cdot \varepsilon^T_0 - E^T_w - E^T_{fuel} - \phi_M \tag{9}$$

$$\overline{E}^R_{o\longrightarrow d} = \left(p^R_b + p^R_{empty} - \frac{p^R_{empty} \cdot s' + F_M}{\overline{S}_{od}}\right) \cdot \overline{V}^R_{od} \cdot \varepsilon^R_0 + p_c \cdot \varepsilon^R_0 - E^R_w - E^R_{fuel} \tag{10}$$

Here, p^T_{empty} represents the empty cruise fee per unit of taxi.

Based on the service profit margins of orders and combined with the geographical information system (GIS) spatial-temporal analysis method to distinguish the HQC and HQC points, the steps are as follows:

(1) Identifying the HQC points: first, we calculate the profit margin of each order of taxis and ride-hailing, respectively. Second, regarding the latitude and longitude grids as the research unit, we calculate the average profit margin for all orders in each grid. Third, we identify the HQC points according to the average profit margin in each grid. Referring to previous research's suggestion [41], we take the top 30%

of the grids with the greatest average profit margins as HQC points.

(2) Identifying the HQC: the HQC are the orders appearing in these HQC points. In other words, the HQC often imply a higher probability of obtaining higher service profit margins for taxis and ride-hailing.

3. Results and Discussion

We analyze the distribution of HQC from both space and time dimensions, resulting from HOC's spatial and temporal dependence. In terms of space, we analyze HQC from the perspective of administrative districts and functional zones, including the comparison of the distribution on workdays with nonworkdays. In terms of time, we study the changing rules of HQC in different times of a day. In particular, we analyze the distribution of HQC in the rush hours of morning and evening. We also compare the service profit margins of HQC in different trip lengths and at different times of a day.

3.1. Spatial Distribution of HQC. In fact, there will be some sampling bias for some suburb administrative regions (grids) might be less sampled. When there are too few orders in a grid, the average profit margin in the grid might suffer a relatively greater error or become insignificant statistically. However, the bias will not affect our results severely. We are distinguishing the HQC to analyze its distribution in space and time dimensions. The service orders in suburb grids are much less than those in city center, so very fewer orders in suburb grids will not affect the overall distribution of HQC, since their average profit margin might carry a large standard error. For example, the number of HQC in Yanqing District only accounts for 0.003% of all the HQC in Beijing city, which will not severely affect the distribution of HQC in the city.

ArcGIS visualizes the distribution of all passengers and HQC, where the HQC are represented by the red grid, as shown in Figure 1. On the whole, the passengers of both taxis and ride-hailing are concentrated mainly on Xicheng District, Chaoyang District, Haidian District, Fengtai District, Changping District, and Daxing District, which account for 71% of the total passengers. The HQC in central urban areas of Beijing (such as Dongcheng district and Xicheng District) are few and scattered. Dongcheng District and Xicheng District, as the traditional inner urban areas of Beijing, have larger flow of passengers and greater traffic demand. Transportation facilities are available there, and the distribution of HQC is relatively small.

In Figure 1, there are significant differences in the distribution of HQC between workdays and nonworkdays. The differences are reflected mainly in the fact that the HQC on the workdays are in a state of aggregation, whereas the HQC on nonworkdays are wider and more evenly distributed. For example, on workdays, HQC are distributed mainly in residential areas, educational institutions, commercial areas, and transportation areas, such as airports, subways, universities, companies, and commercial buildings. That phenomenon of the distribution of HQC represents the aggregation distribution on workdays. On nonworkdays, people usually go out to party (restaurants and KTV), shop (shopping malls and shopping centers), and so on. So, the distribution of HQC on weekends is distributed more widely and more evenly.

There are obvious differences in the distribution of HQC between taxis and ride-hailing. The HQC of taxis are distributed mainly in the urban function extended districts and the new urban development districts, such as Chaoyang District, Haidian District, Shunyi District, and Daxing District. In these districts, the HQC account for 55.76% of the total HQC. Chaoyang District is a gathering place for high-end industrial functional areas. Haidian District is a gathering place for educational institutions. These places have a large population density and a large flow of passengers on workdays. The HQC of ride-hailing services are distributed mainly in the traffic and residential areas of the new urban development districts, including Shunyi District, Fangshan District, and Daxing District. In particular, the ride-hailing's HQC in Shunyi District, Chaoyang District, Daxing District, and Fangshan District account for 58.94% of the total ride-hailing's HQC. The number of taxis in these areas is small, but there are large travel needs, such as from the place of residence to the nearest public transportation facilities. At the same time, the traffic in the outer suburban districts is smooth, and the carpooling business in the ride-hailing service also provides passengers with greater convenience. So, the HQC in ride-hailing service exhibit a phenomenon of peripheral agglomeration distribution.

Next, we analyze the distribution of HQC in different administrative districts. Figure 2 shows, in 16 administrative districts, the proportion of orders in the total orders (hereafter referred to as P-O), the proportion of HQC in the total HQC (hereafter referred to as P-HQC), and the proportion of HQC in the total orders of the corresponding administrative districts (hereafter referred to as P-HQC-OAD). Overall, the P-O of taxis (including ride-hailing cars) in the administrative district is positively correlated with the P-HQC [the correlation coefficient is 0.80 (0.43)]. In other words, the more orders, the more HQC. And, from Figure 2, we can find that the P-O is significantly higher than is the P-HQC in Chaoyang District, Haidian District, Fengtai District, and Dongcheng District. However, the P-HQC in other administrative districts is higher than is the P-O. For example, the P-O of taxis in Chaoyang District and Haidian District accounts for a relatively high proportion, at 34.58% and 19.3%, respectively. And the P-HQC is also relatively high, at 19.68% and 13.83%, respectively. Ride-hailing's P-O in Chaoyang District, Haidian District, and Fengtai District account for 56.12%. And the P-HQC in these districts is also high, at 24.16%. However, taxis and ride-hailing in the ecological preservation development districts (such as Pinggu District, Mentougou District, Miyun District, and Yanqing District) have a small number of orders, accounting for approximately 0.43%. And their P-HQC is relatively low, with the proportion being approximately 1.31%.

The higher the P-HQC-OAD in administrative districts (the red line), the more the high efficiency orders and the more the possibilities that drivers will obtain high profits from orders. Although the order number has a significantly positive correlation with the HQC, the distribution of HQC

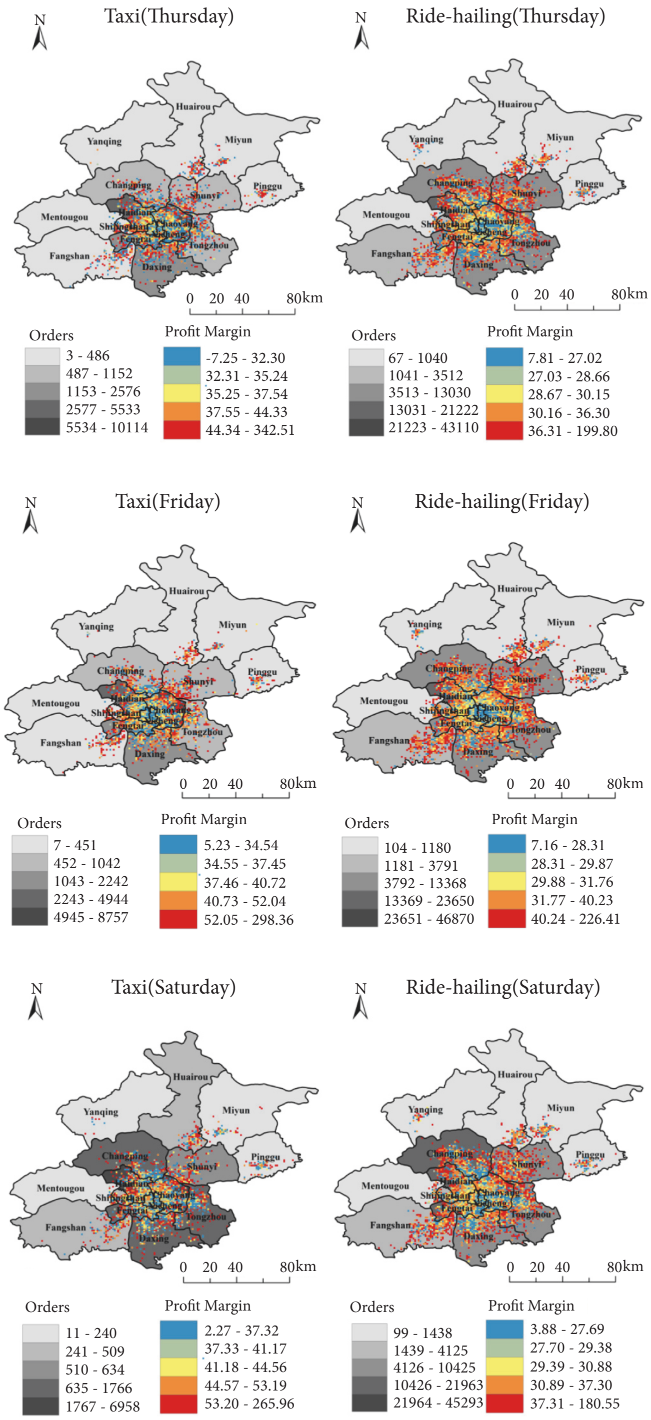


FIGURE 1: The pick-up points of taxis and ride-hailing on Thursday, Friday, and Saturday, respectively. The orders in the legend refer to the service order counts interval in different administrative districts. The background color depth of each administrative district in Beijing is proportional to the number of orders. The red areas are defined as the HQC points in this paper.

in each administrative district is imbalanced. From Figure 2, in each administrative district, the P-HQC-OAD is inversely proportional to the orders and HQC in most cases. The correlation coefficients are -0.60, -0.26 (taxi) and -0.52, 0.086 (ride-hailing), respectively. For example, there are fewer orders and HQC in Yanqing District, but the P-HQC-OAD is relatively high, at 34.97%. Conversely, Chaoyang District has the largest number of orders, accounting for 32.94% of the total orders, and it also has the highest P-HQC. However, its P-HQC-OAD is relatively low, at 4.83%. To a certain

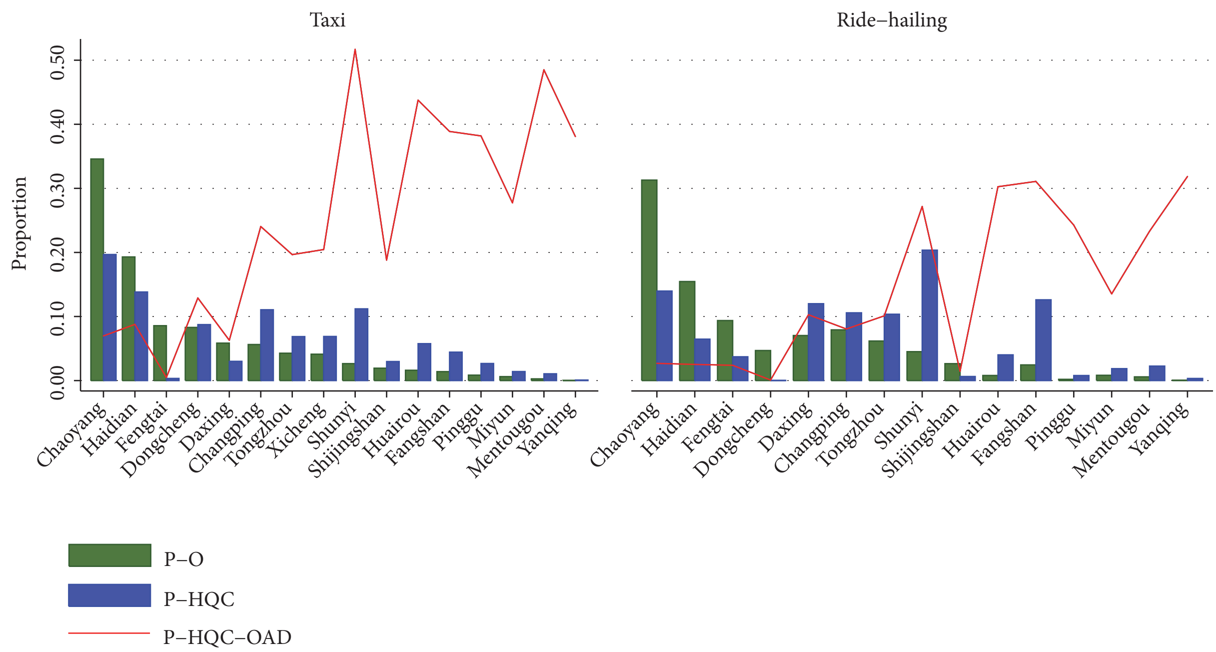


Figure 2: The distribution of orders and HQC in different administrative districts.

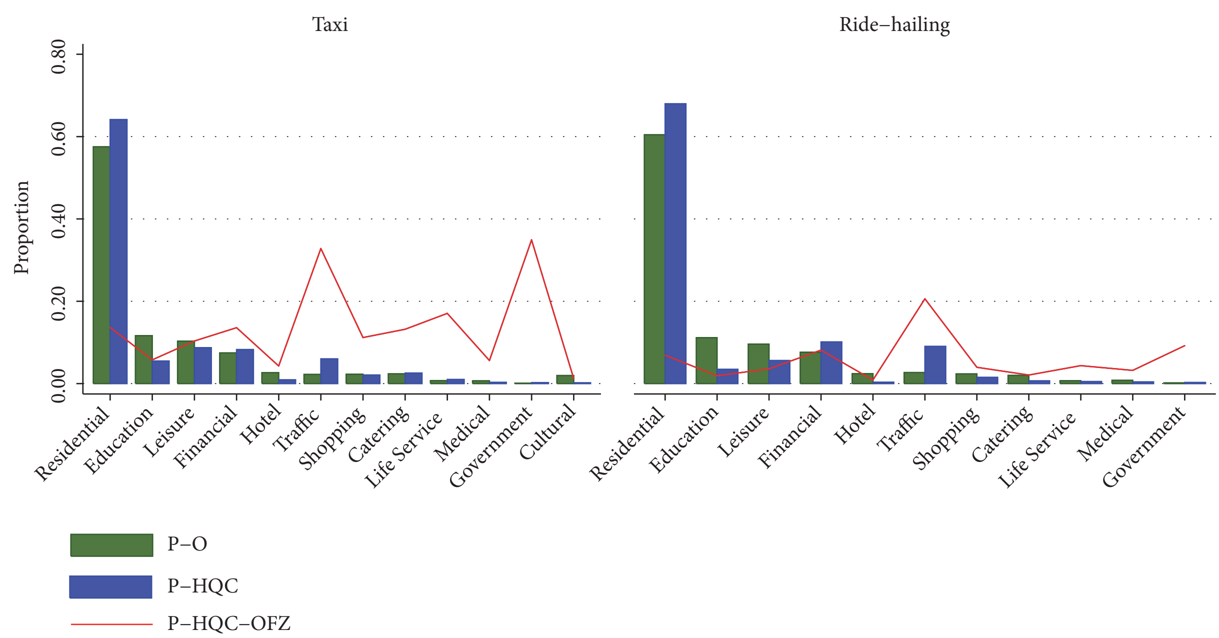


Figure 3: The proportion of orders and HQC in different functional zones.

extent, this indicates that the P-HQC-OAD is not determined directly by either the orders or the HQC.

Next, we will analyze the distribution of HQC in different functional zones. According to the functional attributes in different areas of the city and the classification of Baidu POI industry, all orders are divided into 12 functional zones. The functional zones include shopping, traffic, education, finance, hotels, residence, catering, life service, culture, leisure, medical treatment, and government. Figure 3 shows the P-O, the P-HQC, and the proportion of HQC in the total orders of the corresponding functional zones (hereafter referred to as P-HQC-OFZ). Overall, passengers and HQC in the taxi (ride-hailing cars) are distributed mainly in the five functional zones: residence, leisure, finance, traffic, and education [according to the number of HQC (taxis) from high to low]. The total orders of these functional zones are up to 89.19% (91.52%), and the P-HQC is up to 92.65% (96.23%). Conversely, the orders and HQC in both medical and government functional zones account for less than 1%.

In each functional zone, the higher the P-HQC-OFZ (the red line), the higher the probability of the HQC appearing and the higher the possibility that drivers will make greater profits. Compared with taxis, the HQC of ride-hailing in each functional zone is relatively balanced. Specifically, the P-HQC-OFZ is relatively stable, accounting for approximately 4.43% in each functional zone (the proportion in traffic is relatively high, at 20.60%). However, the HQC of taxis in various functional zones is imbalanced. For example, in government and traffic functional zones, the P-HQC-OFZ is 34.92% and 32.82%, respectively. However, in life service and residential functional zones, the P-HQC-OFZ is 17.06% and 13.65%, respectively. In other words, drivers can get high-efficiency orders easily in the traffic functional zone. In particular, the possibility of drivers receiving orders is low in the government functional zone, but drivers are most likely to receive high-efficiency orders.

3.2. Temporal Distribution of HQC. Next, the distribution of HQC in the temporal dimension is analyzed. Figure 4 shows the P-O per hour, the P-HQC per hour, and the proportion of HQC in total orders per hour (hereafter referred to as P-HQC-O per hour) in respect of taxis and riding-hailing on Thursday, Friday, and Saturday. On the whole, the P-O per hour and the P-HQC per hour of taxis and ride-hailing show a similar daily changing trend. And the P-O per hour and the P-HQC per hour have great value in the morning and evening rush hours, which have a significantly positive correlation. Specifically, the total orders and HQC are in a relatively high state, at 7:00–9:00 and 17:00–19:00. The P-O per hour and the P-HQC per hour are approximately 16.11% and 16.52% in the morning rush hours. And the two proportions are 18.39% and 21.95%, respectively, in the evening rush hours. Due to work, on workdays, residents have high travel demands during the morning and evening rush hours. So, the number of orders and HQC in rush hours is higher than in other periods. On nonworkdays, the morning rush hours are postponed, the P-O per hour and the P-HQC per hour of taxis and ride-hailing services do not increase significantly, and the evening rush hours are basically the same as on workdays.

The greater the value of P-HQC-O per hour, the higher the probability of HQC in this period. From Figure 4, the P-HQC-O per hour of taxis is significantly higher than that of ride-hailing. In other words, drivers of taxis are more likely to accept high-profit orders. The extreme point of the taxis' P-HQC-O per hour appears in the evening rush hours, whereas the ride-hailing's extreme point appears in the morning and evening rush hours. Compared with taxis, the peak change of ride-hailing is not obvious (it rises slightly in the morning and evening peaks), and it is stable on the whole. This indicates that the pricing of ride-hailing services is more reasonable, and ride-hailing's HQC in the temporal dimension presents a balanced distribution. Conversely, the extreme point of the taxi's P-HQC-O per hour appears in the evening rush hours, and the peak value is significantly higher than in other periods. This indicates that taxis' HQC in the temporal dimension are imbalanced. To be specific, the incentives for taxis are insufficient in morning rush hours, but the incentives are oversufficient in evening rush hours. The relevant departments should further coordinate taxi pricing, strengthen market control, and promote balanced development in the taxi and ride-hailing markets.

It is of great significance for not only the taxi and ride-hailing markets but also urban planning departments to understand the residents' travel rules in the morning and evening rush hours. The HQC of taxis (including ride-hailing vehicles) during morning and evening rush hours will be analyzed by selecting 7:00–9:00, 17:00–19:00 on workdays and 10:15–11:45, 14:15–18:45 on nonworkdays. Figure 5 shows the P-O and the P-HQC-OFZ of taxis (including ride-hailing) in different functional zones during morning and evening rush hours.

Figure 5 also shows that the P-O of taxis and ride-hailing is mostly the same in different functional zones, and there are relatively more orders in residence, education, and leisure functional zones than in other functional zones. However, there is a great difference between the P-O and the P-HQC-OFZ of taxis (including ride-hailing). And the P-O reveals a significantly negative correlation with the P-HQC-OFZ. The higher the P-HQC-OFZ is, the more likely the HQC will appear in the functional zones. Figure 5 shows that the P-HQC-OFZ value is relatively high in traffic, catering, life service, and government functional zones, which are the high-profit areas where drivers find it easier to pick up HQC orders. Compared with taxis, the HQC of ride-hailing during the rush hours is relatively balanced in the functional zones. Specifically, the ride-hailing's P-HQC-OFZ is relatively stable in each functional area. However, the HQC of taxis is imbalanced in each functional zone. In particular, the P-HQC-OFZ has a high proportion in government and traffic functional zones.

The service profit margins of trips for the HQC orders in different trip lengths and at different times of a day are shown in Figure 6.

The service profit margins of the short trips (<3 km) and long trips (>15 km) are at a relatively high level, while those of the intermediate trips (3-15 km) are at a relatively low level. For example, the service profit margins of the trips of "0-3 km" and ">15 km" are 1.78 and 1.94 times higher than those of the trips of "3-15 km," respectively. The service profit margin of long trips at night is higher than that in daytime. For instance, on average, their profit margin from night to early morning (such as 0:00-5:00) is 1.21 times more than that in daytime (take 8:00-18:00 as an example). Comparing the profit margins of the short trips with long trips, we can find that the short trips tend to have greater profit margins in daytime while the long trips tend to have greater profit margins at night and in early morning. We speculate that this is because the road condition is much more congested in day time than in night and early morning, for the profit margins of short trips are mainly determined by the base rate, while those of long trips are mainly determined by the mileage fee and empty cruise fee.

Figure 4: The temporal distribution of taxi (ride-hailing) orders and HQC.

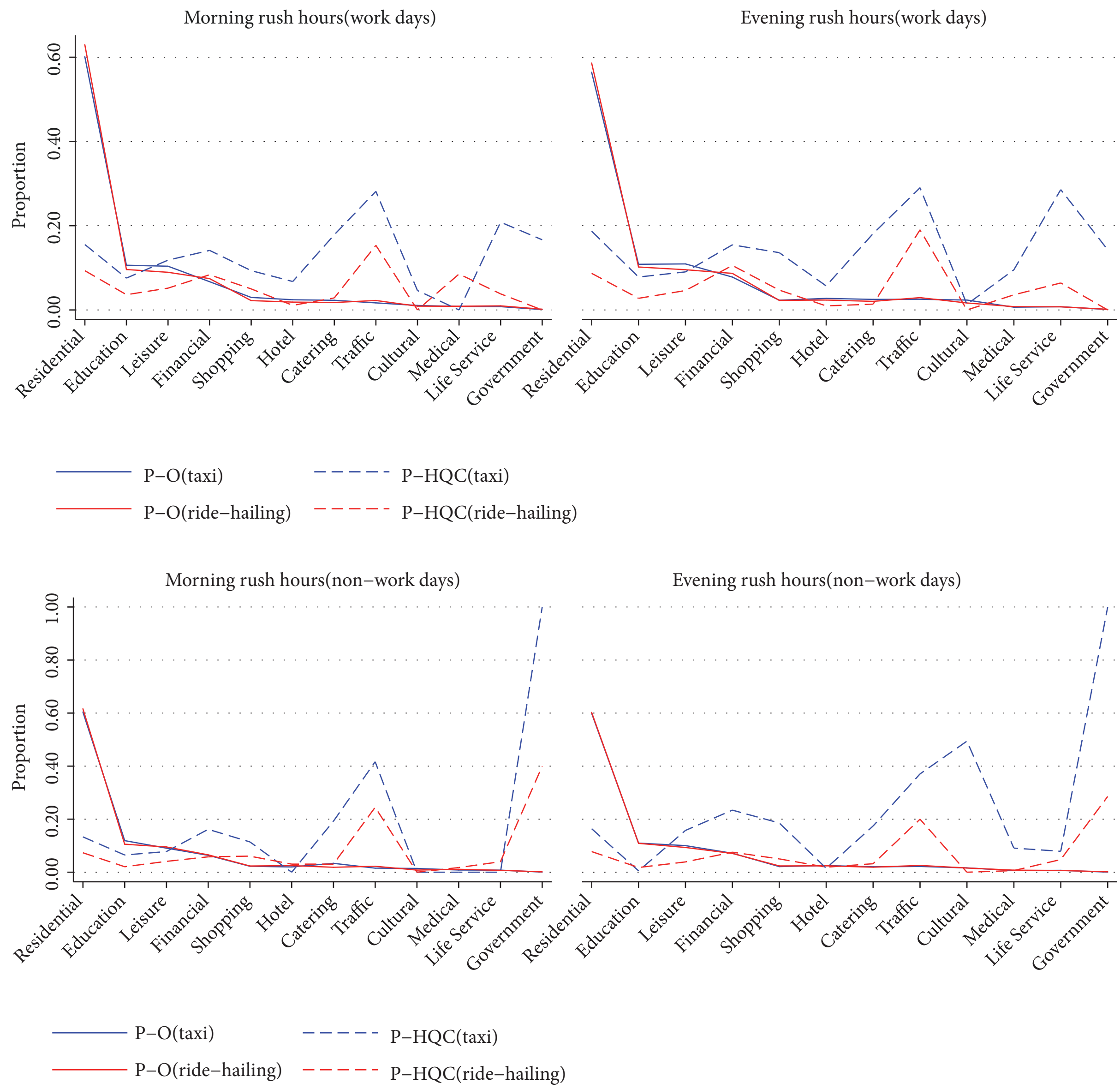
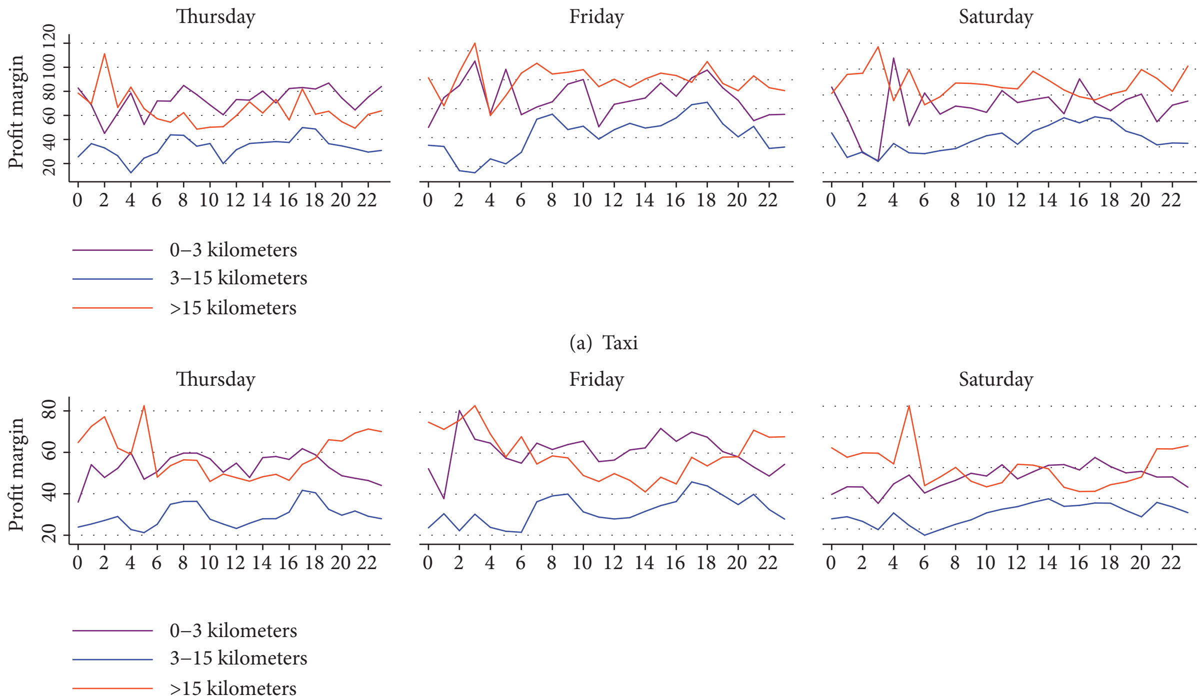


Figure 5: The proportion of orders and HQC in different functional zones on workdays and nonworkdays (taxi and ride-hailing).

4. Conclusions

We distinguish HQC and HQC points from pick-up locations of passengers based on service profit margin of each order and then analyze the distribution of HQC and HQC points. In this paper, we provide a new perspective to analyze the imbalance of taxi market from the distribution of HQC. The balance of the taxi market has an important impact on the sustainable development of the city. Therefore, it is especially important to fully understand the spatial-temporal distribution of taxi's high-quality customers (HQC). Based on global positioning system trajectory data, this paper constructs an evaluation model of passengers from the perspective of profit margins. Then, we present the selection criteria of HQC and discuss the distribution laws of HQC in different areas and different time periods. The conclusions are as follows.

The HQC of taxis and ride-hailing are distributed mainly in the periphery of the main urban area, which indicates that traffic condition is even worse in the urban center because of factors such as congestion. In central urban areas of Beijing (such as Dongcheng District and Xicheng District), there are more passengers and more traffic demands. Due to the well-developed public transportation, HQC in these areas are few and scattered. There are significant differences in the distribution of HQC between workdays and nonworkdays. The HQC on workdays are in a state of aggregation (concentrated in the functional zones of residence, leisure, financial, traffic, and education). On the other hand, the HQC are distributed more widely and more evenly on nonworkdays.



(b) Ride-hailing

Figure 6: The daily trend of HQC in different trip lengths. We categorize all the trips into three classes, namely, short trips (i.e., the trips that are less than 3 km), long trips (i.e., the trips that are more than 15 km), and intermediate trips (i.e., the trips that are more than 3 km and less than 15 km).

There are also obvious differences in the distribution of HQC between taxis and ride-hailing. For example, the traffic in the outer suburban districts is smooth, and the car-pooling service of the ride-hailing also provides passengers with greater convenience, which results in the distribution of ride-hailing's HQC being in aggregation in the outer periphery.

The proportion of orders in the total orders, or P-O, of taxis (including ride-hailing cars) in the administrative district is positively correlated with the proportion of HQC in the total HQC, or P-HQC. However, the proportion of HQC in the total orders of the corresponding administrative districts, or P-HQC-OAD, is inversely proportional to the orders and HQC in most cases, which indicates that the HQC in each administrative district is imbalanced. The distribution of HQC in each functional zone is also imbalanced, especially the HQC of taxis. The government and traffic function zones have the highest proportion in P-HQC-OFZ, with 34.92% and 32.82%, respectively. And the functional zones of the life service and residence are second.

The P-O per hour and the P-HQC per hour of taxis and ride-hailing show a similar daily changing trend. In the morning and evening rush hours, the two proportions account for a relatively high proportion and have a significant positive correlation. The P-HQC-O per hour of taxis is significantly higher than that of ride-hailing, which indicates that drivers of taxis are more likely to accept high-margin orders. The peak change of ride-hailing is stable on the whole, which indicates that ride-hailing pricing is more reasonable. The extreme point of the taxi's P-HQC-O per hour appears in the evening rush hours, and the peak value is significantly higher than in other periods, indicating that taxis' HQC in the temporal dimension exhibits an imbalance. The relevant departments should further coordinate taxi pricing, strengthen market control, and promote balanced development in the taxi and ride-hailing markets.

Understanding the spatiotemporal distribution of HQC can benefit the taxi drivers a lot. For example, it can improve the information asymmetry between taxi drivers and the HQC, which would reduce their seeking time of HQC. In other words, it is in favor of improving the operational efficiency of taxi, which can increase the drivers' income and improve the service quality. In sum, the distribution of HQC of online ride-hailing services is of great significance to improve the operational efficiency of the online ride-hailing services market. To promote sustainable development in the online ride-hailing service market, it is recommended that the relevant departments coordinate the supply of taxis with different service modes to rationally optimize resource allocation according to the market demand in different time periods and different regions. At the same time, they should also coordinate the supply of vehicles in the rush hours of morning and evening. Finally, the relevant departments should utilize big data on the location of urban traffic and strengthen the construction of the big data platform on traffic to grasp the real-time traffic situation accurately and elaborately from a global perspective, which can also

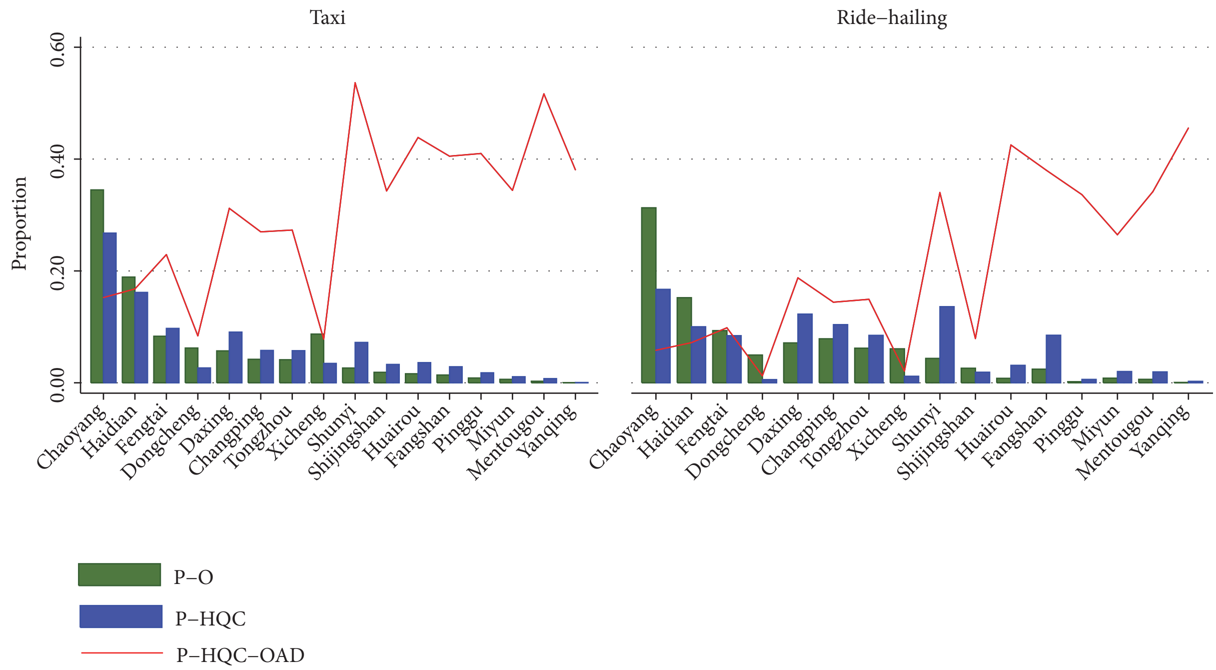


Figure 7: The distributions of orders and HQC in different administrative districts. Note: P-O refers to the proportion of orders in the total orders in the 16 administrative regions. P-HQC refers to the proportion of HQC in the total HQC. P-HQC-OAD refers to the proportion of HQC in the total orders of the corresponding administrative districts.

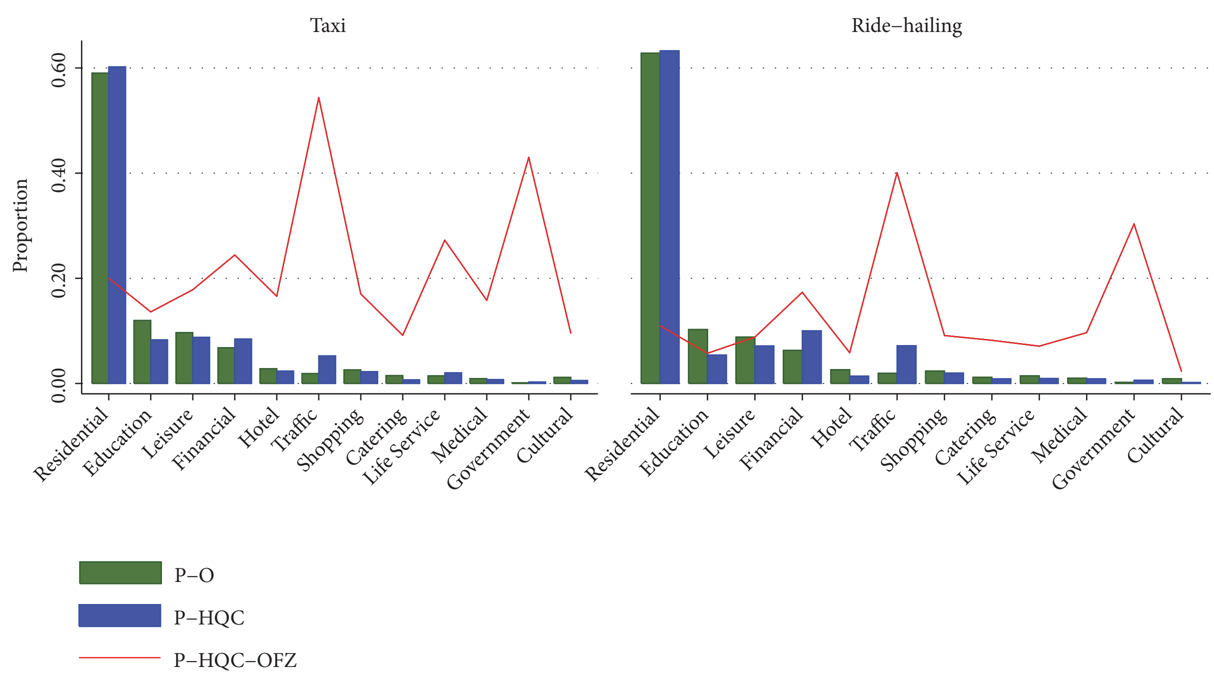


Figure 8: The distributions of orders and HQC in different functional zones. Note: P-HQC-OFZ refers to the proportion of HQC in the total orders of the corresponding functional zones.

provide guidance for methods of management and data-driven decision-making.

Appendix

See Figures 7 and 8.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by the programs of National Social Science Foundation of China (Grant: 16CJY056).

References

[1] D. Zhang, L. Sun, B. Li et al., "Understanding taxi service strategies from taxi GPS traces," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 1, pp. 123–135, 2015.

[2] P. I. R. Institute, Report of Commerical Model Innovation and Investment Opportunities Analysis on China Online Car-hailing Industry (2019-2024), 2018.

[3] J. Allred, A. Saltzman, and S. Rosenbloom, "Factors affecting the use of taxicabs by lower income groups," *Transportation Research Record*, 1978.

[4] Q. Liu, "Model research and policy recommendations of tailored taxi," in *Proceedings of the International Conference on Industrial Economics System & Industrial Security Engineering*, 2017.

[5] B. Hu, Y. Kong, M. Sun, X. Dong, and G. Zong, "Understanding the unbalance of interest in taxi market based on drivers' service profit margins," *PLoS ONE*, vol. 13, no. 6, Article ID e0198491, 2018.

[6] W. Zhang, T. V. Le, S. V. Ukkusuri, and R. Li, "Influencing factors and heterogeneity in ridership of traditional and app-based taxi systems," *Transportation*, pp. 1–26, 2018.

[7] G. Qin, T. Li, B. Yu, Y. Wang, Z. Huang, and J. Sun, "Mining factors affecting taxi drivers' incomes using GPS trajectories," *Transportation Research Part C: Emerging Technologies*, vol. 79, pp. 103–118, 2017.

[8] B. Gilbert, "The nature of occupational violence against taxicab drivers," *Public Health Nursing*, vol. 28, no. 4, pp. 335–348, 2011.

[9] V. Singh and T. Zhu, "Pricing and market concentration in oligopoly markets," *Marketing Science*, vol. 27, no. 6, pp. 1020–1035, 2008.

[10] J. Zhao, Q. Qu, F. Zhang, C. Xu, and S. Liu, "Spatio-temporal analysis of passenger travel patterns in massive smart card data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 11, pp. 1–12, 2017.

[11] H. Wang, K. Zhang, J. Chen, Z. Wang, G. Li, and Y. Yang, "System dynamics model of taxi management in metropolises: Economic and environmental implications for Beijing," *Journal of Environmental Management*, vol. 213, pp. 555–565, 2018.

[12] X. Feng, M. Saito, and Y. Liu, "Improve urban passenger transport management by rationally forecasting traffic congestion probability," *International Journal of Production Research*, vol. 54, no. 12, pp. 1–10, 2015.

[13] K.-C. Chu, R. Saigal, and K. Saitou, "Real-time traffic prediction and probing strategy for lagrangian traffic data," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–10, 2018.

[14] E. Sara, J. P. Romero, J. L. Moura et al., "Minimizing the impact of large freight vehicles in the city: a multicriteria vision for route planning and type of vehicles," *Journal of Advanced Transportation*, vol. 2018, Article ID 1732091, 8 pages, 2018.

[15] Z. Q. Wang, Z. H. Kan, J. Hong, F. Z. Huang, and L. L. Tang, "Path optimization based on city intersection turning analysis from floating car data," *Applied Mechanics and Materials*, vol. 577, pp. 1055–1060, 2014.

[16] R. Sun, M. Li, and Q. Wu, "Research on commuting travel mode choice of car owners considering return trip containing activities," *Sustainability*, vol. 10, no. 10, p. 3494, 2018.

[17] G. Peng, "Passenger flow distribution model and algorithm for urban rail transit network based on multi-route choice," *Journal of the China Railway Society*, 2009.

[18] Y. Liu, C. Kang, S. Gao, Y. Xiao, and Y. Tian, "Understanding intra-urban trip patterns from taxi trajectory data," *Journal of Geographical Systems*, vol. 14, no. 4, pp. 463–483, 2012.

[19] M. I. Jun and Z. Yang, "Research of travel mode choice based on prospect theory," *Journal of Transportation Engineering and Information*, 2015.

[20] Y. Iida, "Basic concepts and future directions of road network reliability analysis," *Journal of Advanced Transportation*, vol. 33, no. 2, pp. 125–134, 1999.

[21] H. Kim, I. Yang, and K. Choi, "An agent-based simulation model for analyzing the impact of asymmetric passenger demand on taxi service," *KSCE Journal of Civil Engineering*, vol. 15, no. 1, pp. 187–195, 2011.

[22] Z. Z. Wang and X. Y. Liu, "Analysis on difference between supply and demand of urban taxi passenger in the case of carpooling," *Applied Mechanics and Materials*, vol. 543-547, pp. 4378–4382, 2014.

[23] X. Chen, L. Zhou, and Y. Yue, "Data-driven method to estimate the maximum likelihood space-time trajectory in an urban rail transit system," *Sustainability*, vol. 10, no. 6, article 1752, 2018.

[24] F. Zong, Y. He, and Y. Yuan, "Dependence of parking pricing on land use and time of day," *Sustainability*, vol. 7, no. 7, pp. 1–21, 2015.

[25] Y. Yang, L. Jia, Y. Qin, S. Han, and H. Dong, "Understanding structure of urban traffic network based on spatial-temporal correlation analysis," *Modern Physics Letters B. Condensed Matter Physics, Statistical Physics, Atomic, Molecular and Optical Physics*, vol. 31, no. 22, Article ID 1750230, 2017.

[26] R. Hwang, Y. Hsueh, and Y. Chen, "An effective taxi recommender system based on a spatio-temporal factor analysis model," *Information Sciences*, vol. 314, pp. 28–40, 2015.

[27] L. Li, S. He, J. Zhang, and B. Ran, "Short-term highway traffic flow prediction based on a hybrid strategy considering temporal–spatial information," *Journal of Advanced Transportation*, vol. 50, no. 8, pp. 2029–2040, 2016.

[28] X. Cao, F. Liang, H. Chen, and Y. Liu, "Circuity characteristics of urban travel based on GPS data: a case study of Guangzhou," *Sustainability*, vol. 9, no. 11, pp. 1–21, 2017.

[29] T. Hagerstraand, "What about people in regional science," *Papers in Regional Science*, vol. 24, no. 1, pp. 7–24, 1970.

[30] Y. Sun, G. Zhang, and H. Yin, "Passenger flow prediction of subway transfer stations based on nonparametric regression model," *Discrete Dynamics in Nature and Society*, vol. 2014, Article ID 397154, 8 pages, 2014.

[31] X. Fu, M.-P. Sun, and H. Sun, "Taxi Commute Recognition And Temporal-Spatial Characteristics Analysis Based On GPS data," *China Journal of Highway and Transport*, vol. 30, no. 7, pp. 134–143, 2017.

[32] J. D. Lees-Miller and R. E. Wilson, "Proactive empty vehicle redistribution for personal rapid transit and taxis," *Transportation Planning and Technology*, vol. 35, no. 1, pp. 17–30, 2012.

[33] Z. Zhou, W. Dou, G. Jia et al., "A method for real-time trajectory monitoring to improve taxi service using GPS big data," *Information and Management*, vol. 53, no. 8, pp. 964–977, 2016.

[34] D. Shao, W. Wu, S. Xiang, and Y. Lu, "Estimating taxi demand-supply level using taxi trajectory data stream," in *Proceedings of the 15th IEEE International Conference on Data Mining Workshop, ICDMW 2015*, pp. 407–413, USA, November 2015.

[35] H. Kim, J.-S. Oh, and R. Jayakrishnan, "Effect of taxi information system on efficiency and quality of taxi services," *Transportation Research Record*, vol. 1903, no. 1903, pp. 96–104, 2005.

[36] Y. Liu, F. Wang, Y. Xiao, and S. Gao, "Urban land uses and traffic 'source-sink areas': evidence from GPS-enabled taxi data in Shanghai," *Landscape and Urban Planning*, vol. 106, no. 1, pp. 73–87, 2012.

[37] B. He, D. Zhao, J. Zhu, A. Darko, and Z. Gou, "Promoting and implementing urban sustainability in China: an integration of sustainable initiatives at different urban scales," *Habitat International*, vol. 82, pp. 83–93, 2018.

[38] J. X. Cui, F. Liu, D. Janssens, S. An, G. Wets, and M. Cools, "Detecting urban road network accessibility problems using taxi GPS data," *Journal of Transport Geography*, vol. 51, pp. 147–157, 2016.

[39] A. Páez, D. M. Scott, and C. Morency, "Measuring accessibility: Positive and normative implementations of various accessibility indicators," *Journal of Transport Geography*, vol. 25, pp. 141–153, 2012.

[40] S. Jiang, W. Guan, Z. He, and L. Yang, "Measuring taxi accessibility using grid-based method with trajectory data," *Sustainability*, vol. 10, no. 9, article 3187, 2018.

[41] F. Sun, X. Zhang, L. Tang, Z. Liu, X. Yang, and K. Dong, "Temporal and spatial distribution of high efficiency passengers based on gps trajectory big data," *Journal of Geo-Information Science*, no. 17, pp. 329–335, 2015.