



Article Uncovering Factors Affecting Taxi Income from GPS Traces at the Directional Road Segment Level

Shuxin Jin ¹, Zhouhao Wu², Tong Shen ³, Di Wang ⁴ and Ming Cai ^{1,*}

- School of Intelligent Systems Engineering, Sun Yat-sen University, Shenzhen 518107, China; jinshx3@mail.sysu.edu.cn
- ² School of Civil Engineering, Tsinghua University, Beijing 100084, China; hero1989wzh@tsinghua.edu.cn
- ³ School of Art and Design, Xi'an University of Technology, Xi'an 710061, China; 2015041002@chd.edu.cn
- ⁴ Department of Civil and Environmental Engineering, Nagoya University, Nagoya 464-8603, Japan; wang.di@e.mbox.nagoya-u.ac.jp
- Correspondence: caiming@mail.sysu.edu.cn

Abstract: Nowadays, the market demand for taxis is still intense. However, there exist lots of issues affecting the healthy development of the taxi industry, such as an increasing difficulty in hailing taxis, detouring behavior etc., and especially, the low incomes of taxi drivers. This paper establishes a multi-layer road index (MRI) system of 7862 directional road segments (DRSs), and collects over 194 million occupied GPS points within a week, revealing the factors affecting taxi drivers' incomes in Shenzhen, China. The income differences has been identified on different DRSs, which accordingly have been categorized into two levels. Four categories of DRS factors, i.e., road attributes, traffic dynamics, points of interest (POIs), and taxi operation strategies, are defined as the impact factors affecting income levels. The selected sample-based binomial logit (SBL) model has been proposed to reveal the significance of these influencing factors. The results indicate that the road segments with different features have different incomes over different time periods. The main factors in income analysis are the factors used to represent taxi operation strategies. Highly rewarding pick-up road segments can be identified, which could contribute to drivers' income improvements, and can further contribute to the development of the taxi analysis.

Keywords: taxi income; GIS; multi-layer road index; map matching; taxi data mining; urban computing; network complexity; GPS

1. Introduction

Taxis play an important role in public transportation provision—effectively plugging the gaps that are left by buses and subways. Generally, they are not considered to be a luxury mode of transport—they are predominantly used by people with mobility problems, and by people who do not own cars, to perform trips that would have otherwise been impossible. As for Shenzhen, in 2015, it has been recorded that there were 16,597 taxis delivering more than 1.1 million passengers per day, which constitute an indispensable part of the city's traffic ecosystem.

In most European countries, taxi fares include a flag-down fare, a fare charged by the ride length, a congestion premium (or so-called low-speed fare, which is the cost of waiting upon the request of the passenger), and other expenses (such as fuel surcharges) [1], which are normally regulated by the taxi associations and regulators. The entry regulation has been formulated in Belgium, Italy, France, and other countries [2]. In other countries, deregulation policies lead to a sharp increase in the number of taxis and have dramatically declined the service quality [3]. Especially with lower incomes, the drivers tend to decline in their service quality. Therefore, establishing a fair, efficient, and comprehensive fare policy and service management system is essential and critical for a boom in the global taxi industry.



Citation: Jin, S.; Wu, Z.; Shen, T.; Wang, D.; Cai, M. Uncovering Factors Affecting Taxi Income from GPS Traces at the Directional Road Segment Level. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 431. https://doi.org/ 10.3390/ijgi11080431

Academic Editors: Wolfgang Kainz and Hartwig H. Hochmair

Received: 6 June 2022 Accepted: 26 July 2022 Published: 31 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).

For the China taxi industry, the income level is still a major concern for taxi drivers and regulators. However, with the growth of the taxi market and changes in travel demands, a lot of problems have accumulated, such as service refusal, cruising, and taxi strikes [4]. For instance, in December 2020, the Shenzhen Municipal Transportation Bureau received 71.43% of complaints against the taxi service. Since 2004, nearly 200 taxi strikes have taken place across the country, involving more than 100 cities nationwide. Especially, since 2015, taxi strikes have been spreading to the big cities, including Shenyang, Changchun, Jinan, Chengdu, and other provincial capitals. These phenomena are ultimately due to low income. Taxi drivers tend to avoid driving during peak hours, because they earn less than expected, and do not know where the high-paying areas are. Therefore, instead of being blocked in a traffic jam, drivers plan for suspension of their services, and they have a rest. In order to maximize profits, in the past few years, taxi drivers always reduce the empty rate by increasing the number of passengers, which is influenced by taxi revenue models and the hunting-cruising service method [5,6]. Nowadays, online driving services provide a new opportunity. Taxi drivers can select the destination across different online taxi service platforms. The destination plays an important role in in determining their income.

The orders of ride-hailing and incomes vary with regions and times. Experienced taxi drivers tend to have higher incomes. These drivers usually spend less time looking for new passengers, because they know where the high-income areas are, according to different time periods and regions [7,8]. Therefore, based on driving experience, they can reasonably choose the cruise route and the order destination. However, for most drivers, they do not have enough experience to determine their range of motion to maximize the average hourly income. As a result, there is an interesting phenomenon where passengers cannot find taxis and drivers cannot find passengers. The reason for this embarrassing, as the driver is blind when searching passengers while using an inefficient ride-hailing operation system [9,10].

To improve the taxi operation efficiency, it is essential to understand the distribution of passengers with high profits. Especially during peak hours, driver income is highly reduced by prolonged periods of being stuck in congestion. Although some research indicates that congestion charges will eliminate the impact of congestion, the effects may not seem like much [1]. Activity hotspots are sometimes being misused by taxi drivers to identify high-income areas, but this method may not be referenced to grasp high-driving profits, instead of the density. Therefore, this paper intends to seek the spatiotemporal distribution law of high taxi income from the inherent attributes of the road sections (including directionality, connection attributes, land use characteristics, etc.) as an initial reference for taxi drivers.

This study aims to identify the road segments of high incomes. The payment for each path is calculated to the road segment where the travel starts. Charging methods can be formulated based on the results to improve the operational efficiency of the taxi market. Further suggestions of cruising can be put up for taxi companies and drivers to increase their daily income, and to improve the efficiency of the taxi industry. The contributions of this paper are as follows: (1) The multi-layer road index (MRI) system has been used for map matching and for calculating the spatiotemporal distribution of taxi revenue in units of directional road segments. (2) A large sample of urban road sections is obtained, with the income of each direction, and the tendency of taxi income in different time periods are analyzed. (3) Except for the driver service strategy, this paper fully considers the road characteristics, with directions in impacting the taxi income.

The remainder of this article is organized as follows: Section 2 reviews the methods and factors for identifying taxi incomes; Section 3 presents the data collection methods. Section 4 illustrates the results of the high-income areas and their causes. Section 5 proposes appropriate conclusions and policy recommendations with regard to increasing the taxi incomes.

2. Literature Review

In the past two decades, research into taxi income can be classified by whether taxi GPS data is being used. On the one hand, the research is mainly focused on theoretical investigations without using GPS data. On the other hand, GPS data-based research tends to be problem-solving oriented, and is the mainstream academic research method nowadays. Previous research has evidenced that a certain intelligence lies in taxi drivers' waiting, hunting, and cruising behaviors [11,12]. Following Castro and Gao, etc., to some extent, such behavior indicates the urban wisdom and experience possessed by taxi GPS data [13–15]. The mining of taxi trajectories is also widely applied in path recommendation [16,17], mobility pattern detection [18,19], spatiotemporal dynamics analysis [20,21], and congestion estimation [22] etc., nowadays. Therefore, GPS data-based income research is the trend of future works in taxi investigation.

GPS data-based research has already influenced the income of the taxi drivers. These related researches are mainly focused on revealing the crowd-intelligence of the taxi income, which are affected by the driving model [23], likelihood of pick ups [24], working time [25,26], speed changing [24,27], taxi service strategy, etc. For instance, in the research of taxi service strategy, the main aspects include space-time distribution of passenger sources [28], space-time patterns on cruising trips and stopping spots [29], hot-spot regions [30], mileage or time utilization ratio [31], and region selection [14,32,33], etc. In addition to these factors, there are also some external authors added indicators concerning as the impact factor of the taxi income, such as ticket price [24,34], taxi market supply and demand ratio [35], weather factors [36], etc.

Recently, scholars have started to pay attention to investigating the spatial features of the income. From the behavioral aspects, the spatial distribution of the top rank of taxi drivers (with a higher efficiency then the average) is heterogeneous, while ordinary taxi drivers (with lower efficiency then the average) have a homogenous spatial distribution [10]. Yuan also confirmed that the distribution of high-income travels and high-income regions are imbalanced at the spatial dimension, while the high-income spots and regions are varied at different periods [37]. Liu did not directly confirm that road design and land use affected the income, but they have already affected the demands of taxis [38]. In [29], the authors classified taxi drivers according to income. The Mash map matching and the DBSCAN (density-based clustering of applications with noise) clustering method have been used to compare the time spent on cruise trips and the parking spots of drivers according to income levels.

Previous studies have analyzed the difference between high-income and low-income taxi drivers from different perspectives. However, most of the research has only focused on the time dimension. Form the spatial dimension, there are a few studies involving the quantitative impact of income. For instance, researchers have little knowledge of when and which road sections have a positive impact on improving the income of the taxi driver. Moreover, the spatial analysis focuses more on rough regional analysis instead of refining the road sections of attributes at the city scale. Thus, this research focuses on finding the significant factors that influence incomes, which uses the directional road segment (DRS) as a statistical analysis unit.

In this paper, we first establish a multi-layer road index (MRI) system and map matching based on taxi tracing data. Then, the temporal distribution of the taxi income at the regular DRS perspective is investigated in explaining the time features of the income. The corresponding influencing factors of DRSs and taxi operation strategies are also put forward based on the MRI system. Thirdly, a binary logit model is selected to analysis the sample-based income level and the impact factors. Fourthly, results are presented to explain the contribution of each factor at different periods. This research provides valuable advice and insights for taxi agencies in developing the charging rules. Especially for taxi companies and taxi drivers, the conclusion helps to increase their daily income and to improve the efficiency of the taxi industry.

3. Materials and Methods

The Materials and Methods section includes data collection and data extraction processes, the method of income calculation, the definition of indicators affecting income,

Objectives of the study Point-of-interest Trace Data Input Map Data Input (POI) data Input Multi-layer road index (MRI) system Trip For each Map eparating matching occupied trip £ Model **DRS** income correlated index calculating Sample selection calculating Model correlated index set DRS income set Selected-sample-based binary logit (SBL) model **Dicussion and conclusion**

and the analysis method for causes of income. An overview of the step-by-step study methodology is illustrated in Figure 1.

Figure 1. Study methodology flow chart.

3.1. Multi-Layer Road Index (MRI) System

Most of the current map-matching approaches use algorithms repeatedly, and passively collect and handle local geographic and topological information for each track point; and urban road networks such as multi-layer roads can be particularly complex [39]. Instead of dividing the whole research area into regular grids or suitable TAZs where different types of road segments will be inevitably mixed up, we suggest building a MRI system [40] by extending a road indexing method called "intersection continuity negotiation", proposed by [41,42].

The MRI system is a technique of pre-processing GIS road networks. It classifies all links in the GIS map of road networks into different groups. Each link belongs to only one group. The links in one group are connected and combined into a road unit that is much longer than any link in the group. A GIS map of road networks pre-processed using the MRI system is called an MRI map, which can help researchers to unveil the features of road networks more deeply, because it supports the analysis at DRS levels.

Given an original topological map of urban road networks G (N, L), where N is the set of road network nodes, and L is the set of the road network links. One or more links constitutes a road segment that connects two adjacent intersections, with attributes such as road name, road direction, and road function classification.

Figure 2a shows the basic component of MRI: a link map layer (the lower layer), a directional road segment layer (the middle layer), and an undirected road unit layer (the upper layer). An MRI system is generated by connecting the three layers, with an illustrative example shown in Figure 2b. The link map layer is derived from the original GIS map. Links, road units (RU), and directional road segments (DRSs) are three different analysis units in the MRI system.



Figure 2. Multi-layer road index (MRI) system. (**a**) Three layers in an MRI system. (**b**) An example of the MRI system.

Definition 1 (RU: Road Unit). For a map G (N, L), a RU represents an undirected virtual traffic corridor that is formed using a series of sequentially connected links. The method in constructing the RU set can be referred to in previous studies.

Definition 2 (DRS: Directional Road Segment). In the RU networks of map G, the RU junction is denoted as the intersection of two different RUs. For any two adjacent RU junctions, if their distance is no more than a specific value (e.g., 200 m), they are considered as a "virtual intersection", called the virtual RU junction. The virtual RU junctions and the original RU junctions form the new RU junctions set. Each RU is divided by some of the new RU junctions in the set into a sequence of directional road segment units called DRSs. The DRSs of the same road in the MRI map are paired; each paired DRS represents different directional road segments of the same road.

The MRI system implements the Cross, Continuity and Condition (3C) rules in merging the original map links into DRSs, which makes the system more conducive to revealing the regularity and continuity of the road network, as well as obtaining the driving direction of the taxi driver. Therefore, this proposed MRI system can be extended successfully to the map-matching algorithm, and it is expected to become an invaluable basis for systematic research at the DRS level in the future.

3.2. Study Area and Data

This study selects the main urban area of Shenzhen, China as the case study area. Shenzhen is located at the center of the Pearl River Delta in the south of China, with a land area of 1996.8 km² and an urban population of over 17 million in 2021. It has six districts, including Luohu, Nanshan, Futian, Baoan, Longgang, and Yantian. It is the first Special Economic Zone (SEZ) city after the institution of reform and the Open-Door Policy in China, in 1979. In the past 30 years, the operation of a market economy has made Shenzhen's economy develop rapidly, bringing with it a dramatic population increase and spatial expansion. Shenzhen has experienced rapid urbanization at a rate of more than 78% [33]. Its gross domestic product (GDP) increased from RMB 0.2 billion in 1979 to RMB 3066.48 billion in 2021. Shenzhen Metro is the seventh largest system in China, and embraces additional investments. In addition, Shenzhen has established trans-it-supportive policies to promote urban development along transit corridors and to facilitate TOD (transitoriented development) around transit stations [43]. However, taking Luohu District for example, the coverage of rail stations within 500 m was only 48.6% in the built-up area before 2021 [44]. The mismatch between the demand and supply of rail services is obvious. The traffic condition in Shenzhen has not been improved since the rapid development of rail transit. The worsening of traffic congestion shows that public transportation is still not

attractive, or it is maybe losing its appeal. Therefore, research into enhancing cab services as a link between buses and subways is still worthwhile.

The map used in this research is based on the road network of Shenzhen Transportation Planning, which is fitted with an MRI grid consisting of 2476 road units (RU) and 7862 directional road segments (DRSs) derived from 21,115 map links. Figure 3 is the topological road network of the study area, and describes a general view of the different classifications of the roads. The DRSs of the MRI map of Shenzhen were generated based on the layout of the road network. Each DRS is labeled as a number, as described in Figure 4. The basic statistics of the MRI structure of Shenzhen are shown in Table 1.



Figure 3. Road network structure of Shenzhen (the source of the geographical background is Mapbox).



Figure 4. Example of directional road segments (DRSs) of Shenzhen (the source of the geographical background is Mapbox).

Table 1. Basic statistics of the MRI system of Shenzhen.

Road Level -	Link]	RU		DRS	
	Amount	Length (km)	Amount	Length (km)	Amount	Length (km)	Ratio
Highway	290	1.93	19	18.27	148	2.35	7.79
Arterial road	7940	0.28	319	7.05	2286	0.98	7.17
Secondary road	5354	0.22	643	1.15	1920	0.39	2.99
Branch road	7531	0.24	1495	0.74	3508	0.31	2.35
Total	21115	0.27	2476	1.73	7862	0.59	3.18

3.3. Taxi Trace Data

The taxi GPS trajectory data were collected every 30 s during a week from the City of Shenzhen Taxi (COST) database, between 15 January 2015 and 21 January 2015, containing a total of consecutive 194 million passenger track points produced by 15,726 taxis. In order to facilitate the analysis, the trajectory data were recognized as 4.85 million passenger travel paths [45], and each pathway contained a set of continuous tracing points, while the latitude, longitude, and timestamps were associated.

3.4. Point-of-Interest (POI) Data

Points of interest (POIs) provides rich semantics with the mixture of human movement [46], and have been commonly used in various location-based services tools, such as Uber, DiDi, etc. Concerning the taxi market demand, the distribution of POIs in the DRS represent the features of transportation, such as travel purpose, potential travel distance, time consumption, and even travel mode. This research models the features of the taxi income, using the 248,162 POI data points of seven categories in Shenzhen, including the realty/company (3.66%), store (27.28%), transportation (0.04%), hotel (1.93%), entertainment (21.47%), hospital and clinic (0.04%), and parking (0.83%). All of the POIs will be aggregated in the view of the DRSs, and discretized as the model input factors.

3.5. Calculation of Income

According to the taxi charging standards, we counted the payment of each route on the DRSs, based on the MRI system and taxi trace data. The charging standards were classified by the Guan Nei (Nanshan District, Luohu District, Futian District) and Guan Wai (Longgang District, Longhua District, Baoan District) districts, as shown in Figure 5a. The specifics of the charging standards are seen in Figure 5b. Each DRS's total revenue was calculated by the sum of the route revenue with starting point on this segment.



Figure 5. Shenzhen district partition and taxi charge list (the source of the geographical background is Mapbox).

3.6. Defining Correlated Variables on Each DRS

Based on the established MRI system, four categories of DRS-influencing factors are included, including the road network structure, road section dynamic information, POI distribution, and taxi market operation indicators. These categories are explained as follows.

3.6.1. DRS-Correlated Variables

As a promising application system for map coding, MRI not only provides a complete set of map statistics at the DRS scale, but also makes it possible to analyze the structural characteristics of the road network in depth.

DRS Attributes

For a map G, each map link has a unique DRS label, which in turn is subordinated to a unique RU in representing a continuous traffic corridor. Clearly, the larger the range of the RU to which the DRS belongs (namely, having a larger degree), the more freedom of the path that a trip may choose. Therefore, we use the degree value of the RU as an influencing factor in defining the DRS to which it belongs.

Definition 3 (DRS degree). DRS degree: For a given RUx of a map G, the DRS degree value is the total number of other RUs directly connected to it, corresponding to the DRS degree of that RU. For a single DRS, it may also be directly connected to other DRSs, so that the upstream and downstream DRSs can be defined.

Definition 4 (Number of upstream/downstream DRSs). For a given RUx of a map G, the DRS degree value is the total number of other RUs directly connected to it, corresponding to the DRS degree of that RU. For a single DRS, it may also be directly connected to other DRSs, so that the upstream and downstream DRSs can be defined.

Definition 5 (DRS grade). *DRSy describes the yth route of the DRS. For a given DRSy in a map G, a grade value has been applied to indicate the road class, e.g., 1—highway, 2—arterial, 3—secondary, and 4—branch road. The lower the rank value, the higher its road condition and design speed.*

Definition 6 (DRS length). For a given pair of virtual intersections for each DRS, the length is calculated as the Euclidean distance from the center of one end to the other. As shown in Figure 2, the lengths of DRS1 and DRS2 are the Euclidean distances from the virtual intersection A to B.

DRS Dynamics

This type of DRS feature indicates the market dynamics of the DRS. To be more specific, as we mentioned before, the roads become more congested during peak hours, which reduces the market revenue. However, when the congestion dissipates, the revenue increases. Therefore, we use the average speed of the road segment to reflect the actual state of traffic, as no matter what the conditions are, the congestion will affect the traffic volume on the road networks.

Definition 7 (DRS average travel speed). For a special time period, assuming that *Q* vehicles passes through the DRSy without intentionally suspending movement (for example, waiting for a passenger), the average travel speed of the DRSy is measured as follows:

$$avgSpd_y = \frac{\sum_{q=1}^{Q} d_q^y}{\sum_{q=1}^{Q} t_q^y}$$
(1)

where, d_q^y is the total distance of the car q driving on DRSy within the t_q^y travel time period. According to the four time periods, e.g., the morning peak, evening peak, daytime, and nighttime,

the average travel speed of each DRS is generated respectively, using all of the valid occupied taxi trace data.

The average traffic speed of DRS represents the level of road infrastructure supply that adapts to the variety of traffic activities. Considering the taxi demand, it is also important to reveal the dynamic travel characteristics of passengers on each DRS. We notice that, the travel distance not only reflects these characteristics, but also has a direct impact on the service revenue of each trip. Therefore, the proportion of long-distance trips is included in the analysis as follows.

Definition 8 (Long-distance trip ratio of DRS). For a special time period, a total of P valid trips are supposed staring from DRSy, and the ratio of the long-distance trip DRSy is:

$$LongDist_{y} = \sum_{p=1}^{P} MAX(\frac{dist_{p}^{y} - \gamma}{\left|dist_{p}^{y} - \gamma\right|}, 0)/P$$
(2)

where $dist_p^y$ represents the travel distance of the trip p, where the starting point is on DRSy. γ is the assumed borderline between short distances and long distances, where the value is taken as 10 km.

Aggregating the POI Data

The distribution of different types of POIs near pick-up points may also reveal some potential taxi demand characteristics, such as possible travel purposes [47,48], the expected travel time or distance [49] and even the mobility pattern [46]. The POIs within 250 m of each DRS were statistically calculated, in order to describe the distribution of POIs by each DRS precisely. It is assumed that the more specific the distribution of POIs spread on the DRSs, the more possible travel demands associated with the POIs.

3.6.2. The Index of DRS Taxi Operation Strategies

In addition to obtaining the attributes of DRSs, the indictors of operation strategies are also obtained, in representing the rules of a taxi picking up passengers and passing through the DRS. This indictor suggests the differences of the service conditions of each DRS.

In the age of ride-hailing, the right operating strategies will boost the earnings of taxi drivers [50]. The literature on the operation strategy has covered passenger searching, passenger delivering, and service area selection [27]. The objectives of the passenger-searching strategy include maximizing profit and maximizing demand coverage [51]. In terms of passenger-delivery strategies and service region preference, choosing similar spatiotemporal areas and dropping off customers as quickly as possible can also increase the profit. Additionally, there are operating strategies with descriptive indicators such as occupied distance, occupied time, capacity utilization [14], etc. In the paper, we defined a new concept, "driver operation strategy", from the spatial-temporal perspective.

Definition 9 (Ratio of search distance). This indictor refers to the ratio of deadhead cruising distance to the total cruising distance (deadhead or occupied) generated on DRSy in a certain time period:

$$RSD_y = dist_e^y / dist_e^y + dist_o^y$$
(3)

where, $dist_e^y$ indicates the total deadhead cruising distance on DRSy, and $dist_o^y$ indicates total occupied distance on DRSy.

The higher the RSD, the more likely taxis would move through the road section without passengers; however, this does not always imply a lower travel demand.

Definition 10 (ratio of noninitial price trips). *This indicator is the ratio of the trips that exceed the starting distance on DRSy during a certain period of time:*

$$RNPT_{y} = trip_{e}^{y} / trip_{o}^{y}$$
(4)

where $trip_e^y$ is the number of the daily trips, except for the trips within the distance of the staring price on DRSy. $trip_o^y$ is the number of occupied trips on DRSy.

Definition 11 (Ratio of long service distance). For a special time period, this refers to the ratio of the count of occupied passes longer than 10 km on DRSy, to the total count of occupied passes on that segment. A 10 km distance is also taken as the threshold of the long service distance.

$$RLSD_y = dist_f^y / 8000 * trip_e^y$$
(5)

where dist^y denotes the total follow distance of picked-up customers on DRSy, which is the part that takes away the initial distance (the initial distance value is 2 km in Guan Nei district). For a long service distance of 10 km, the follow distance is 8 km.

Definition 12 (Ratio of occupied trips). This indicator represents the ratio of the occupied trips to the overall trips passing through the DRS in a certain period of time.

$$ROT_{y} = trip_{o}^{y} / trip^{y} y$$
(6)

where trip^y is the total trips that pass through the DRSy.

The correlated variables are described and discretized in Table 2 for the model fitting in the following section.

Variable		Description	Value Type	
	DegreeL	The level of road degree of current DRS. If the road degree value of the RU containing current DRS is less than 12, DegreeL = 1; not less than 25, DegreeL = 3; otherwise, DegreeL = 2.		
	Grade	The road grade of current DRS. For highway or expressway, Grade = 1; arterial road, Grade = 2; secondary road, Grade = 3; branch road, Grade = 4.	Fixed	
Level of DRS Attributes	LengthL	The level of length of current DRS. For DRS with a length of less than 0.7 km, LengthL = 1; not less than 1.3 km, LengthL = 3; otherwise, LengthL = 2.		
	DownstreamNumL	The level of outgoing DRS number of current DRS. For DRS with a downstream DRS number of less than 4, DownstreamNumL = 1; not less than 6, DownstreamNumL = 3; otherwise, DownstreamNumL = 2.		
	UpstreamNumL	The level of incoming DRS number of current DRS. For DRS with a upstream DRS number of less than 4, UpstreamNumL = 1; not less than 6, UpstreamNumL = 3; otherwise, UpstreamNumL = 2.		
Level of DRS	LongDistL	The level of long-distance trip (>10 km) ratio of current DRS. For DRS with a long-distance trip ratio of less than 10%, LongDistL = 1; not less than 30%, LongDistL = 3; otherwise, LongDistL = 2.	Changed for different	
Dynamics	AvgSpdLThe level of average travel speed of current DRS. For DRS with an average travel speed of less than 20 km/h, AvgSpdL = 1; not less than 35 km/h, AvgSpdL = 3; otherwise, AvgSpdL = 2.		time periods	

Table 2. DRS-correlated variable description and classified list.

Variable		Description	Value Type	
	POIL.Realty/ Company	The level of number of realty/company entities on current DRS. For DRS with a number of realty/company entities of less than 20, POIL.Realty/Company = 1; not less than 40, POIL.Realty/Company = 3; otherwise, POIL.Realty/Company = 2.		
	POIL.Store	The level of store number on current DRS. For DRS with a store number of less than 120, POIL.Store = 1; not less than 240, POIL.Store = 3; otherwise, POIL.Store = 2.		
	POIL.Transportation	The level of transportation enterprise numbers on current DRS. For DRS with no transportation enterprise, POIL.Transportation = 1; not less than 1, POIL.Transportation= 2.		
Level of POI	POIL.Hotel	The level of hotel number on current DRS. For DRS with a hotel number of less than 8, POIL.Hotel = 1; not less than 20, POIL.Hotel = 3; otherwise, POIL.Hotel = 2.	Fixed	
	POIL.Entertainments	The level of number of entertainment entities on current DRS. For DRS with a number of entertainment entities of less than 100, POIL.Entertainments = 1; not less than 200, POIL.Entertainments = 3; otherwise, POIL.Entertainments = 2.		
	POIL.Hospital/ Clinic	The level of number of hospital/clinic entities on current DRS. For DRS with a number of hospital/clinic entities of less than 12, POIL.Hospital/Clinic = 1; not less than 24, POIL.Hospital/Clinic = 3; otherwise, POIL.Hospital/Clinic = 2.		
	POIL.Park	The level of park number on current DRS. For DRS with a park number of less than 2, POIL.Park = 1; not less than 4, POIL.Park = 3; otherwise, POIL.Park = 2.		
	RSDL	Top 20% range of RSD, RSDL = 1; bottom 20% of RSD, RSDL = 3; otherwise, RSDL = 2.		
Level of driver	RNPTL	Top 20% range of RNPT, RNPTL = 1; bottom 20% of RNPT, RNPTL = 3; otherwise, RNPTL = 2.	Changed for different	
operation strategy	RLSDL	Top 20% range of RLSD, RLSDL = 1; bottom 20% of RLSD, RLSDL = 3; otherwise, RLSDL = 2.	time periods	
	ROTL	Top 20% range of ROT, ROTL = 1; bottom 20% of ROT, ROTL = 3; otherwise, ROTL = 2.		

Table 2. Cont.

3.7. Applying the Selected Sample-Based Binary Logit (SBL) Model

The binary logit (BL) model [52] has strong interpretative power for mathematical analysis and the odds ratio. This model is one of the most widely used methods in dealing with multi-scene analysis, from a practical statistical point of view. In general, the dependent variable can be viewed as a dichotomous variable to reflect the positive and negative aspects of a phenomenon. We divide the incomes of all directional road segments into two categories with sample selection. Since the sample size is large enough, the top and bottom 20 percent of DRSs (the sample size is 1865) can be selected to exclude abnormal samples.

Model evaluation: p < 0.05 shows the significance threshold of the explanatory variable. In addition, we used the -2 Log likelihood and Pearson's Chi-squared statistics to measure the goodness of fit of the model. Then, pseudo-R2 was used to evaluate the proportion of variance explained by the model to the total variance. Finally, the effectiveness and forecast accuracy of the DRS income model was analyzed from the statistical and logical judgment aspects.

4. Experimental Results and Discussion

4.1. Distribution of Taxi Income on Directional Road Segments (DRSs)

The total incomes of the DRSs where the taxi journey starts were aggregated. In this research, the sample size was 4661 among the 7862 full-scaled road networks in Shenzhen, where the DRS of the daily travel time was less than 5, namely, a weekly travel time of less than 35, this was excluded. The upper part of Figure 6 illustrates the fluctuation of

the income within 24 h, and the lower part shows the hourly income range. To be more specific, the median hourly income for a week is USD 15331 (United States dollars), while the total income is at its lowest level at 4 a.m., suggesting the lowest travel demand at this time. For the morning hours, there is a significant increase in market income, starting at 7 a.m., and with a peak at 9 a.m., which may be largely related to the increase in commuter traffic by taxi in the morning peak. In the evening, the income pinnacle appears at 11 pm. The potential explanation would be the lack of public transportation at that time, because Shenzhen has a large number of leisure events at the end of the day, while the majority of public transportation has been closed.



Figure 6. Trend diagram for income on DRSs.

Figure 7 indicates the regularity of the average speed varying with time at different DRS levels. As can be seen, the average speed exhibits phase characteristics of bimodal variation within the speed restriction of the China's road traffic safety regulations. It is obvious that there are intensive trip demands during the morning and evening peaks, leading to a significant drop in average speeds. At night, the traffic speed starts to increase, which improves the incomes of taxi drivers with a faster speed and efficient operations. We further divided the data by following four time periods, morning peak (7:00–10:00), daytime (10:00–17:00), evening peak (17:00–22:00), and nighttime (22:00–7:00). Additionally, the income frequency of each time period is calculated. Meanwhile, the spatial distribution of income was more divergent over the whole urban area, as Figure 8 shows, in different periods.

Figure 9 is the distribution histogram of DRS income by time period. This plot indicates that the frequency of the DRS daily income follows an exponential distribution, with a small number of DRSs of high income, and certain DRSs of extremely high income. This shows a good exponential distribution of income for the rest of the time period, except for nighttime, when there is less variation within the DRSs for medium income.



Figure 7. Trend diagram for speed on DRSs.



Figure 8. Spatial distributions of incomes (the source of the geographical background is Mapbox).



Distribution of Income

Figure 9. Distribution histogram for DRS income at different time periods.

4.2. Result of the SBL Models and Significant Factors

A selected sample-based binary logit (SBL) was used to evaluate the percentage contributions of DRS incomes during the four time periods and all-day. The top and bottom 20/100 rankings of the DRS incomes were selected in representing the income levels, with 1 =high and 0 =low.

Collinearity can be explained as the excessive correlation between independent variables. As the most commonly used methods for the collinearity diagnostics, variance inflation factors (VIF) are estimated before the model fitting. In general, VIF will be not less than 1, and a bigger VIF means that the explanatory variables have a larger collinearity; while the VIF exceeds 4, the variables need to be re-selected or disregarded. Table 3 shows the results of the VIFs, with a majority value of between 1.1 and 3.4. Obviously, there is multicollinearity in the result of the SBL model of all-day, respecting the daytime and evening peak average speeds. With the variables for POIs of Entertainments multicollinearity being serious in all of these five models, this may be closely related to the POI of Store, in representing the consumption characters. Therefore, the POIL. Entertainments was excluded from the input factors for the SBL model fitting, while in the model for all-day, the variables of daytime and evening peak were disregarded.

Table 3. Collinearity detection of the factors affecting DRS incomes.

	VIF for Different Time Periods						
Variable	All-Day	Morning Peak	Daytime	Evening Peak	Nighttime		
DegreeL	1.637	1.611	1.605	1.566	1.456		
Grade	1.667	1.683	1.611	1.646	1.505		
LengthL	1.279	1.283	1.239	1.286	1.215		
DownstreamNumL	2.360	2.464	2.207	2.193	2.302		
UpstreamNumL	2.365	2.539	2.224	2.200	2.280		
AvgSpdL(Morning peak)	3.329	1.293	n.a.	n.a.	n.a.		
AvgSpdL(Daytime)	4.933	n.a.	1.276	n.a.	n.a.		
AvgSpdL(Evening peak)	4.659	n.a.	n.a.	1.325	n.a.		
AvgSpdL(Nighttime)	2.444	n.a.	n.a.	n.a.	1.278		
LongDistL	2.107	1.787	2.105	1.889	1.987		
POIL.Realty/Company	2.209	2.232	2.229	2.203	1.988		
POIL.Store	2.917	2.862	3.103	2.947	2.809		
POIL.Transportation	1.116	1.112	1.109	1.113	1.125		
POIL.Hotel	2.618	2.662	2.644	2.546	2.578		
POIL.Entertainments	5.088	5.223	5.432	5.248	4.899		
POIL.Hospital/Clinic	3.167	3.108	3.273	3.265	3.194		
POIL.Park	1.559	1.526	1.521	1.511	1.544		
RSDL	2.520	1.184	2.292	2.292	2.135		
RNPTL	1.249	1.130	1.174	1.199	1.161		
RLSDL	2.146	1.777	2.145	1.910	2.028		
ROTL	2.257	1.042	2.227	2.124	2.024		

n.a. = not applicable or not available.

Table 4 shows the performances of the five SBL models. For each model, the *p* value is less than 0.05, which means that at least one input factor has been detected for having a significant odds ratio. In addition, the pseudo-R-squared value indicates that the variation degree of dependent variable can be explained by the input factors. Most of the values of pseudo-R2 are over 0.6, representing a good explanatory power of the SBL models and selected variables. Additionally, the accuracy of each model for DRS income is around 85 percent, with a narrow fluctuation of less than 3 percent, except for the accuracy of the nighttime model, which is 70.5%. The nighttime DRS income may be slightly weaker in prediction and explanation, because the sample size is not balanced enough, and the working hours of taxi drivers are characterized by fluctuation.

		A			
Period	Log Likelihood	Pearson's X ²	p Value	Pseudo R ²	Accuracy (%)
All-day	1115.242	10.568	0.000	0.727	86.5
Morning peak	1202.063	12.042	0.000	0.624	82.8
Daytime	1046.861	24.963	0.000	0.695	86.2
Evening peak	1152.677	10.024	0.000	0.646	83.1
Nighttime	1850.375	5.397	0.000	0.274	70.5

 Table 4. Overall model performance statistics.

Table 5 shows the SBL model's results of different time periods. The insignificant factors (p > 0.05) have been excluded. Explanations of each variable are shown in Table 2. It is obviously that the DRS attributes (i.e., DegreeL, Grade, Length, DownstreamNumL, AvgSpdL, LongDistL, and POIL) and the taxi operation strategies (i.e., RSDL, RLSDL, ROTL, and RNPTL, respectively) are relevant to the incomes. Specifically, Grade, DownstreamNumL, AvgSpdL, RSDL, RLSDL, and ROTL are the common impact factors correlating with DRS incomes. In general, concerning the taxi operation strategies, RSDL, representing the efficiency of customer search, as well as RLSDL, representing the long-distance ratio of delivery passengers, had highly negative impacts on the incomes, representing the ROTL of the passenger turnover ratio boost incomes. In the all-day model, the indictor of ODDs indicates that drivers should certainly be more hard working or flexible, and cruise to increase their incomes (positive odd ratio (OR) of ROTL = 3.730; negative OR of RLSDL = 0.323; negative OR of RSDL = 0.087).

The degree of the DRS dominates the attractiveness of the road segment, while the grade of the DRS determines its accessibility. The number of the downstream DRS increases the probability that passengers choose the current DRS as the starting point, and the average travel speed also reflects the efficiency of the taxi transport. For example, in the all-day model, positive OR of DownstreamNumL = 2.612, positive OR of DegreeL = 1.833, negative OR of Grade = 0.611, and negative OR of AvgSpdL = 0.570. This shows that the importance of the road in the road network, the connection form, and the degree of accessibility will all have an impact on the DRS income.

It is worth noting though that POI is also an important factor impacting on incomes, but with different specific types in each model. Specifically, five types of POI appeared to be the significant factors distributed throughout various periods, including "Hotel" (only nighttime), "Scenic" (all periods except nighttime), "Realty&Company" (all periods except nighttime), "transportation" (only nighttime), and "Hospital&Clinic" (all periods except daytime).

General analysis: Grade, AvgSpdL, RSD, and RLSD are negatively correlated with the DRS income. The lower Grade indicates the lower level of activities, which leads lower taxi market demands and incomes. Nevertheless, the lower AvgSpdL is related to a higher building density and taxi market demand, which refers to a higher income. Interestingly, a higher RLSD does not lead to an increase in DRS income, but rather, to a decrease. This result suggests that distant travel will decrease the incomes because of its time consuming nature. The number of downstream DRS, Degree, and the ROT are positively correlated with the incomes. This result indicates that more downstream DRSs, a greater degree, and a higher incidence of carrying passengers, with more potential taxi demands directly causes a higher revenue concentration effect in the market.

The only difference is that RNPT is positively correlated with the all-day (coefficient 0.537, p < 0.05) and daytime models (coefficient 0.451, p < 0.05), and negatively correlated with the nighttime model (coefficient -0.364, p < 0.05). This finding implies that income is reduced as a result of more long-distance trips taken at night.

Table 5.	Results	of the	SBL	models	5.
lable 5.	Results	of the	SBL	models	5.

Period	Variable	Coefficient	Std.err.	p Value	Odds Ratio	95% Cont	. Interval
	DegreeL	0.606	0.132	0.000	1.833	1.416	2.373
	Grade	-0.493	0.104	0.000	0.611	0.499	0.748
	LengthL	0.336	0.140	0.016	1.400	1.064	1.841
	DownstreamNumL	0.960	0.126	0.000	2.612	2.039	3.346
	AvgSpdL(Morning peak)	-0.540	0.178	0.002	0.583	0.411	0.827
	AvgSpdL(Nighttime)	-0.562	0.183	0.002	0.570	0.399	0.815
	LongDistL	0.437	0.190	0.021	1.548	1.068	2.245
All-day	POIL.Realty/Company	0.626	0.122	0.000	1.870	1.473	2.373
7 m-uay	POIL.Hospital/Clinic	0.472	0.127	0.000	1.604	1.251	2.056
	POIL.Park	0.425	0.116	0.000	1.529	1.218	1.919
	RSDL	-2.445	0.211	0.000	0.087	0.057	0.131
	RNPTL	0.537	0.142	0.000	1.710	1.295	2.258
	RLSDL	-1.130	0.196	0.000	0.323	0.220	0.474
	ROTL	1.316	0.180	0.000	3.730	2.624	5.304
	Constant	0.796	0.919	0.387	2.216	n	a.
	DegreeL	0.556	0.125	0.000	1.744	1.367	2.227
	Grade	-0.492	0.097	0.000	0.611	0.505	0.740
	DownstreamNumL	0.757	0.158	0.000	2.133	1.564	2.909
	UpstreamNumL	0.353	0.155	0.023	1.423	1.049	1.930
	AvgSpdL(Morning peak)	-0.666	0.119	0.000	0.514	0.407	0.649
	LongDistL	0.331	0.143	0.021	1.392	1.052	1.843
Morning peak	POIL.Realty/Company	0.579	0.116	0.000	1.784	1.420	2.242
	POIL.Hospital/Clinic	0.394	0.119	0.001	1.483	1.174	1.873
	POIL.Park	0.250	0.109	0.022	1.284	1.037	1.589
	RSDL	-2.516	0.151	0.000	0.081	0.060	0.108
	RLSDL	-0.678	0.158	0.000	0.508	0.372	0.693
	ROTL	0.409	0.121	0.001	1.505	1.187	1.908
	Constant	2.764	0.646	0.000	15.869	n	a.
	DegreeL	0.663	0.133	0.000	1.940	1.495	2.518
Daytime	Grade	-0.492	0.105	0.000	0.611	0.498	0.751
	DownstreamNumL	1.089	0.125	0.000	2.971	2.327	3.793
	AvgSpdL(Daytime)	-0.674	0.130	0.000	0.509	0.395	0.658
	LongDistL	0.575	0.182	0.002	1.778	1.245	2.538
	POIL.Realty/Company	0.883	0.111	0.000	2.419	1.945	3.008
	POIL.Park	0.509	0.114	0.000	1.664	1.332	2.080
	RSDL	-2.191	0.194	.000	0.112	0.076	0.164
	RNPTL	0.451	0.137	.001	1.570	1.200	2.054
	RLSDL	-1.069	0.193	.000	0.343	0.235	0.501
	ROTL	1.166	0.179	.000	3.209	2.261	4.554
	Constant	-0.138	0.899	0.878	0.871	n	a.

Period	Variable	Coefficient	Std.err.	p Value	Odds Ratio	95% Con	f. Interval
	DegreeL	0.628	0.128	0.000	1.875	1.457	2.411
	Grade	-0.558	0.102	0.000	0.572	0.469	0.698
	DownstreamNumL	0.915	0.121	0.000	2.496	1.969	3.163
	AvgSpdL(Evening peak)	-0.926	0.128	0.000	0.396	0.308	0.509
	LongDistL	0.459	0.158	0.004	1.583	1.162	2.156
Evening peak	POIL.Realty/Company	0.412	0.117	0.000	1.509	1.199	1.900
01	POIL.Hospital/Clinic	0.525	0.124	0.000	1.690	1.326	2.155
	POIL.Park	0.517	0.114	0.000	1.676	1.342	2.095
Nighttime	RSDL	-1.660	0.177	0.000	0.190	0.134	0.269
	RLSDL	-0.838	0.167	0.000	0.433	0.312	0.601
	ROTL	1.500	0.177	0.000	4.482	3.166	6.346
	Constant	-0.363	0.863	0.674	0.695	n.a.	
	Grade	-0.506	0.074	0.000	0.603	0.522	0.697
	LengthL	0.430	0.092	0.000	1.538	1.285	1.840
	DownstreamNumL	0.406	0.092	0.000	1.501	1.255	1.796
	AvgSpdL(Nighttime)	-0.872	0.099	0.000	0.418	0.344	0.508
	POIL.Transportation	0.358	0.169	0.034	1.431	1.027	1.993
	POIL.Hotel	0.350	0.108	0.001	1.419	1.149	1.752
	POIL.Hospital/Clinic	0.215	0.102	0.036	1.239	1.014	1.515
	RSDL	-0.527	0.171	0.002	0.591	0.422	0.826
	RNPTL	-0.364	0.092	0.000	0.695	0.580	0.832
	RLSDL	-0.652	0.091	0.000	0.521	0.436	0.623
	ROTL	0.321	0.170	0.059	1.378	0.988	1.921
	Constant	3.027	0.812	0.000	20.633	n	.a.

Table 5. Cont.

n.a. = not applicable or not available.

4.3. Discussion and Implications

Although the OLS model has indicated the main factors determining DRS income, the significance of each indicator varies over time. More specifically, the differences between the results of these five models are as follows:

- (1) The DRS length only impacts on the nighttime income. This result may be attributed to the sparse taxi demands, where the longer the length of the DRS, the higher the incomes.
- (2) Degree and LongDist have no impact on the nighttime model, which are due to the high driving speed and the dispersions of the travel destination at nighttime.
- (3) The number of upstream DRSs only has a positive impact on market revenue during the morning peak. This phenomenon can be explained as the greater number of upstream and downstream DRSs, which contribute to alleviating the traffic congestion, as well as increasing the incomes. This result also indicates that the travel demand is more intensive in the morning rush hours compared with other time periods.
- (4) The POIL of Realty/Company, Hospital/Clinic, and Park has a greater contribution to the DRS income in the peak hours and all-day models, comparing to the other POI type. The Hospital/Clinic-type POI is not related to the DRS income in the daytime model, as it is consistent with the travel characteristics of patients. The nighttime model has a distinctive feature, in that the POI of Transportation (OR 1.431; p < 0.05) has a positive impact on DRS incomes, in accordance with the land use of Hotel (OR 1.239; p < 0.05) and Hospital/Clinic (OR 1.419; p < 0.05) at night.
- (5) RNPT only has an impact in the model of peak traffic hours, owing to the traffic congestion causing incomplete trips within the starting distance.

The key contribution of this study is that, instead of using the driver as the statistical unit, the branch road section is utilized to integrate taxi income and to analyze the elements that affect income level. The taxi GPS trajectory is accurately matched by direction, using a new map-matching technique. At the same time, it increases the effect of the characteristics of the branch road itself on income, in addition to taking traditional taxi demand, driver service strategy indicators, and driving speed into account. The behavioral motive indicating driver intelligence was left out of the influencing elements due to the lack of sufficient driver information. The impacts of significant influencing factors on the income levels of DRSs are summarized as follows:

- (1) Taxi operation strategies are important aspects in determining DRS income, among which, the RSD is the most indicative factor in describing the service efficiency. These results are consistent with findings in similar research [35], which indicates that high-income DRSs often have shorter search distances.
- (2) The number of downstream DRSs is the determinant factor affecting incomes. From the angle of a complex network, more travel options coinciding with the taxi driving direction contribute more to the DRS income than the degree of interconnection.
- (3) In the nighttime model, AvgSpdL is the only significant indicator affecting income, and it is also the only controllable factor that can increase the efficiency of the distribution performance.
- (4) In the evening peak model, AvgSpdL is also a key factor in influencing the option of travel path. During the evening peak, this indictor is particular important in avoiding the congested section of the road, to drive up the speed and increase the income. These results are consistent with findings in similar research [37], which found that the income is greatly affected by traffic conditions in the evening peak hours.
- (5) POI types have different effects on DRS income at different time periods, but "Scenic" and "Realty&Company" are constant factors that affect income. As with more "Realty&Company" being more likely to form larger crowds, the surrounding roads with "Scenic" will also generate a lot of taxi demand for foreign tourists, thereby increasing the income. These results are consistent with findings in similar research [38].

5. Conclusions

This research uses the SBL model in investigating the impact factors of the DRS incomes. We have made up for the blankness of analyzing taxi drivers' incomes from the perspective of DRSs. Four kinds of influencing factors have been classified according to the impacts of different time periods, to improve the accuracy of the analysis of DRS income. The major findings are included as follows:

- (1) There exists a marked difference in DRS incomes. The average hourly incomes within the study hours have a mean of USD 15,331 and a standard deviation of USD 3952. The gap between the lowest average hourly income of DRS and the highest average hourly income of DRS, which approaches USD 17,000, is larger.
- (2) The main factors in income analysis are the factors used to represent taxi operation strategies and the number of downstream DRSs. RSD (coefficient –2.445), RLSD (coefficient –1.13), and ROT (coefficient 1.316) are significant operational measures of the taxi market, according to the SBL all-day model, which was tested over time. The daytime, nighttime, and all-day models are all possible with RNPT. In addition, DownstreamNumL is a very important element of its positive effect in the five models, which were found to have ORs of 2.612, 2.133, 2.971, 2.496, and 1.501 during the all-day, morning peak, daytime, evening peak, and nighttime, respectively. This conclusion can be used as a starting point for further research into the taxi market revenue, from both the driver and DRS perspectives.
- (3) The factors that influence incomes in different time periods are completely different. DRSs with more real estate/companies, hospitals with many upstream roads, degrees, and high road grades are more high-income DRSs during the morning peak. The DRSs with several realty/companies, hospitals, and parks nearby, as well as more downstream roads, more degrees, and higher road ranks, are high-income areas during the evening peak. During the daytime, high-income DRSs congregate in areas

with a lot of real estate/companies and parks, as well as a lot of downstream roads, a lot of degrees, and a lot of long-distance travel. DRS income distribution is more scattered at night, with fewer impact factors, but a higher grade, more downstream, longer road length, and adjacent hotels, traffic stations, and hospitals corresponding to high-income road sections were also identified.

These findings quantitatively demonstrate the influence of road attributes on the spatiotemporal distribution of the income, from a novel directional perspective. For different time periods, a summary of the characteristics of the high-income DRSs can provide new ideas for taxi industry managers to balance the income levels of drivers. A selected order delivery process by the managers can reduce the proportion of low-income drivers, as well as improving the overall efficiency of taxi drivers. Two suggestions are being recommended for the taxi industry managers: (1) Taxi charge rules should take traffic circumstances into account, especially in locations where the average speed is excessively slow during the morning and evening peak hours. (2) The vehicle dispatch system should offer a fair opportunity to drivers traveling to high-income areas.

In this research, a feasible model for large-scale DRS data mining operation is provided. As far as we are aware, this is the first demonstration of how to investigate taxi revenue with a procedure of map preprocessing, and the trajectory matching method. For future study, this method could be utilized in analyzing the impact of taxis on urban carbon emissions, finding restricted areas, road sections, and directions for taxi drivers, controlling the total amount of taxis, etc. Although this article only focuses on the physical road elements that drive DRS incomes, the fundamental method and research structure can be utilized for relevant factors, such as human characteristics. In the following phase, we plan to gather private information on drivers, such as their gender, age, and whether or not they have children, etc., and analyze how these factors may affect taxi revenues.

Author Contributions: Conceptualization, Ming Cai; methodology, Shuxin Jin; data curation, Zhouhao Wu; writing—original draft preparation, Shuxin Jin; writing—review and editing, Tong Shen; visualization, writing—review, Di Wang. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Fundamental Research Funds for the Central Universities, Sun Yat-sen University (Grant number 22dfx08), the National Key R&D Program of China (Grant number 2020YFB1600400), the Special Scientific Research Program of the Education Department of Shaanxi Province of China (Grant number 19JK0382).

Data Availability Statement: The data presented in this study are available on request to the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Lozano, J.A.; Macías, J.A.G.; Chávez, E. Crowd location forecasting at points of interest. Int. J. Ad Hoc Ubiquitous Comput. 2015, 18, 191–204. [CrossRef]
- Ashkrof, P.; Correia, G.H.D.A.; Cats, O.; van Arem, B. Understanding ride-sourcing drivers' behaviour and preferences: Insights from focus groups analysis. *Res. Transp. Bus. Manag.* 2020, *37*, 100516. [CrossRef]
- Castro, P.S.; Zhang, D.; Chen, C.; Li, S.; Pan, G. From taxi GPS traces to social and community dynamics. ACM Comput. Surv. 2013, 46, 1–34. [CrossRef]
- Chen, Y.; Fu, Q.; Zhu, J. Finding next high-quality passenger based on spatio-temporal big data. In Proceedings of the 2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA), Chengdu, China, 10–13 April 2020; pp. 447–452.
- 5. Cramer, J.; Krueger, A.B. Disruptive Change in the Taxi Business: The Case of Uber. Am. Econ. Rev. 2016, 106, 177–182. [CrossRef]
- 6. Duong, H.L.; Chu, J.; Yao, D. Taxi Drivers' Response to Cancellations and No-Shows: New Evidence for Reference-Dependent Preferences. *Manag. Sci.* 2022. [CrossRef]
- Gao, Y.; Xu, P.; Lu, L.; Liu, H.; Liu, S.; Qu, H. Visualization of Taxi Drivers' Income and Mobility Intelligence. In *International Symposium on Visual Computing*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 275–284. [CrossRef]
- 8. He, Z.; Xi-Wei, S.; Zhuang, L.; Nie, P. On-line map-matching framework for floating car data with low sampling rate in urban road networks. *IET Intell. Transp. Syst.* **2013**, *7*, 404–414. [CrossRef]

- 9. Hu, B.; Xia, X.; Sun, H.; Dong, X. Understanding the imbalance of the taxi market: From the high-quality customer's perspective. *Phys. A Stat. Mech. Its Appl.* **2019**, *535*. [CrossRef]
- 10. Liu, X.; Sun, L.; Sun, Q.; Gao, G. Spatial Variation of Taxi Demand Using GPS Trajectories and POI Data. J. Adv. Transp. 2020, 2020, 7621576. [CrossRef]
- 11. Lai, X.; Fu, H.; Li, J.; Sha, Z. Understanding drivers' route choice behaviours in the urban network with machine learning models. *IET Intell. Transp. Syst.* **2018**, *13*, 427–434. [CrossRef]
- 12. Liu, L.; Andris, C.; Ratti, C. Uncovering cabdrivers' behavior patterns from their digital traces. *Comput. Environ. Urban Syst.* 2010, 34, 541–548. [CrossRef]
- 13. Liu, Q.; Ding, C.; Chen, P. A panel analysis of the effect of the urban environment on the spatiotemporal pattern of taxi demand. *Travel Behav. Soc.* **2019**, *18*, 29–36. [CrossRef]
- 14. Lv, Q.; Qiao, Y.; Ansari, N.; Liu, J.; Yang, J. Big Data Driven Hidden Markov Model Based Individual Mobility Prediction at Points of Interest. *IEEE Trans. Veh. Technol.* 2016, 66, 5204–5216. [CrossRef]
- 15. Maruthasalam, A.P.; Roy, D.; Venkateshan, P. Refuse or Accept?: Analysis of Taxi Driver Operating Strategies in E-Hailing Platforms. *Soc. Sci. Electron. Publ.* **2018**. [CrossRef]
- Menard, S.W. Quantitative Applications in the Social Sciences. In *Applied Logistic Regression Analysis*; Sage Pubns: Thousand Oaks, CA, USA, 2013.
- Naji, H.; Wu, C.; Hui, Z.; Li, L. Towards understanding the impact of human mobility patterns on taxi drivers' income based on GPS data: A case study in Wuhan—China. In Proceedings of the 2017 4th International Conference on Transportation Information and Safety (ICTIS), Banff, AB, Canada, 8–10 August 2017; pp. 1152–1160. [CrossRef]
- 18. Nie, Y. How can the taxi industry survive the tide of ridesourcing? Evidence from Shenzhen, China. *Transp. Res. Part C Emerg. Technol.* **2017**, *79*, 242–256. [CrossRef]
- 19. Oleyaei-Motlagh, S.Y.; Vela, A. Inferring demand from partially observed data to address the mismatch between demand and supply of taxis in the presence of rain. *arXiv* **2019**, arXiv:1903.06619.
- Ou, G.; Wu, Y.; Wang, G.; Guo, Z. Big-data-based analysis on the relationship between taxi travelling patterns and taxi drivers' incomes. In Proceedings of the 2019 16th International Conference on Service Systems and Service Management, ICSSSM, Shenzhen, China, 13–15 July 2019; pp. 1–6. [CrossRef]
- Porta, S.; Crucitti, P.; Latora, V. The network analysis of urban streets: A dual approach. *Phys. A Stat. Mech. Its Appl.* 2006, 369, 853–866. [CrossRef]
- Qin, G.; Li, T.; Yu, B.; Wang, Y.; Huang, Z.; Sun, J. Mining factors affecting taxi drivers' incomes using GPS trajectories. *Transp. Res.* Part C Emerg. Technol. 2017, 79, 103–118. [CrossRef]
- 23. Qu, M.; Zhu, H.; Liu, J.; Liu, G.; Xiong, H. A cost-effective recommender system for taxi drivers. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 24–27 August 2014. [CrossRef]
- Rong, H.; Zhou, X.; Yang, C.; Shafiq, Z.; Liu, A. The rich and the poor: A Markov decision process approach to optimizing taxi driver revenue efficiency. In Proceedings of the International Conference on Information and Knowledge Management, Beijing, China, 24–28 October 2016; pp. 2329–2334. [CrossRef]
- Salanova, J.M.; Estrada, M.; Aifadopoulou, G.; Mitsakis, E. A review of the modeling of taxi services. *Procedia-Soc. Behav. Sci.* 2011, 20, 150–161. [CrossRef]
- Scellato, S.; Musolesi, M.; Mascolo, C.; Latora, V.; Campbell, A.T. NextPlace: A spatio-temporal prediction framework for pervasive systems. In Proceedings of the International Conference on Pervasive Computing, San Francisco, CA, USA, 12–15 June 2011; pp. 152–169. [CrossRef]
- 27. Sirisoma, R.M.N.T.; Wong, S.C.; Lam, W.H.K.; Wang, D.; Yang, H.; Zhang, P. Empirical evidence for taxi customer-search model. *Proc. Inst. Civ. Eng.-Transp.* 2010, 163, 203–210. [CrossRef]
- Sun, L.; Zhang, D.; Chen, C.; Castro, P.S.; Li, S.; Wang, Z. Real Time anomalous trajectory detection and analysis. *Mob. Netw. Appl.* 2012, 18, 341–356. [CrossRef]
- 29. Tang, L.; Sun, F.; Kan, Z.; Ren, C.; Cheng, L. Uncovering distribution patterns of high performance taxis from big trace data. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 134. [CrossRef]
- Tang, L.; Zheng, W.; Wang, Z.; Hong, X.U.; Hong, J.; Dong, K. Space Time Analysis on the Pick-up and Drop-off of Taxi Passengers Based on GPS Big Data. J. Geo-Inf. Sci. 2015, 17, 1179–1186. [CrossRef]
- 31. Tu, J.; Duan, Y. Detecting Congestion and Detour of Taxi Trip via GPS Data. In Proceedings of the 2017 IEEE 2nd International Conference on Data Science in Cyberspace, DSC 2017, Shenzhen, China, 26–29 June 2017; pp. 615–618. [CrossRef]
- 32. Wang, D.; Miwa, T.; Morikawa, T. Interrelationships between traditional taxi services and online ride-hailing: Empirical evidence from Xiamen, China. *Sustain. Cities Soc.* **2022**, *83*. [CrossRef]
- 33. Wang, D.; Miwa, T.; Morikawa, T. Comparative Analysis of Spatial–Temporal Distribution between Traditional Taxi Service and Emerging Ride-Hailing. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 690. [CrossRef]
- 34. Wang, W.; Pan, L.; Yuan, N.; Zhang, S.; Liu, D. A comparative analysis of intra-city human mobility by taxi. *Phys. A Stat. Mech. Its Appl.* **2015**, *420*, 134–147. [CrossRef]
- Wang, Y.; Wu, Z.; Li, C. The Complexity of Large-scale Urban Networks: A Comparative Study in China. In Proceedings of the Transportation Research Board Annual Meeting, Washington, DC, USA, 11–15 January 2015; pp. 15–4997.

- Wong, R.; Szeto, W.; Wong, S.C.; Yang, H. Modelling multi-period customer-searching behaviour of taxi drivers. *Transp. B Transp. Dyn.* 2013, 2, 40–59. [CrossRef]
- Wu, Z.; Xie, J.; Wang, Y.; Nie, Y.M. Map matching based on multi-layer road index. Transp. Res. Part C Emerg. Technol. 2020, 118, 102651. [CrossRef]
- Yang, J.; Su, P.; Cao, J. On the importance of Shenzhen metro transit to land development and threshold effect. *Transp. Policy* 2020, 99, 1–11. [CrossRef]
- Ye, Y.; Zheng, Y.; Chen, Y.; Feng, J.; Xie, X. Mining Individual Life Pattern Based on Location History. In Proceedings of the 2009 Tenth International Conference on Mobile Data Management: Systems, Services and Middleware, Taipei, Taiwan, 18–20 May 2009; pp. 1–10. [CrossRef]
- 40. Yuan, C.; Geng, X.; Mao, X. Taxi High-Income Region Recommendation and Spatial Correlation Analysis. *IEEE Access* 2020, *8*, 139529–139545. [CrossRef]
- 41. Yuan, C.; Wu, D.; Wei, D.; Liu, H. Modeling and Analyzing Taxi Congestion Premium in Congested Cities. *J. Adv. Transp.* 2017, 2017, 2619810. [CrossRef]
- 42. Yuan, J.; Zheng, Y.; Zhang, L.; Xie, X.; Sun, G. Where to find my next passenger. In Proceedings of the 13th international Conference on Ubiquitous Computing, Beijing, China, 17–21 September 2011; pp. 109–118. [CrossRef]
- Yuan, N.J.; Zheng, Y.; Zhang, L.; Xie, X. T-Finder: A Recommender System for Finding Passengers and Vacant Taxis. *IEEE Trans. Knowl. Data Eng.* 2012, 25, 2390–2403. [CrossRef]
- 44. To Create a "Main Position" for Social Integration between Shenzhen and Hong Kong, Shenzhen Luohu Strives to Build a Pioneer Area for the Shenzhen-Hong Kong Port Economic Belt. Available online: http://www.tzb.sz.gov.cn/xwzx/gzdt/gqgz/content/ post_825145.html (accessed on 16 July 2022).
- 45. Yu, X.; Gao, S.; Hu, X.; Park, H. A Markov decision process approach to vacant taxi routing with e-hailing. *Transp. Res. Part B Methodol.* **2019**, *121*, 114–134. [CrossRef]
- Zhang, D.; Sun, L.; Li, B.; Chen, C.; Pan, G.; Li, S.; Wu, Z. Understanding Taxi Service Strategies From Taxi GPS Traces. *IEEE Trans. Intell. Transp. Syst.* 2014, 16, 123–135. [CrossRef]
- Zhang, W.; Huang, B.; Luo, D. Effects of land use and transportation on carbon sources and carbon sinks: A case study in Shenzhen, China. *Landsc. Urban Plan.* 2014, 122, 175–185. [CrossRef]
- 48. Zheng, Y. Trajectory Data Mining. ACM Trans. Intell. Syst. Technol. 2015, 6, 1–41. [CrossRef]
- 49. Zheng, Y.; Capra, L.; Wolfson, O.; Yang, H. Urban Computing. ACM Trans. Intell. Syst. Technol. 2014, 5, 1–55. [CrossRef]
- 50. Zhang, Y.; Zhong, M.; Jiang, Y. A data-driven quantitative assessment model for taxi industry: The scope of business ecosystem's health. *Eur. Transp. Res. Rev.* 2017, *9*, 23. [CrossRef]
- 51. Zheng, Z.; Rasouli, S.; Timmermans, H. Modeling taxi driver search behavior under uncertainty. *Travel Behav. Soc.* 2020, 22, 207–218. [CrossRef]
- Zhou, B.; Ma, L.; Hu, J.; Wu, S.; He, G. Extraction of Urban Hotspots and Analysis of Spatial interaction Based on Trajectory Data Field: A Case Study of Shenzhen City. *Trop. Geogr.* 2019, *39*, 117–124. [CrossRef]