Sequential Logit Approach to Modeling the Customer-search Decisions of Taxi Drivers

Ryan C.P. WONG [a], W.Y. SZETO [b], S.C. WONG [c]

[a,b,c] *Department of Civil Engineering, The University of Hong Kong, Hong Kong*
[a] *E-mail: ryancpw@hku.hk*
[b] *E-mail: ceszeto@hku.hk*
[c] *E-mail: hhecwsc@hku.hk*

Abstract: This paper develops an enhanced sequential logit model to depict the customer-search behavior of vacant-taxi drivers. This model considers that vacant-taxi drivers can change the choices they make on their way to a designated district. Trip data from the global positioning systems (GPSs) of 460 urban taxis were extracted to calibrate the model and verify the factors underlying the drivers' search decisions. The findings reveal that the proposed sequential logit model is capable of predicting the search paths of vacant-taxi drivers. This model form is considered more informative for policymakers who aim to study search paths and the associated traffic congestion contributed by taxis in each district.

Keywords: Sequential Logit Model, Urban Taxis, Customer-search, Search Path, Global Positioning System Data.

1. INTRODUCTION

In Hong Kong, Taiwan, mainland China and many other parts of the world, taxis tend to circulate in search of customers, consuming road space and worsening traffic congestion in the process. Regulation policies have been implemented to tackle these problems, often in the forms of price controls and entry restrictions. Various economic studies have been conducted (Fernández *et al*., 2006; Manski and Wright, 1976; Schaller, 1998; Yang *et al*., 2005) to investigate the consequences of such taxi regulation policies. However, these studies have been developed based on an idealized market and conventional economic analysis, in which the spatial structure of the market is ignored.

In an attempt to capture the spatial structure of the market, Yang and Wong (1998) developed a model to determine the taxi movements on a given road network for a given customer origin-destination (OD) demand pattern. In this model, the customer-search behavior of taxi drivers is based on the assumption that each vacant-taxi driver attempts to minimize his or her expected search time to find a customer. However, the solution method it proposes is a fixed-point algorithm that is not guaranteed to converge. Wong and Yang (1998) reformulated the model to develop a more efficient and convergent iterative balancing algorithm. However, neither model considers congestion effects, customer demand elasticity, multiple user classes or multiple taxi modes.

Wong *et al*. (2001) extended the work of Yang and Wong (1998) to develop a model incorporate congestion effects and customer demand elasticity, and proposed a new solution algorithm based on a quasi-Newtonian approach to solve their model. Wong *et al*. (2002) further developed a sensitivity-based solution algorithm for the taxi network model proposed by Wong *et al*. (2001) and showed it to be more efficient. This taxi network model was calibrated and validated by Yang *et al*. (2001) using Hong Kong data, and was then applied

by Yang *et al*. (2002) to investigate the nature of the demand and supply equilibrium under a prescribed fare structure and fleet size regulation in regulated, competitive and monopolistic markets for taxi services in the urban area of Hong Kong. The model was later extended by Wong *et al*. (2008) to consider multiple user classes, multiple taxi modes and customers' hierarchical modal choice.

Considering that the profit taxi drivers could gain by picking up customers in these remote areas would be considerably higher than that earned in an urban area, Wong *et al*. (2003) explicitly examined the effects of perceived profitability on the customer-search behavior of vacant taxis and the profit taxi drivers could be expected to gain by picking up customers in particular zones. Yang *et al*. (2010a) further extended this concept to "profit per unit time" by incorporating the operational cost and time consumption involved in taxi trips to maximize profits from customer searches.

All of these taxi network models ignore the taxi-search behavior of customers in response to local variations in the level of taxi services. In response to this lack, Wong *et al*. (2005) developed a mathematical model to consider the bilateral search behavior of vacant taxis and customers using the absorbing Markov chain approach. Yang *et al*. (2010b) developed another mathematical model to characterize the same bilateral search behavior on road networks by incorporating a macroscopic search-and-meet model into the network model proposed by Wong *et al*. (2003).

Among the existing taxi network models, logit-based search models are often used to describe the customer-search behavior of vacant-taxi drivers based on the hypotheses that the drivers are attempting to either minimize their search time or maximize their perceived profits. However, few studies have validated the hypothesized customer-search behavior of taxi drivers. Sirisoma *et al*. (2010) determined the significant factors affecting vacant-taxi drivers' decisions to search for customers across districts based on the data collected from a stated preference survey. Wong *et al*. (2013) extended that study and developed a destination-based multinomial logit (MNL) model based on the global positioning system (GPS) data from 460 urban taxis to predict the drivers' zonal choices in searching for customers at various times of day. The findings indicated that around 80% of taxi drivers intended to circulate within their current zones during peak hours when the overall passenger demand was high, but preferred cross-zonal travel to high-demand districts during the mid-night period.

The MNL model developed by Wong *et al*. (2013) assumes that vacant-taxi drivers make their individual zonal decisions only once during each customer-search cycle. Consequently, the model does not consider that the drivers could meet and pick up a customer on the way to the original designated district, or that the zonal decision could be changed along the way. The destination-based approach may be accurate enough to predict the strategic decisions made on the search-location choices of vacant-taxi drivers at the zonal level. However, it may also ignore the possibility of the drivers changing their eventual destination choices while traveling to a distant (not adjacent) district, for reasons such as picking up a customer in an intermediate district along the way. This paper builds and tests an enhanced sequential logit (ESL) model to address this potential discrepancy and improve upon the MNL model. The proposed model is inspired by the traditional sequential logit (SL) model, which allows the modeling of sequential decision making. However, because the traditional SL model cannot suit our purposes directly, we improve upon it to do so.

This paper also attempts to apply our proposed model to predict the search paths of vacant-taxi drivers and vacant-taxi distribution in Hong Kong, and illustrates the properties and performances of the model based on a case study. *The contributions of this paper are (1) its proposal of a new model that depicts the behavior of vacant-taxi drivers who are searching for customers and (2) its presentation of a case study that illustrates the application*

of that model to real-world situations.

The remainder of the paper proceeds as follows. Section 2 describes the data collection method. Section 3 presents the model structure of our ESL model and variables and introduces the model calibration process. Section 4 examines the drivers' customer-search strategies at different periods and discusses the model's results. Section 5 concludes the paper and suggests a future research direction.

2. DATA

2.1 Data Collection

At present, there are 18,138 taxis in Hong Kong, of which 15,250 are urban taxis, 2,838 serve the New Territories and 50 serve Lantau. In this study, field survey data were collected from GPS devices previously installed in 460 urban taxis that tracked the drivers' daily activities. Using satellite communication, the database recorded the taxis' locations in terms of longitude and latitude, spot travel speeds and occupational statuses every 30 seconds. The activities recorded included waiting at taxi stands, circulating on roads for customers, responding to call requests, waiting at call-request locations, driving customers to their destinations, stopping for breaks and traveling to get fuel. However, the information available in the dataset to distinguish between these activities was insufficient. Although no information on the fares was collected, the location and speed information were used to deduce the vehicle trajectories and wait times and hence the fares collected. Moreover, because HK$5 is charged for telephone bookings and it is not difficult for a passenger to catch a vacant taxi along an urban roadside in Hong Kong, telephone bookings are uncommon. We consider such bookings as having a negligible effect on our proposed customer-search model. Hence, we believe that the dataset provided sufficient information to track the urban taxis' daily operations in general and to develop our model.

The sample of 460 taxis represents approximately 3% of the entire Hong Kong-area taxi population, and adequately represents the travel behavior of the region's drivers. The GPS survey data were collected during the week spanning August 16-23, 2009. The taxis' customer-search data were extracted from three periods, i.e., the morning peak (07:30-09:30), evening peak (17:00-19:00) and mid-night (01:00-05:00) periods. Trips that occurred outside of these periods were excluded from the data analysis, and only entire journeys completed within these periods were considered. The periods were selected based on the results of Wong *et al.* (2013), which found that the hourly variation in passenger demand was small during each of these periods and that cross-district customer-search behavior remained unchanged over a single day. The average values of travel distance and time, revenue and operational cost were determined to develop a discrete choice model.

We suspected that some of the extracted occupied or vacant trips were actually non-existent, such as trips of exceptionally long or short travel distances and durations (i.e., occupied trips traveling less than 1 minute, or vacant trips traveling farther than 30 km). The recordings of these trips could have been extracted due to GPS device malfunctions, poor connectivity to satellites in urban areas surrounded by high-rise buildings or human error by the taxi drivers operating the devices. Therefore, a data-screening process was carried out to eliminate such trips. However, some non-search-related unoccupied trips, such as trips or roadside stops taken for short breaks or trips to a gas station, could not be eliminated. Nevertheless, based on the information obtained from taxi drivers, the number of these trips was small. Hence, we consider these trips to have had a negligible effect on the accuracy of

our customer-search model. We also observed that the number of records for certain OD pairs did not achieve a reasonable frequency (i.e., occurring fewer than five times during the relevant period). These pairs were considered insufficient for giving a representative average value, and were thus excluded.

2.2 Taxi Travel Fare and Operational Cost

Taxi fares in Hong Kong are based on a non-linear fare structure according to the type of taxis. In August 2009, the initial charge for the first 2 km in all urban taxis was set at HK$18. Every subsequent 200 m or any part thereof, or every 1 minute of waiting time or any part thereof, cost an additional HK$1.5. This unit charge decreased to HK$1 after the chargeable amount reached HK$70.5. Passengers also had to pay for any toll charges incurred on the journey, and surcharges on return tolls involving the use of the three cross-harbor tunnels or the Lantau Link (connecting the airport on Lantau Island with the city center). An additional HK$5 was also charged for each piece of baggage, each animal and each hiring arranged through telephone booking. The taxi fare structure was the same throughout the day, with no surcharges applied at night or during peak hours.

According to a survey conducted by the Hong Kong Transport and Housing Bureau (2008), the monthly taxi rental cost for a driver was about HK$9,009, drivers operated about 25 days per month. With urban taxis running a normal shift of 12 hours, the rental cost was approximately HK$0.50 per minute of operation.

The fuel cost is directly related to the cost of liquefied petroleum gas (LPG), the fuel used by all taxis in Hong Kong. Based on a trial carried out by the Hong Kong Planning, Environment and Lands Bureau (1998) on the fuel consumption of taxis, a normal taxi consumed approximately 0.14 liter of LPG per kilometer traveled. The unit cost of LPG in August 2009 was HK$3.26 per liter, which was also the mean ceiling price of LPG controlled by the Hong Kong Electrical and Mechanical Services Department (2009). This put fuel at a cost of HK$0.46 per kilometer.

The GPS data offered us the coordinates and associated times of the relevant taxis every 30 seconds. These data were used to estimate the traveled distance (by summing all of the Euclidean distances defined by two consecutive points) and the travel time (by determining the end time minus the start time) of each single trip. Furthermore, the coordinates could be used to deduce the vehicle trajectories and determine which toll roads had been used. To determine the waiting time, we simply assumed that the occupied taxis with spot speeds of less than 5 km/h recorded every 30 seconds had been waiting for 30 seconds. This implies that waiting times were rounded up to the nearest 30 seconds. Knowing the extracted travel distance, estimated waiting time, toll roads used, unit costs for traveling and waiting, toll charges and nonlinear fare structure, the associated taxi fare and profit could be determined.

2.3 Study Area

In this paper, we analyze the sequential-zone decisions made by vacant-taxi drivers as they search for customers over time. To facilitate this purpose, we divided the Hong Kong territory into 18 zones according to their administrative districts. While these districts were set based on population size, each retained a certain degree of land use homogeneity, as shown in Figure 1. Nine of the zones were in rural areas (New Territories, Zones 1-9) and the remaining zones were in urban areas (Kowloon and Hong Kong Island, Zones 10-18). Whereas urban taxis are permitted to operate in all districts in Hong Kong, New Territories taxis are only allowed to cruise for customers in Tuen Mun, Yuen Long, Tai Po, the North

District and parts of the Sha Tin, Tsuen Wan, Kwai Tsing, Sai Kung, and Islands Districts. Lantau taxis are only authorized to operate in the Islands District and part of the Tsuen Wan District.
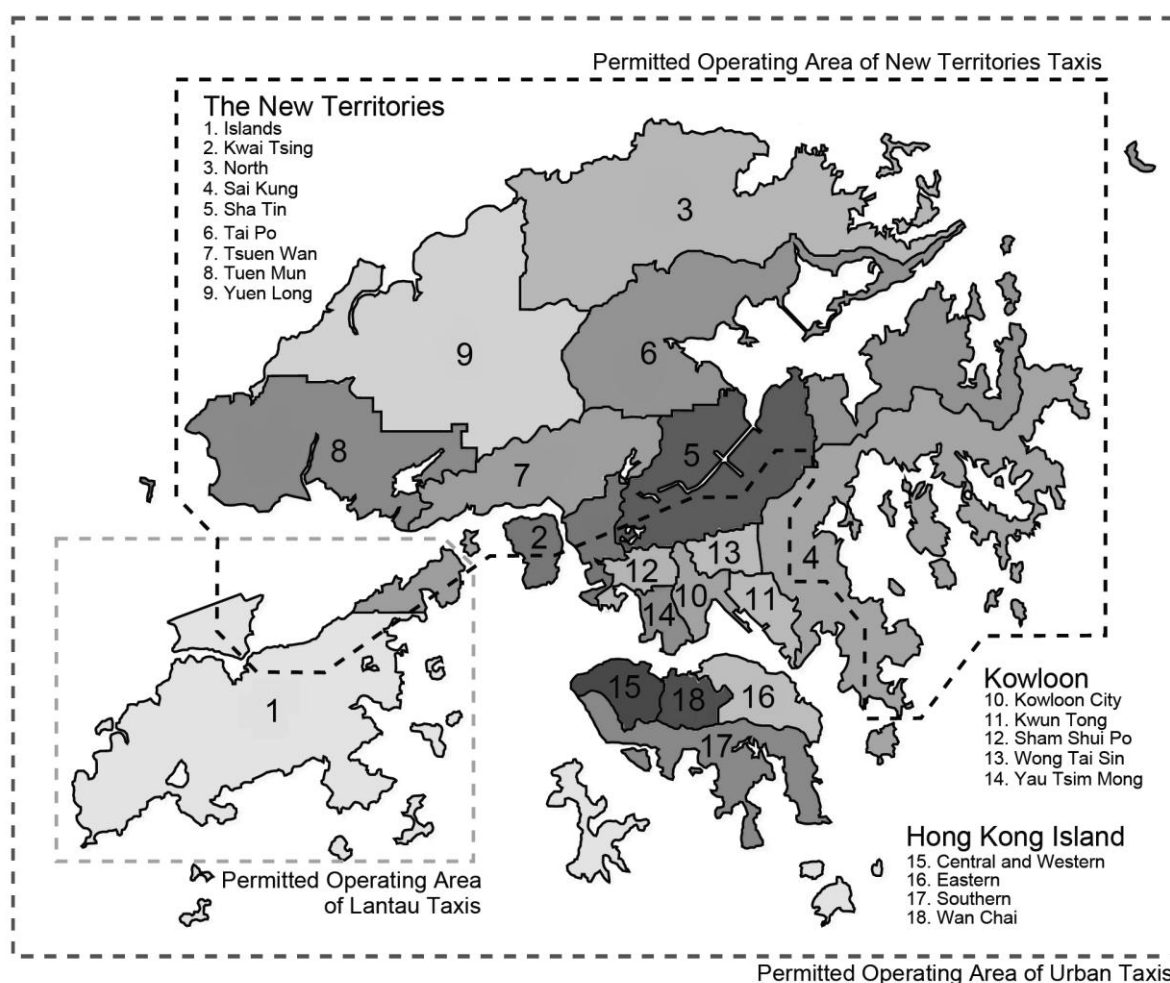

Figure 1.   District map and permitted operating areas of taxis

3. METHOD

3.1 ESL Model Structure

The MNL model is commonly used in discrete choice modeling to predict the probability of selection between alternatives, and has been used to model the decisions of vacant-taxi drivers in the literature. If vacant-taxi drivers decide on strategic districts for finding customers at the beginning of the customer-search cycle and do not want to pick up (due to a lower expected profit) or finding it impossible to meet a customer (through having to travel across districts via expressways), then the MNL approach is appropriate. However, in most cases, vacant-taxi drivers do not have perfect foresight, and may change their plans along the way after meeting a customer or facing serious congestion. To model this change in plans, we assume that the drivers make sequential decisions on the way to distant designated districts.

At first glance, the SL model seems appropriate for our purposes. While it is more complicated, it can represent hierical or sequential decisions in many applications. Whereas the MNL model is based on the assumption that an individual chooses an alternative after

considering the entire choice set simultaneously, the SL model is based on the assumption that an individual's choice process consists of sequential and independent choices. Several studies have compared simultaneous- and sequential-choice models and discussed their properties (Gudishala and Wilmot, 2012; Kaha and Morimune, 1979; Nagakura and Kobayashi, 2009; Ophem and Schram, 1997).

However, we found that we could not directly apply the SL model to our application for several reasons. First, determining the average number of decisions made (i.e., the number of levels required) by taxi drivers must be calibrated from the real data. However, the number of levels in the traditional SL model is known in advance. Second, in the traditional SL model, while the choices are independent, the choices between different levels can be the same (i.e., the choices are not independent). A zone could be adjacent to two zones that are also adjacent to each other. Such a zone could be reached directly from one of the two zones or after visiting both zones. Hence, the SL model required improvement to address these cases.

Figure 2 illustrates the decision tree of the ESL model developed in this study. The ESL model assumes that all of the zonal customer-search decisions made at each level are made independently without depending on previous and subsequent decisions, and that the decisions are made sequentially. Moreover, while the choices on one level are independent, they could be dependent or the same between levels. In each district, the driver decides whether to search for customers in the current district or in one of the adjacent districts, and this decision is modeled by the MNL model. The driver then similarly decides whether to search locally or in the adjacent zones. This sequential process is repeated M times. This parameter M is calibrated from the GPS data. It is used to reflect the *average* number of decisions made before reaching a customer. The model allows us to capture the case in which a taxi driver picks up a customer on the way to a distant district and hence abandons the route to the distant district. It allows us to trace the zones to be passed through before a customer is reached, or the search paths taken to reach the customer.
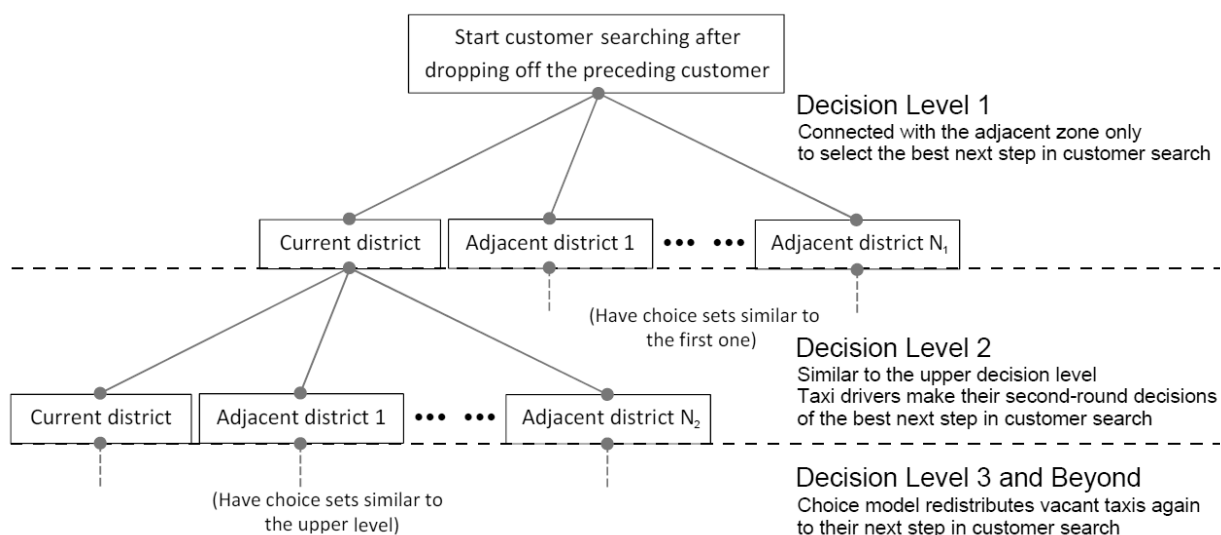

Figure 2.   Decision tree of our proposed sequential logit model

3.2 Variable Definitions

The zonal decisions at each level are affected by the following factors.

Relative Passenger Demand: Higher relative passenger demand for taxi services in particular zones usually attracts more vacant taxis to those zones, where the probability of

finding a customer is higher. Hence, relative passenger demand is used to capture this factor. The relative passenger demand (E_j) of zone j is defined as the number of customers (O_j) picked up in that zone compared with the total number of customers picked up in the entire study area, and is mathematically expressed by

$$E_j = \frac{O_j}{\sum_m O_m}. \tag{1}$$

This definition assumes that the pickup rate is the same in each zone for each period during the day, and that each vacant taxi takes one customer. In that case, even though the pickup rate is not 100%, the relative passenger demand is equal to the relative *actual* passenger demand. The actual passenger demand multiplied by the pickup rate gives the number of customers picked up by taxis, and the pickup rate in each zone is cancelled out when the relative actual passenger demand is calculated.

Intra-zonal Circulation Distance and Time: These factors can be interpreted as the investment taxi drivers make when searching for their next customer within a designated zone, which may be the destination zone of the previous trip. Intra-zonal circulation distance and time represent how far and how long a vacant-taxi driver travels in a designated zone, respectively. The intra-zonal circulation distance depends on the driver's route choices, which in turn depend on his or her experience in picking up customers. Meanwhile, intra-zonal circulating time (also known as waiting time) is indirectly related to passenger demand. The intra-zonal circulating distance D_j^{c} and time T_j^{c} of zone j are calculated by averaging all of the vacant trips' travel distances and lengths of time spent searching for customers within the boundaries of zone j, respectively.

Cross-zonal Travel Distance and Time: These factors can also be interpreted as the investment that vacant-taxi drivers make in searching for their next customer when they travel to a different designated zone. A lengthy travel distance to a designated zone implies a high fuel consumption cost, which discourages vacant-taxi drivers from traveling to that zone. Hence, cross-zonal travel distance affects the search choices of vacant-taxi drivers. A lengthy travel time to a zone also discourages drivers. However, unlike travel distance, travel time captures the congestion effect, which depends on the time of day. While there is no congestion effect during the mid-night period, the congestion effect causes longer travel times during peak hours. In this paper, D_{ij}^{t} and T_{ij}^{t} denote the travel distance and time vacant taxis spend traveling from zone i to zone j, respectively, which are set to equal the corresponding means of all of the vacant trips from the starting point in zone i to the boundary of its designated zone j. When the destination zone of the previous trip is the same as the designated zone (i.e., $i = j$) and the vacant trips do not involve cross-district travel, the travel time and distance are set at zero, and the circulating time and distance are used to capture the distance and time factors, respectively.

Costs of Intra-zonal Circulation and Cross-zonal Travel: These cost factors affect the search choices of vacant-taxi drivers, as a high cost of traveling toward or circulating within a designated zone generally discourages taxi drivers from searching for customers in that zone. The distance and time costs for both cross-zonal travel and intra-zonal circulation include the fuel cost per unit of distance (C^{f}), rental cost per minute (C^{r}) and the associated toll charges. The cost of intra-zonal circulation within a designated zone j can be estimated by

$$C_j^{\text{c}} = C^{\text{f}} D_j^{\text{c}} + C^{\text{r}} T_j^{\text{c}}. \tag{2}$$

The cost of cross-zonal travel from zone i to zone j to search for a new customer can

similarly be expressed as

$$C_{ij}^{\mathrm{t}} = C^{\mathrm{f}} D_{ij}^{\mathrm{t}} + C^{\mathrm{r}} T_{ij}^{\mathrm{t}} + \tau_{ij}, \tag{3}$$

where τ_{ij} is the toll charge associated with the trip from zone i to zone j.

Expected Profit: A higher expected profit from an occupied trip starting from a particular zone can attract more vacant-taxi drivers to that zone. Hence, expected profits also affect the zone choices of vacant-taxi drivers. The expected profit from an occupied trip starting from a particular zone is calculated by the average taxi fare paid by passengers, excluding the corresponding average operational cost of the occupied trips. The average operational cost of all of the occupied trips starting from zone j ($\bar{C}_j^{\mathrm{t}}$) is given by

$$\bar{C}_j^{\mathrm{t}} = C^{\mathrm{f}} \bar{D}_j^{\mathrm{t}} + C^{\mathrm{r}} \bar{T}_j^{\mathrm{t}} + \bar{\tau}_j, \tag{4}$$

where $\bar{D}_j^{\mathrm{t}}$, $\bar{T}_j^{\mathrm{t}}$ and $\bar{\tau}_j$ are the average travel distance and time, respectively, and the associated toll charge of all of the occupied trips starting from zone j. The expected profit ($\bar{P}_j$), which represents the in-pocket profit to taxi drivers, can then be determined by

$$\bar{P}_j = \bar{F}_j - \bar{C}_j^{\mathrm{t}}, \tag{5}$$

where $\bar{F}_j$ is the average taxi fare of all of the occupied trips starting from zone j, according to the non-linear taxi fare structure mentioned in Section 2.2.

Rate of Return: The rate of return includes factors such as the cost of cross-zonal travel, the cost of intra-zonal circulation, the expected profit, the cross-zonal travel and intra-zonal circulation time of a vacant taxi and the expected travel time of an occupied trip. The rate of return is based on a search cycle that consists of the vacant trip from the destination zone of the previous occupied trip to the designated zone, and the subsequent occupied trip starting from the designated zone. The rate of return is equal to the expected profit obtained from the search cycle over the expected time taken to obtain that profit.

The expected profit ($\bar{P}_{ij}$) obtained from a search cycle involving taxis traveling from zone i to zone j to search for new customers is determined by

$$\bar{P}_{ij} = -C_{ij}^{\mathrm{t}} - C_j^{\mathrm{c}} + \bar{P}_j. \tag{6}$$

The expected time spent on the search cycle ($\bar{T}_{ij}$) is obtained by

$$\bar{T}_{ij} = T_{ij}^{\mathrm{t}} + T_j^{\mathrm{c}} + \bar{T}_j^{\mathrm{t}}. \tag{7}$$

The (average) rate of return (R_{ij}) for a vacant taxi traveling from zone i to zone j to search for its next customer can be expressed as

$$R_{ij} = \frac{\bar{P}_{ij}}{\bar{T}_{ij}}. \tag{8}$$

3.3 Model Form

Based on the preceding variables, the ESL search model is developed to depict the sequential customer-search decisions of vacant-taxi drivers. The trip distribution probability at each level generally follows that estimated by the MNL model (McFadden, 1974). The probability $P_q^l(j|i)$ of an individual vacant-taxi driver starting trip q at zone i at level l and selecting zone j (at the same level) is formulated as follows:

$$P_q^l(j|i) = \begin{cases} 0 & \text{if } j \notin A_i \cup \{i\} \\ \exp(V_{ijq}) \Big/ \sum_{n \in A_i \cup \{i\}} \exp(V_{inq}) & \text{if } j \in A_i \cup \{i\} \end{cases}, \qquad (9)$$

where A_i denotes the set of zones adjacent to zone i. V_{ijq} is a deterministic utility that captures the factors influencing a taxi driver currently in zone i to travel to zone j on trip q to search for a new customer. According to Equation (9), when $j \notin A_i$ or when zone j is not adjacent to zone i at any level l, clearly no trip will be made from zones i to j, and consequently the associated probability is zero. Otherwise, the probability is determined based upon the MNL model, with the alternatives comprising every zone adjacent to zone i and zone i itself.

Let $\hat{P}^l(j|i)$ be the *cumulative* probability of a vacant-taxi driver at zone i at level 1 eventually selecting zone j at level l. When the optimal number of levels M is equal to 1, we have $\hat{P}^M(j|i) = P_q^1(j|i)$. This implies that the trip distribution of vacant taxis is obtained simply using an MNL model.

When $M = 2$, we have $\hat{P}^1(j|i) = P_q^1(j|i)$ and $\hat{P}^2(j|i) = \sum_{k \in A_j \cup \{i\}} \hat{P}^1(k|i) P_q^2(j|k)$.

That is, when the average number of search levels M is equal to 2, $\hat{P}^1(j|i)$ is obtained as in the case when $M = 1$, and the probability $\hat{P}^2(j|i)$ is equal to the probability of vacant-taxi drivers moving from zone i to an intermediate zone k times the probability of the subsequent decision to travel to zone j. Zone k must be adjacent to both zones i and j. Because more than one possible search path could link zones i and j, a summation form is introduced to aggregate the probability with different search paths.

When $M = 3$, we have $\hat{P}^1(j|i) = P_q^1(j|i)$, $\hat{P}^2(j|i) = \sum_{k \in A_j \cup \{i\}} \hat{P}^1(k|i) P_q^2(j|k)$ and $\hat{P}^3(j|i) = \sum_{k \in A_j \cup \{i\}} \hat{P}^2(k|i) P_q^3(j|k)$. Hence, when $M \geq 2$, we have

$$\hat{P}^l(j|i) = \begin{cases} P_q^l(j|i) & \text{if } l = 1 \\ \sum_{k \in A_j \cup \{i\}} \hat{P}^{l-1}(k|i) P_q^l(j|k) & \text{if } l = 2,...,M \end{cases}. \qquad (10)$$

The deterministic utility V_{ijq} is expressed mathematically as

$$V_{ijq} = \beta^E E_j + \beta^t D_{ij}^t + \beta^c D_j^c + \beta^R R_{ij}, \qquad (11)$$

where β^E is the coefficient associated with the relative passenger demand in zone j, β^t is the coefficient associated with the cross-zonal travel distance from the starting zone i to the district boundary of zone j, β^c is the coefficient associated with the intra-zonal circulation distance of the customer search conducted within zone j and β^R is the coefficient associated with the rate of return for the search cycle.

It is important to clarify that the attributes in these utility functions are indeed the perceived values of a vacant-taxi driver before he or she makes the decision to conduct a

customer search in a particular district. Vacant-taxi drivers are assumed to take the average values calculated from the trips in our samples, which can be considered as unbiased estimates of the long-term average by an individual driver based on his or her experience. According to this assumption, every taxi driver has the same perception of the utility function attributes. Therefore, for the sake of simplicity, the subscript q is omitted in each attribute.

3.4 Model Calibration

The calibration process has two steps. The first step is to calibrate the coefficient of each variable, and the second is to determine the optimal number of levels. According to the structure of the ESL model, decisions are made independently and sequentially. Each search option level is indeed described by an MNL model form. The zone choice to be made in a particular district is *the same at any level*. Hence, the variable coefficients are calibrated under the maximum log-likelihood principle at the first search-decision level ($l=1$) by considering the following objective function:

$$\text{Maximize LL} = \sum_i \sum_j \ln P_q^1(j|i), \tag{12}$$

where LL is the log-likelihood of the estimated SL model at the first search-decision level, in which $P_q^1(j|i)$ is the calculated probability from Equation (9).

Taking the probabilities of trip distributions to the adjacent zones at any level into account, the total number of predicted vacant trips terminating in each district can be determined. The total number of predicted vacant trips (T_j^l) ending at zone j at level l is calculated as follows:

$$T_j^l = \sum_m \left[F_m \hat{P}^l(j|m) \right], \tag{13}$$

where F_m is the total number of trips generated from zone m (i.e., the total number of occupied taxis dropping off their customers at zone m). According to Equation (13), the total number of predicted trips ending at zone j after l decisions is equal to the sum of the product of the total number of trips generated at each zone times the corresponding probability that the trips will end at zone j.

In the second step, the determination of the optimal number of search-decision levels M is based on a root mean square error (RMSE) evaluation. The RMSE for a particular search-decision level l is defined as the square root of the average of the sum of the square difference between the vacant-taxi zonal distributions (formed by T_j^l for all j) estimated by the model with the number of search-decision levels l and that observed from the GPS data. The lower the RMSE, the better the fit of the model to the taxi trip data. Hence, the optimal number of search-decision levels is the number that gives the lowest RMSE. M can be expressed mathematically as

$$M = \arg \min_l \text{RMSE} = \sqrt{\frac{\sum_{m=1}^{N} \left(T_m^o - T_m^l \right)^2}{N}}, \tag{14}$$

where T_m^o is the observed number of vacant taxis terminating at zone m during customer searches and N is the number of zones.

4. RESULTS AND DISCUSSION

4.1 Cross-zonal Customer-search Pattern

Table 1 illustrates the customer-search strategies of vacant-taxi drivers during the relevant periods. The drivers' cross-district searches were easily observed during the mid-night period, but not during the peak periods. During the morning and evening peak periods, when passenger demand was high, most of the drivers (79% and 82%, respectively) easily found new customers within the districts where their previous customers ended their trips. Consequently, they had a low motivation to cross districts in search of new customers. However, during the mid-night period, when passenger demand was not high, the drivers had to spend more time and travel longer distances on average, including across districts, to search for customers. About 14% of the drivers reached customers in distant districts. These findings provide a general concept of the cross-zonal search decisions of vacant-taxi drivers and the average number of districts reached to meet customers during different periods.

Table 1. Search decisions of vacant-taxi drivers during different periods

Search Decision		Number of Vacant Taxi Trips (%)*		
		Morning peak	Evening peak	Mid-night
Cruising customers in current district		8,730 (79%)	9,554 (82%)	5,508 (58%)
Traveling to other districts	Adjacent districts	1,932 (17%)	1,741 (15%)	2,699 (28%)
	Distant districts (not adjacent district)	392 (4%)	307 (3%)	1,325 (14%)

* The values in brackets represent percentages of the total.

4.2 Model Results

The typical modeling software NLOGIT, which uses the maximum likelihood method, was adopted to determine the coefficients of each variable in the choice models at each level. Table 2 shows the variable coefficients for the customer-search models in terms of sequential-zone decisions.

Table 2. Coefficients of the utility functions and their t-statistics for the various periods

Explanatory variables	Coefficients* [t-statistics]**		
	Morning peak	Evening peak	Mid-night
Relative passenger demand (%)	0.08 [17.2]	0.05 [28.0]	0.06 [83.7]
Cross-zonal travel distance (km)	-0.19 [-14.7]	-0.26 [-17.8]	-0.30 [-31.9]
Intra-zonal circulating distance (km)	-0.13 [-4.7]	-0.88 [-24.0]	-0.86 [-32.0]
Rate of return (HK$/min)	2.38 [43.2]	2.29 [36.1]	0.52 [12.8]

* All of the parameters are significant at the 1% level.
** The values in brackets represent the t-statistics of the explanatory variables.

None of the coefficients in the three models equals zero at the 1% significance level. Hence, the relative passenger demand, cross-zonal travel distance, intra-zonal circulating distance and rate of return are considered significant factors affecting the sequential-zone decisions of vacant-taxi drivers when searching for customers.

The results of the preceding model allow us to draw the following conclusions. First, the coefficient signs were logical in terms of representing drivers' search decisions. For example, the coefficient of cross-zonal travel distance was negative, meaning that the lengthy

cross-zonal travel distance to an adjacent zone indeed discouraged the drivers from searching for customers in that zone. Moreover, the rate of return had a positive coefficient in each modeling period. This agrees with the expectation that a zone's higher rate of return would attract more drivers to that zone. Second, the coefficients of the relative passenger demand in each period were positive and fairly close to each other, suggesting that the drivers had similar perceptions of passenger demand throughout the day. Third, the drivers' perceptions of intra-zonal circulating distance fluctuated more across the periods, indicating that they were more willing to circulate within a zone to search for customers during the morning peak period and preferred to wait for customers during the evening peak period. Fourth, the magnitude of the coefficient of the rate of return for the mid-night period was obviously smaller than that for each of the rush-hour periods. This could be explained by the high profit variation during the mid-night period, which made it difficult for the drivers to estimate their perceived profits.

A key point worth discussing is the differences in the coefficients of cross-zonal travel distance and intra-zonal circulating distance during the different periods, as shown in Table 2. In theory, 1 km of travel is equal to 1 km of circulation, and the coefficients of these two variables should be more or less identical. However, the results show that for the morning peak period, the drivers preferred circulating within the same district rather than traveling to other districts. However, for the evening peak and mid-night periods, they preferred to travel to other districts. This illustrates that the drivers' perceptions of the profitability involved in traveling to other districts and circulating within the same district changed throughout the day. In both cases, 1 km of travel was *not* equal to 1 km of circulation.

4.3 Model Accuracy and Applications

To evaluate the prediction accuracy, we determined the optimal number of search-decision levels required by the model for each period and the corresponding RMSE. We found that both the morning and evening peak models achieved their minimum RMSEs of 100 and 116, respectively, when the number of levels was equal to one. These findings are reasonable and agree with the results in Table 1 related to the customer-search strategies of vacant-taxi drivers over time. As illustrated in Table 1, more than 95% of the drivers met their next customers in the current district or an adjacent district during peak hours. Hence, we expected and found that the optimal number of search-decision levels is equal to one. In this way, the ESL model behaves like the MNL model. The MNL model was good enough to estimate the trip distributions of vacant taxis during the peak periods and could be used to provide valuable information to policymakers on peak-period operations. However, our more general ESL model could be used for the same purpose.

For the mid-night period, Figure 3 demonstrates the actual zonal distribution of vacant taxis deduced directly from the GPS data and the estimated zonal distribution under various numbers of search-decision levels. The RMSEs at each level are also shown in this figure to illustrate how they were affected by the numbers of search-decision levels. It is notable that the RMSE was 276 at one search level, dropped to its minimum amount of 211 when the number of search levels increased to two and grew to 225 when the number further increased to three or higher. Moreover, the estimated zonal distribution of vacant taxis during the mid-night period fit the actual zonal distribution most when two search levels were involved. This implies that the optimal number of search levels is two (i.e., $M = 2$).
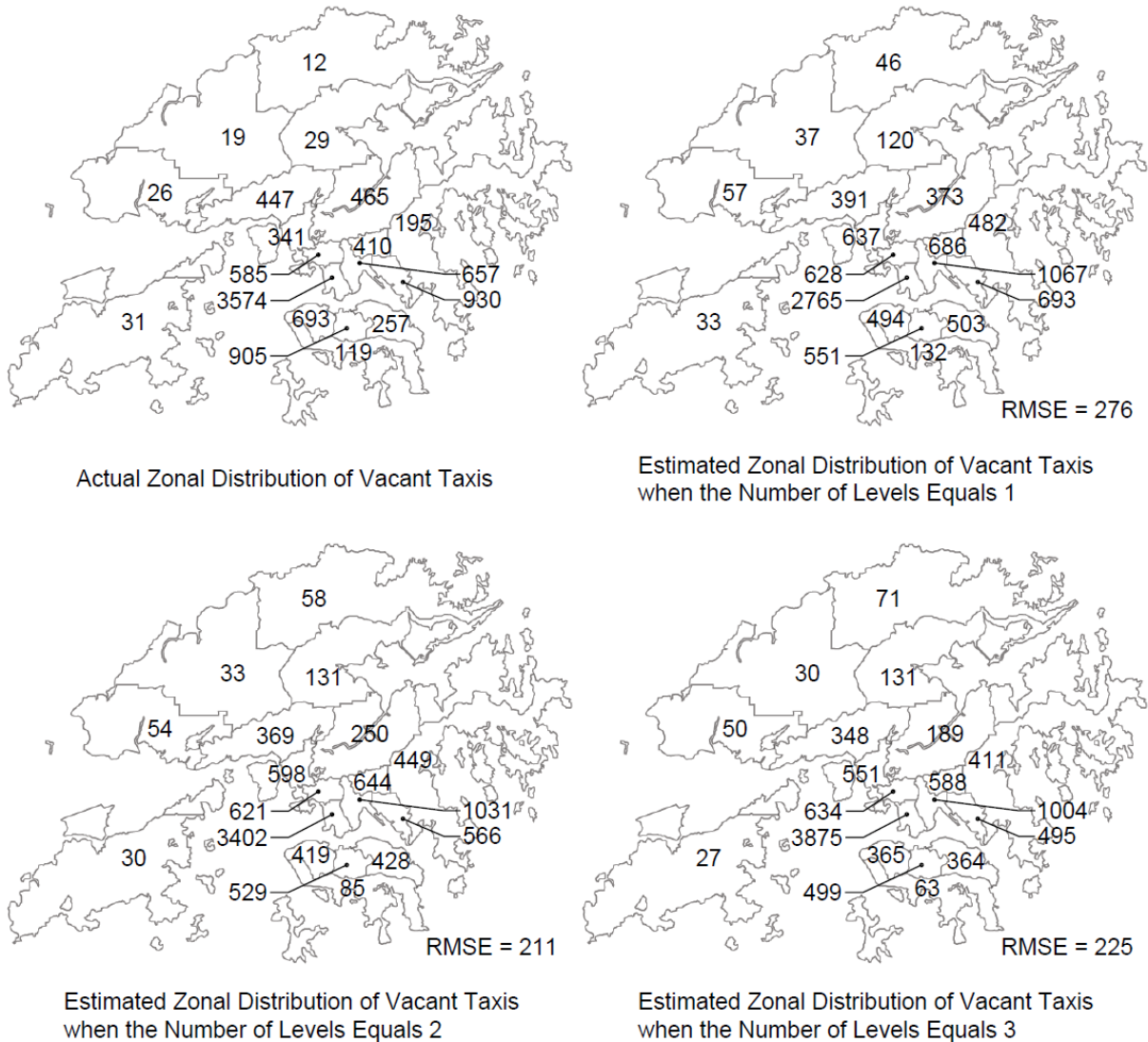
Figure 3. Comparison between actual and estimated zonal distributions of vacant taxis during the mid-night period under various numbers of search-decision levels

Unlike the two peak-period cases, the optimal number of levels for the mid-night period model is larger than one because the overall passenger demand is low during the mid-night period, leading to a higher proportion of drivers required to travel across multiple districts to meet a customer. That $M = 2$ for the mid-night period model also has several implications. First, using a simple MNL model (i.e., $l = 1$) will not obtain the best prediction for the mid-night period, as the RMSE value is not at its smallest. Second, compared with the MNL model, it is more appropriate to use our proposed model to determine the sequential decisions of vacant-taxi drivers conducting spatial searches for customers during this period. Our proposed model could also provide better information to policymakers on mid-night taxi operations. Third, it is better to illustrate the applications of the mid-night period model and discuss its properties in the remainder of this section, as more search levels are involved in that model compared with the peak-period models with $M = 1$.

Figure 3 also illustrates that the number of vacant taxis allocated to the Yau Tsim Mong District substantially increased at each level and reached 3,402 at two search levels. The outcomes indicate that this area was the most attractive to surrounding vacant-taxi drivers for conducting customer searches. The Yau Tsim Mong District offered more overnight activities

and served as a hub of public transport interchanges at night, resulting in a higher passenger demand compared with the neighboring districts. However, the Sha Tin District presented an opposing situation, as observed in the preceding spatial graph. The number of vacant taxis allocated to the Sha Tin District dropped continuously from 373 at each search level. Vacant-taxi drivers kept leaving the district and moving to areas with higher passenger demand, such as the Yau Tsim Mong District. The Tai Po District offered a comparatively small growth from 120 to 131 when the number of search levels increased from one to two.

Benefitting from the ESL model's structure, the customer-search paths of vacant-taxi drivers can be reproduced. Table 3 presents the top 10 origin districts and the top 10 search paths of the vacant-taxi drivers who terminated at the Yau Tsim Mong District during the mid-night period based on $M = 2$, and illustrates the model's performance.

Table 3. Top 10 origin districts and route choices reaching Yau Tsim Mong during the mid-night period

Rank	Origin District (District Index)*	Proportion	(Search Path)*	Proportion
1	Yau Tsim Mong (14)	41.6%	(14)	36.8%
2	Sham Shui Po (12)	16.2%	(12-14)	14.8%
3	Kowloon City (10)	14.8%	(10-14)	12.9%
4	Wan Chai (18)	8.1%	(18-14)	7.7%
5	Central and Western (15)	6.3%	(15-14)	5.1%
6	Sha Tin (05)	3.1%	(11-10-14)	3.0%
7	Kwun Tong (11)	3.0%	(13-10-14)	2.5%
8	Wong Tai Sin (13)	2.5%	(16-18-14)	1.9%
9	Eastern (16)	1.9%	(14-10-14)	1.9%
10	Kwai Tsing (02)	1.6%	(10-12-14)	1.9%

* The values in brackets represent the district indexes.

The predicted trip patterns on the left-hand side show that most of the trips to the Yau Tsim Mong District were generated from the district itself (meaning that those taxi drivers were intra-zone circulating within the previous trip's destination zone). The origin districts ranked 2 to 5 were indeed adjacent to the Yau Tsim Mong District. The predictions are considered reasonable, as vacant-taxi drivers are more likely to search nearby rather than travel long distances to search for customers. In total, the current and adjacent districts contributed to about 86.9% of the trips. The Wan Chai and Central and Western Districts were classified as adjacent districts, as they connected with the Yau Tsim Mong District by the Cross Harbor Tunnel and the Western Harbor Crossing, respectively. Nevertheless, the trip proportions were obviously lower than those of other districts on the Kowloon side, as comparatively fewer taxi drivers were willing to cross the tunnels and pay the toll charges to customers.

The top 10 predicted search paths toward the Yau Tsim Mong District, shown on the right-hand side of Table 3, provide a result that is very consistent with the popular origin districts shown on the left. The origin zones of the paths were very often the popular origin districts. In particular, if we only consider the origin zones of the paths on the right, we find that the top five rankings of the two lists are identical. For those ranked 6 to 10, the route choices tell us that the vacant-taxi drivers reached three districts and terminated at the Yau Tsim Mong District to meet customers. The most interesting search path is the No. 9 ranking Yau Tsim Mong – Kowloon City – Yau Tsim Mong (14-10-14). It reveals that about 1.9% of the drivers initially moved to the Kowloon City District and returned back to the Yau Tsim Mong District when circulating for customers.

5. CONCLUSION

This paper builds and tests an ESL model to address the limitations of the MNL model in capturing the sequential decisions made by vacant-taxi drivers in intermediate districts while heading to designated districts. The trip data from the GPSs of 460 urban taxis were extracted to calibrate the model and verify the factors underlying the drivers' search decisions. This paper attempts to apply the model and review its properties and performances. The findings reveal that the proposed ESL model is capable of predicting the search paths of vacant-taxi drivers.

Whereas the MNL model can only offer accurate estimations of vacant-taxi distribution, the ESL model can additionally provide the predicted search paths of vacant-taxi drivers. The ESL model is considered more informative for policymakers who aim to study search paths and the associated traffic congestion contributed by taxis in each district. The ESL model was not developed to replace the MNL model in predicting the strategic zonal choices of vacant-taxi drivers searching for customers, but rather as an alternative model. Which model should be selected depends on the scenario and application. Nevertheless, the ESL model should be adopted when the sequential-decision process is considered or found to be important and search-path information is required.

In this paper, we examine the zonal choices made by vacant-taxi drivers searching for customers during different periods, using aggregated average data extracted from the corresponding periods and assuming that the trip patterns within each study period are in a steady state. This steady state assumption is reasonable when the modeling period is long enough and the variations in passenger demand and customer-search patterns are small. When the modeling period is short and the variation in passenger demand is large or there is a significant change in customer-search patterns over time, it may be important to capture how taxis move as time advances. How to extend our model to capture and confirm the importance of the dynamic movement of taxis over time and space offers a future research direction.

ACKNOWLEDGEMENTS

REFERENCES

Fernández, J.E., De Cea, J., Briones, J. (2006) A diagrammatic analysis of the market for cruising taxis. *Transportation Research Part E*, 42, 498-526.

Government of the Hong Kong Special Administrative Region, Electrical and Mechanical Services Department (2009) *August adjustment in auto LPG ceiling prices*.

Government of the Hong Kong Special Administrative Region, Planning, Environment and Lands Bureau (1998) *A proposal to introduce LPG taxis: a consultation paper*.

Government of the Hong Kong Special Administrative Region, Transport and Housing Bureau (2008) *Average monthly operating cost for urban and Lantau taxis*.

Gudishala, R., Wilmot, C. (2012) A comparison of time-dependent sequential logit and nested logit for modeling hurricane evacuation demand. *Transportation Research Record: Journal of the Transportation Research Board*, 2312, 134-140.

Kahn, L.M., Morimune, K. (1979) Unions and employment stability: a sequential logit approach. *International Economic Review*, 20, 217-236.

Manski, C.F., Wright, J.D. (1976) Nature of equilibrium in the market for taxi services. *Transportation Research Record: Journal of the Transportation Research Board*, 619, 296-306.

McFadden, D. (1974) Conditional logit analysis of qualitative choice behavior. In: Zarembka, P. (ed.), *Frontiers in Econometrics*. New York: Academic Press, 105-142.

Nagakura, D., Kobayashi, M. (2009) Testing the sequential logit model against the nested logit model. *The Japanese Economic Review*, 60, 345-361.

Ophem, H., Schram, A. (1997) Sequential and multi-nominal logit: a nested model. *Empirical Economics*, 22, 131-152.

Schaller, B. (1998) Issues in fare policy: case of the New York taxi industry. *Transportation Research Record: Journal of the Transportation Research Board*, 1618, 139-142.

Sirisoma, R.M.N.T., Wong, S.C., Lam, W.H.K., Wang, D., Yang, H., Zhang, P. (2010) Empirical evidence for taxi customer-search model in Hong Kong. *Proceedings of the Institution of Civil Engineers, Transport*, 163, 203-210.

Wong, K.I., Wong, S.C., Yang, H. (2001) Modeling urban taxi services in congested road networks with elastic demand. *Transportation Research Part B*, 35, 819-842.

Wong, K.I., Wong, S.C., Bell, M.G.H., Yang, H. (2005) Modeling the bilateral micro-searching behavior for urban taxi services using the absorbing Markov chain approach. *Journal of Advanced Transportation*, 39, 81-104.

Wong, K.I., Wong, S.C., Yang, H., Tong, C.O. (2002) A sensitivity-based solution algorithm for the network model of urban taxi services. In: Taylor, M.A.P. (ed.), *Proceedings of the 15th International Symposium on Transportation and Traffic Theory*. Amsterdam: Elsevier Science, 23-42.

Wong, K.I., Wong, S.C., Yang, H., Tong, C.O. (2003) The effect of perceived profitability on the level of taxi service in remote areas. *Journal of the Eastern Asia Society for Transportation Studies*, 5, 79-94.

Wong, K.I., Wong, S.C., Yang, H., Wu, J.H. (2008) Modeling urban taxi services with multiple user classes and vehicle modes. *Transportation Research Part B*, 42, 985-1007.

Wong, R.C.P., Szeto, W.Y., Wong, S.C., Yang, H. (2013) Modeling multi-period customer-searching behavior of taxi drivers. *Journal of Advanced Transportation*, in press.

Wong, S.C., Yang, H. (1998) Network model of urban taxi services: improved algorithm. *Transportation Research Record: Journal of the Transportation Research Board*, 1623, 27-30.

Yang, H., Wong, S.C. (1998) A network model of urban taxi services. *Transportation Research Part B*, 32, 235-246.

Yang, H., Fund, C.S., Wong, K.I., Wong, S.C. (2010a) Nonlinear pricing of taxi services. *Transportation Research Part A*, 44, 337-348.

Yang, H., Leung, C.W.Y., Wong, S.C., Bell, M.G.H. (2010b) Equilibria of bilateral taxi-customer searching and meeting on networks. *Transportation Research Part B*, 44, 1067-1083.

Yang, H., Wong, K.I., Wong, S.C. (2001) Modeling urban taxi services in road networks: progress, problem and prospect. *Journal of Advanced Transportation*, 35, 237-258.

Yang, H., Wong, S.C., Wong, K.I. (2002) Demand-supply equilibrium of taxi services in a network under competition and regulation. *Transportation Research Part B*, 36, 799-819.

Yang, H., Ye, M., Tang, W.H.C., Wong, S.C. (2005) Regulating taxi services in the presence of congestion externality. *Transportation Research Part A*, 39, 17-40.