The HKU Scholars Hub   The University of Hong Kong   香港大學學術庫

Title	A time-dependent logit-based taxi customer-search model
Author(s)	Szeto, WY; Wong, RCP; Wong, SC; Yang, H
Citation	International Journal of Urban Sciences, 2013, v. 17 n. 2, p. 184-198
Issued Date	2013
URL	http://hdl.handle.net/10722/185768
Rights	This is an Accepted Manuscript of an article published by Taylor & Francis Group in International Journal of Urban Sciences on 19 Mar 2013, available online at: http://www.tandfonline.com/doi/abs/10.1080/12265934.2013.776292

A TIME-DEPENDENT LOGIT-BASED TAXI CUSTOMER-SEARCH MODEL

Wai Yuen SZETO (corresponding author)
Department of Civil Engineering, The University of Hong Kong, Hong Kong
E-mail: ceszeto@hku.hk
Phone: (852) 28578552
Fax: (852) 25595337

Ryan Cheuk Pong WONG
Department of Civil Engineering, The University of Hong Kong, Hong Kong
E-mail: ryancpw@hku.hk
Phone: (852) 28592662
Fax: (852) 25595337

Sze Chun WONG
Department of Civil Engineering, The University of Hong Kong, Hong Kong
E-mail: hhecwsc@hku.hk
Phone: (852) 28591964
Fax: (852) 25595337

Hai YANG
Department of Civil and Environmental Engineering,
The Hong Kong University of Science and Technology, Hong Kong
E-mail: cehyang@ust.hk
Phone: (852) 23587178
Fax: (852) 23581534

ABSTRACT

In this study, Global Positioning System (GPS) data from 460 urban taxis is used to develop a time-dependent logit model. The rate of return (ROR, also known as profit per unit time) is used as a factor underlying taxi drivers' searching behavior for customers. The data also reveal that the search behavior across districts as well as the decisions towards a particular district in customer-search is strongly related to the daily profile of passenger demand, and that when the overall passenger demand is high, vacant taxi drivers tend to circulate within or wait at the area where their preceding customers got off to find their next customer. The results also show that the ROR is a significant factor that affects the customer-searching strategies of vacant taxi drivers over a day, and is inversely related to the percentage of taxi idling time. More importantly, this paper illustrates that there is a change in searching behavior over time of day.

Keywords: Time-dependent; Urban taxi; Customer-search; Global Positioning System; Discrete choice model; Rate of return

1. INTRODUCTION

The taxi has become a popular mode in urban areas because of its convenience, speediness, door-to-door attribute, privacy, comfort, long-time operation, and lack of parking fees (Chang and Chu, 2009). It can also be used as a complement mode to existing public transit services (Austin and Zegras, 2012; King et al., 2012). However, there are at least two major problems associated with taxi service provisions. First, in cities where population density is high, taxis are often allowed to circulate to search for customers but this worsens the congestion and air pollution problems. Second, there is always a mismatch between demand and supply of taxis over time and space. How to improve the customer satisfaction and system performance is the key question. Hence, in practice, advanced technology and taxi policies are introduced to address these problems.

Various studies related to using advanced technology to improve the system performance have been conducted. Lee et al. (2004) proposed a taxi-dispatch system based on real-time traffic conditions to increase customer satisfaction. Lee et al. (2005) presented a taxipooling dispatching system to allow taxis to be a feeder service of mass transit system. Tao (2007) presented algorithms for assigning passengers to taxis so that the distance/time traveled by both passengers and taxis can be minimized. Lee and Cheng (2008) investigated a new taxi business model that utilized smart phones to improve operational effectiveness. Seow et al. (2007) proposed a collaborative taxi-dispatch system to increase customer satisfaction, by concurrently dispatching multiple taxis to the same number of customers in the same geographical region, and vis-à-vis human driver satisfaction. Chang et al. (2010) proposed mining historical data to predict demand distributions with respect to contexts of time, weather, and taxi location to help the management of taxi fleets. Chang et al. (2012) proposed a mathematical model for optimizing the fleet size of global positioning system (GPS) taxis. Conway et al. (2012) presented a case study to discuss the challenges in managing centralized taxi dispatching at high volume airports. da Costa and de Neufville (2012) provided a practical procedure for designing efficient taxi pick-up operations at airports. All these studies aim to increase the customer satisfaction or improve overall system efficiency from the perspective of taxi operators.

To investigate the consequences of taxi policies such as price controls or entry restrictions policies, various economic studies have been performed (Fernández et al., 2006; Manski and Wright, 1976; Schaller, 1998; Yang et al., 2005). However, these studies have been developed based on the idealized market of 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 taxi movements and customer origin-destination (OD) demand patterns on a given road network. In this model, the customer-searching behavior of taxi drivers was based on the assumption that each vacant taxi driver tries to minimize his/her expected searching time to find a customer. Wong and Yang (1998) reformulated the model to develop a more efficient and convergent iterative balancing algorithm. Wong et al. (2001) extended the work of Yang and Wong (1998) to incorporate congestion effects and customer demand elasticity, and developed a new solution algorithm based on a quasi-Newtonian approach to solve the proposed taxi network model. Wong et al. (2002) further developed a sensitivity-based solution algorithm for the taxi network model of Wong et al. (2001) and showed this algorithm 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 monopoly 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 customer hierarchical modal choices. Kim et al. (2005) proposed a taxi network model that considers both demand and supply uncertainties. Hu et al. (2012) developed a probabilistic dynamic programming model to depict the route choice behavior of vacant taxi drivers.

Nevertheless, the aforementioned taxi network models still rely on the assumption that vacant taxi drivers seek to minimize their expected searching time for their next customer. This is realistic when modeling urban areas, where the trip length of customers may not significantly vary from zone to zone and taxi drivers may not have much information about the profitability of going to particular zones to search for customers. However, this searching behavior may not reflect the real situation in more remote areas with considerable potential demand (e.g., an airport or satellite town) for taxi services to go to the urban area of a city. Although these zones are distant from the urban area, large numbers of vacant taxis still go to these areas to wait for customers. The drivers do this because their profit from picking up customers in these remote areas is considerably higher than that earned in the urban area. The effect of this profitability consideration on customer-searching decisions of taxi drivers was explicitly considered by Wong et al. (2003). Yang et al. (2010) further extended this concept to 'profit per unit time' by incorporating the operational cost and time consumption of taxi trips into profit.

All of these taxi network models ignore the taxi searching 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 searching behavior of vacant taxi drivers and customers using the absorbing Markov chain approach. Yang et al. (2010) developed another mathematical model to characterize the same bilateral searching behavior on road networks by incorporating a macroscopic searching-and-meeting model into the network model of Wong et al. (2003).

Not only theoretical works but also empirical studies on the taxi searching behavior of customers have been conducted. For example, Hwang et al. (2006) presented a case study in Taiwan that modeled consumer preference for a GPS-based taxi dispatching service. Sheu et al. (2010) studied taxi hailing in vehicular ad-hoc networks. Tao and Wu (2008) conducted a field trial of dynamic taxi-sharing services to examine the factors that affect the acceptance of customers with the services. Liu et al. (2010) classified cabdrivers into two classes based on their daily income, and compared different cabdrivers' behaviors and operation patterns.

In most existing taxi network models (e.g., Wong and Yang, 1998; Yang and Wong, 1998; Wong et al., 2001, 2003, 2005; Yang et al., 2001, 2002, 2010), a logit-based search model were often used to describe the customer-searching behavior of vacant taxi drivers based on the hypothesis of either searching time minimization or perceived profit maximization. However, there has been no empirical evidence to validate the hypothesized customer-searching behavior of taxi drivers until Sirisoma et al. (2010) carried out a stated preference survey of 400 taxi drivers in Hong Kong to demonstrate the validity of logit-based search models developed under the assumption of searching time minimization. Chiu and Lin (2011) also conducted an onboard interview in Taipei to find out the customer-searching behavior of

vacant taxi drivers. However, these two studies did not consider the temporal variations of passenger demand and the expected profit, which may affect the customer-searching strategies of vacant taxi drivers.

In fact, the temporal variation of demand has been considered in other problems such as traffic assignment (Gudishala and Wilmot, 2012; Lam et al., 1999; Lam and Yin, 2001), and transit assignment (Hamdouch et al., 2011; Schmöcker et al., 2008) but has rarely been considered in taxi models. Moreover, whether taxi passenger demand varies over time of day has not been verified and how this variation affects the searching behavior of taxi drivers has not been examined yet. As such, it is essential to obtain survey data to verify that the demand varies over time of day and to analyze the customer-searching behavior of vacant taxi drivers in different times of day. Since the profitability or the ROR is highly related to demand, and is a concern of taxi drivers, it is also important to examine how the ROR varies over time of day.

For these purposes, this study develops time-dependent logit-based search models using GPS data obtained from 460 Hong Kong urban taxis operating around the clock. The concept of 'profit per unit time' (Yang et al., 2010) is adopted in the proposed logit-based choice model. This profit per unit time is shown to be inversely related to the percentage of idling time and to be a significant factor that affects the customer-searching strategies of vacant taxi drivers over a day. The relationship between passenger demand and cross-district searching strategies over time is also illustrated and discussed. The study further investigates whether drivers' searching strategies change across hourly periods by developing combined logit models for the two adjacent periods and carrying out Watson and Westin pooling test (Watson and Westin, 1975).

The remainder of this paper proceeds as follows. The second section describes the data collection method, the taxi travel fare, the operational cost, the zonal choice, and the study area. The third section presents the methodology, the model formulation, and the pooling test. The fourth section discusses the results. Finally, the last section concludes the paper and suggests some future research directions.

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 are New Territories taxis and 50 are Lantau taxis. In this study, the survey data were collected from GPS devices previously installed in 460 urban taxis, which provided field survey data on their daily activities. Using satellite communication, the database recorded the location in terms of longitude and latitude, spot traveling speed, and occupation status of the taxis at every 30 seconds. These recorded activities 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, or traveling to get fuel. However, we have insufficient information from the dataset to distinguish them. Although there was no information on the fare collected, the location and speed information were used to deduce the vehicle trajectory and waiting time and hence the fare collected. Moreover, because HK$5 is charged for telephone bookings and it is not difficult for a passenger to catch a vacant taxi along the urban roadside in Hong Kong, telephone bookings are uncommon. We consider

4

such bookings as having a negligible impact 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 it adequately represents the travel behavior of the region's drivers. The GPS survey data were collected in a full day on 18 August 2009 (Tuesday). It is a typical weekday without abnormal traffic conditions. To facilitate the evaluation of the changes in the customer-searching behavior of vacant taxi drivers in different hourly periods in a day, survey data were extracted and categorized into 24 one-hour period accordingly. To develop the discrete choice model, the average values of travel distance and time, revenue, and operational cost for each OD pair were determined.

We suspected that some of the extracted occupied or vacant trips were actually non-existent, such as trips of exceptionally long or short travel distance and time (i.e., occupied trips traveling less than 1 minute, or vacant trips traveling more than 30 km). The recordings of these extracted trips could be due to malfunctions of GPS devices, poor connectivity to satellites in the urban areas surrounded by high-rise buildings, or human error by the taxi drivers who operated the devices Therefore, a data screening process was carried out to eliminate those trips. However, some non-searching unoccupied trips, such as trips or roadside stops taken for a short break or to trips to a gas station, could not be eliminated. Nevertheless, based on information obtained from taxi drivers, the number of these trips is small. Hence, we consider these trips to have a negligible impact on the accuracy of the customer-search models. We also observed that the number of records for certain OD pairs did not achieve a reasonable frequency (i.e., the number of occurrences less than 3 in each hourly period). These pairs were considered as insufficient to give a representative average value, and were thus excluded. After data screening, more than 36,000 taxi trips were extracted in the concerned day. The numbers of occupied and vacant taxi trips were 18,053 and 18,012 respectively, and were approximately equal in each hourly period. It suggests that the number of occupied and vacant taxis were about the same within one hour. Each of these trips was considered as an individual data for model development.

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. For all the urban taxis as of August 2009, HK$18 was the initial charge for the first 2 km. Every subsequent 200 m, or any part thereof, or every period of 1 minute waiting time or part thereof, cost an additional HK$1.5. This unit charge decreased to HK$1 after the chargeable amount had reached HK$70.5. The passengers also had to pay for any toll charges incurred on the journey, and surcharges of return tolls involving use of three harbor crossing tunnels and 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, without surcharges 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, and the number of operating days of a driver was about 25 per month. With a normal shift of 12 hours of urban taxis, the rental cost was approximately HK$0.50/min of operation.

The fuel cost is directly related to the cost of Liquefied Petroleum Gas (LPG) since all taxis in Hong Kong were running on LPG. 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 gives a fuel cost of HK$0.46/km.

The GPS data offered us the coordinates and the associated times of the concerned taxis at each 30 seconds. These data were used to estimate the traveled distance (by summing up all 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 can be used to deduce the vehicle trajectories and determine which toll roads have been used. To determine the waiting time, we simply assumed that the occupied taxi with spot speed less than 5 km/h recorded at each 30 seconds has been waiting for 30 seconds. This implies that the waiting time was rounded up to the nearest 30-second. With the extracted traveling distance, the estimated waiting time, the toll roads used, the unit costs for traveling and waiting, the toll charges, and the nonlinear fare structure, the associated taxi fare and profit could be determined.

2.3   Zonal Choice and Study Area

In this study, we analyze the zonal decisions of vacant taxi drivers in finding customers to investigate their changes in customer-searching strategies over time. To facilitate this purpose, Hong Kong territory is divided into 18 zones according to administrative districts, which are set based on population size but each district retains certain degree of homogeneity of land use as shown in Figure 1.

[Please insert Figure 1 here]

3   METHOD

3.1   Definition of the Variable

Taxi driver's customer-searching involves a decision-making process that incorporates real-time traffic conditions and the driver's past experience. In this study, we consider that an urban taxi driver has two categories of search options in searching for potential customers:

Search option (1):   Intra-zonal circulating in the destination zone i of the previous trip (circulating in the current district), and

Search options (2) – (18):   Cross-zonal traveling towards another zone j and intra-zonal circulating within that zone (which can be any zone other than the destination zone i of the previous trip)

These 18 search options are assumed to be independent, and predetermined to be affected by the ROR.

Rate of Return (ROR): The ROR, which is also known as 'profit per unit time' (Yang et al., 2010), includes the cost of cross-zonal travel, the cost of intra-zonal circulating, expected profit, the cross-zonal travel and intra-zonal circulating time of a vacant taxi, and the expected travel time of an occupied trip.

The cost of intra-zonal circulating and cost of cross-zonal travel comprise the distance traveled and time required for both cross-zonal travel and intra-zonal circulating, fuel cost per unit of distance (C^{f}) and rental cost per minute (C^{r}), and the associated toll charges. Mathematically, the cost of intra-zonal circulating within designated zone j in search of the next customer, C_j^{c}, can be estimated by:

$$C_j^{\mathrm{c}} = C^{\mathrm{f}} D_j^{\mathrm{c}} + C^{\mathrm{r}} T_j^{\mathrm{c}}, \tag{1}$$

where D_j^{c} and T_j^{c} are the intra-zonal circulating distance and time in customer-search at zone j respectively.

Similarly, the cost of cross-zonal traveling from zone i to zone j to search for the next customer, C_{ij}^{t}, can be expressed as:

$$C_{ij}^{\mathrm{t}} = C^{\mathrm{f}} D_{ij}^{\mathrm{t}} + C^{\mathrm{r}} T_{ij}^{\mathrm{t}} + \tau_{ij}, \tag{2}$$

where D_{ij}^{t}, T_{ij}^{t} and τ_{ij} are the cross-zonal travel distance and time, and the associated toll charge respectively, in customer-search from zone i to zone j. When the destination zone of the previous trip is the same as the designated zone (i.e., $i = j$) or when vacant trips do not involve cross-district travel, the travel time and distance are set to zero, and circulating time and distance are used to capture the distance and time factors, respectively. The miscellaneous costs, such as the insurance cost and the vehicle maintenance cost, were not shown in the equation, since these costs were normally borne by the taxi owners and had probably been incorporated into the rental cost already.

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.

$$\overline{P}_j = \overline{F}_j - C^{\mathrm{f}} \overline{D}_j^{\mathrm{t}} - C^{\mathrm{r}} \overline{T}_j^{\mathrm{t}} - \overline{\tau}_j, \tag{3}$$

where $\overline{D}_j^{\mathrm{t}}$, $\overline{T}_j^{\mathrm{t}}$ and $\overline{\tau}_j$ are the average travel distance and time, and the associated toll charge, respectively, of all occupied trips starting from zone j. $\overline{P}_j$ is the expected profit from an occupied trip starting from zone j. $\overline{F}_j$ is the average taxi fare of all occupied trips starting from zone j, according to the non-linear taxi fare structure mentioned in Section 2.2.

The ROR is based on a searching cycle that consists of the vacant taxi trip from the destination zone of the previous occupied trip to the designated zone and the subsequent occupied trip starting from the designated zone. Given that the destinations of next customer could not be assured by the vacant taxi driver while making his/her instant decisions towards a zone in customer-searching. Therefore, in this study, we assumed that vacant taxi drivers made their customer-searching decisions for the following trip only, but without considering the potential subsequent trips.

The ROR is equal to the expected profit obtained from the searching cycle over the expected time taken to obtain that profit. The ROR describes the effect of perceived profit in the decision-making process of the search mechanism, which also partially depicts the effects of perceived demand, the separation between customers' drop-off and pick-up locations, and the level of circulation in the destination zone based on profit considerations. A higher ROR for a particular OD pair defined by the destination zone of the previous occupied trip to the designated zone results in a higher probability of a vacant taxi driver going from the destination zone to the designated zone. For a given destination zone of the previous trip, the designated zone determines the value of ROR of the trip. Hence, it is expected that the ROR is a factor that affects the designated zone choice of vacant taxi drivers.

The (average) ROR (R_{ij}) for a vacant taxi driver traveling from zone i to zone j to search for the next customer can be expressed as follows:

$$R_{ij} = \frac{-C_{ij}^{t} - C_{j}^{c} + \overline{P}_{j}}{T_{ij}^{t} + T_{j}^{c} + \overline{T}_{j}^{t}} \tag{4}$$

3.2 Time-dependent Logit-based Search Model

A time-dependent logit-based search model is developed to depict the customer-searching behavior of taxi drivers in 24 hourly periods in the day. The multi-nominal logit model takes the following form (McFadden, 1974), which is commonly used in discrete choice modeling to predict the probability of selection between alternatives:

$$P_{qt}(j|i) = \frac{\exp\left(\beta_t^{R} R_{ijt}\right)}{\sum_m \exp\left(\beta_t^{R} R_{imt}\right)}, \tag{5}$$

where $P_{qt}(j|i)$ is the probability that an individual vacant taxi driver starts trip q in zone i in time period t and selects zone j to search for the next customer over all possible destination zones. R_{ijt} and R_{imt} are both the rates of return for a vacant taxi driver starting in zone i in time period t. The former considers the driver searching at destination zone j and the latter considers the driver searching at possible destination zone m. Their meanings are similar to the rate of return in Equation (4) but additionally consider the time dimension. β_t^{R} is the coefficient associated with the ROR for time period t. It can be calibrated by using maximum likelihood estimation with the observed proportion of vacant taxi drivers' zonal selections as an input. According to the above model, a higher ROR for a particular OD pair results in a higher probability of a vacant taxi driver going from the origin to the destination.

It is important to clarify that the inputted attribute indeed is the perceived ROR to a taxi driver before he/she actually makes his/her decision towards a particular district in customer-search. It is set to be the average value calculated from the trips in our samples, which can be considered as an unbiased estimate of the long-term average for an individual driver in his/her mindset based on his/her past experience. According to this assumption, every taxi driver has the same perception to the above attribute. Therefore, for simplicity, the subscript q is omitted in the ROR.

3.3 Watson and Westin Pooling Test

In developing the individual logit models in each hourly period, it was assumed that the customer-searching strategies of vacant taxi drivers vary in different hourly periods in a day according to the ROR. This study combines individual logit models in adjacent time periods to verify this assumption. The Watson and Westin pooling test (Watson and Westin, 1975) was then applied. The test is based on the log-likelihood ratio (LR), which is calculated as:

$$LR = -2(L_R - L_U),\qquad(6)$$

where L_R is the log-likelihood for a combined model of two adjacent hourly time periods and L_U is the sum of the log-likelihoods of the two individual models developed for the two adjacent hourly time periods. The null hypothesis that there is no intervention in period segmentation is rejected, as the test statistic exceeds the threshold value that is specified for the chi-squared distribution at the chosen level of significance. The degree of freedom is calculated as the difference between the number of variables of the combined models and the sum of the number of individual models.

4 RESULTS AND DISCUSSION

4.1 Trip Characteristics

[Please insert Figure 2 here]

Figure 2 provides the daily profiles of the average customer-searching time of all vacant taxis and the mean profit of all occupied taxi trips respectively. It is observed that the customer-searching time between 01:00 and 08:00 was obviously high, which was about a double of those in day-time. It reflects the variation in passenger demand over time. It was more difficult for vacant taxi drivers to successfully pick up a customer after mid-night because of low passenger demand. Reversely, when the passenger demand was high, vacant taxi drivers were capable to meet a customer within 10 minutes of searching. On the other hand, the mean profit earned in each occupied taxi trip is observed with less fluctuation comparatively and ranging from HK$30 to HK$45 in different hourly periods. It is noticed that the day-time profits were slightly lower, which explains that more passengers took short trips during these periods.

Instead of incorporating either the customer-searching time or the profit individually as the factor that influences the customer-searching strategies of vacant taxi drivers, it is more reasonable to combine both the concepts of time minimization and profit maximization into an attribute of the ROR. A model focuses on time minimization or profit maximization in customer-search may wrongly interpret vacant taxi drivers over-concentrating at city centers with relatively shorter searching time or at remote areas (e.g., an airport or satellite town) with relatively higher profit. These approaches are considered as over-simplifying the complicated decision-making of vacant taxi drivers in customer-search.

Figure 3 shows the mean ROR for each OD pair with an adequate sample size in the concerned study time periods. It is observed that the mean value steadily maintained in the range between HK$0.6 and HK$1.0 per minute at most of the time in a day, except the obvious jump at around 09:00 to HK$1.5 per minute. It can be explained by the variation of taxi customer OD demand patterns due to different trip purposes. For example, a lot of passengers traveled by taxis at the morning peak for work or school activities, whereas a few

passengers traveled by taxis during the off-peak for non-routine activities. Although vacant taxi drivers had to spend more time in customer-searching after mid-night on average, more customers in these hours took longer occupied journeys to their destinations. Hence, the resulting ROR is not low. Passenger demand at day-time was higher as reflected by the mean searching time, but customers traveled in short trips. Therefore, the ROR was similar to that after mid-night.

Figure 3 also illustrates the changes in the percentage of taxi idling time over a day, which was calculated as the percentage of time consumed for customer-searching. This percentage is shown to be inversely related to the mean ROR. A small percentage of taxi idling time results in a high ROR in an hourly period because a low percentage of taxi idling time implies both a high percentage of time on-hire and short searching time, which both contribute to a high ROR. In particular, the percentage of taxi idling time in the morning peak between 08:00 and 09:00 was extremely small, which leads to the obvious jump of the ROR during that period and the highest ROR of HK$1.5/min.

[Please insert Figure 3 here]

To sum up, the distributions of the survey data in different hourly periods shows that the taxi market changed significantly across the day.

4.2   Passenger demand and Customer-searching Strategies

Figure 4 shows the changes in behaviors of vacant taxi drivers to travel across districts in customer-search over time of day in response to the variations of hourly (observed) passenger demand. The observed passenger demand was obtained from the number of occupied trips started in the particular hourly period. Although it does not exactly equal to the real demand with consideration of unserved passengers, it roughly offers us the trend of variation in passenger demand over a day. The cross-district searching phenomenon was easily observed after mid-night until 09:00. It is because the hourly passenger demand was high compared with the number of vacant urban taxis in day-time (which is reflected by the average hourly passenger demand of 4.0 per vacant urban taxi). Hence, most of the vacant taxi drivers could easily find their customers within the district where the last customer got off. They could spend 10 minutes on average to find their customers, and consequently they had a low motivation and seldom needed to cross districts to search for customers. However, after mid-night, the passenger demand was not high (i.e., reflected by the average hourly passenger demand of 1.7 per vacant taxi) and hence vacant taxi drivers needed to spend more time and travel longer distance on average to search for customers across districts.

For a broader view of the whole day, about 28% of the vacant trips traveled across districts in customer-searching and only 5% of the vacant trips involved the usage of toll roads. It shows that the majority of vacant taxi drivers refused traveling for a long trip or using toll roads in customer searching to lower the associated operational cost.

[Please insert Figure 4 here]

Figure 4 also demonstrates that during the daytime period from 07:00 to 23:00, the percentage of vacant urban taxis picking up customers by cross-district searching was only around 20-30% and the hourly passenger demand per vacant taxi was around 4.0. However,

the opposite situation was observed at periods between 01:00 and 06:00. The hourly passenger demand during this period dropped markedly to 1.0-2.5 passenger trips per vacant urban taxi, but cross-district customer-searching climbed to 55%. In a matter of fact, the correlation coefficient between cross-district vacant trips and overall hourly passenger demand per vacant urban taxi was 0.94, implying that the drivers' cross-district customer-searching strategies are strongly correlated with hourly demand, in which higher passenger demand leads to less intention to a vacant urban taxi driver for cross-district customer-searching, or vice versa.

To summarize, the customer-searching strategies of vacant taxi drivers vary with time, and depend on the changing passenger demand over time. Due to the high demand for taxi services during peak hours, taxi drivers tend to search customers locally. Reversely, after mid-night, most vacant taxi drivers search for customers across districts, particularly in high demand districts, to have a better opportunity of finding the next customer.

4.3 Calibration Results

To determine the coefficient of the ROR in each individual hourly period model, a typical modeling software package NLOGIT was adopted, which uses the maximum likelihood estimation. The coefficients are shown in Figure 5. It is found that the coefficients are not equal to zero at all hourly periods at the 1% significance level. Moreover, the coefficients are positive, meaning that the OD pair with a higher ROR has more vacant taxi trips, which is reasonable. The result confirms that the ROR concept proposed by Yang et al. (2010) can be solely used in the utility function in the simple logit model to depict the customer-searching behavior of vacant taxi drivers.

[Please insert Figure 5 here]

The ROR ranges from 0.38 to 4.10, in which the magnitude of the coefficients at the periods between mid-night and 07:00 are obviously smaller than those in day-time. Higher calibrated coefficients reflect that the zonal choices of vacant taxi drivers in finding customers were more dependent to the ROR. In day-time, more vacant taxi drivers stayed at the district where preceding passengers dropped-off to search for the next customers because local search most likely offered a shorter searching time and hence a higher ROR. However, during the periods after mid-night, it was difficult for vacant taxi drivers to reach a customer by local search, and therefore the drivers preferred traveling towards a district with a higher passenger demand, even the vacant traveling time was slightly longer and sometimes the trip led to a negative ROR. The revenue earned in the designated zone cannot compensate for the preceding operational cost of making a long vacant trip. It may be due to the habitual zone choices of individual taxi drivers, especially during the midnight when the demand is highly concentrated on certain zones. These issues are left for future studies.

A deeper joint analysis of Figures 4 and 5 reveals that the patterns of the calibrated coefficients of the ROR and the hourly passenger demand per vacant taxi are similar. The correlation coefficient between them was 0.84, implying that the changes in customer-searching strategies are strongly correlated with the passenger demand as well.

4.4 Time-dependent Customer-Search Behavior

11

A series of LR tests was conducted between models for adjacent hourly periods. Given that the degree of freedom is 1, the chi-square critical value at the 1% level of significance is 6.64, which is significantly lower than each of the log-likelihood ratios (i.e., the log-likelihoods of individual model in the time periods of 01:00 to 02:00 and 02:00 to 03:00 are -405.49 and -230.28 respectively, and the log-likelihood of combined model of these periods is -761.18.) Based on Equation (6), LR could be determined and equals 160.82, and is much larger than the critical value of 6.64. Therefore, the null hypothesis that there is no intervention in period segmentation is rejected. We conclude that different individual models cannot be pooled. The implication is that the drivers' searching strategies do vary across different hourly time periods, and hence we need one model for each period.

5   CONCLUSION

This study develops a time-dependent logit-based taxi customer-search model using GPS data obtained from 460 urban taxis in Hong Kong, to investigate the changes in zonal decisions of vacant taxi drivers in customer-search in different hourly periods in a day. The ROR is found as a significant factor that affects taxi drivers' customer-searching strategies at the 1% significance level throughout the day and hence this paper validates the concept of 'profit per unit time' proposed by Yang et al. (2010) or the ROR in relation to vacant taxi customer-searching. The ROR is also shown to be inversely related to the percentage of taxi idling time. The pooling test results illustrate that the drivers' customer-searching behavior varies with the time of day. Taxi drivers have similar, but not identical, searching strategies in terms of zonal choices, in that they tend to circulate within or wait at the area of the preceding destination to find their next customer when the overall passenger demand was high. Therefore, assuming that the trip characteristics of vacant taxi drivers searching for customers are homogenous throughout the day is considered to be oversimplified. The data also reveal that the search behavior across districts as well as the decisions towards a particular district in customer-search is strongly related to the daily profile of passenger demand. All these findings are useful to help evaluate the effectiveness of time-dependent taxi regulation policies, such as travel fare premium at peak hours and night.

During the periods after midnight, we observed that vacant taxi drivers preferred traveling towards a district with a higher passenger demand, even the vacant traveling time was slightly longer and sometimes the trip led to a negative ROR. The revenue earned in the designated zone cannot compensate for the preceding operational cost of making a long vacant trip. A possible reason is that some taxi drivers prefer circulating within familiar areas. This habitual zone effect may influence the drivers' customer-searching behavior. A comprehensive analysis of travel patterns to investigate personal preferences of taxi drivers is hence suggested.

In our study, the cost of cross-zonal traveling is expressed as a linear combination of the fuel cost, rental cost, and the toll charges. However, these costs could be combined based on different functional forms (not necessarily linear) and with different weights to reflect the perception on direct and indirect costs. For example, the toll charge is a direct cost while the unit taxi rental cost is a perceived value. While searching for customers, vacant taxi drivers may be more reluctant to pay any direct cost (e.g. toll) rather than a perceived cost. These perception and functional form issues were left for future research.

In this study, we assumed that vacant taxi drivers made their customer-searching decisions for the following trip only, but without considering the potential consecutive trips afterwards. In reality, they may consider also the potential trip or the potential destinations after the following trip to ensure that they can pick up customers. This can be examined in the future research.

ACKNOWLEDGEMENTS

REFERENCES

Austin, D. and Zegras, P.C. (2012) Taxicabs as public transportation in Boston, Massachusetts. Transportation Research Record, No. 2277, pp. 65-74. .

Chang, J.S.K. and Chu, C.H. (2009) Taxi vacancy rate, fare, and subsidy with maximum social willingness-to-pay under log-linear demand function. Transportation Research Record, No. 2111, pp. 90-99.

Chang, H.W., Tai, Y.C. and Hsu, Y.J. (2010) Context-aware taxi demand hotspots prediction. International Journal of Business Intelligence and Data Mining, Vol. 5, pp. 3-18.

Chang, J.S.K., Wu, C.H. and Lin, C.H. (2012) Optimization of fleet size for GPS taxi. Proceedings of the 91st Transportation Research Board Annual Meeting.

Chiu, Y.C. and Lin, D.Y. (2011) Onboard interview of taxi drivers travel decisions. Taipei.

Conway, A., Kamga, C., Yazici, A. and Singhal, A. (2012) Challenges in managing centralized taxi dispatching at high-volume airports: Case study of John F. Kennedy International Airport. Transportation Research Record, No. 2300, pp. 83-90.

da Costa, D.C.T. and de Neufville, R. (2012) Designing efficient taxi pickup operations at airports. Transportation Research Record, No. 2300, pp. 91-99.

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

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

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, Electrical and Mechanical Services Department. (2009) August adjustment in auto LPG ceiling prices.

Gudishala, R. and Wilmot, C. (2012) A comparison of time-dependent sequential logit and nested logit for modeling hurricane evacuation demand. Proceedings of the 91st Transportation Research Board Annual Meeting.

Hamdouch, Y., Ho, H.W., Sumalee, A. and Wang, G. (2011) Schedule-based transit assignment model with seat availability. Transportation Research Part B, Vol. 45, pp. 1805-1830.

Hu, X., Gao, S., Chiu, Y.C. and Lin, D.Y. (2012) Modeling routing behavior for vacant taxicabs in urban traffic networks. Transportation Research Record, No. 2284, pp. 81-88.

Hwang, K.P., Wu, H.C. and Jian, R.J. (2006) Modeling consumer preference for Global Positioning System-based taxi dispatching service: Case study of Taichung City, Taiwan. Transportation Research Record, No. 1971, pp. 99-106.

Kim, H., Oh, J.S. and Jayakrishnan, R. (2005) Effect of taxi information system on efficiency and quality of taxi services. Transportation Research Record, No. 1903, pp. 96-104.

King, D.A., Peters, J.R. and Daus, M.W. (2012) Taxicabs for improved urban mobility: Are we missing an opportunity? Proceedings of the 91st Transportation Research Board Annual Meeting.

Lam, W.H.K., Zhang, Y.P. and Yin, Y.F. (1999) Time-dependent model for departure time and route choices in networks with queues. Transportation Research Record, No. 1667, pp. 33-41.

Lam, W.H.K. and Yin, Y.F. (2001) An activity-based time-dependent traffic assignment model. Transportation Research Part B, Vol. 35, pp. 549-574.

Lee, D.H., Wang, H., Cheu, R.L. and Teo, S.H. (2004) Taxi dispatch system based on current demands and real-time traffic conditions. Transportation Research Record, No. 1882, pp. 193-200.

Lee, K.T., Lin, D.J. and Wu, P.J. (2005) Planning and design of a taxipooling dispatching system. Transportation Research Record, No. 1903, pp. 86-95.

Lee, W.C. and Cheng, B.W. (2008) Incorporating e-technology to advantage in a greener taxi industry and its impact on driving performance and safety. Transportation Planning and Technology, Vol. 31, pp. 569-588.

Liu, L., Andris, C. and Ratti, C. (2010) Uncovering cabdrivers' behavior patterns from their digital traces. Computer, Environment and Urban Systems, Vol. 34, pp. 541-548.

Manski, C.F. and Wright, J.D. (1976) Nature of equilibrium in the market for taxi services. Transportation Research Record, No. 619, pp. 296-306.

McFadden, D. (1974) Conditional logit analysis of qualitative choice behavior. Frontiers in econometrics, pp. 105-142.

Schaller, B. (1998) Issues in fare policy: Case of the New York taxi industry. Transportation Research Record, No. 1618, pp. 139-142.

Schmöcker, J.D., Bell, M.G.H. and Kurauchi, F. (2008) A quasi-dynamic capacity constrained frequency-based transit assignment model. Transportation Research Part B, Vol. 42, pp. 925-945.

Seow, K.T., Dang, N.H. and Lee, D.H. (2007) Towards an automated multiagent taxi-dispatch system. Automation Science and Engineering, IEEE International Conference, pp. 1045-1050.

Sheu, J.P., Chang, G.Y. and Chen, C.H. (2010) A distributed taxi hailing protocol in vehicular ad-hoc networks. Vehicular Technology Conference, pp. 1-5.

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

Tao, C.C. (2007) Dynamic taxi-sharing service using intelligent transportation system technologies. International Conference on Wireless Communications, Networking and Mobile Computing, pp. 3209-3212.

Tao, C. and Wu, C. (2008) Behavioral responses to dynamic ridesharing services - The case of taxi-sharing project in Taipei. Service Operations and Logistics, and Informatics, IEEE International Conference, pp. 1576-1581.

Watson, P.L. and Westin, R.B. (1975) Transferability of disaggregate mode choice models. Regional Science and Urban Economics, Vol. 5, pp. 227-249.

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

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

Wong, K.I., Wong, S.C., Yang, H. and Tong, C.O. (2002) A sensitivity-based solution algorithm for the network model of urban taxi services. Proceedings of the 15th International Symposium on Transportation and Traffic Theory, pp. 23-42.

Wong, K.I., Wong, S.C., Yang, H. and 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, Vol. 5, pp. 79-94.

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

Wong, S.C. and Yang, H. (1998) Network model of urban taxi services: improved algorithm. Transportation Research Record, No. 1623, pp. 27-30.

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

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

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

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

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

LIST OF FIGURES

The New Territories
1. Islands
2. Kwai Tsing
3. North
4. Sai Kung
5. Sha Tin
6. Tai Po
7. Tsuen Wan
8. Tuen Mun
9. Yuen Long

Kowloon
10. Kowloon City
11. Kwun Tong
12. Sham Shui Po
13. Wong Tai Sin
14. Yau Tsim Mong

Hong Kong Island
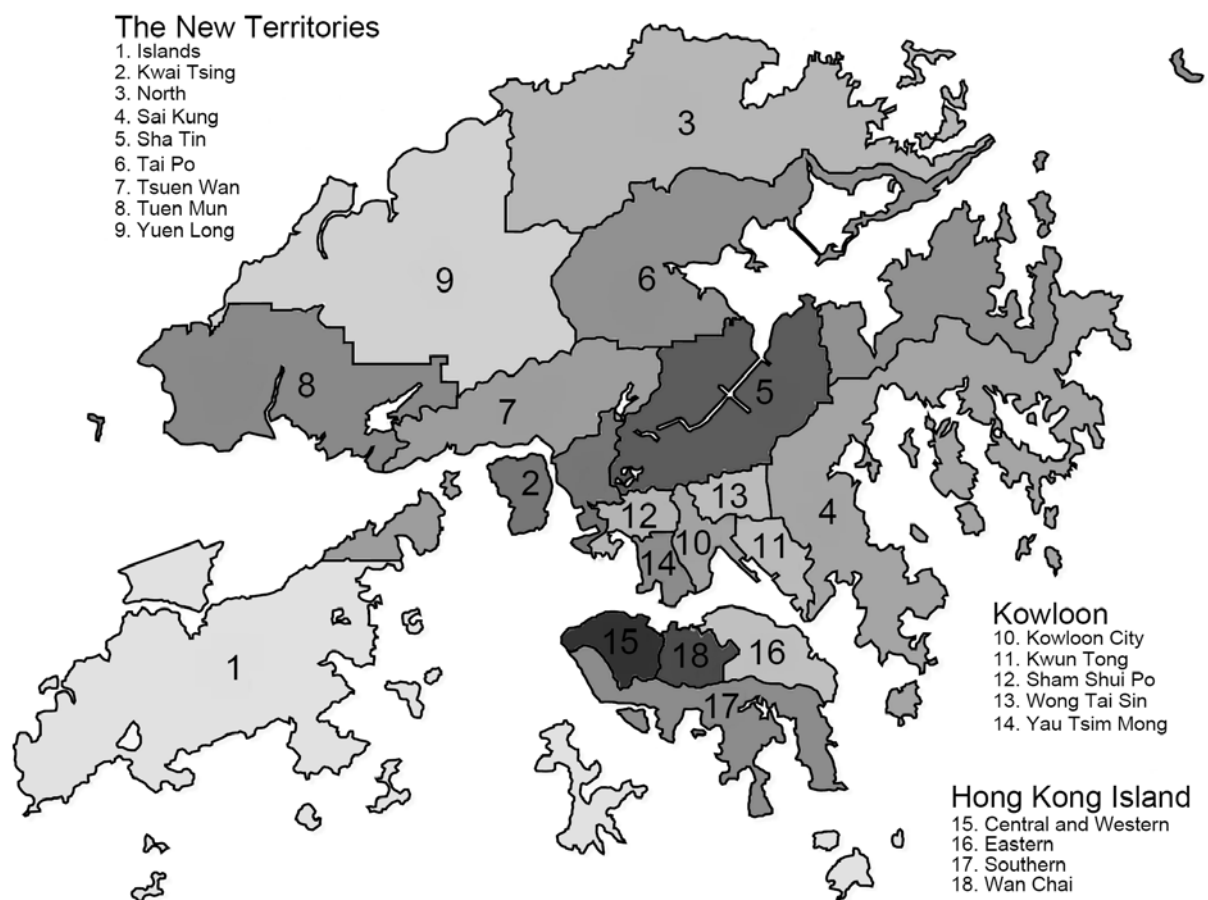15. Central and Western
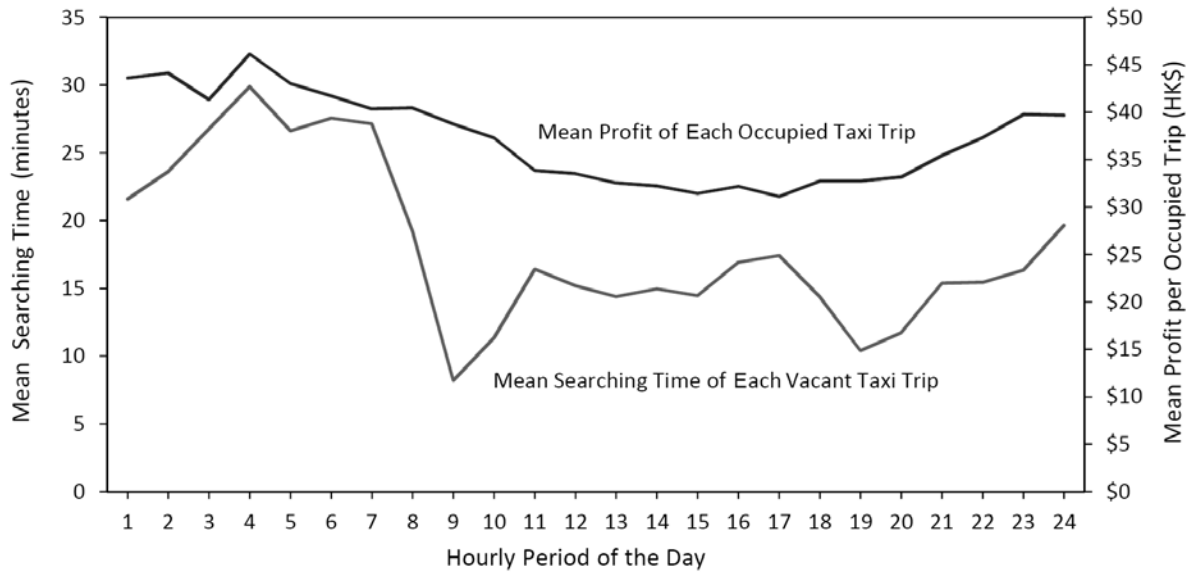16. Eastern
17. Southern
18. Wan Chai

Figure 1 – District Map

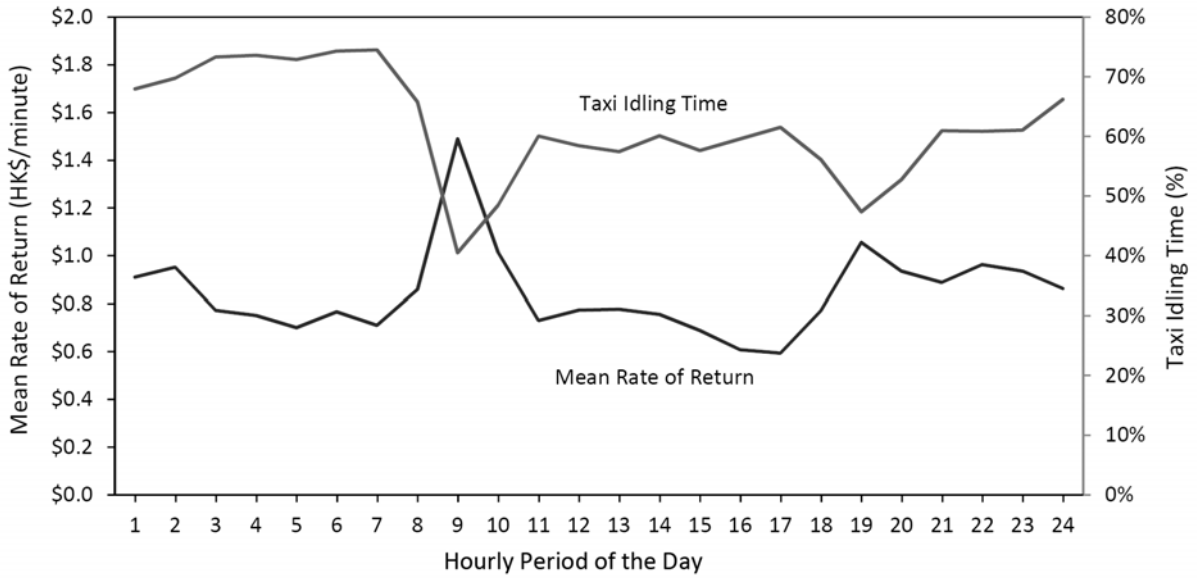Figure 2 – Daily Profiles of Mean Searching Time of Each Vacant Taxi and Mean Profit of Each Occupied Taxi Trip

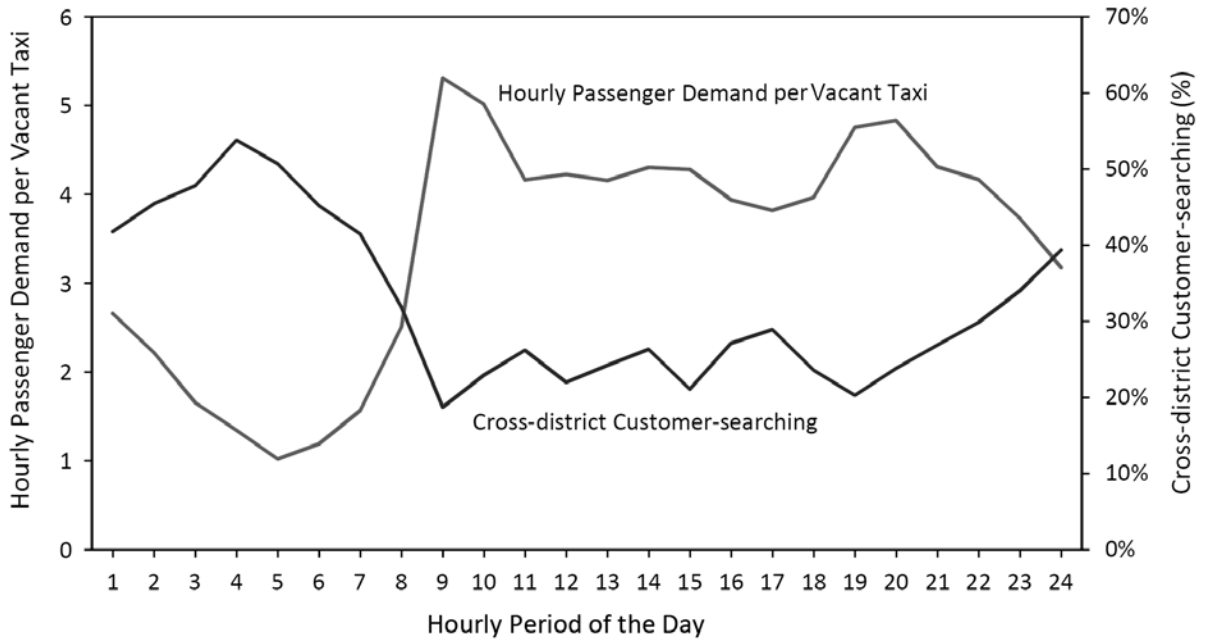Figure 3 – Daily Profiles of Mean ROR and the Percentage of Taxi Idling Time

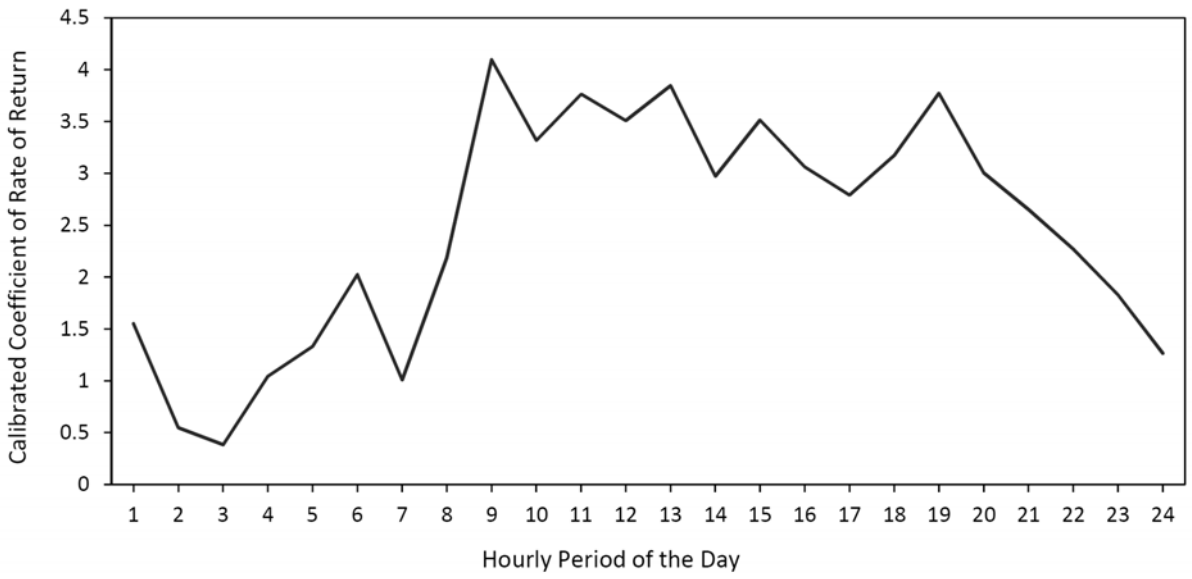Figure 4 – All-day Cross-district Customer-searching Profile

Figure 5 – Changes in the Calibrated Coefficient of ROR over Time