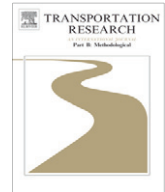




ELSEVIER

Contents lists available at ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

Modeling urban taxi services with multiple user classes and vehicle modes

K.I. Wong^a, S.C. Wong^{b,*}, Hai Yang^c, J.H. Wu^d^a Department of Transportation Technology and Management, National Chiao Tung University, 1001 Ta Hsueh Road, Hsinchu 30050, Taiwan^b Department of Civil Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong^c Department of Civil Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong^d TJKM Transportation Consultants, 5960 Inglewood, Suite 100, Pleasanton, CA 94588, USA

ARTICLE INFO

Article history:

Received 5 September 2007

Received in revised form 13 March 2008

Accepted 13 March 2008

Keywords:

Network model

Urban taxi services

Multiple user classes

Multiple vehicle modes

Hierarchical modal choice

ABSTRACT

This paper extends the model of urban taxi services in congested networks to the case of multiple user classes, multiple taxi modes, and customer hierarchical modal choice. There are several classes of customers with different values of time and money, and several modes of taxi services with distinct combinations of service area restrictions and fare levels. The multi-class multi-mode formulation allows the modeling of both mileage-based and congestion-based taxi fare charging mechanisms in a unified framework, and can more realistically model most urban taxi services, which are charged on the basis of both time and distance. The introduction of multiple taxi modes can also be used to model the differentiation between luxury taxis and normal taxis by their respective service areas and customer waiting times. We propose a simultaneous mathematical formulation of two equilibrium sub-problems for the model. One sub-problem is a combined network equilibrium model (CNEM) that describes the hierarchical logit mode choice model of occupied taxis and normal traffic, together with the vacant taxi distributions in the network. The other sub-problem is a set of linear and nonlinear equations (SLNE), which ensures the satisfaction of the relation between taxi and customer waiting times, the relation between customer demand and taxi supply for each taxi mode, and taxi service time constraints. The CNEM can be formulated as a variational inequality program that is solvable by means of a block Gauss–Seidel decomposition approach coupled with the method of successive averages. The SLNE can be solved by a Newtonian algorithm with a line search. The CNEM is formulated as a special case of the general travel demand model so that it is possible to incorporate the taxi model into an existing package as an add-on module, in which the algorithm for the CNEM is built in practice. Most of the parameters are observable, given that such a calibrated transport planning model exists. A numerical example is used to demonstrate the effectiveness of the proposed methodology.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

In most large cities the taxi industry is subject to various types of regulation, such as entry restriction and price control, and the economic consequences of regulatory restraints have been examined in different ways (Douglas, 1972; De Vany, 1975; Shreiber, 1975; Manski and Wright, 1976; Foerster and Gilbert, 1979; Beesley and Glaister, 1983; Schroeter, 1983; Frankena and Pautler, 1986; Cairns and Liston-Heyes, 1996; Arnott, 1996; Yang et al., 2005a,b; Loo et al., 2007). The general objective of these studies has been to understand the manner in which demand and supply are equilibrated in the presence of such regulations, thereby providing information for government decision making (Beesley and Glaister, 1983). It is

* Corresponding author. Tel.: +852 2859 1964; fax: +852 2559 5337.

E-mail address: hhecwsc@hkucc.hku.hk (S.C. Wong).

commonly realized that two principal characteristics distinguish the taxi market from the idealized market of conventional economic analyses: the role of customer waiting time and the complex relationship between the users (customers) and suppliers (firms) of the taxi services.

In the taxi market, the equilibrium quantity of the service supplied (total taxi-hours) is greater than the equilibrium quantity demanded (occupied taxi-hours) by a certain amount of slack (vacant taxi-hours). This amount of slack governs the average customer waiting time. The expected customer waiting time is generally considered to be an important value or quality of the services that are received by customers. This variable affects customer decisions about whether to take a taxi, and thus plays a crucial role in the determination of the price level and the resultant equilibrium of the market (De Vany, 1975; Abe and Brush, 1976; Manski and Wright, 1976; Foerster and Gilbert, 1979). A reduction in the expected waiting time increases the demand for taxi services. However, from the point of view of each taxi firm, the expected customer waiting time is different from the quality of the typical product. In most markets in which quality is a variable, each firm decides what quality to offer. In the taxi market, the expected waiting time is usually not amenable to differentiation, but depends on the total number of vacant taxi-hours. An individual firm cannot offer customers an expected waiting time that is different from that offered by other firms, although a large firm may be able to reduce the expected waiting time (Frankena and Pautler, 1986).

All of the above studies use an aggregate approach, and do not take into account the spatial structure of the taxi market. Although taxis constitute an important transportation mode that offers a speedy, comfortable, and direct transportation service, they make considerable demands on limited road space and significantly contribute to traffic congestion, including when empty and cruising for customers. In view of the necessity and importance of the network modeling of taxi traffic, Yang and Wong (1998) made an initial attempt to characterize taxi movements in a road network for a given customer origin–destination (O–D) demand pattern. They proposed a simultaneous system of equations to describe the movements of both empty and occupied taxis, and solved the problem with a fixed-point algorithm. The model explicitly deals with the effects of taxi fleet size and the degree of taxi driver uncertainty about customer demand and service conditions through various system performance measures such as taxi utilization and taxi availability at equilibrium, and thus helped to provide information for government decision making about taxi regulations. Wong and Yang (1998) reformulated the taxi service problem in networks as an optimization problem, which led to a more efficient and convergent iterative balancing algorithm for the case without traffic congestion.

Wong et al. (2001) extended the simple network model of urban taxi services proposed by Yang and Wong (1998) by incorporating congestion effects and customer demand elasticity, reformulating the problem as a simultaneous optimization of two sub-problems, and developing a solution algorithm. Wong et al. (2002) developed a sensitivity-based solution algorithm for a taxi model with congested effects and elastic demand. The potential applications of the model have been demonstrated by several case studies of the urban area of Hong Kong. Yang et al. (2001) conducted a case study of the calibration and validation of the simple network model for the urban area of Hong Kong. Yang et al. (2002) then investigated the nature of the demand-supply equilibrium in a regulated market for taxi service in the urban area of Hong Kong. They concentrated on the effects of alternative regulatory restraints on market equilibrium by investigating the social surplus, firm profit, and customer demand at various levels of taxi fare and fleet size in regulated, competitive, and monopoly markets. Wong et al. (2005) studied the bilateral micro-searching behavior for urban taxi services using the absorbing Markov chain approach.

Although a network equilibrium approach for urban taxi services has been developed, some important problems remain unsolved. One of the main issues is the provision of several modes of taxi services in large cities. For instance, there has been concern about the provision of accessible taxis for the handicapped, whose travel characteristics are very different from those of other customers, and demand from affluent customers for luxury taxis (Hong Kong Transport Advisory Committee, 1992, 1998). Another difficulty is how to model the modal similarity of different kinds of taxi services for traveler decision-making processes. In addition, operational considerations have led to the imposition of certain service area restrictions on taxi operations. For example, in response to the rapid urban expansion in Hong Kong, nine new towns were developed in the New Territories (rural areas) in the past three decades (Loo and Chow, 2008). To cope with the new demand in these new towns, rural taxis were introduced to ensure service quality and taxi availability. While rural taxis are only allowed to operate in the New Territories, urban taxis can serve the entire Hong Kong (see Fig. 1 of Loo et al. (2007) for the geographical operating areas). As urban and rural taxis make different contributions to traffic congestion, different government regulations such as license fees and unit charges are applied. Finally, the coexistence of mileage-based and congestion-based taxi fare charging mechanisms is not uncommon in most large cities.

To tackle these difficult problems, this paper extends the single-class network model of urban taxi services to incorporate multiple user classes, multiple taxi modes, and the hierarchical modal choice of customers for taxi services. We consider taxis and normal traffic to analyze the structure of the taxi model, in which normal traffic is assumed to capture all traffic other than taxis in the network. We assume that there are several classes of customers with different values of time and money, and several modes of taxi services with distinct combinations of service area restrictions and fare levels. For service area restrictions, the taxi services are geographically divided into several modes by restricting certain types of taxis to picking up or setting down customers in certain areas (i.e., they operate only within specified zones). For taxi fares, the model allows different modes of taxis to charge at different levels, but the fares are independent of the types of customers who are using the service. Both mileage-based and congestion-based charging mechanisms are considered for the general taxi fare structure. This more realistically models most urban taxi services, which are charged on the basis of both time and distance. The introduction of multiple taxi modes can also be used to model the differentiation between luxury taxis and normal taxis

by their respective service areas and customer waiting times. This extension has important implications for modeling taxi services with service area regulations such as taxi services in Hong Kong, where rural taxis are restricted to operating in rural areas, whereas urban taxis can provide service in both urban and rural areas. To model taxi traffic, it is assumed that a customer, having taken a taxi (similar to normal traffic), will try to minimize his or her individual travel cost from origin to destination; and a vacant taxi, having set down a customer, will try to minimize its individual expected search cost to meet the next customer. The probability that a vacant taxi meets a customer in a particular zone is specified by a logit model by assuming that the expected search time in each zone is an identically distributed random variable due to variations in the perceptions of taxi drivers and the random arrival of customers.

A simultaneous mathematical formulation of two sub-problems is proposed for the model. One sub-problem is a combined network equilibrium model (CNEM) that describes the hierarchical logit mode choice model of occupied taxis and normal traffic, together with the distribution of vacant taxis in the network that are searching for customers. Two common approaches for this CNEM are the optimization approach (Sheffi, 1985; Lam and Huang, 1992; Wong et al., 2004) and the variational inequality (VI) approach (Dafermos, 1982; Wu, 1997; Florian et al., 2002; Wu et al., 2006). When a congestion-based taxi charge is neglected, the CNEM can be formulated as an optimization problem and solved by means of a partial linearization approach (Wong et al., 2004). However, when both mileage-based and congestion-based fares are considered, as in this paper, the problem can no longer be formulated as an optimization problem. Hence, we adopt the variational inequality (VI) approach developed by Florian et al. (2002) for a general multi-class multi-mode network equilibrium problem with a hierarchical (nested) logical structure. There are several multi-class combined models for travel forecasting, notably those of Oppenheim (1995), Boyce and Bar-Gera (2001), and de Cea and Fernandez (2001). For a comprehensive review, readers may refer to a recent paper by Boyce and Bar-Gera (2004). The other sub-problem is a set of linear and nonlinear equations (SLNE), which ensures that the relation between taxi and customer waiting times, the relation between customer demand and taxi supply for each taxi mode, and taxi service time constraints are satisfied. The VI program for the CNEM is solved by means of a block Gauss–Seidel decomposition approach coupled with the method of successive averages. The SLNE is solved by a Newtonian algorithm with a line search. On the solution algorithm, the block Gauss–Seidel decomposition algorithm is a widely used approach. We formulate the CNEM as a special case of the general travel demand model so that it is possible to incorporate our taxi model into an existing package as an add-on module, in which the algorithm for the CNEM is built in practice. The assumptions in the model are reasonable and can be determined in practice, and most of the parameters are observable, given that such a calibrated transport planning model exists. The additional effort in calibrating and validating the model with the inclusion of taxi flows has been shown in Yang et al. (2001, 2002), using the city of Hong Kong as an example. The model can be used for the assessment of the interrelationships between customer demand and taxi utilization for different modes of taxis that provide services in the network. A numerical example is used to demonstrate the effectiveness of the proposed methodology.

2. Model assumptions

2.1. Taxi movements in a road network

Consider a road network $G(V, A)$ in which V is the set of nodes and A is the set of links in the network. Let I and J be the sets of customer origin and destination zones, respectively, and P and Q be the sets of user classes and taxi modes, respectively. In the following, the superscript “n” denotes normal traffic, “o” denotes occupied taxis, and “v” denotes vacant taxis. We also use $\psi \in \Psi = ((n, p), (o, pq), (v, q))$ to represent the combination of user classes and transportation modes for $p \in P$ and $q \in Q$. The notation that is used throughout this paper is described as follows, and is also summarized in Appendix.

In any given hour, the total number of travelers of class p from origin zone i to destination zone j is \bar{D}_{ij}^p , which is assumed to be fixed and known, and is expressed in vehicles per hour. The demand matrix can be split into two sub-matrices that are associated with non-taxi and taxi mode choices,

$$\bar{D}_{ij}^p = T_{ij}^{n,p} + T_{ij}^{o,p}, \quad p \in P, j \in J, i \in I, \quad (1)$$

where $T_{ij}^{n,p}$ and $T_{ij}^{o,p}$ are the numbers of non-taxi (normal traffic) users and taxi users (customers), respectively. The matrix for taxi traffic can be further split into several matrices for each taxi mode:

$$T_{ij}^{o,p} = \sum_{q \in Q} T_{ij}^{o,pq}, \quad p \in P, j \in J, i \in I, \quad (2)$$

where $T_{ij}^{o,pq}$ is the number of users who choose taxis of mode $q \in Q$. A description of the mode split structure between normal traffic and taxis is shown in Fig. 1, and the hierarchical mode choice is further described in Section 2.6.

For each taxi mode, to meet customer demand, the occupied taxis carry customers from their origins to destinations. For each taxi mode $q \in Q$, we have the following trip end constraints:

$$O_i^q = \sum_{j \in J} \sum_{p \in P} T_{ij}^{o,pq}, \quad q \in Q, i \in I \subset V, \quad (3)$$

$$D_j^q = \sum_{i \in I} \sum_{p \in P} T_{ij}^{o,pq}, \quad q \in Q, j \in J \subset V, \quad (4)$$

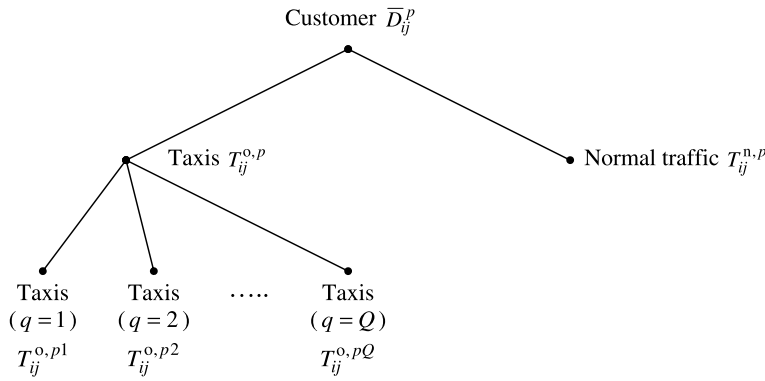


Fig. 1. The hierarchical mode choice structure for normal traffic and taxis.

where O_i^q and D_j^q are the total customer demand from origin zone $i \in I$ and total customer demand to destination zone $j \in J$, respectively, for taxis of mode $q \in Q$. In addition to occupied taxis, vacant taxis are also traveling in the network and searching for customers. The number of vacant taxis of mode $q \in Q$ that are traveling from zone j to zone i and searching for customers is denoted as $T_{ji}^{v,q}$.

Traffic in network T_{ij}^ψ as described above will choose paths $R_{ij}^\psi \subseteq R$, $\psi \in \Psi$, where $R = \cup_{p \in P} R_{ij}^{n,p} \cup \cup_{q \in Q} R_{ij}^{o,pq} \cup \cup_{q \in Q} R_{ji}^{v,q}$ is the set of all routes. The corresponding path flow variables are denoted as f_r^ψ , $r \in R_{ij}^\psi$. Let t_a be the travel time on link $a \in A$, and v_a^ψ be the total vehicle flow of the corresponding class-mode combination on link $a \in A$. Further, let $\delta_{ij,ar}^\psi$ be the link route incidence matrix, which is equal to 1 when route r between the O–D pair (i,j) of the corresponding users traverses link a and 0 otherwise. The total flow on a link can then be obtained by

$$v_a = \sum_{p \in P} v_a^{n,p} + \sum_{p \in P} \sum_{q \in Q} v_a^{o,pq} + \sum_{q \in Q} v_a^{v,q}, \quad a \in A, \tag{5}$$

where $v_a^\psi = \sum_{i \in I} \sum_{j \in J} \sum_{r \in R_{ij}^\psi} \delta_{ij,ar}^\psi f_r^\psi$.

2.2. Customer and taxi waiting times

Let W_i^q be the customer waiting time for taxis of mode q in zone i , and w_i^q be the taxi waiting/searching time for taxis of mode q in zone i . The customer waiting time, which is an endogenous variable of the model, varies across zones. It is expected that an additional customer waiting for a taxi of a particular mode will only increase the expected waiting time for taxis of that mode, and will not be related to customers who are waiting for taxis of other modes. Thus, within each zone, the expected waiting time of customers who are waiting for taxis of mode $q \in Q$ can be described as a function of the density of vacant taxis of that mode in the zone $W_i^q = W_i^q(N_i^q, w_i^q)$, $i \in I$, $q \in Q$, where N_i^q is the number of vacant taxis of mode q per hour that meet customers at zone $i \in I$ (note that $N_i^q = O_i^q$ at equilibrium). The function does not limit the value of customer waiting time. Our stationary model represents an equilibrium state over time; the modeling period has neither a beginning nor an end, and it is possible for the waiting time to be more than 1 h (the modeling period).

The specification of the customer waiting time function depends on the distribution of taxi stands over individual zones. In the case with a continuous taxi stand distribution (taxis can pick up customers anywhere), we can assume that vacant taxis move randomly through the streets, and that for each taxi mode the expected average customer waiting time is proportional to the area of the zone and inversely proportional to the (cruising) vacant taxi hours:

$$W_i^q = \eta \frac{Z_i}{N_i^q w_i^q} = \eta \frac{Z_i}{O_i^q w_i^q}, \quad q \in Q, \quad i \in I, \tag{6}$$

where Z_i is the area of zone $i \in I$ and η is a model parameter that is common to all zones and all taxi modes. This approximate distribution of waiting times can be derived theoretically (Douglas, 1972; Yang et al., 2002). Annual taxi service surveys (at sampled taxi stands and roadside observation points) have been conducted in Hong Kong since 1986 to gather information on these customer and taxi waiting times (Transport Department, 1986–1998), and these surveys have also been used for case studies of the validity of taxi models (Yang et al., 2001, 2002).

2.3. Cost structures

Let the generalized costs of travel on link $a \in A$ as perceived by different users be c_a^ψ for the corresponding class-mode combination, which is assumed to be a linear combination of the link travel time t_a and link length d_a . For simplicity, we assume that the travel time $t_a(v_a)$ on link $a \in A$ is separable and an increasing function of total vehicular flow v_a on link

$a \in A$. Let b_0^p be the value of time of users in class p (irrespective of whether they are taking taxis or normal traffic), b^n be the mileage costs that are charged to customers who are taking normal traffic, $b_1^{o,q}$ and $b_2^{o,q}$ be the mileage and congestion-based costs that are charged to customers who are taking a taxi of mode $q \in Q$, and b_0^q and $b^{v,q}$ be the hourly and mileage (e.g., fuel) operating costs of a taxi of mode $q \in Q$. We have

$$c_a^{n,p} = b_0^p t_a(v_a) + b^n d_a, \quad a \in A, p \in P, \quad (7)$$

$$c_a^{o,pq} = b_0^p t_a(v_a) + b_1^{o,q} d_a + b_2^{o,q} t_a(v_a) + p_a^q, \quad a \in A, q \in Q, p \in P, \quad (8)$$

$$c_a^{v,q} = b_0^q t_a(v_a) + b^{v,q} d_a + p_a^q, \quad a \in A, q \in Q, \quad (9)$$

where p_a^q is a cost equivalent user-specified charge that is applied to taxis of mode q that are entering link $a \in A$. This user-specified charge can serve several purposes. For example, in some cities, when a taxi enters a terminal (e.g., an airport) searching for customers, the taxi may have to pay a terminal charge. In the case of multiple taxi classes in which some of the classes are subject to area restrictions, this charge is a penalty to prevent the restricted mode of taxis from entering the restricted links. These charges are user specified according to the network characteristics and are zero when the taxi mode is not restricted.

We assume that each potential customer knows (from day-to-day experience) the expected waiting times for the different modes of taxi services that are available at the origin location, and can estimate the fares of different taxi modes for his or her trip. The customer's choice among various modes of taxis and normal traffic depends on the corresponding disutility values, including the minimum cost of travel and other factors such as the waiting time and whether the user prefers a particular mode of vehicle. The total generalized cost of travel (disutility) for users of class p who are taking normal traffic or taxis of mode q from zone i to zone j on path r can be defined as

$$C_r^{n,p} = \sum_{a \in A} \delta_{ij,ar}^{n,p} c_a^{n,p}, \quad r \in R_{ij}^{n,p}, p \in P, j \in J, i \in I, \quad (10)$$

$$C_r^{o,pq} = \sum_{a \in A} \delta_{ij,ar}^{o,pq} c_a^{o,pq} + b_1^p W_i^q - \rho^{pq}, \quad r \in R_{ij}^{o,pq}, q \in Q, p \in P, j \in J, i \in I, \quad (11)$$

respectively, where b_1^p is the value of customer waiting time (for taxis) as perceived by user class p , and ρ^{pq} is the inherent inertia, which indicates that users of class p prefer taxis of mode q in relation to normal traffic. The total searching cost for a vacant taxi of mode q that leaves zone j and goes to zone i on path r can be defined as

$$C_r^{v,q} = \sum_{a \in A} \delta_{ji,ar}^{v,q} c_a^{v,q}, \quad r \in R_{ji}^{v,q}, q \in Q, i \in I, j \in J. \quad (12)$$

With the above defined path cost C_r^ψ , the corresponding minimum cost via the shortest route from origin i to destination j is defined as C_{ij}^ψ , where $C_{ij}^\psi = \min(C_r^\psi, r \in R_{ij}^\psi, \psi \in \Psi)$.

To facilitate our presentation, let the fare of a taxi ride be the monetary cost that is charged to a customer, which is a function of the fare level, time, and total distance that is traveled. The fare for taking a taxi of mode $q \in Q$ that is traveling from zone i to zone j along path r by a user of class $p \in P$ can be defined as

$$F_r^{pq} = \sum_{a \in A} \delta_{ij,ar}^{o,pq} F_a^{o,q} = \sum_{a \in A} \delta_{ij,ar}^{o,pq} (b_1^{o,q} d_a + b_2^{o,q} t_a(v_a)). \quad (13)$$

As the travel time on each path between each origin–destination (OD) pair may not necessarily be identical, we further define the average travel times of users who are taking various modes of transportation as h_{ij}^ψ , where

$$h_{ij}^\psi = \sum_{r \in R_{ij}^\psi} \left(f_r^\psi \sum_{a \in A} \delta_{ij,ar}^\psi t_a \right) / \sum_{r \in R_{ij}^\psi} f_r^\psi, \quad \psi \in \Psi, j \in J, i \in I. \quad (14)$$

These are the average travel times along all of the used paths between origin i and destination j , and they are weighted by the volume of flows on the paths.

2.4. Taxi service time constraints

For each taxi mode $q \in Q$, suppose that N^q cruising taxis are operating in the network, and consider one unit period (1 h) of taxi operations. The total occupied time of all taxis of mode $q \in Q$ is the taxi hours that are required to complete all $T_{ij}^{o,pq}$, $i \in I, j \in J, p \in P$ trips, and is thus given by $\sum_{p \in P} \sum_{i \in I} \sum_{j \in J} T_{ij}^{o,pq} h_{ij}^{o,pq}$. In addition, the total unoccupied time consists of the moving (searching) times of vacant taxis from zone to zone and the waiting (searching) times within zones, and is given by $\sum_{j \in J} \sum_{i \in I} T_{ji}^{v,q} \{h_{ji}^{v,q} + w_i^q\}$. The sum of total occupied taxi hours and total vacant taxi hours should be equal to the total taxi service time. Therefore, for each taxi mode $q \in Q$, the following taxi service time constraint must be satisfied in the (1 h) modeled period:

$$\sum_{i \in I} \sum_{j \in J} \sum_{p \in P} T_{ij}^{o,pq} h_{ij}^{o,pq} + \sum_{j \in J} \sum_{i \in I} T_{ji}^{v,q} \{h_{ji}^{v,q} + w_i^q\} = N^q, \quad q \in Q. \quad (15)$$

2.5. Behavior of vacant taxi drivers

The destination choice of empty taxis depends on the behavior of taxi drivers. We assume that all customers who are assigned to a particular mode of taxi services are indistinguishable to the vacant taxis of that mode. That is, the taxi drivers of this mode will consider these potential customers to be identical when searching, and the customer search behavior and characteristics of a particular mode of vacant taxis are assumed to be independent of those for other modes of vacant taxis. We assume here that once a customer ride is completed, a taxi becomes vacant and cruises either in the same zone or goes to other zones in search of the next customer. During the search process, each taxi driver is assumed to attempt to minimize the individual expected search time/cost (including the operating cost) that is required to meet the next customer. The expected search time/cost in each zone is a random variable because of variations in the perceptions of taxi drivers and the random arrival of customers. We assume that this random variable is identically distributed, with a Gumbel density function. With these behavioral assumptions, the probability that a vacant taxi of mode $q \in Q$ that originated in zone $j \in J$ will eventually meet a customer in zone $i \in I$ is specified by the following logit model:

$$P_{ij}^{v,q} = \frac{\exp\{-\theta^q(C_{ji}^{v,q} + b_0^q w_i^q)\}}{\sum_{m \in I} \exp\{-\theta^q(C_{jm}^{v,q} + b_0^q w_m^q)\}}, \quad q \in Q, j \in J, i \in I, \tag{16}$$

where θ^q ($1/h$) is a non-negative parameter that reflects the degree of uncertainty in customer demand and taxi services of mode q in the whole market from the perspective of individual taxi drivers.

In a stationary equilibrium state, the movements of vacant taxis in the network should meet customer demand in all origin zones, i.e., every customer will eventually receive taxi service after a certain waiting time. As there are D_j^q taxis of mode $q \in Q$ to operate in destination zone $j \in J$ per hour, we have $\sum_{i \in I} T_{ji}^{v,q} = D_j^q, q \in Q, j \in J$ and $\sum_{j \in J} T_{ji}^{v,q} = \sum_{j \in J} D_j^q$. $P_{ij}^{v,q} = O_i^q, q \in Q, i \in I$.

2.6. Hierarchical logit mode choice

To determine the customer demand matrices for taxis, the modal split between normal traffic and taxis is determined in such a way that the selection of taxis is conditioned by a choice between taking a taxi and non-taxi traffic, based on the total generalized costs (disutility values) of the alternatives. We assume a hierarchical logit-type mode choice between taxis and normal traffic, where the selection of a taxi or normal traffic is in the upper choice level, and the choice of taxi mode is in the lower choice level. We also assume a logit-type mode choice function, which gives the proportion of trips taken by normal traffic $P_{ij}^{n,p}$ and the proportion of trips taken by taxis $P_{ij}^{o,p}$:

$$P_{ij}^{n,p} = \frac{\exp(-\beta_1^p C_{ij}^{n,p})}{\exp(-\beta_1^p L_{ij}^{o,p}) + \exp(-\beta_1^p C_{ij}^{n,p})}, \quad p \in P, i \in I, j \in J, \tag{17}$$

where $P_{ij}^{o,p} = 1 - P_{ij}^{n,p}$, and $L_{ij}^{o,p} = -(1/\beta_2^p) \ln \sum_{q' \in Q} \exp(-\beta_2^p C_{ij}^{o,pq'})$, $p \in P, j \in J, i \in I$, is the logsum of the disutility of travel as perceived by users of class p who are using taxis as their mode of transportation from zone i to zone j . β_1^p and β_2^p are the dispersion coefficients for the upper-level and lower-level logit mode choice model, respectively, for user-class p .

Given that taxis have been selected as the transportation mode, to determine the proportion of trips $P_{ij}^{o,pq}$ of taxis from origin zone i to destination zone j in taxis of mode q as chosen by users of class p , we introduce the following lower-level logit mode choice function:

$$P_{ij}^{o,pq} = \frac{\exp(-\beta_2^p C_{ij}^{o,pq})}{\sum_{q' \in Q} \exp(-\beta_2^p C_{ij}^{o,pq'})} = \frac{\exp(-\beta_2^p C_{ij}^{o,pq})}{\exp(-\beta_2^p L_{ij}^{o,p})}, \quad q \in Q, p \in P, j \in J, i \in I. \tag{18}$$

The service area restrictions for different modes of taxis have been defined implicitly in their (minimum) travel cost matrix, in which the travel cost will be very high (due to the penalty on the links based on restricted entry, as in Eqs. (8) and (9)) if the corresponding origin or destination is not accessible by that taxi mode, and the proportion of users who choose that mode of taxi will be very small (and negligible). Thus, for a given number of travelers of class p who are moving from origin zone i , to destination zone $j \in J$, the number of users of class p who are taking normal traffic is $T_{ij}^{n,p} = \bar{D}_{ij}^p \cdot P_{ij}^{n,p}$, and the number of users of class p who are taking taxi services of mode q is $T_{ij}^{o,pq} = \bar{D}_{ij}^p \cdot P_{ij}^{o,p} \cdot P_{ij}^{o,pq}$. A description of the hierarchical mode choice structure for normal traffic and taxis is shown in Fig. 1. Note that the dispersion coefficient must be chosen such that $\beta_2^p \geq \beta_1^p \forall p \in P$ for the consistency of the hierarchical logit model (Ortuzar and Willumsen, 1996).

3. Simultaneous mathematical formulation

In this section, we assume that two types of essential constraints can be temporarily relaxed: the relationship (Eq. (6)) between the customer and taxi waiting time at each origin zone (with $O_i^q = \sum_{j \in J} \sum_{p \in P} T_{ij}^{o,pq}$),

$$w_i^q W_i^q \sum_{j \in J} \sum_{p \in P} T_{ij}^{o,pq} - \eta Z_i = 0, \quad q \in Q, i \in I \tag{19}$$

and the conservation of flows between the zonal origin and destination subtotals of customer demands and occupied taxi movements (Eqs. (3) and (4)). Now, W_i^q , O_i^q , and D_j^q are temporarily treated as exogenous variables, i.e., Eqs. (3), (4) and (6) are not restricted to be satisfied at this stage. The internal inconsistency of these equations will be rectified later by a Newtonian algorithm for solving the SLNE.

3.1. Combined network equilibrium model (CNEM)

With the temporary relaxation of two types of essential constraints described in the previous section, we introduce the following mathematical program, which represents the combined network equilibrium model (CNEM) for multiple user classes and multiple vehicle modes. The problem is formulated using the VI approach, which can handle asymmetric interactions of network flows, especially for the problem of multiple user classes and vehicle modes. Note that the model presented here is a special case of that which Florian et al. (2002) presented. Our model considers only a road network in which customers (in the taxi or non-taxi mode) and vacant taxis are traveling. Whereas the general model of Florian et al. (2002) uses a “trip distribution-mode choice-assignment” hierarchy, our model uses a simultaneous “trip distribution-assignment” and “mode choice-assignment” hierarchy in which vacant taxis follow trip distribution and customers follow mode choice, and all of their movements are on the same road network. Therefore, it is a special case of the multiple user classes and vehicle modes equilibrium model in which the vacant taxi movements are not involved in mode choice and customer choices do not include trip distribution.

The feasible region Ω of the demands and path flows of the equilibrium model is as follows, where the dual variables are stated in brackets:

$$\sum_{j \in J} T_{ji}^{v,q} = O_i^q, \quad q \in Q, \quad i \in I, \quad (\alpha_i^q) \tag{20a}$$

$$\sum_{i \in I} T_{ji}^{v,q} = D_j^q, \quad q \in Q, \quad j \in J, \quad (\beta_j^q) \tag{20b}$$

$$\sum_{q \in Q} T_{ij}^{o,pq} = T_{ij}^{o,p}, \quad p \in P, \quad j \in J, \quad i \in I, \quad (L_{ij}^{o,p}) \tag{20c}$$

$$T_{ij}^{o,p} + T_{ij}^{n,p} = \bar{D}_{ij}^p, \quad p \in P, \quad j \in J, \quad i \in I, \quad (L_{ij}^{D,p}) \tag{20d}$$

$$\sum_{r \in R_{ij}^\psi} f_r^\psi = T_{ij}^\psi, \quad \psi \in \Psi, \quad j \in J, \quad i \in I, \quad (u_{ij}^\psi) \tag{20e}$$

$$f_r^\psi \geq 0, \quad r \in R_{ij}^\psi, \quad \psi \in \Psi, \quad j \in J, \quad i \in I, \quad (\zeta_r^\psi) \tag{20f}$$

$$T_{ij}^\psi > 0, \quad \psi \in \Psi, \quad j \in J, \quad i \in I. \tag{20g}$$

The constraints equations (20a) and (20b) are the conservation of flows for vacant taxis; Eqs. (20c) and (20d) are the conservation of flows for the total demand for taxis and the total demand for travel, respectively; Eq. (20e) defines that the sum of all path flows between each O–D pair should be equal to the OD matrix; and Eqs. (20f) and (20g) are the non-negativity constraints for the variables of path flow and OD flow, respectively.

Let $\mathbf{f} = \{f_r^\psi, r \in R_{ij}^\psi, \psi \in \Psi, j \in J, i \in I\}$ and $\mathbf{T} = \{T_{ij}^\psi, \psi \in \Psi, j \in J, i \in I\}$. The VI program is given as follows. Find $(\mathbf{f}, \mathbf{T}) \in \Omega$ such that

$$\begin{aligned} & \sum_{i \in I} \sum_{j \in J} \left(\sum_{\psi \in \Psi} \sum_{r \in R_{ij}^\psi} C_r^\psi(\mathbf{f}^*) (f_r^\psi - f_r^{\psi*}) + \sum_{p \in P} \left(\frac{1}{\beta_1^p} \ln T_{ij}^{n,p*} (T_{ij}^{n,p} - T_{ij}^{n,p*}) + \frac{1}{\beta_1^p} \ln T_{ij}^{o,p*} (T_{ij}^{o,p} - T_{ij}^{o,p*}) \right) \right. \\ & \left. + \frac{1}{\beta_2^p} \sum_{q \in Q} (\ln(T_{ij}^{o,pq*} / T_{ij}^{o,p*}) (T_{ij}^{o,pq} - T_{ij}^{o,pq*})) \right) + \sum_{q \in Q} \left(\frac{1}{\beta^q} \ln T_{ji}^{v,q*} (T_{ji}^{v,q} - T_{ji}^{v,q*}) \right) \geq 0 \quad \forall (\mathbf{f}, \mathbf{T}) \in \Omega \end{aligned} \tag{21}$$

where C_r^ψ is defined by Eqs. (10)–(12).

The Karush–Kuhn–Tucker (KKT) conditions of the VI program (Eq. (21)) are given in the following:

$$f_r^\psi : C_r^\psi - u_{ij}^\psi - \zeta_r^\psi = 0, \quad r \in R_{ij}^\psi, \quad \psi \in \Psi, \quad j \in J, \quad i \in I, \tag{22}$$

$$T_{ij}^{n,p} : \frac{1}{\beta_1^p} \ln T_{ij}^{n,p} + u_{ij}^{n,p} - L_{ij}^{D,p} = 0, \quad p \in P, \quad j \in J, \quad i \in I, \tag{23}$$

$$T_{ij}^{o,p} : \frac{1}{\beta_1^p} \ln T_{ij}^{o,p} + L_{ij}^{o,p} - L_{ij}^{D,p} = 0, \quad p \in P, \quad j \in J, \quad i \in I, \tag{24}$$

$$T_{ij}^{o,pq} : \frac{1}{\beta_2^p} \ln \frac{T_{ij}^{o,pq}}{T_{ij}^{o,p}} + u_{ij}^{o,pq} - L_{ij}^{o,p} = 0, \quad q \in Q, \quad p \in P, \quad j \in J, \quad i \in I, \tag{25}$$

$$T_{ji}^{v,q} : \frac{1}{\beta^q} \ln T_{ji}^{v,q} + u_{ji}^{v,q} + \alpha_i^q + \beta_j^q = 0, \quad q \in Q, \quad i \in I, \quad j \in J. \tag{26}$$

The complementarity conditions are

$$f_r^\psi \zeta_r^\psi = 0, \quad r \in R_{ij}^\psi, \quad \psi \in \Psi, \quad j \in J, \quad i \in I, \tag{27}$$

$$\zeta_r^\psi \geq 0, \quad r \in R_{ij}^\psi, \quad \psi \in \Psi, \quad j \in J, \quad i \in I \tag{28}$$

and Eqs. (20a)–(20g).

At equilibrium, u_{ij}^ψ is interpreted as the minimum generalized cost and $C_{ij}^\psi = u_{ij}^\psi$. From Eq. (25) and the fact that $T_{ij}^{\psi,pq} > 0$, we have

$$\frac{T_{ij}^{\psi,pq}}{T_{ij}^{\psi,p}} = \exp(-\beta_2^p u_{ij}^{\psi,pq}) \exp(\beta_2^p L_{ij}^{\psi,p}), \quad q \in Q, \quad p \in P, \quad j \in J, \quad i \in I. \tag{29}$$

The definition of the conservation of flows (Eq. (20c)) and the summation of Eq. (29) over $q \in Q$ implies the logsum of the generalized cost of various taxi modes be $L_{ij}^{\psi,p} = -\frac{1}{\beta_2^p} \ln \sum_{q \in Q} \exp(-\beta_2^p u_{ij}^{\psi,pq})$. As $T_{ij}^{\psi,p} > 0$, we can show that

$$\frac{T_{ij}^{\psi,pq}}{T_{ij}^{\psi,p}} = \frac{\exp(-\beta_2^p u_{ij}^{\psi,pq})}{\sum_{q' \in Q} \exp(-\beta_2^p u_{ij}^{\psi,pq'})}, \quad q \in Q, \quad p \in P, \quad j \in J, \quad i \in I, \tag{30}$$

which satisfies the lower-level logit mode choice proportion of taxi trips by users of class $p \in P$ from origin zone i to destination zone j who are taking taxis of mode $q \in Q$. From Eq. (24), we have

$$T_{ij}^{\psi,p} = \exp(-\beta_1^p L_{ij}^{\psi,p}) \exp(\beta_1^p L_{ij}^{\psi,p}), \quad p \in P, \quad j \in J, \quad i \in I. \tag{31}$$

From Eq. (23) and the fact that $T_{ij}^{\psi,p} > 0$, we have

$$T_{ij}^{\psi,p} = \exp(-\beta_1^p u_{ij}^{\psi,p}) \exp(\beta_1^p L_{ij}^{\psi,p}), \quad p \in P, \quad j \in J, \quad i \in I. \tag{32}$$

Now, from the definition of the conservation of total flows over the network (Eq. (20d)), we have

$$\frac{T_{ij}^{\psi,p}}{D_{ij}^p} = \frac{\exp(-\beta_1^p u_{ij}^{\psi,p})}{\exp(-\beta_1^p L_{ij}^{\psi,p}) + \exp(-\beta_1^p u_{ij}^{\psi,p})}, \quad p \in P, \quad j \in J, \quad i \in I, \tag{33}$$

$$\frac{T_{ij}^{\psi,p}}{D_{ij}^p} = \frac{\exp(-\beta_1^p L_{ij}^{\psi,p})}{\exp(-\beta_1^p L_{ij}^{\psi,p}) + \exp(-\beta_1^p u_{ij}^{\psi,p})}, \quad p \in P, \quad j \in J, \quad i \in I, \tag{34}$$

which satisfies the upper-level logit mode choice proportion of normal traffic and taxi trips for each user class $p \in P$. Combining Eqs. (31) and (34) gives the logsum of the generalized cost of travel in the network taking either normal traffic or taxis as $L_{ij}^{\psi,p} = -\frac{1}{\beta_1^p} \ln(\exp(-\beta_1^p L_{ij}^{\psi,p}) + \exp(-\beta_1^p u_{ij}^{\psi,p}))$.

From Eq. (26) and the fact that $T_{ji}^{\psi,q} > 0$, we have

$$T_{ji}^{\psi,q} = \exp(-\theta^q (u_{ji}^{\psi,q} + \alpha_i^q + \beta_j^q)), \quad q \in Q, \quad i \in I, \quad j \in J. \tag{35}$$

Substituting Eq. (35) into Eq. (20b) gives rise to

$$\sum_{i \in I} T_{ji}^{\psi,q} = \sum_{i \in I} \exp(-\theta^q (u_{ji}^{\psi,q} + \alpha_i^q + \beta_j^q)) = D_j^q, \quad q \in Q, \quad j \in J. \tag{36}$$

Thus,

$$\exp(-\theta^q \beta_j^q) = \frac{D_j^q}{\sum_{i \in I} \exp(-\theta^q (u_{ji}^{\psi,q} + \alpha_i^q))}, \quad q \in Q, \quad j \in J. \tag{37}$$

Substituting Eq. (37) into Eq. (35) leads to

$$T_{ji}^{\psi,q} = D_j^q \frac{\exp(-\theta^q (u_{ji}^{\psi,q} + \alpha_i^q))}{\sum_{m \in I} \exp(-\theta^q (u_{jm}^{\psi,q} + \alpha_m^q))}, \quad q \in Q, \quad i \in I, \quad j \in J. \tag{38}$$

Comparing Eq. (38) with the logit-based searching behavioral assumption (16), α_i^q can now be related to the taxi waiting time in zone i for the taxis of mode q . As there is a set of dependent equations in the set of Eq. (20a) and (20b), the taxi waiting times are related to the taxi mode-specific variables by $w_i^q = \alpha_i^q + c^q$, where $\alpha_i^q = \alpha_i^q / b_0^q$, and

$$c^q = \frac{N^q - \sum_{i \in I} \sum_{j \in J} \sum_{p \in P} T_{ij}^{\psi,pq} h_{ij}^{\psi,pq} - \sum_{j \in J} \sum_{i \in I} T_{ji}^{\psi,q} h_{ji}^{\psi,q} - \sum_{i \in I} O_i^q \bar{\alpha}_i^q}{\sum_{i \in I} O_i^q}, \quad q \in Q, \tag{39}$$

which can be derived from the taxi service time constraint (Eq. (15)), are the taxi mode-specific variables (for details, see Wong and Yang, 1998).

We showed that the solution of the VI program (Eq. (21)) satisfies all of the functional relationships that are required by the taxi model as defined in Section 2, except the set of temporarily relaxed constraints. It is noteworthy that the VI program can be formulated as an optimization program when the congestion-based taxi charge is neglected (Wong et al., 2004), in

which case more efficient algorithms are available. The existence of the solution for the VI program is demonstrated by the fact that the VI problem admits at least one solution if its feasible region is a compact convex set and the function of the VI problem is continuous in the feasible region. In our case, the feasible region Ω as defined in Eq. (20) is a set of linear constraints, positive and non-negative constraints, and the function defined in Eq. (21) is continuous with the region of non-negative flows and positive OD matrices. Thus, we can show that there is at least one solution for the VI program. Other analytical properties of this VI program can be found in Wu (1997) and Florian et al. (2002).

The final output of the CNEM includes a complete set of link flow patterns, the minimum travel cost and the corresponding average travel time between each OD pair, the customer demand for taxis and the distribution of vacant taxis within the network, and the corresponding Lagrange multipliers (α_i^q) in the gravity model for the distribution of vacant taxis.

3.2. A set of linear and nonlinear equations (SLNE)

Solving the CNEM does not, in general, ensure the satisfaction of the relaxed constraints, Eqs. (3), (4) and (6), and the taxi service time constraint, Eq. (15). This creates internal inconsistency in the modeling and solution approach. To ensure that the desirable characteristics of the taxi model are satisfied, this inconsistency can be rectified by formulating the constraints as a set of equations, and the rectification of the set of equations is designated as the SLNE. Note that there is one dependent equation for each taxi mode $q \in Q$ in the system of Eqs. (3) and (4) in view of the following fact. $\sum_{i \in I} \sum_{j \in J} \sum_{p \in P} T_{ij}^{o,pq} = \sum_{i \in I} O_i^q = \sum_{j \in J} \sum_{i \in I} \sum_{p \in P} T_{ij}^{o,pq} = \sum_{j \in J} D_j^q, q \in Q$. Therefore, one equation for each taxi mode can be dropped in this set of flow conservation equations, which is associated with an origin zone $z \in I$ or a destination zone $z \in J$, where z is an arbitrarily selected zone. Assume that D_z^q is an arbitrarily chosen dependent variable in the deleted constraint that satisfies $D_z^q = \sum_{i \in I} O_i^q - \sum_{j \in \{J-z\}} D_j^q, q \in Q$.

The control variables of the SLNE are $\mathbf{W} = \text{Col}(W_i^q, i \in I, q \in Q)$, $\mathbf{O} = \text{Col}(O_i^q, i \in I, q \in Q)$, $\bar{\mathbf{D}} = \text{Col}(D_j^q, j \in \{J-z\}, q \in Q)$, and $\mathbf{c} = \text{Col}(c^q, q \in Q)$. The results that are obtained from the CNEM include $\bar{\alpha} = (\bar{\alpha}_i^q(\mathbf{W}, \mathbf{O}, \bar{\mathbf{D}}), i \in I, q \in Q)$, $\mathbf{h}^o = (h_{ij}^{o,pq}(\mathbf{W}, \mathbf{O}, \bar{\mathbf{D}}), i \in I, j \in J, p \in P, q \in Q)$, $\mathbf{h}^v = (h_{ji}^{v,q}(\mathbf{W}, \mathbf{O}, \bar{\mathbf{D}}), j \in J, i \in I, q \in Q)$, $\mathbf{T}^o = (T_{ij}^{o,pq}(\mathbf{W}, \mathbf{O}, \bar{\mathbf{D}}), i \in I, j \in J, p \in P, q \in Q)$, and $\mathbf{T}^v = (T_{ji}^{v,q}(\mathbf{W}, \mathbf{O}, \bar{\mathbf{D}}), j \in J, i \in I, q \in Q)$.

Considering the relaxed constraints, Eqs. (3), (4) and (6), and the taxi service time constraint, Eq. (15), and dropping one equation for each taxi mode $q \in Q$ in the set of flow conservation equations, the SLNE can be formulated as follows:

$$R_{1i,q}(\mathbf{W}, \mathbf{O}, \bar{\mathbf{D}}, \mathbf{c}) = (\bar{\alpha}_i^q(\mathbf{W}, \mathbf{O}, \bar{\mathbf{D}}) + c^q) W_i^q \sum_{j \in J} \sum_{p \in P} T_{ij}^{o,pq}(\mathbf{W}, \mathbf{O}, \bar{\mathbf{D}}) - \eta Z_i = 0, \quad q \in Q, \quad i \in I, \quad (40a)$$

$$R_{2i,q}(\mathbf{W}, \mathbf{O}, \bar{\mathbf{D}}, \mathbf{c}) = O_i^q - \sum_{j \in I} \sum_{p \in P} T_{ij}^{o,pq}(\mathbf{W}, \mathbf{O}, \bar{\mathbf{D}}) = 0, \quad q \in Q, \quad i \in I, \quad (40b)$$

$$R_{3j,q}(\mathbf{W}, \mathbf{O}, \bar{\mathbf{D}}, \mathbf{c}) = D_j^q - \sum_{i \in I} \sum_{p \in P} T_{ij}^{o,pq}(\mathbf{W}, \mathbf{O}, \bar{\mathbf{D}}) = 0, \quad q \in Q, \quad j \in \{J-z\}, \quad (40c)$$

$$R_{4,q}(\mathbf{W}, \mathbf{O}, \bar{\mathbf{D}}, \mathbf{c}) = \sum_{i \in I} \sum_{j \in J} \sum_{p \in P} T_{ij}^{o,pq}(\mathbf{W}, \mathbf{O}, \bar{\mathbf{D}}) h_{ij}^{o,pq}(\mathbf{W}, \mathbf{O}, \bar{\mathbf{D}}) + \sum_{j \in J} \sum_{i \in I} T_{ji}^{v,q}(\mathbf{W}, \mathbf{O}, \bar{\mathbf{D}}) \{h_{ji}^{v,q}(\mathbf{W}, \mathbf{O}, \bar{\mathbf{D}}) + \bar{\alpha}_i^q(\mathbf{W}, \mathbf{O}, \bar{\mathbf{D}}) + c^q\} - N^q = 0, \quad q \in Q. \quad (40d)$$

Denote $\mathbf{X} = \text{Col}(\mathbf{W}, \mathbf{O}, \bar{\mathbf{D}}, \mathbf{c})$ as the solution vector and $\mathbf{R} = \text{Col}(\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3, \mathbf{R}_4)$ as the column residual vector that contains all of the residues, where $\mathbf{R}_1 = \text{Col}(R_{1i,q}, i \in I, q \in Q)$, $\mathbf{R}_2 = \text{Col}(R_{2i,q}, i \in I, q \in Q)$, $\mathbf{R}_3 = \text{Col}(R_{3j,q}, j \in \{J-z\}, q \in Q)$, and $\mathbf{R}_4 = \text{Col}(R_{4,q}, q \in Q)$. The problem becomes one of finding a solution vector \mathbf{X} such that the corresponding residual vector \mathbf{R} vanishes. This is a fixed-point problem within the solution space of \mathbf{X} .

4. Solution algorithm

4.1. Algorithm for the CNEM

An extensive review of the solution algorithms for solving traffic equilibrium models is given by Patriksson (2004). To solve the VI program (Eq. (21)) for the CNEM, the block Gauss–Seidel decomposition approach coupled with the method of successive averages (MSA) is adopted (see Florian et al. (2002) for details). The conditions for convergence are that the demand and link travel cost functions are strongly monotonic and that the Jacobian matrix of the cost functions with respect to the link flow of each mode is only mildly asymmetric. This is true in this multi-class multi-mode model because the link cost functions are separable for the total flow on each link. In reality, even if it is not clear whether or not this condition is satisfied, this method can be used to find an approximate solution. Readers are referred to Wu et al. (2006) for a discussion of the conditions for convergence in a multi-class network equilibrium model.

Step 1. Initialization

Set the iteration number $r = 0$. Select an initial feasible solution $\mathbf{T}^{(r)}$ and the corresponding $\mathbf{f}^{(r)}$. Generally, we can select the initial solution that corresponds to the free-flow network.

Step 2. Computation of generalized costs

Set $r = r + 1$.

Compute the minimum generalized costs u_{ij}^ψ and then the logsums $L_{ij}^{o,p}$ and $L_{ij}^{D,p}$.

Step 3. *Computation of demand flows in the combined model*

Compute the demand flows $T_{ij}^{n,p}$, $T_{ij}^{o,p}$, and $T_{ij}^{o,p,q}$ based on the generalized costs that were obtained in Step 2. Compute the vacant taxi flow $T_{ji}^{v,q}$ by solving the doubly constrained gravity sub-model, i.e., Eq. (35) subject to Eqs. (20a), (20b) and (20g).

Step 4. *Equilibrium assignment of demand flows on the network*

With the demand matrices $\mathbf{T}^{(r)}$ obtained in Step 3 above, solve the fixed demand network equilibrium problem below to obtain the auxiliary flow pattern $\bar{\mathbf{f}}$, such that

$$\sum_{i \in I} \sum_{j \in J} \sum_{\psi \in \Psi} \sum_{r \in R_{ij}^{\psi}} C_r^{\psi}(\bar{\mathbf{f}}) (f_r^{\psi} - \bar{f}_r^{\psi}) \geq 0 \quad \forall \mathbf{f} \in \Omega(\mathbf{T})$$

subject to Eqs. (20e) and (20f).

Step 5. *Convergence test*

If $\|\bar{\mathbf{f}} - \mathbf{f}^{(r)}\| \leq \epsilon^f$, then go to Step 7.

Step 6. *Method of successive averages*

Obtain the flow pattern of the next iteration by $\mathbf{f}^{(r+1)} = \frac{r}{r+1} \mathbf{f}^{(r)} + \frac{1}{r+1} \bar{\mathbf{f}}$; return to Step 2.

Step 7. *Computation of final results*

Compute the final \mathbf{f} and \mathbf{C} , where \mathbf{C} is a collection of the path costs.

4.2. Sensitivity analysis of the CNEM

Based on the implicit function theorem and the work of Tobin and Friesz (1988), the sensitivity analysis of the equilibrium results in the CNEM with respect to the perturbation vector (small changes in the SLNE control variables) can be formulated as a system of linear equations, in which the number of equations equals the number of variables, and the perturbed CNEM results can be obtained by solving this system of equations. The variables in the system of equations include the equilibrium path flows and the OD flows for normal traffic and (occupied and vacant) taxis, the Lagrange multipliers that are associated with the hierarchical logit modal split and vacant taxi distribution, and the minimum travel cost between each OD pair.

However, in practical applications, the number of paths is enormous for large networks. This leads to several computational difficulties in the sensitivity analysis. First, the memory requirement for storing the arc/path incidence matrix for each OD pair of a large network is huge. Second, the sensitivity analysis for the network equilibrium problems that are discussed above requires inverting a very large matrix. The computational effort is extremely demanding when the number of paths in the network is large.

To overcome these difficulties we adopt a diagonalization approach, in which the travel time of each link in the network is assumed to be less sensitive to the change in the flow on that link when the network is perturbed from the equilibrium solution, i.e., the derivative of the travel time with respect to link flow is neglected from one SLNE iteration to the next. In other words, we conduct sensitivity analysis of the VI program (Eq. (21)) in the region $\forall \mathbf{T} \in \Omega(\mathbf{f})$ (with Eq. (20e) relaxed) rather than sensitivity analysis of the whole network equilibrium problem. The path flow variables, and thus the travel cost matrix, are fixed when sensitivity analysis is conducted to determine the descent direction for the next SLNE iteration. Although without proof, this sensitivity analysis at the restricted region is very useful and sufficiently accurate for the determination of the Jacobian matrix in the SLNE. This is because in the SLNE, from computational experience, the feasible region that is governed by the set of equations is largely dominated by the OD demand \mathbf{T} .

Assuming negligible changes in travel times on all links, the derivatives of the minimum travel costs for normal traffic, occupied taxi flow, and vacant taxi flow between each OD pair vanish. Thus, the derivatives of the link flow are unconstrained in the system of equations, and the derivatives of the path flow variables are no longer needed in the analysis. The derivatives of the hierarchical modal split between normal traffic and taxis and the derivatives of the vacant taxi distribution become independent of the path flow variables, and can be solved separately. In contrast, the average travel times of users in the SLNE depend on the path flow variables and the travel times along the paths. Although we make the assumption of negligible marginal travel times for all links, the average travel times between the OD pairs are not necessarily constant. It is tedious and difficult to determine the actual changes of the average travel times with respect to the perturbation parameters. Hence, we make a further approximation, that the average travel time is fixed in the iteration. This approximation is equivalent to the requirement that the change of path flows due to the change of total flows between the OD pairs is distributed among all paths on a pro-rata basis, so that the average travel times do not change.

Given the above assumptions in the diagonalization approach, the following Jacobian matrix for the SLNE can be derived by using an approach that is similar to that of Wong et al. (2002):

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} & \mathbf{B}_{13} & \mathbf{B}_{14} \\ \mathbf{B}_{21} & \mathbf{B}_{22} & \mathbf{B}_{23} & \mathbf{B}_{24} \\ \mathbf{B}_{31} & \mathbf{B}_{32} & \mathbf{B}_{33} & \mathbf{B}_{34} \\ \mathbf{B}_{41} & \mathbf{B}_{42} & \mathbf{B}_{43} & \mathbf{B}_{44} \end{bmatrix} = \begin{bmatrix} \nabla_{\mathbf{w}} \mathbf{R}_1 & \nabla_{\mathbf{o}} \mathbf{R}_1 & \nabla_{\bar{\mathbf{d}}} \mathbf{R}_1 & \nabla_{\mathbf{c}} \mathbf{R}_1 \\ \nabla_{\mathbf{w}} \mathbf{R}_2 & \nabla_{\mathbf{o}} \mathbf{R}_2 & \nabla_{\bar{\mathbf{d}}} \mathbf{R}_2 & \nabla_{\mathbf{c}} \mathbf{R}_2 \\ \nabla_{\mathbf{w}} \mathbf{R}_3 & \nabla_{\mathbf{o}} \mathbf{R}_3 & \nabla_{\bar{\mathbf{d}}} \mathbf{R}_3 & \nabla_{\mathbf{c}} \mathbf{R}_3 \\ \nabla_{\mathbf{w}} \mathbf{R}_4 & \nabla_{\mathbf{o}} \mathbf{R}_4 & \nabla_{\bar{\mathbf{d}}} \mathbf{R}_4 & \nabla_{\mathbf{c}} \mathbf{R}_4 \end{bmatrix}, \tag{41}$$

where $\mathbf{B}_{11}, \mathbf{B}_{22}$ are $|I \times Q| \times |I \times Q|$ matrices, \mathbf{B}_{33} is a $\{|J - z\} \times Q\} \times \{|J - z\} \times Q\}$ matrix, \mathbf{B}_{44} is a $|Q| \times |Q|$ matrix, and \mathbf{B} is a $\{|I + I + J\} \times Q\} \times \{|I + I + J\} \times Q\}$ matrix. The elements of the Jacobian matrix can be obtained by substituting the results of the partial sensitivity analysis of the CNEM into the direct derivative of the SLNE, with respect to the perturbation parameters in the SLNE. Explicit expressions for the elements in the Jacobian matrix are shown in [Appendix](#).

4.3. Solution procedure

This section presents a Newtonian solution procedure to solve the problem. At the first iteration and starting from an initial guess solution for the SLNE $\mathbf{X}^{(0)}$, we compute the residual vector $\mathbf{R}^{(0)}$ using the CNEM results. By applying the results of sensitivity analysis, we determine the Jacobian matrix $\mathbf{B}^{(0)}$. Let the current iteration be k . The improved solution can be obtained by

$$\mathbf{X}^{(k+1)} = \mathbf{X}^{(k)} - \lambda^{(k)} \mu^{(k)} \mathbf{A}^{(k)} \mathbf{R}^{(k)}, \tag{42}$$

where $\mathbf{A}^{(k)} = [\mathbf{B}^{(k)}]^{-1}$, $\mu^{(k)}$ is a control factor between 0 and 1 so that $\mathbf{X}^{(k)} - \lambda^{(k)} \mu^{(k)} \mathbf{A}^{(k)} \mathbf{R}^{(k)} > 0$, which guarantees that the updated control variables will not fall into the negative side, and $\lambda^{(k)}$ is the optimal step size between 0 and 1 so that the residual error is minimized. We adopt the golden section method for the line search. Here, the residual error is defined as the weighted norm of the residual vector $\mathbf{R}^{(k+1)}(\mathbf{X}^{(k)} - \lambda^{(k)} \mu^{(k)} \mathbf{A}^{(k)} \mathbf{R}^{(k)})$, where a diagonal matrix of weighting factors is adopted to normalize the different dimensions that appear in the residual vector. Denote $\underline{\omega} = \text{Diag}(\underline{\omega}_1, \underline{\omega}_2, \underline{\omega}_3, \underline{\omega}_4)$ as the weighting vector, where $\underline{\omega}_1 = \text{Diag}(1/\eta\Omega_i, i \in I, q \in Q)$, $\underline{\omega}_2 = \text{Diag}(1/\sum_{j \in J} \sum_{p \in P} \bar{D}_{ij}^p, i \in I, q \in Q)$, $\underline{\omega}_3 = \text{Diag}(1/\sum_{i \in I} \sum_{p \in P} \bar{D}_{ij}^p, j \in \{J - z\}, q \in Q)$, and $\underline{\omega}_4 = \text{Diag}(1/N^q, q \in Q)$, and $E(\mathbf{R}^{(k+1)}) = \|\underline{\omega} \cdot \mathbf{R}^{(k+1)}(\mathbf{X}^{(k)} - \lambda^{(k)} \mu^{(k)} \mathbf{A}^{(k)} \mathbf{R}^{(k)})\|$ as the total residual error. With the new residual vector, the improved solution can be determined using Eq. (42) for the next iteration. The solution process continues until $E(\mathbf{R}^{(k+1)}) < \varepsilon^R$, which is an acceptable error. The procedure can be summarized as follows:

- Step 1: Set $k = 0$. Select an initial guess solution, $\mathbf{X}^{(k)}$, and compute the residual vector $\mathbf{R}^{(k)}$ using the CNEM results.
- Step 2: Determine the Jacobian matrix $\mathbf{B}^{(k)}$ at the current solution $\mathbf{X}^{(k)}$.
- Step 3: Solve for the descent direction $\mathbf{Y}^{(k)}$ from the set of simultaneous equations $\mathbf{B}^{(k)} \mathbf{Y}^{(k)} = \mathbf{R}^{(k)}$. This is more computationally efficient than directly inverting the matrix $\mathbf{B}^{(k)}$.
- Step 4: Compute the maximum control factor $\mu^{(k)}$ in the range (0, 1) so that $\mathbf{X}^{(k)} - \mu^{(k)} \mathbf{Y}^{(k)} > 0$.
- Step 5: Perform a line search to determine the optimal step size $\lambda^{(k)}$ in the range (0, 1) so that the total residual error $E(\mathbf{R}^{(k+1)})$ is minimized, where $\mathbf{R}^{(k+1)}$ is evaluated at $\mathbf{X}^{(k+1)} = \mathbf{X}^{(k)} - \lambda^{(k)} \mu^{(k)} \mathbf{Y}^{(k)}$.
- Step 6: If $E(\mathbf{R}^{(k+1)}) < \varepsilon^R$, then stop; otherwise, set $k = k + 1$ and go to Step 2.

5. Numerical example

Consider an 8×8 square grid network with bidirectional links between each adjacent node, as shown in [Fig. 2](#). We have designed a scenario with two classes of customers and three modes of taxis. The customers are divided into a high-income group and a low-income group, where the high-income customers place a higher value on time and prefer to take luxury taxis, and the low-income customers place a lower value on time and prefer to take normal taxis. In the taxi market, there are normal taxis, luxury taxis, and restricted area taxis. Normal taxis provide the main means of transportation in the network and have a reasonable fare level; luxury taxis provide a better transportation service through less waiting time and more comfortable seats or extra facilities and have a higher fare level; and restricted area taxis provide services with a lower fare level, but pick up or set down customers only in certain areas. The normal taxis and luxury taxis can operate anywhere in the network, whereas the restricted area taxis can provide services only in the upper triangular portion of the network, as shown in [Fig. 2c](#). In the one-hour study period, the total number of customers that is generated in each zone is 1000 veh/h, of which 20% is in the high-income group and 80% is in the low-income group, with their destinations being evenly distributed among all of the zones in the network. To compare the differences between the multiple-class multiple-mode model and models in previous studies, we compute another scenario with a single user class and single taxi mode. In this scenario, the total number of customers generated in each zone is 1000 veh/h, and the taxis can travel anywhere in the network without area restrictions.

The travel impedance functions for the links are given as $t_a = t_a^0(1 + 0.5(v_a/s_a)^2)$, where the free-flow travel times of all links, t_a^0 , are 0.04 h, and the capacities, s_a , of all links are 3000 veh/h. The lengths of all links, d_a , are 3 km. The generalized link cost functions and disutility functions take the forms that are shown in Eqs. (7)–(12), and the relationship between the customer and taxi waiting times takes the form that is shown in Eq. (6). The other input data in the example are given in [Table 1](#). The parameters for the single-class single-mode scenario are simply taken to be the weighted average of the parameters in the multiple-class multiple-mode scenario.

To illustrate the numerical convergence characteristics of the solution algorithm at different congestion levels, all O–D demand matrices are multiplied by a “scaling factor” to generate various levels of demand, and hence congestion, in the network. The initial solution for the SLNE is taken as follows. For all taxi classes, the customer waiting time \mathbf{W} is taken as 0.05 h for all zones; the taxi customer demands at all origins and destinations, \mathbf{O} and $\bar{\mathbf{D}}$, respectively, are taken as the corresponding

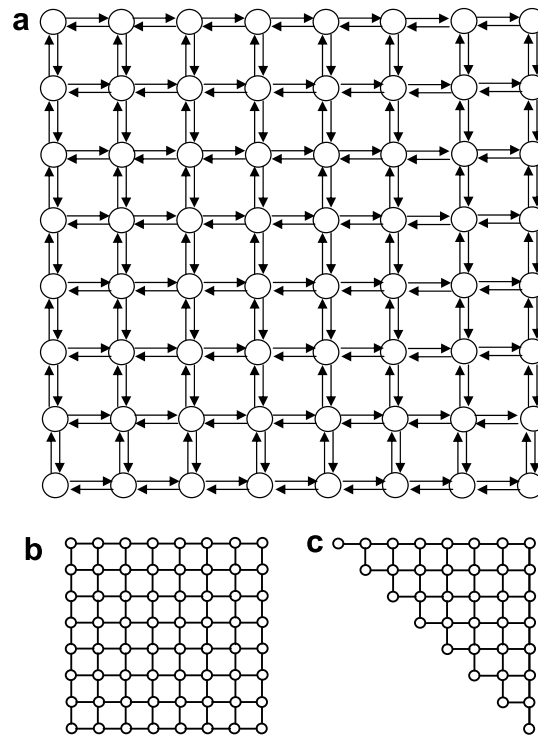


Fig. 2. The example network: (a) an 8 × 8 grid network; (b) service area for normal taxis and luxury taxis; and (c) service area for restricted area taxis.

Table 1
The input parameters for the example problems

Input parameters	Scenario 1. Two user classes and three taxi modes Values	Scenario 2. Single class single mode Values
Users' value of time	$(b_0^p, p \in P) = (100, 50)$ (\$/h)	$(b_0^p, p \in P) = (60)$ (\$/h)
Value of customer waiting time for taxis	$(b_1^p, p \in P) = (200, 100)$ (\$/h)	$(b_1^p, p \in P) = (120)$ (\$/h)
Mileage cost to a user in normal traffic	$(b^N) = 3$ (\$/km)	$(b^N) = 3$ (\$/km)
Mileage charge to a taxi customer	$(b_1^{q,a}, q \in Q) = (3, 4, 2)$ (\$/km)	$(b_1^{q,a}, q \in Q) = (3)$ (\$/km)
Congestion-based charge to a taxi customer	$(b_2^{q,a}, q \in Q) = (60, 80, 40)$ (\$/h)	$(b_2^{q,a}, q \in Q) = (60)$ (\$/h)
Bias coefficients for the preference to take a taxi	$(\rho_{pq}, p = 1, q \in Q) = (0, 40, 0)$ (\$) and $(\rho_{pq}, p = 2, q \in Q) = (20, 0, 20)$ (\$)	$(\rho_{pq}, p = P, q \in Q) = (0)$ (\$) [Not applicable in this scenario]
Hourly operating costs of taxis	$(b_3^q, q \in Q) = (80, 100, 80)$ (\$/h)	$(b_3^q, q \in Q) = (85)$ (\$/h)
Mileage operating costs of taxis	$(b^{v,q}, q \in Q) = (0.5, 0.5, 0.5)$ (\$/km)	$(b^{v,q}, q \in Q) = (0.5)$ (\$/km)
Dispersion coefficient for the upper-level logit mode choice	$(\beta_1^p, p \in P) = (0.01, 0.03)$ (1/\$)	$(\beta_1^p, p \in P) = (0.026)$ (1/\$)
Dispersion coefficient for the lower-level logit mode choice	$(\beta_2^p, p \in P) = (0.02, 0.06)$ (1/\$)	$(\beta_2^p, p \in P) = (0.052)$ (1/\$) [Not applicable in this scenario]
Dispersion coefficient for vacant taxi search behavior	$(\theta^q, q \in Q) = (0.2, 0.2, 0.2)$ (1/\$)	$(\theta^q, q \in Q) = (0.2)$ (1/\$)
Taxi fleet size	$(N^q, q \in Q) = (10,000, 5000, 5000)$ (veh)	$(N^q, q \in Q) = (20,000)$ (veh)
Parameter for the relationship of customer and taxi waiting times	$(\eta\Omega_i, i \in I) = 2$ (veh · h)	$(\eta\Omega_i, i \in I) = 2$ (veh · h)

demands that are evaluated at the free-flow travel time in the network; and the slack variable **c** is evaluated at the corresponding demand flows. The problem is solved by the sensitivity-based solution algorithm, and the solution is said to have converged when the residual error $E(\mathbf{R})$ is less than 1%. Fig. 3 shows the convergence characteristics of the problem, in which all cases at different demand levels converge quickly in a few iterations. Nevertheless, the required number of iterations increases with the level of congestion due to the approximation in the evaluation of the Jacobian matrix for the SLNE. Similar convergence characteristics are observed in all of the analyses that are discussed in the following paragraphs.

To compare the multiple-class multiple-mode scenario with the single-class single-mode scenario, the service level and performance of each taxi mode of the two scenarios are investigated with the total customer demand matrix being uniformly scaled up or down. Fig. 4 displays the average taxi utilization versus the value of the scaling factor on demand, where there

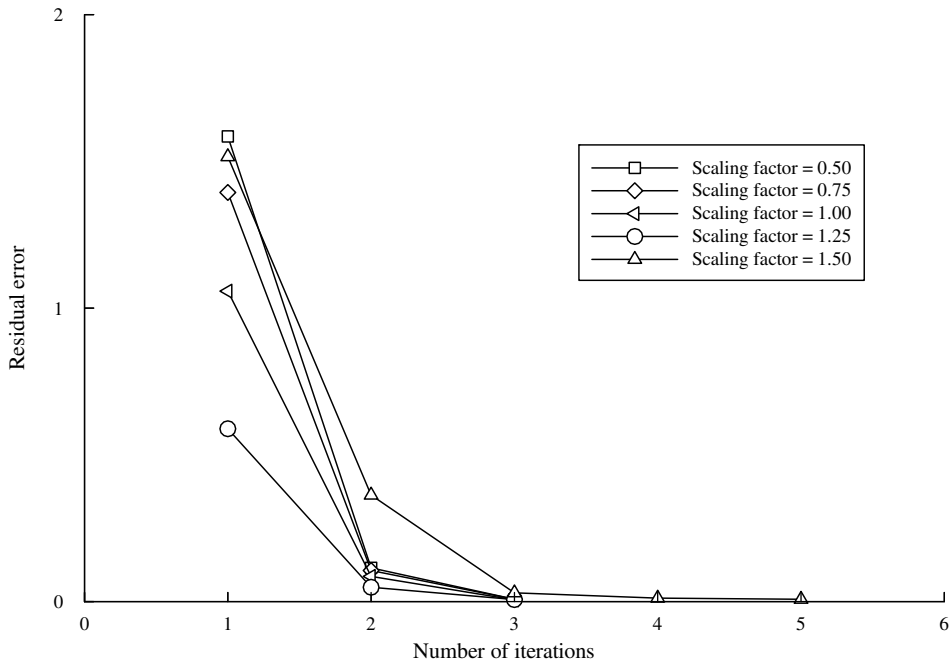


Fig. 3. The convergence characteristics of the SLNE with different scaling factors on the demand matrix.

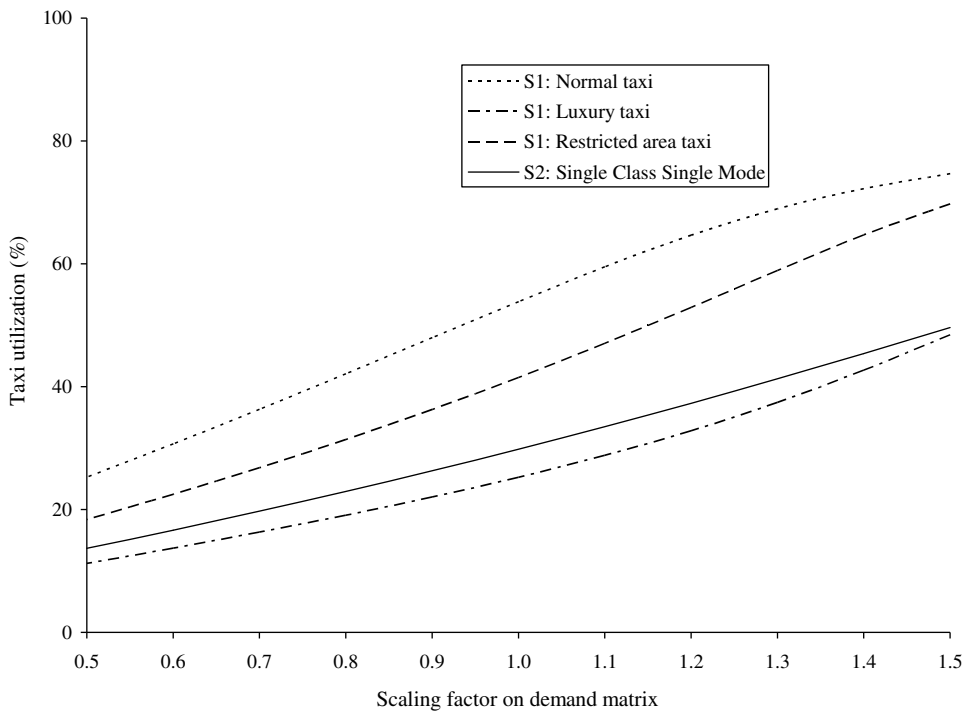


Fig. 4. Taxi utilizations against the scaling factor on the demand matrix.

are three taxi modes for scenario 1 (S1) and one taxi mode for scenario 2 (S2). The average taxi utilization is the ratio of the total occupied taxi time to the total taxi service time. For all cases, the average utilization increases with the scaling factor, and the rate of increase may be greater or less than the rate of the scaling factor because of two effects. As the scaling factor increases, the total demand for taxis increases, but traffic congestion will suppress the demand for taxis because of the congestion-based charges, hence the demand for taxis may increase or decrease, and the occupied taxi time for each customer ride should increase because of traffic congestion. For the first scenario, the utilization of luxury taxis and restricted area

taxi increases sharply with the scaling factor, but for normal taxis the utilization increases at a lower rate than the scaling factor when the scaling factor exceeds 1. This can be explained by the shift in the increased demand from normal taxis to luxury taxis, restricted area taxis, and normal traffic. Although the utilization of normal taxis increases at a lower rate than the scaling factor, it does not decrease (and is expected to approach the maximum utilization rate, or the total taxi service time), because the increasing effect on the total trip time due to increased total demand and traffic congestion should always dominate the decreasing effect because of decreased demand through traffic congestion and thus a longer customer waiting time. As discussed in Wong et al. (2001), taxi utilization should always increase with traffic congestion for a wide range of parameters. For the second scenario, the case of a single taxi mode also shows increasing utilization with the scaling on customer demand, but the first scenario can provide more insightful analysis into the multi-class nature of the problem.

Figs. 5 and 6 plot, respectively, the average waiting times of taxis and customers for each taxi mode versus the scaling factor on the total customer demand. For both scenarios, the average taxi waiting times decrease sharply with the scaling factor with an increase in the total customer demand for travel. This is because when the taxi utilization and total occupied taxi hours increase, the vacant taxi hours decrease, and thus the taxi waiting times decrease. However, the average customer waiting times increase with the scaling factor, as shown in Fig. 6. For the first scenario, the average customer waiting times of normal taxis and restricted area taxis increase more sharply with the scaling factor than does the average customer waiting time of luxury taxis. This is because most taxi service hours for normal taxis and restricted area taxis have already been occupied by customers because of their high utilization rates, and any additional occupied taxi hours will significantly decrease taxi availability and hence increase customer waiting times. The second scenario shows the customer waiting time of the single taxi model operating as that of the normal taxis of S1 but with a fleet size of 20,000, the total number of taxis in S1. As expected, the customer waiting time is relatively short compared to the customer waiting times of the taxi modes in the first scenario, as there is no area restriction and the taxi utilization is lower than that of the normal or restricted area taxis in S1.

The rest of the section focuses on the multiple-class multiple-mode scenario. To show the effect of congestion-based charges in the fare structure of taxi services, the congestion-based charges for all of the taxi modes are uniformly scaled and the mileage costs are fixed at a level, as shown in Table 1. Fig. 7 portrays the average taxi utilizations for the normal, luxury, and restricted area taxis versus the ratio of the congestion-based charge to the mileage cost. The taxi utilizations for all taxi modes generally decrease as the ratio increases, and those of the normal taxis and luxury taxis decrease more rapidly than the taxi utilization of the restricted area taxis. This is because as the ratio increases, the fare levels of the taxi services increase, and thus the demand for a taxi mode generally decreases and may shift to other modes of transport (i.e., normal traffic here) or other modes of taxis. In this example, because the congestion level of the restricted area (which is geographically remote from the city centre) is lower, the increase in the congestion-based charge has a lesser effect on the restricted area taxis. In addition, because the fare level of the restricted area taxis is generally lower, the increase in

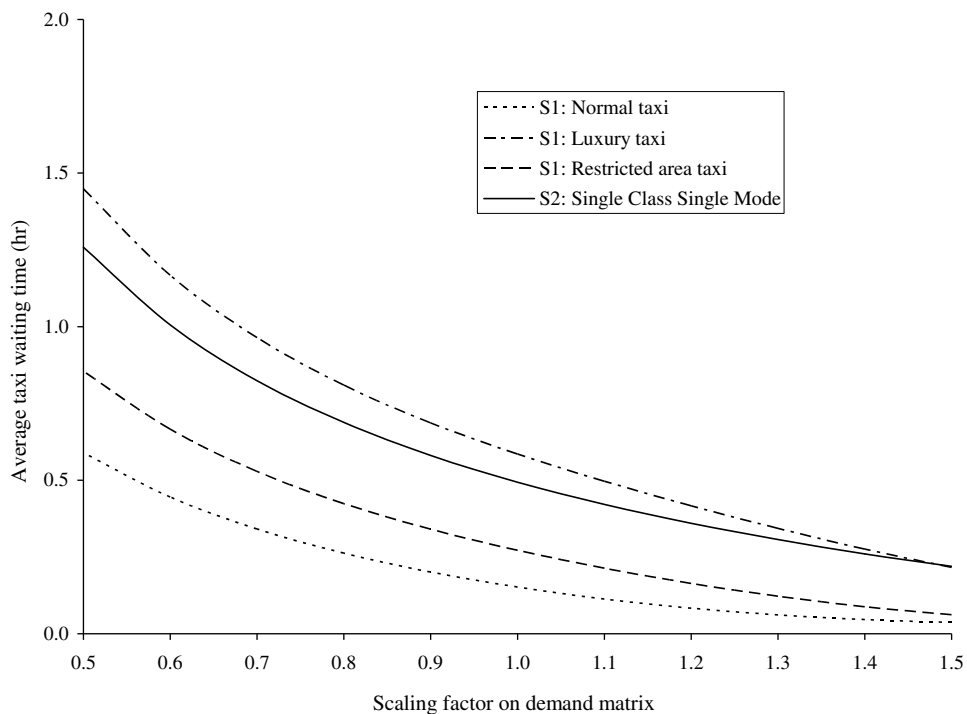


Fig. 5. Average taxi waiting times against the scaling factor on the demand matrix.

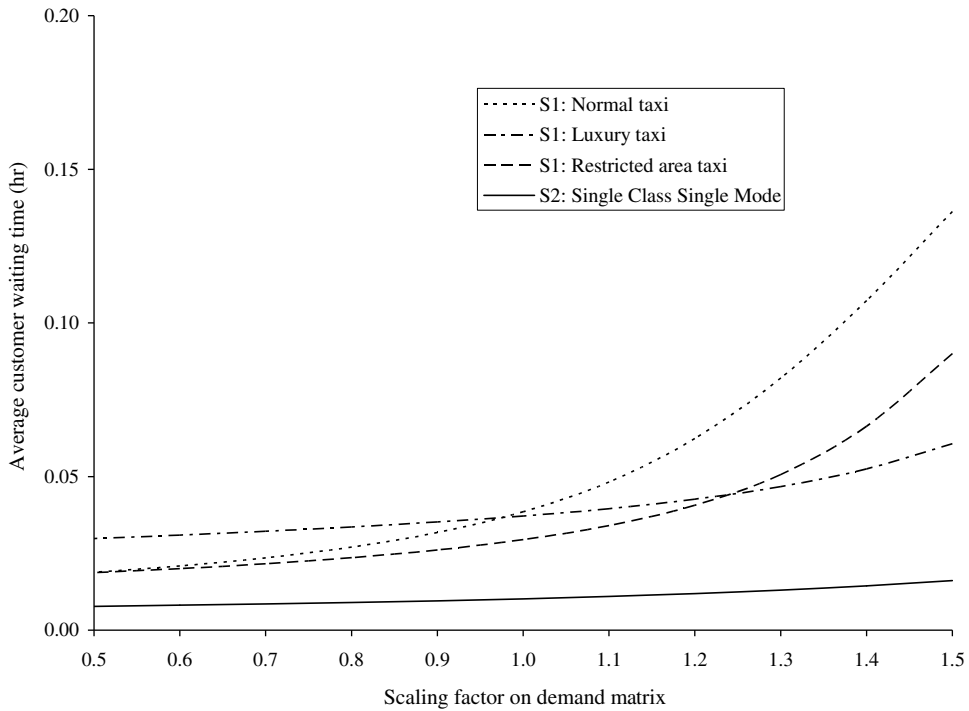


Fig. 6. Average customer waiting times against the scaling factor on the demand matrix.

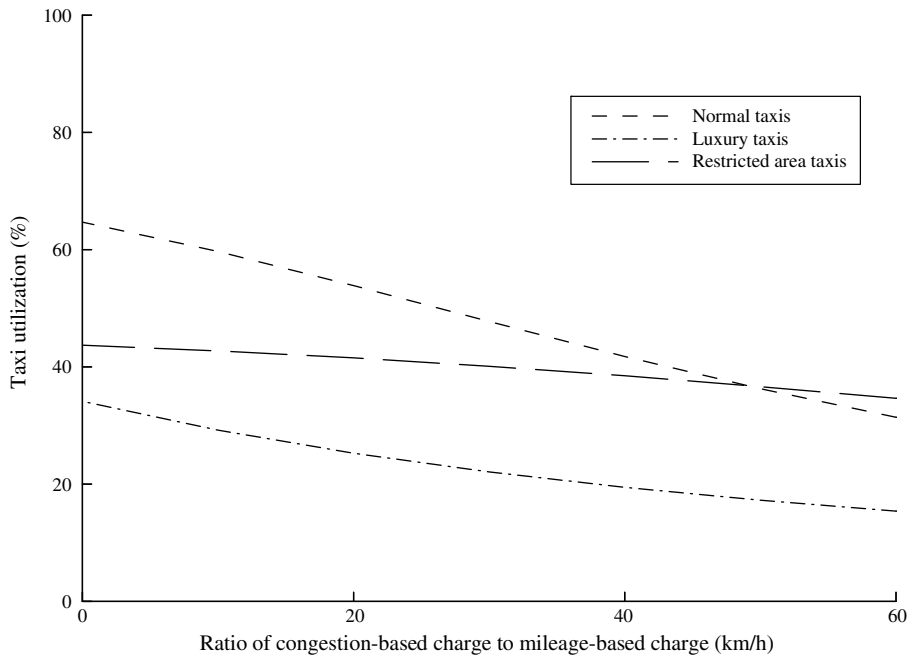


Fig. 7. Taxi utilizations against the ratio of the congestion-based charge to the mileage-based charge.

the congestion-based charge has a lower impact on this taxi mode in accordance with the hierarchical mode split function. At the point at which the ratio equals zero, the congestion in the network has no effect on the taxi charges and does not affect the split in demand between taxi modes. However, when the ratio increases beyond 50 (km/h) here (although this value is too large and unreasonable in a real-world situation), the taxi utilization, which also represents the taxi market share, of restricted area taxis becomes larger than that of normal taxis. This shows that the level of congestion-based charges in a multi-class taxi market can also play an important role in the operational performance of taxi services.

To investigate the effects of taxi fleet size and fare level on the demand for and supply of taxi services, the performance measures of the different taxi services are calculated for various numbers of taxis in service and fare levels. Figs. 8 and 9 depict the level of demands for normal traffic and taxis versus the ratio of taxi fleet sizes for luxury taxis, N2, and restricted area taxis, N3, to the taxi fleet size of normal taxis, N1. Fig. 8 shows the demands for normal traffic and taxi services as the number of luxury taxis varies. The level of demand for luxury taxis initially grows with the ratio N2/N1 because luxury taxis provide better transportation service through a larger fleet size and decrease in the customer waiting time. Therefore, luxury

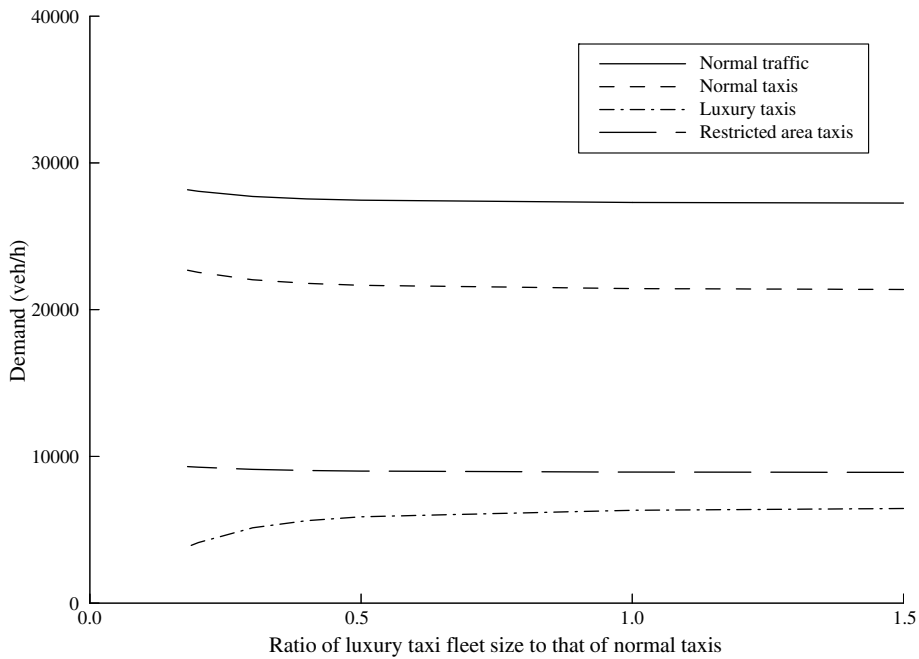


Fig. 8. Demands for normal traffic and taxis against the ratio of the luxury taxi fleet size to the normal taxi fleet size.

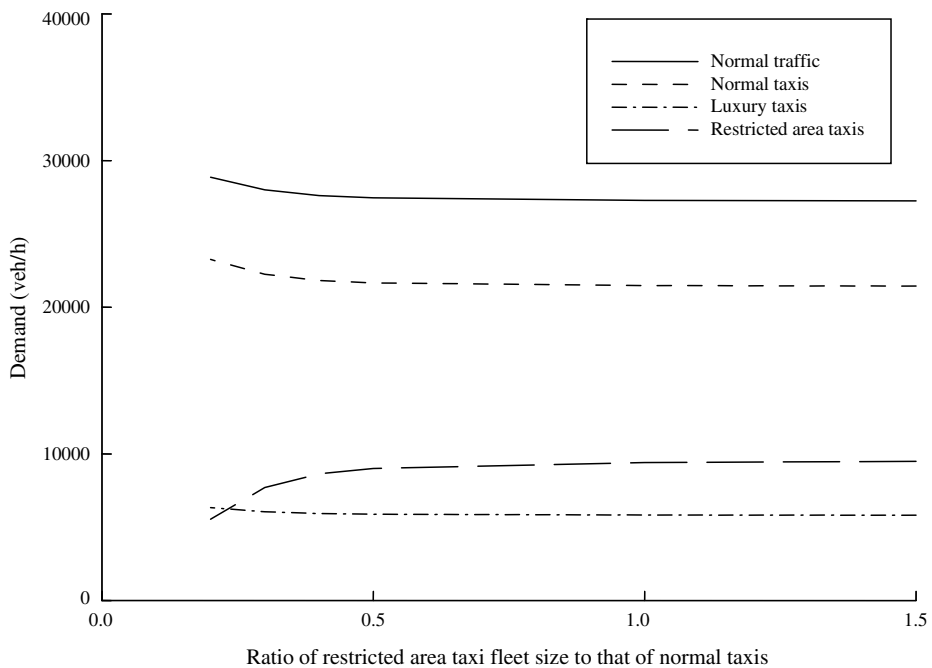


Fig. 9. Demands for normal traffic and taxis against the ratio of the restricted area taxi fleet size to the normal taxi fleet size.

taxis can attract more customers from other taxis or non-taxi traffic. The demand for luxury taxis becomes steady as the ratio exceeds a certain value (approximately 1 here) because for this example problem, most potential customers of luxury taxis have already been exploited through the provision of better service by a large fleet size, and any additional taxis that are put into service may not substantially increase the demand. A similar trend is also observed when the number of restricted area

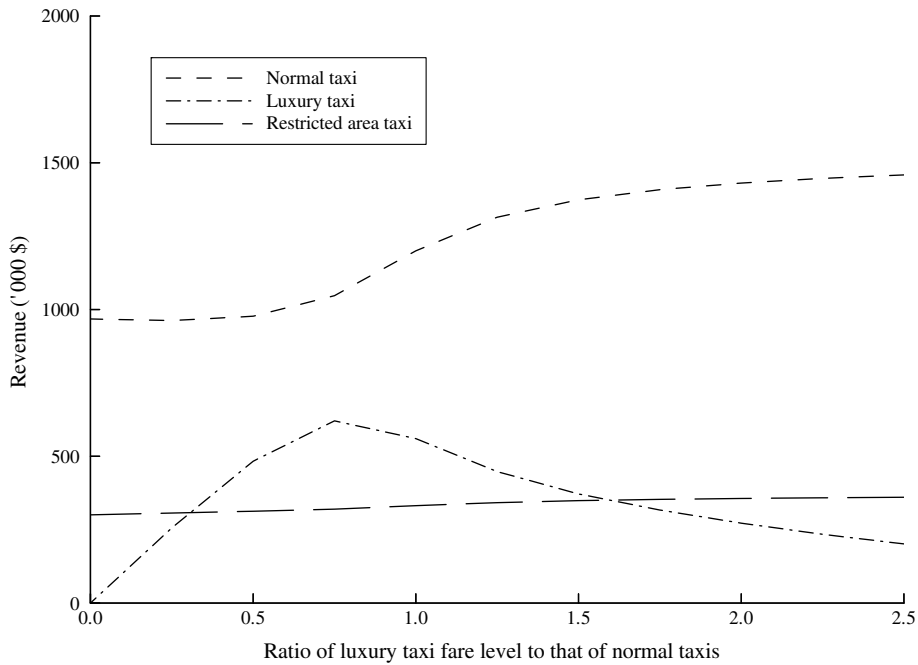


Fig. 10. Taxi revenues against the ratio of the luxury taxi fare level to the normal taxi fare level.

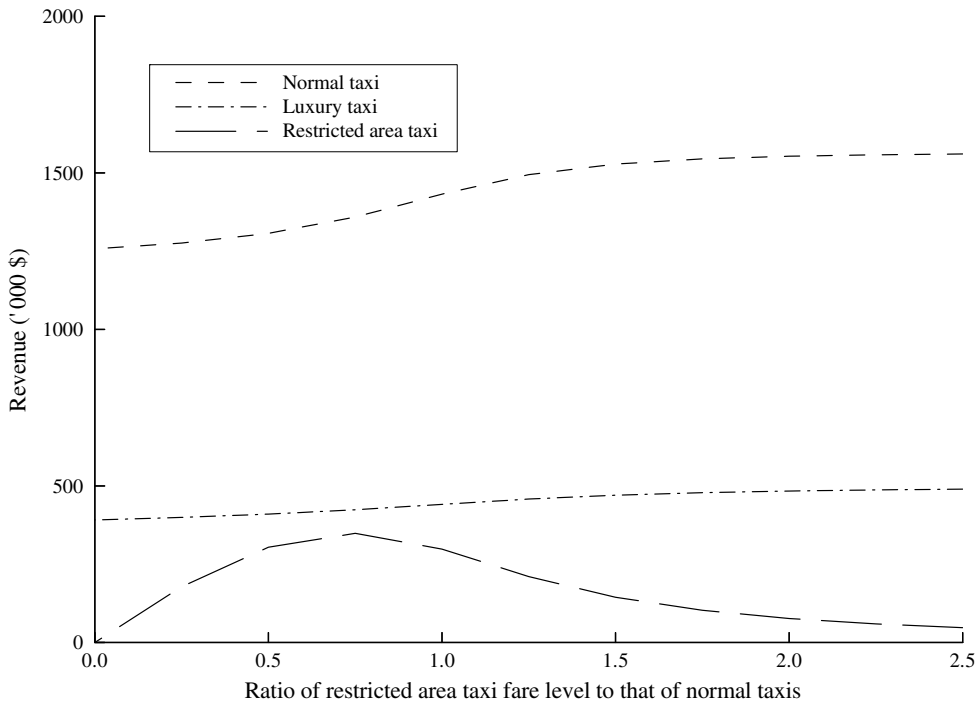


Fig. 11. Taxi revenues against the ratio of the restricted area taxi fare level to the normal taxi fare level.

taxi varies, as shown in Fig. 9. When the ratio of the restricted area taxi fleet size to the fleet size of normal taxis, $N3/N1$ exceeds 1, although the restricted area taxis provide better transportation service through a large fleet size with a lower fare level, the total level of demand for restricted area taxis becomes steady, and is much lower than that for normal taxis. This occurs because in the example network, restricted area taxis cannot pick up or set down customers in certain areas.

Figs. 10 and 11 show the revenues of taxis of various modes against the ratio of the fare levels of luxury taxis, $F2$, and restricted area taxis, $F3$, to the fare level of normal taxis, $F1$. The number of customers, the distance of each customer trip, and the fare level affect the revenues of the taxis. The revenue of luxury taxis initially increases with the corresponding fare level with the ratio $F2/F1$, up to approximately 0.75, and decreases thereafter, as shown in Fig. 10. The reason is that for this example problem, when the fare level is at the low end, the gain in revenue that is generated by increasing the fare level is greater than the loss of revenue due to the loss of customer demand. However, when the fare increases beyond a certain level, the demand decreases at a higher rate and hence the revenue drops. This observation is characterized by the logit functional form that we adopt for the hierarchical mode choice model, in which the demand initially remains at a high level when the fare is low and decays exponentially when the fare exceeds a certain level (an inverse “S” shape). As restricted area taxis cannot capture the long trips outside of their service areas that have shifted from the luxury taxis, normal taxis benefit the most and show a sharp increase in revenue when the revenue of luxury taxis drops. A similar trend is observed for the revenues of taxis against the ratio of the fare level of restricted area taxis to that of normal taxis, $F3/F1$, as shown in Fig. 11. In this case, normal taxis still derive the greatest benefit and capture the main demand shift from restricted area taxis, because of their fare competitiveness in relation to luxury taxis.

6. Conclusions

We have proposed a taxi model with multiple user classes, multiple taxi modes, and customer hierarchical modal choice in a congested road network. The contributions of this paper include the explicit consideration of multiple user classes that model user heterogeneity and multiple taxi modes that model both time-based and distance-based fare charges. The simultaneous consideration of time-based and distance-based taxi fares at the network level is original, and can more realistically model most urban taxi services, which are charged on the basis of both time and distance. This extension has important implications for modeling taxi services with service area regulations such as taxi services in Hong Kong, where rural taxis are restricted to operating in rural areas, but urban taxis can provide service anywhere.

In the simultaneous mathematical formulation of two equilibrium sub-problems, one sub-problem is a combined network equilibrium model (CNEM) that is formulated as a VI program and solved by the widely used block Gauss-Seidel decomposition approach coupled with the method of successive averages, and the other sub-problem is a set of linear and nonlinear equations (SLNE) and is solved by a Newtonian solution algorithm. In the solution algorithm, the Jacobian matrix in the SLNE is obtained as a function of the solution from the CNEM results, and is renewed at each SLNE iteration to maintain the best quality, most likely descent direction. The proposed Newtonian solution algorithm makes the taxi model applicable to large-scale networks. We formulated the CNEM as a special case of the general travel demand model so that it is possible to incorporate our taxi model into an existing package as an add-on module, in which the algorithm for the CNEM is built in practice. We have presented a numerical example to illustrate the proposed model and algorithm. Interesting results have been obtained for the example problem, and phenomena that may not hold for the abstract aggregate demand-supply model, in which all customers and taxis are assumed to have homogenous values of time and money, can be observed from our model with multiple user classes and taxi modes.

The taxi model that has been presented in this paper provides a very useful tool for the planning and evaluation of different policy options and scenarios for urban taxi services. Its ability to model multiple user classes and multiple taxi modes makes the model applicable to a wide range of taxi problems, such as the modeling of accessible taxis for providing special services to handicapped passengers, and luxury taxis with better services and facilities for affluent customers. For example, accessible taxis, which are equipped with special facilities and equipment such as wheelchairs and/or are operated by specially trained taxi drivers, can generally serve both regular customers and customers with disabilities, and disabled customers will have a high preference for them (although they may still use normal taxis). It is possible that those taxis will charge at a higher fare level to ensure availability, and that handicapped passengers, in some cities, may receive various forms of transportation subsidization from the government or charity organizations. In the case of luxury taxis, more affluent customers (such as those on business trips or tourists) will prefer to take a luxury taxi, which is equipped with better seats or entertainment facilities, rather than a normal taxi, despite the higher fare. This can be reflected in the model by adopting a higher value in the corresponding inertia parameter. The values of the inertia parameters can be obtained by the calibration of the cost functions, or user-specified according to the situation.

Acknowledgements

The work that is described in this paper was jointly supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project Nos: HKU7019/99E and HKUST6212/07E), the William Mong Young Researcher Award in Engineering 2002–2003 from the University of Hong Kong, and the National Basic Research Program of China (2006CB705503).

Appendix A1. Nomenclature

Summarized below is the notation used in the paper.

V	set of nodes
A	set of links
I	set of origin zones
J	set of destination zones
P	set of user classes
Q	set of taxi modes
\bar{D}_{ij}^p	total customer demand of user class $p \in P$ from origin zone $i \in I$ to destination zone $j \in J$
$T_{ij}^{n,p}$	normal (non-taxi) traffic movements of user class $p \in P$ from origin zone $i \in I$ to destination zone $j \in J$
$T_{ij}^{o,p}$	occupied taxi movements of user class $p \in P$ from origin zone $i \in I$ to destination zone $j \in J$
$T_{ij}^{o,pq}$	occupied taxi movements of user class $p \in P$ from origin zone $i \in I$ to destination zone $j \in J$ who choose taxis of mode $q \in Q$
$T_{ji}^{v,q}$	vacant taxi movements of mode $q \in Q$ from origin zone $i \in I$ to destination zone $j \in J$
O_i^q	total customer demand from origin zone $i \in I$ for taxis of mode $q \in Q$
D_j^q	total customer demand to destination zone $j \in J$ for taxis of mode $q \in Q$
Ψ	combination of user classes and transportation modes with $\Psi = ((n,p), (o,pq), (v,q))$
R	set of all routes
R_{ij}^ψ	set of shortest paths for the demand for class-mode combination $\psi \in \Psi$ from origin zone $i \in I$ to destination zone $j \in J$
f_r^ψ	traffic flow on route $r \in R_{ij}^\psi$.
$\delta_{ij,ar}^\psi$	link route incidence matrix for route $r \in R_{ij}^\psi$ on link $a \in A$, where $\delta_{ij,ar}^\psi = 1$ if route r uses link a , and 0 otherwise
v_a^ψ	vehicular flow for class-mode combination $\psi \in \Psi$ on link $a \in A$
v_a	total vehicular flow on link $a \in A$
s_a	practical capacity on link $a \in A$
d_a	length of link $a \in A$
t_a	travel time on link $a \in A$
t_a^0	free-flow travel time on link $a \in A$
h_{ij}^ψ	average travel time from origin zone $i \in I$ to destination zone $j \in J$ by class-mode combination $\psi \in \Psi$
w_i^q	taxi waiting/search time of mode $q \in Q$ at zone $i \in I$
W_i^q	customer waiting time for taxis of mode $q \in Q$ at zone $i \in I$
N^q	fleet size of taxis of mode $q \in Q$ operating in the network
$C_r^{n,p}$	total generalized cost of travel (disutility) for users of class $p \in P$ who are taking normal traffic from zone $i \in I$ to zone $j \in J$ on path $r \in R_{ij}^{n,p}$
$C_r^{o,pq}$	total generalized cost of travel (disutility) for users of class $p \in P$ who are taking taxis of mode q from zone i to zone j on path $r \in R_{ij}^{o,pq}$
$C_r^{v,q}$	total searching cost for a vacant taxi of mode q on path $r \in R_{ij}^{v,q}$
C_{ij}^ψ	minimum generalized cost from zone $i \in I$ to zone $j \in J$ by class-mode combination $\psi \in \Psi$
c_a^ψ	generalized cost of travel on link $a \in A$ for the corresponding class-mode combination $\psi \in \Psi$
b_0^p	value of time of users in class $p \in P$
b^n	mileage costs that are charged to customers who are taking normal traffic
$b_1^{o,q}$	mileage costs charged to customers who are taking a taxi of mode $q \in Q$
$b_2^{o,q}$	congestion-based costs charged to customers who are taking a taxi of mode $q \in Q$
b_0^q	hourly operating cost of a taxi of mode $q \in Q$.
$b^{v,q}$	mileage operating cost of a taxi of mode $q \in Q$.
p_a^q	cost equivalent user-specified charge applied to taxis of mode q entering link $a \in A$
b_1^p	value of customer waiting time (for taxis) as perceived by user class $p \in P$
ρ_{pq}	inherent inertia, which reflects that users of class $p \in P$ prefer taxis of mode $q \in Q$ in relation to normal traffic
F_r^{pq}	fare for taking a taxi of mode $q \in Q$ that is traveling from zone $i \in I$ to zone $j \in J$ along path $r \in R_{ij}^{o,pq}$ by a user of class $p \in P$
$p_{ij}^{n,p}$	probability that a user of class $p \in P$ traveling from zone $i \in I$ to zone $j \in J$ chooses normal traffic
$p_{ij}^{o,p}$	probability that a user of class $p \in P$ traveling from zone $i \in I$ to zone $j \in J$ chooses to take a taxi
$p_{ij}^{o,pq}$	probability that a user of class $p \in P$ traveling from zone $i \in I$ to zone $j \in J$ chooses to take a taxi of mode $q \in Q$
$p_{i/j}^{v,q}$	probability that a vacant taxi of mode $q \in Q$ originating at zone $j \in J$ eventually meets a customer at zone $i \in I$
β_1^p	dispersion coefficient for the upper-level logit mode choice model for user-class $p \in P$
β_2^p	dispersion coefficient for the lower-level logit mode choice model for user-class $p \in P$
θ^q	non-negative model parameter for taxis of mode $q \in Q$ in reflecting the degree of uncertainty in searching for customers
η	model parameter for the customer waiting time function

Z_i	area of zone $i \in I$
α_i^q	Lagrange multiplier associated with the customer demand constraint for taxi mode $q \in Q$ at origin zone $i \in I$
β_j^q	Lagrange multiplier associated with the customer demand constraint for taxi mode $q \in Q$ at destination zone $j \in J$
$L_{ij}^{o,p}$	Lagrange multiplier associated with the conservation constraint of demand for taxis of user class $p \in P$ from zone $i \in I$ to zone $j \in J$; it is also the logsum of the disutility of travel as perceived by users of class $p \in P$ who are using taxis as their mode of transportation from zone $i \in I$ to zone $j \in J$
$I_{ij}^{D,p}$	Lagrange multiplier associated with the conservation constraint of the total customer demand for travel of user class $p \in P$ from zone $i \in I$ to zone $j \in J$
u_{ij}^ψ	Lagrange multiplier associated with the flow conservation constraint from origin zone $i \in I$ to destination zone $j \in J$ of class-mode combination $\psi \in \Psi$
ε_r^ψ	Lagrange multiplier associated with the flow non-negativity constraint from origin zone $i \in I$ to destination zone $j \in J$ of class-mode combination $\psi \in \Psi$
$\bar{\alpha}_i^q$	normalized value of Lagrange multiplier α_i^q
c^q	slack variable in calculating the taxi waiting times of taxi mode $q \in Q$
w	vector of taxi waiting/search times
W	vector of customer waiting times
O	vector of total customer demand from origin zones
D	vector of total customer demand to destination zones
$\bar{\mathbf{D}}$	vector of total customer demand to destination zones, excluding an arbitrarily chosen zone
c	vector of slack variables in calculating the taxi waiting times
$\bar{\alpha}$	vector of normalized Lagrange multipliers
h^o	vector of occupied taxi travel times
h^v	vector of vacant taxi travel times
T^o	vector of occupied taxi movements
T^v	vector of vacant taxi movements
$R_{1,i,q}$	residual value associated with the waiting time relationship of taxi mode $q \in Q$ at zone $i \in I$
$R_{2i,q}$	residual value associated with the constraint of total customer demand for taxi mode $q \in Q$ at origin zone $i \in I$
$R_{3j,q}$	residual value associated with the constraint of total customer demand for taxi mode $q \in Q$ at destination zone $j \in J$
$R_{4,q}$	residual value associated with the taxi service time constraint of taxi mode $q \in Q$
X	solution vector
R	residual vector
B	Jacobian matrix
A	inverse of Jacobian matrix
f	vector of the flow pattern
μ	control factor to ensure non-negative waiting times and demands
λ	optimal step size of line search
$\underline{\omega}$	vector of weighting factors to the residual values
$E(\mathbf{R})$	total residual error
ε^f	acceptable error of the flow pattern in the CNEM
ε^R	acceptable error of the total residual value in the SLNE

Appendix A2. Explicit expressions for the Jacobian matrix of the SLNE

This appendix shows the explicit expressions for the Jacobian matrix of the SLNE, given all of the assumptions in the diagonalization approach that were discussed in Section 4.2.

Let the perturbation parameters for the customer waiting time at origin W_i^q , the total demand from origin O_i^q , and the total demand to destination D_j^q be $\underline{\varepsilon}^W = \text{Col}(\varepsilon_i^{W,q}, i \in I, q \in Q)$, $\underline{\varepsilon}^O = \text{Col}(\varepsilon_i^{O,q}, i \in I, q \in Q)$, and $\underline{\varepsilon}^D = \text{Col}(\varepsilon_j^{D,q}, j \in \{J - z\}, q \in Q)$, respectively. Denote the vector of perturbation parameters as $\underline{\varepsilon} = \text{Col}(\underline{\varepsilon}^W, \underline{\varepsilon}^O, \underline{\varepsilon}^D)$. For each taxi class $q \in Q$, we can form the following equation in a matrix form, which is derived from Eqs. (20a), (20b) and (26), assuming that $u_{ji}^{v,q}$ is constant throughout the iteration,

$$-\theta^q \begin{bmatrix} \sum_{j \in J} T_{j,1}^{v,q*} & 0 & T_{1,1}^{v,q*} & \cdots & T_{|J-z|,1}^{v,q*} \\ & \ddots & \vdots & & \vdots \\ 0 & \sum_{j \in J} T_{j,|I|}^{v,q*} & T_{1,|I|}^{v,q*} & \cdots & T_{|J-z|,|I|}^{v,q*} \\ T_{1,1}^{v,q*} & \cdots & T_{1,|I|}^{v,q*} & \sum_{i \in I} T_{1,i}^{v,q*} & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ T_{|J-z|,1}^{v,q*} & \cdots & T_{|J-z|,|I|}^{v,q*} & 0 & \sum_{i \in I} T_{|J-z|,i}^{v,q*} \end{bmatrix} \begin{bmatrix} \nabla_{\underline{\varepsilon}^q} \\ \vdots \\ \nabla_{\underline{\varepsilon}^q} \alpha_{|I|}^q \\ \nabla_{\underline{\varepsilon}^q} \beta_1^q \\ \vdots \\ \nabla_{\underline{\varepsilon}^q} \beta_{|J-z|}^q \end{bmatrix} = \begin{bmatrix} \nabla_{\underline{\varepsilon}}(O_1^q) \\ \vdots \\ \nabla_{\underline{\varepsilon}}(O_{|I|}^q) \\ \nabla_{\underline{\varepsilon}}(D_1^q) \\ \vdots \\ \nabla_{\underline{\varepsilon}}(D_{|J-z|}^q) \end{bmatrix}, \tag{43}$$

where $|I|$ and $|J|$ are the sizes of the sets of origins and destinations, respectively, and α_i^q and β_j^q are the Lagrange multipliers associated with constraints Eqs. (20b) and (20c), respectively. Partition the matrix on the left-hand side of the system of Eq.

(43) (including the common factor of $-\theta^q$) as $\mathbf{M}^q = \begin{bmatrix} \mathbf{P}_{11}^q & \mathbf{P}_{12}^q \\ \mathbf{P}_{21}^q & \mathbf{P}_{22}^q \end{bmatrix}$, $q \in Q$, where \mathbf{P}_{11}^q is an $|I| \times |I|$ diagonal matrix and \mathbf{P}_{22}^q is an $|J - z| \times |J - z|$ diagonal matrix. Let $[\mathbf{M}^q]^{-1} = \begin{bmatrix} \mathbf{Q}_{11}^q & \mathbf{Q}_{12}^q \\ \mathbf{Q}_{21}^q & \mathbf{Q}_{22}^q \end{bmatrix}$, $q \in Q$, then this becomes the sensitivity of α_i^q and β_j^q with respect to perturbation parameters ε^O and ε^D .

Each element matrix in the Jacobian matrix (Eq. (41)) can be further divided into matrices for each taxi class $q \in Q$ in the residual vector and $n \in Q$ in the perturbation parameters, so that $\mathbf{B}_{11}^{q,n}, \mathbf{B}_{22}^{q,n}$ are $|I| \times |I|$ matrices, $\mathbf{B}_{33}^{q,n}$ is a $|J - z| \times |J - z|$ matrix, and $\mathbf{B}_{44}^{q,n}$ is a 1×1 element, etc., where

$$\mathbf{B}_{lm} = \begin{bmatrix} \ddots & \vdots & \ddots \\ \cdots & \mathbf{B}_{lm}^{q,n} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}. \tag{44}$$

Recall that $\bar{\alpha}_i^q = \alpha_i^q / b_0^q$. We can show that

$$\mathbf{B}_{11}^{q,n}[i, k] = \frac{\partial R_{1i,q}}{\partial W_k^n} = \begin{cases} (\bar{\alpha}_i^q + c^q) \sum_{j \in J} \sum_{p \in P} T_{ij}^{o,pq} + (\bar{\alpha}_i^q + c^q) W_i^q \sum_{j \in J} \sum_{p \in P} S_{ij}^{pq}, & \text{if } k = i, n = q, \\ (\bar{\alpha}_i^q + c^q) W_i^q \sum_{j \in J} \sum_{p \in P} \bar{S}_{ij}^{pqn}, & \text{if } k = i, n \neq q, \quad i \in I, q \in Q, k \in I, n \in Q, \\ 0, & \text{if } k \neq i, \end{cases} \tag{45}$$

$$\mathbf{B}_{12}^{q,n}[i, k] = \frac{\partial R_{1i,q}}{\partial O_k^n} = \begin{cases} \frac{1}{b_0^q} \mathbf{Q}_{11}^q[i, k] W_i^q \sum_{j \in J} \sum_{p \in P} T_{ij}^{o,pq}, & \text{if } n = q, \\ 0, & \text{if } n \neq q, \end{cases} \quad i \in I, q \in Q, k \in I, n \in Q, \tag{46}$$

$$\mathbf{B}_{13}^{q,n}[i, k] = \frac{\partial R_{1i,q}}{\partial D_k^n} = \begin{cases} \frac{1}{b_0^q} \mathbf{Q}_{12}^q[i, k] W_i^q \sum_{j \in J} \sum_{p \in P} T_{ij}^{o,pq}, & \text{if } n = q, \\ 0, & \text{if } n \neq q, \end{cases} \quad i \in I, q \in Q, k \in \{J - z\}, n \in Q, \tag{47}$$

$$\mathbf{B}_{14}^{q,n}[i] = \frac{\partial R_{1i,q}}{\partial c^n} = \begin{cases} W_i^q \sum_{j \in J} \sum_{p \in P} T_{ij}^{o,pq}, & \text{if } n = q, \\ 0, & \text{if } n \neq q, \end{cases} \quad i \in I, q \in Q, n \in Q, \tag{48}$$

$$\mathbf{B}_{21}^{q,n}[i, k] = \frac{\partial R_{2i,q}}{\partial W_k^n} = \begin{cases} - \sum_{j \in J} \sum_{p \in P} S_{ij}^{pq}, & k = i, n = q, \\ - \sum_{j \in J} \sum_{p \in P} \bar{S}_{ij}^{pqn}, & k = i, n \neq q, \quad i \in I, q \in Q, k \in I, n \in Q, \\ 0, & k \neq i, \end{cases} \tag{49}$$

$$\mathbf{B}_{22}^{q,n}[i, k] = \frac{\partial R_{2i,q}}{\partial O_k^n} = \begin{cases} 1, & \text{if } k = i, n = q, \\ 0, & \text{otherwise,} \end{cases} \quad i \in I, q \in Q, k \in I, n \in Q, \tag{50}$$

$$\mathbf{B}_{23}^{q,n}[i, k] = \frac{\partial R_{2i,q}}{\partial D_k^n} = 0, \quad i \in I, q \in Q, k \in \{J - z\}, n \in Q, \tag{51}$$

$$\mathbf{B}_{24}^{q,n}[i] = \frac{\partial R_{2i,q}}{\partial c^n} = 0, \quad i \in I, q \in Q, n \in Q, \tag{52}$$

$$\mathbf{B}_{31}^{q,n}[j, k] = \frac{\partial R_{3j,q}}{\partial W_k^n} = \begin{cases} - \sum_{p \in P} S_{kj}^{pq}, & \text{if } n = q, \\ - \sum_{p \in P} \bar{S}_{kj}^{pqn}, & \text{if } n \neq q, \quad j \in \{J - z\}, q \in Q, k \in I, n \in Q, \end{cases} \tag{53}$$

$$\mathbf{B}_{32}^{q,n}[j, k] = \frac{\partial R_{3j,q}}{\partial O_k^n} = 0, \quad j \in \{J - z\}, q \in Q, k \in I, n \in Q, \tag{54}$$

$$\mathbf{B}_{33}^{q,n}[j, k] = \frac{\partial R_{3j,q}}{\partial D_k^n} = \begin{cases} 1, & \text{if } k = j, n = q, \\ 0, & \text{otherwise,} \end{cases} \quad j \in \{J - z\}, q \in Q, k \in \{J - z\}, n \in Q, \tag{55}$$

$$\mathbf{B}_{34}^{q,n}[j] = \frac{\partial R_{3j,q}}{\partial c^n} = 0, \quad j \in \{J - z\}, q \in Q, n \in Q, \tag{56}$$

$$\mathbf{B}_{41}^{q,n}[k] = \frac{\partial R_{4,q}}{\partial W_k^n} = \begin{cases} \sum_{j \in J} \sum_{p \in P} \{S_{kj}^{pq} h_{kj}^{o,pq}\}, & \text{if } n = q, \\ \sum_{j \in J} \sum_{p \in P} \{\bar{S}_{kj}^{pqn} h_{kj}^{o,pq}\}, & \text{if } n \neq q, \end{cases} \quad q \in Q, k \in I, n \in Q, \tag{57}$$

$$\mathbf{B}_{42}^{q,n}[k] = \frac{\partial R_{4,q}}{\partial O_k^n} = \begin{cases} \sum_{j \in J} \sum_{i \in I} \{-\theta^q T_{ji}^{v,q} \mathbf{Q}_{11}^q [i, k] \{h_{ji}^{v,q} + \bar{\alpha}_i^q + c^q\}\} \\ + \sum_{j \in \{J-z\}} \sum_{i \in I} \{-\theta^q T_{ji}^{v,q} \mathbf{Q}_{21}^q [j, k] \{h_{ji}^{v,q} + \bar{\alpha}_i^q + c^q\}\} \\ + \sum_{j \in J} \sum_{i \in I} \{T_{ji}^{v,q} \frac{1}{b_0^q} \mathbf{Q}_{11}^q [i, k]\}, & \text{if } n = q, \quad q \in Q, \quad k \in I, \quad n \in Q, \\ 0, & \text{if } n \neq q, \end{cases} \quad (58)$$

$$\mathbf{B}_{43}^{q,n}[k] = \frac{\partial R_{4,q}}{\partial D_k^n} = \begin{cases} \sum_{j \in J} \sum_{i \in I} \{-\theta^q T_{ji}^{v,q} \mathbf{Q}_{12}^q [i, k] \{h_{ji}^{v,q} + \bar{\alpha}_i^q + c^q\}\} \\ + \sum_{j \in \{J-z\}} \sum_{i \in I} \{-\theta^q T_{ji}^{v,q} \mathbf{Q}_{22}^q [j, k] \{h_{ji}^{v,q} + \bar{\alpha}_i^q + c^q\}\} \\ + \sum_{j \in J} \sum_{i \in I} \{T_{ji}^{v,q} \frac{1}{b_0^q} \mathbf{Q}_{12}^q [i, k]\}, & \text{if } n = q, \quad q \in Q, \quad k \in \{J-z\}, \quad n \in Q, \\ 0, & \text{if } n \neq q, \end{cases} \quad (59)$$

$$\mathbf{B}_{44}^{q,n} = \frac{\partial R_{4,q}}{\partial c^n} = \begin{cases} \sum_{j \in J} \sum_{i \in I} T_{ji}^{v,q}, & \text{if } n = q, \\ 0, & \text{if } n \neq q, \end{cases} \quad q \in Q, \quad n \in Q, \quad (60)$$

where $\mathbf{B}_{lm}^{q,n}[j, k]$ is the j th row and k th column element of the matrix $\mathbf{B}_{lm}^{q,n}$:

$$S_{ij}^{pq} = b_1^p T_{ij}^{o,pq} \left(-\beta_2^p + \beta_2^p \frac{T_{ij}^{o,pq}}{T_{ij}^{o,p}} - \beta_1^p \frac{T_{ij}^{n,p} T_{ij}^{o,pq}}{D_{ij}^p T_{ij}^{o,p}} \right), \quad i \in I, \quad j \in J, \quad p \in P, \quad q \in Q, \quad (61)$$

$$\bar{S}_{ij}^{pqn} = b_1^p T_{ij}^{o,pq} \left(\beta_2^p \frac{T_{ij}^{o,pn}}{T_{ij}^{o,p}} - \beta_1^p \frac{T_{ij}^{n,p} T_{ij}^{o,pn}}{D_{ij}^p T_{ij}^{o,p}} \right), \quad i \in I, \quad j \in J, \quad p \in P, \quad q \in Q, \quad n \in Q \quad (62)$$

and the sub-matrices \mathbf{Q}_{11}^q , \mathbf{Q}_{12}^q , \mathbf{Q}_{21}^q , and \mathbf{Q}_{22}^q are obtained by inverting the matrix \mathbf{M}^q . However, as the two principal sub-matrices, \mathbf{P}_{11}^q and \mathbf{P}_{22}^q , are diagonal matrices, the inversion is very efficient. In fact, Eqs. (61) and (62) are the sensitivity of the demand matrices $T_{ij}^{o,pq}$ with respect to the perturbation parameter $\underline{\varepsilon}^W$.

References

- Abe, M.A., Brush, B.C., 1976. On the regulation of price and service quality: the taxicab problem. *Quarterly Review of Economics and Business* 16 (3), 105–111.
- Arnott, R., 1996. Taxi travel should be subsidized. *Journal of Urban Economics* 40 (3), 316–333.
- Beesley, M.E., Glaister, S., 1983. Information for regulation: the case of taxis. *The Economic Journal* 93 (371), 594–615.
- Boyce, D., Bar-Gera, H., 2001. Network equilibrium models of travel choices with multiple classes. In: Lahr, M.L., Miller, R.E. (Eds.), *Regional Science Perspectives in Economic Analysis*. Elsevier Science, Amsterdam, pp. 85–98.
- Boyce, D., Bar-Gera, H., 2004. Multiclass combined models for urban travel forecasting. *Network and Spatial Economics* 4 (1), 115–124.
- Cairns, R.D., Liston-Heyes, C., 1996. Competition and regulation in the taxi industry. *Journal of Public Economics* 59 (1), 1–15.
- Dafermos, S.C., 1982. The general multimodal network equilibrium problem with elastic demands. *Networks* 12 (1), 57–72.
- de Cea, J., Fernandez, J.E., 2001. ESTRAUS: a simulation equilibrium model to analyze and evaluate multimodal urban transportation systems with multiple user classes. In: *Proceedings of the Ninth World Conference on Transport Research*. Seoul, Korea.
- De Vany, A.S., 1975. Capacity utilization under alternative regulatory constraints: an analysis of taxi markets. *Journal of Political Economy* 83 (1), 83–94.
- Douglas, G.W., 1972. Price regulation and optimal service standards: the taxicab industry. *Journal of Transport Economics and Policy* 6 (2), 116–127.
- Florian, M., Wu, J.H., He, S.G., 2002. A multi-class multi-mode variable demand network equilibrium model with hierarchical logit structures. In: Marcotte, P., Gendreau, M. (Eds.), *Transportation and Network Analysis: Current Trends: Miscellanea in Honor of Michael Florian*. Kluwer Academic, London, pp. 119–133.
- Foerster, J.F., Gilbert, G., 1979. Taxicab deregulation: economic consequences and regulatory choices. *Transportation* 8 (4), 371–387.
- Frankena, M.W., Pautler, P.A., 1986. Taxicab regulation: an economic analysis. *Research in Law and Economics* 9, 129–165.
- Hong Kong Transport Advisory Committee, 1992. Consultative paper on taxi policy review. Hong Kong Transport Advisory Committee, Hong Kong.
- Hong Kong Transport Advisory Committee, 1998. Review of taxi licensing system. Hong Kong Transport Advisory Committee, Hong Kong.
- Lam, W.H.K., Huang, H.J., 1992. A combined trip distribution and assignment model for multiple user classes. *Transportation Research Part B* 26 (4), 275–287.
- Loo, B.P.Y., Chow, A.S.Y., 2008. Changing urban form in Hong Kong: what are the challenges on sustainable transportation? *International Journal of Sustainable Transportation* 2 (3), 177–193.
- Loo, B.P.Y., Leung, B.S.Y., Wong, S.C., Yang, H., 2007. Taxi license premiums in Hong Kong: can their fluctuations be explained by taxi as a mode of public transport? *International Journal of Sustainable Transportation* 1 (4), 249–266.
- Manski, C.F., Wright, J.D., 1976. Nature of equilibrium in the market for taxi services. *Transportation Research Record* 619, 296–306.
- Oppenheim, N., 1995. *Urban Travel Demand Modeling*. John Wiley & Sons, New York.
- Ortuzar, J.D., Willumsen, L.G., 1996. *Modelling Transport*, second ed. John Wiley & Sons, New York.
- Patriksson, M., 2004. Algorithms for computing traffic equilibria. *Networks and Spatial Economics* 4 (1), 23–38.
- Schroeter, J.R., 1983. A model of taxi service under fare structure and fleet size regulation. *Bell Journal of Economics* 14 (1), 81–96.
- Sheffi, Y., 1985. *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*. Prentice-Hall, New Jersey.
- Shreiber, C., 1975. The economic reasons for price and entry regulation of taxicabs. *Journal of Transport Economics and Policy* 9 (3), 268–279.
- Tobin, R.L., Friesz, T.L., 1988. Sensitivity analysis for equilibrium network flows. *Transportation Science* 22 (4), 242–250.
- Transport Department, 1986–1998. The level of taxi services. TTSD Publication Series, Hong Kong Government, Hong Kong.
- 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 (9), 819–842.
- 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*. Elsevier Science, The Netherlands, pp. 23–42.

- Wong, K.I., Wong, S.C., Wu, J.H., Yang, H., Lam, W.H.K., 2004. A combined distribution, hierarchical mode choice, and assignment network model with multiple user and mode classes. In: Lee, D.H. (Ed.), *Urban and Regional Transportation Modeling*. Edward Elgar Publishing Inc., Northampton, USA, pp. 25–42.
- 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 (1), 81–104.
- Wong, S.C., Yang, H., 1998. Network model of urban taxi services: improved algorithm. *Transportation Research Record* 1623, 27–30.
- Wu, J.H., 1997. A modified primal path following scheme for the monotone variational inequality problems. *Journal of Optimization Theory and Applications* 95 (1), 189–208.
- Wu, J.H., Florian, M., He, S.G., 2006. An algorithm for multi-class network equilibrium problem in PCE of trucks: application to the SCAG travel demand model. *Transportmetrica* 2 (1), 1–9.
- Yang, H., Wong, S.C., 1998. A network equilibrium model of urban taxi services. *Transportation Research Part B* 32 (4), 235–246.
- 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 (3), 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 (9), 799–819.
- Yang, H., Ye, M., Tang, W.H., Wong, S.C., 2005a. Regulating taxi services in the presence of congestion externality. *Transportation Research Part A* 39 (1), 17–40.
- Yang, H., Ye, M., Tang, W.H.C., Wong, S.C., 2005b. A multi-period dynamic model of taxi services with endogenous service intensity. *Operations Research* 53 (3), 501–515.