The Optimal Transit Fare Structure under Different Market Regimes with Uncertainty in the Network

Zhi-Chun Li · William H. K. Lam · S. C. Wong

Abstract This paper proposes a network-based model for investigating the optimal transit fare structure under monopoly and oligopoly market regimes with uncertainty in the network. The proposed model treats the interaction between transit operators and transit passengers in the market as a two-level hierarchical problem with the transit operator sub-model at the upper-level and the transit passenger sub-model at the lower-level. The upper-level problem is to determine the fare structure so as to optimize the objective function of the transit operators, whereas the lower-level problem represents the path choice equilibrium of the transit passengers. In order to consider the uncertainty effects on transit network, the proposed model incorporates the unreliability component of transit services into the passenger disutility function, which is mainly due to variations of the in-vehicle travel time and the dwelling time of transit vehicles at stops. With the use of the proposed model, a numerical example is given to assess the impacts of the market regimes and the unreliability of the transit services on the optimal transit fare structure.

Keywords Transit fare · Market regime · Network equilibrium · Reliability · Monopoly · Social optimum · Oligopolistic competition

Notation
G modified transit network with $G=(N, S)$
N set of nodes representing centroids and transit stops, in which passengers can board, alight or change vehicles

Z.-C. Li · W. H. K. Lam (✉)
Department of Civil and Structural Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China
e-mail: cehklam@polyu.edu.hk

Z.-C. Li
School of Management, Fudan University, Shanghai 200433, China

S. C. Wong
Department of Civil Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong, China

S	set of links in the transit network G; $S=S_1 \cup S_2$
S_1	set of transit links which connect two transit stops
S_2	set of walking links including the access links from origin to transit stops or the egress links from alighting points to destination
W	set of network origin-destination (OD) pairs, $w \in W$
R_w	set of paths connecting OD pair $w \in W$ in the transit network
P_r^w	probability that path r is chosen for a trip between OD pair $w \in W$
θ	parameter representing the perception variation of passengers on travel disutility
u_r	expected travel disutility on path r
u_s	expected travel disutility on link s
δ_{sr}	indicator variable; it equals to 1 if link s is on path r, and 0 otherwise
h_r	passenger flow on path r
v_s	passenger flow on link s
S_w	expected minimum disutility between OD pair $w \in W$
g_w	total resultant passenger demand between OD pair $w \in W$; $g_w = G_w(S_w)$
g_w^0	potential (or latent) passenger demand between OD pair $w \in W$
π_w	parameter of demand sensitivity to travel disutility between OD pair $w \in W$
L	set of transit lines in the transit network
p_s^l	fare of line l on link s
N_l	number of vehicles or fleet size on line l
C_l^0	fixed operating cost of line l
C_l^1	operating cost per vehicle-hour on line l
K	set of transit operators in the transit network
L_k	set of transit lines operated by operator k
Φ_k	profit of transit operator k
A_s	set of attractive lines on link s
T_s	actual travel time on link s; a random variable with mean t_s [i.e. $t_s=E(T_s)$] and standard deviation σ_s
T_{s1}	actual in-vehicle travel time on link s; a random variable with mean t_{s1} [i.e. $t_{s1}=E(T_{s1})$] and standard deviation σ_{s1}
T_{s2}	actual waiting time on link s; a random variable with mean t_{s2} [i.e. $t_{s2}=E(T_{s2})$] and standard deviation σ_{s2}
g_s	in-vehicle crowding discomfort cost on transit link s
$f(\sigma_s)$	unreliability cost of transit services on link s; a function of the standard deviation σ_s of the travel time on link s
τ_1, τ_2	parameters for converting the different quantities to the same unit
ρ_s, σ_s	parameters for measuring the relationship between mean and variance of travel time
t_s^0	free-flow travel time on link s
t_s^l	mean in-vehicle travel time of line l passing through link s
x_s^l	probability of passengers on link s choosing line l
f_l	frequency of line l; a random variable with mean $E(f_l)$ and standard deviation $\sigma(f_l)$
f_s	total frequency on link s; a random variable with mean $E(f_s)$ and standard deviation $\sigma(f_s)$; $f_s = \sum_{l \in A_s} f_l$
g_s^l	in-vehicle discomfort cost of line l passing through link s

g_s^{l0} baseline discomfort level or riding quality of line l passing through link s

v_s^l passenger flow of line l passing through link s

$\bar{v}_s^l$ passenger flow competing with v_s^l for the same common capacity of line l on link s

$\bar{v}_s$ passenger flow competing with v_s for the same common capacity on link s

κ_l capacity of transit vehicle on line l

K_l capacity of line l; $K_l = \kappa_l f_l$

K_s total vehicle capacity on link s; $K_s = \sum_{l \in A_s} K_l$

$i(s)$ tail node of link s

$A_{i(s)}^{l+}$ set of links going out from node $i(s)$ on which line l is included as an attractive line but link s is excluded

$\overline{A}_{i(s)}^l$ set of links on which line l is included as an attractive line, with origin node before $i(s)$ and end node after $i(s)$

λ_1 parameter for measuring the degree of unreliability of transit services

$t_{s^{+(-)}}$ walking time in direction $+$ $(-)$ on walkway s with bi-directional flows

C_s capacity of physical walkway s under unidirectional flow conditions

Γ_l cycle journey time of a transit vehicle on line $l \in L$; a random variable with mean $E(\Gamma_l(\mathbf{v}))$ and standard deviation $\sigma(\Gamma_l(\mathbf{v}))$

$dt_l^n(v)$ dwelling time for the transit vehicle at node n on line l; a random variable with mean $\overline{dt}_l^n(v)$ and standard deviation $\sigma(dt_l^n(v))$

$\mathbf{h}$ vector of path passenger flow; $\mathbf{h} = (h_r, r \in R_w, w \in W)$

$\mathbf{v}$ vector of link passenger flow; $\mathbf{v} = (v_s, s \in S)$

$\mathbf{p}$ vector of transit fare; $\mathbf{p} = (p_s^l, s \in S_1, l \in L)$

$\mathbf{g}$ vector of OD demand; $\mathbf{g} = (g_w, w \in W)$

1 Introduction

Recently, optimizing transit fare structure has been advocated as an efficient means of coordinating transit passenger demand and alleviating traffic congestion in Hong Kong (Zhou et al. 2005). This can be attributed to the fact that Hong Kong, a city of more than 6.9 million people with a land area of only 1,095 km^2, has a well-established and extensive transit network with a variety of transit services. A large proportion of the population uses transit services as the main mode of transportation. Over 90% of the 11 million daily trips in Hong Kong are made on privately operated public transit services (Transport Department 2003). In this highly competitive and profitable environment, it is, therefore, crucial for transit operators to optimize and design their fare structures carefully to compete with other operators to maximize their own profits.

    The optimization and design of transit fare structure have been extensively studied from the perspective of transportation economics (Kocur and Hendrickson 1982; Yang and Woo 2000; Yang et al. 2001) and network equilibrium (Lam and Zhou 2000; Zhou et al. 2005). However, most of these existing studies have considered only time and cost-related attributes as explanatory variables of the quality of transit services. Service reliability has often been ignored in describing the quality of transit services. The empirical studies that have been conducted by Allen et al. (1985) and Frima et al. (1998), however, have found that reliability is one of

the most important factors for measuring the quality of transit services. Transit passengers consider not only time and cost-related factors but also the reliability of the services in their travel decision making.

The unreliability of transit services can be caused by uncertainty in line-haul (or in-vehicle) travel time and/or transit vehicle dwelling time at stops. The uncertainty of the line-haul travel time, particularly for the bus mode, may be due to variations of the congestion level on the road network as a result of the time of day, day of the week, and/or season. The uncertainty of the dwelling time of transit vehicles at stops may be due to the fluctuations in passenger volumes boarding and alighting at stops. The sources of these uncertainties in transit services can be referred to as recurrent congestion. Non-recurrent congestion, which is caused by unexpected irregular and random incidents, such as traffic accidents, vehicle breakdowns, road work, signal failure, or adverse weather, is another important source that induces the unreliability of transit services (Yin et al. 2004). In this paper, attention is given to the effects of recurrent congestion on transit services and the effects of non-recurrent congestion are left for future studies.

The aforementioned sources of transit service uncertainty lead to variability of the transit service frequency and subsequently the fluctuation of passenger waiting times at transit stops. Thus, transit passengers have to incorporate more redundancy or an additional time margin into their travel schedules to accommodate the possible uncertainty of transit services (Jackson and Jucker 1981; Hall 1983).

Although the existing reliability studies mainly focus on the auto (or highway) network, the issues of reliability of transit services have recently attracted considerable attention. Tisato (1998) studied the effects of unreliable bus services on the subsidy requirements. Carey (1999) proposed various measures of schedule reliability for improvements of transit services. Bell et al. (2002) adopted an absorbing Markov chain model for the transit assignment problem so as to analyze the reliability of transit networks. Yin et al. (2004) introduced the concept of waiting time reliability, i.e., the probability that the average waiting time of passengers is less than a given threshold, as a complement of the schedule reliability measure. Chien et al. (2007) developed a probabilistic model to optimize disseminated real-time bus arrival information for pre-trip passengers.

The previous related studies on transit reliability mainly stemmed from the operator's point of view and the key focus was on the assessment of the reliability of transit services. Little attention has been paid to assessing the impacts of the reliability factor on the travel behavior of passengers, i.e., modeling the effects of transit service reliability from the passenger's perspective. This is a very important consideration indeed to improve the design and planning of transit services to accommodate the needs of passengers.

This paper incorporates the uncertainty effects of transit services into an investigation of the optimal transit fare structure under the monopoly and oligopoly market regimes. The three objectives of this paper are (1) to propose a network-based model that explicitly incorporates the unreliability of transit services in the route choice decisions of passengers, (2) to determine the optimal transit fare structure under different market regimes and compare the efficiency of different market regimes in terms of total system profit and social welfare, and (3) to ascertain

the effects of the unreliability of transit services on the optimal market fares and on the market shares among the competing transit services.

In the proposed model, there are two types of agents: transit operators and transit passengers. Transit operators compete for passengers by adjusting their fare structures. In response to the fare structures that are determined by transit operators, transit passengers select the transit route that minimizes their perceived disutility of travel. This results in the interplay between operator and passenger equilibrium problems. To consider the uncertainty effects of transit services, the model that is proposed in this paper explicitly incorporates the reliability factor of transit services into the travel decisions of passengers. To enhance the realism of the modeling framework, the transit vehicle dwelling time at the stops due to passenger boarding and alighting and the walking delays on walkways are explicitly represented by flow-dependent functions. In addition, passenger demand elasticity and passenger discomfort due to crowding within transit vehicles are also considered in the model.

The remainder of this paper is organized as follows. In Section 2, some basic concepts and assumptions are described. Section 3 presents the model formulation. Section 4 provides a numerical example to illustrate the application of the proposed model. Finally, conclusions are given in Section 5 together with recommendations for further studies.

2 Basic considerations

2.1 Some useful concepts

A *transit network* is composed of a set of transit lines and stations (nodes) at which passengers can board, alight, or change vehicles. A *transit line* is a group of vehicles that run back and forth between two transit stops, and can be described by the frequency of the vehicles and the vehicle types (e.g., bus or metro). A *line segment* is any portion of a transit line between two consecutive stations. All vehicles in the same line going through the same sequence of transit stops are on the same *itinerary*. A *line section* is any portion of a transit line between two (not necessarily consecutive) transit stops of its itinerary.

In a transit network, different transit lines can run in parallel for parts of their itineraries and have stations in common. In other words, there exist the overlapping routes that share the same transit stops while running on common segments. The existence of the common lines or attractive lines problem poses a challenge to transit network modeling. De Cea and Fernandez (1993) adopted the concept of route sections to simplify the calculation of common lines. A *transit route* is defined as a path that a transit passenger can follow in the transit network to travel between any two nodes. In general, a transit route can be identified by a sequence of nodes, with the first node as the origin of the trip and the final node as the destination. All of the intermediate nodes represent the transfer points on this route. The section between two consecutive transfer nodes is referred to as a *route section* or *link*. Therefore, a transit route, which is also referred to as a *path* in this paper, can be described as a series of route sections or links.

Figure 1 gives a simple example to illustrate these concepts. In Fig. 1(a), a transit network with one origin–destination (OD) pair and four lines is represented in terms of lines and itineraries. Figure 1(b) shows the modified transit network, which is represented in terms of route sections or links. With the concept of the route section or link, a primitive transit network is transformed into a simplified network that is described by a set of nodes and a set of links. Figure 1(b) also shows that there are two types of links in the transit network: the links (route sections) between two stops or nodes, which are referred to as the *transit links* in this paper, and the *walking links*, i.e., the access links from origin centroids to transit stops or the egress links from alighting points to destination centroids. For example, links s_1, s_2, s_3, s_4, s_5, and s_6 in Fig. 1(b) are transit links, and links (O, N_1) and (N_4, D) are walking links.

2.2 Assumptions

To facilitate the presentation of the essential ideas without loss of generality, the following basic assumptions are made in this paper.

A1 There are two types of agents in the transit network: transit passengers and transit operators. Accordingly, two interrelated competitive equilibria exist, i.e., the competitive equilibrium among transit operators and the competitive travel choice equilibrium of transit passengers.

A2 Transit passengers make their route choice decisions in a stochastic manner based on the tradeoff between the service quality or travel disutility of different transit services. The disutility of travel is measured by a weighted combination of the travel time, monetary cost, in-vehicle crowding discomfort, and unreliability cost of transit services. The unreliability of transit services is affected by variations of the in-vehicle travel time and transit vehicle dwelling time at transit stops, which lead to variability of the transit service frequency and thus the fluctuation of the passenger waiting times at transit stops. The
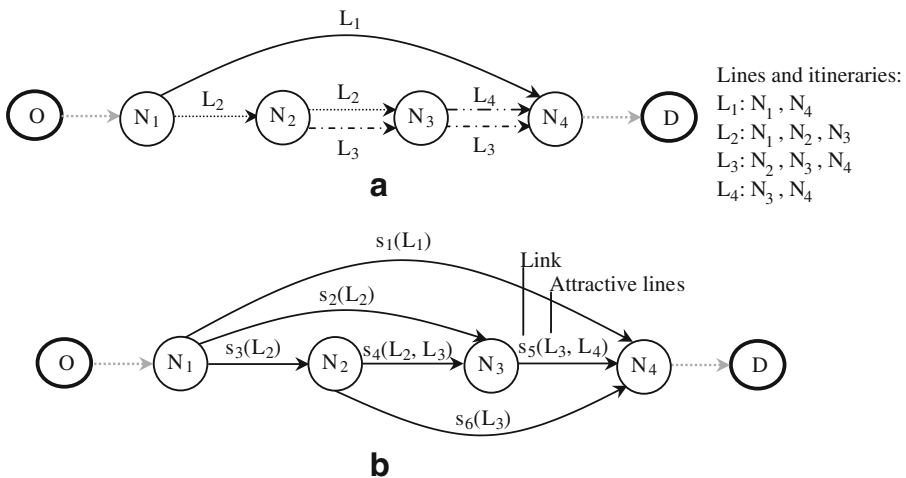

Fig. 1 (a) The original transit network. (b) The modified transit network

passenger discomfort measure and waiting time are mainly due to passenger crowding within transit vehicles and transit stops, respectively.

A3   The common line problem in the transit network leads passengers at a transit stop to face multiple alternative transit lines. Hence, passengers cannot decide beforehand which line to board between any pair of two consecutive transfer stops on their journeys. Passengers can choose not to board the vehicle that arrives first and wait for the next vehicle with express service to minimize their total journey time. In this paper, it is assumed that the passengers who are boarding or waiting on a transit link will be allocated to each line in the set of attractive lines that is associated with that link in the proportion of their mean frequencies (Nguyen and Pallottino 1988; De Cea and Fernandez 1993; Uchida et al. 2005).

A4   It is assumed that the uncertainty of transit services is caused by recurrent congestion. The effects of non-recurrent congestion are not considered in this paper. The uncertainty of transit services is modeled on the basis of the empirical studies of Richardson and Taylor (1978) and Taylor (1982), i.e., the actual link travel time follows the independent normal distribution, and the relationship between the mean, t_s, and standard deviation, s_s, of the link travel time satisfies $\sigma_s = \rho_s \left(t_s / t_s^0 - \varphi_s \right) \sqrt{t_s}$, where t_s^0 is the free-flow travel time on link s, and ρ_s and φ_s are link-dependent parameters that can be calibrated by observed data.

A5   An elastic demand function is used to represent the responses of passengers to changes in the disutility of travel (or generalized costs of travel) due to traffic congestion and/or transit operator strategies. The crowding effect within the transit vehicles can be modeled by a discomfort cost function (Spiess and Florian 1989; Wu et al. 1994; Lo et al. 2003).

3 Model formulation

According to A1, there are two types of agents in the transit market; namely transit passengers and transit operators. The transit operators optimize their fare structures to achieve their own objective, whereas the transit passengers choose the transit service that minimizes their perceived disutility of travel. This induces two interrelated competitive equilibria, i.e., the competitive travel choice equilibrium of transit passengers, which is described as transit passenger path flow sub-model, and the competitive equilibrium among transit operators, which is described as transit operator optimal fare sub-model. In the following, we in turn formulate these two sub-models.

3.1 Transit passenger path flow sub-model

3.1.1 Transit passenger path choice equilibrium

Consider a transit network $G=(N, S)$, where N is the set of all nodes (transit stations) and S is the set of all links, including the transit links between two transit stops and the walking links from origin centroids to transit stops or alighting points to

destination centroids. Let W be the set of OD pairs, $w \in W$, and R_w be the set of paths (routes) between OD pair $w \in W$. In a multinomial logit-based stochastic user equilibrium (SUE), the probability P_r^w that path r is chosen between OD pair $w \in W$ is given by (see, e.g., Sheffi 1985)

$$P_r^w = \frac{\exp(-\theta u_r)}{\sum\limits_{r \in R_w} \exp(-\theta u_r)}, \quad \forall r \in R_w, w \in W \tag{1}$$

where the parameter θ represents the variation of passenger perception on travel disutility. The higher the value of θ, the smaller the variation of passenger perception, and vice versa. The (expected) travel disutility u_r on path r can be expressed as the sum of all of the (expected) travel disutilities on links along this path, i.e.,

$$u_r = \sum_{s \in S} u_s \delta_{sr}, \quad \forall r \in R_w, w \in W \tag{2}$$

where u_s is the expected travel disutility on link s. $\delta_{sr} = 1$ if link s is on path r between OD pair w, and 0 otherwise.

As stated in A2, the travel disutility u_s on link s is composed of the in-vehicle travel time, waiting time, crowding discomfort, fare, and access/egress time (i.e., walking time). In order to consider the uncertainty effects, which is mainly due to variations of the in-vehicle travel time and the dwelling time of transit vehicles at stops, the unreliability cost of transit services is explicitly incorporated into the disutility function. For the detailed definition of the link travel disutility in the transit network, readers may refer to Appendix A.

Therefore, the passenger flow, h_r, on path r can be computed by

$$h_r = g_w P_r^w, \quad \forall r \in R_w, w \in W \tag{3}$$

where g_w is the total resultant passenger demand between OD pair w. Hence, the passenger flow, v_s, on link s can be expressed as the sum of the flows on the paths that are related to link s, i.e.,

$$v_s = \sum_{w \in W} \sum_{r \in R_w} h_r \delta_{sr}, \quad \forall s \in S \tag{4}$$

According to the random utility theory, the expected minimum disutility, S_w, between OD pair w can be measured by the following log-sum formula (Oppenheim 1995),

$$S_w = -\frac{1}{\theta} \ln \sum_{r \in R_w} \exp(-\theta u_r), \quad \forall w \in W \tag{5}$$

Consequently, the total resultant passenger demand g_w between OD pair w can be defined by a continuous and monotonically decreasing function $G_w(\cdot)$ of the travel disutility S_w for the OD pair concerned, i.e.,

$$g_w = G_w(S_w), \quad \forall w \in W \tag{6}$$

3.1.2 Equivalent fixed-point formulation

According to Eqs. (1)–(6), it can be inferred that the transit passenger path flow h_r, $\forall r \in R_w, w \in W$ is the function of the expected path travel disutility u_r, $\forall r \in R_w, w \in W$, which is the function of the expected link travel time t_s, $\forall s \in S$ in terms of the equations that are defined in Appendix A, which is in turn the function of the link flow v_s, $\forall s \in S$, and thus the function of the path flow h_r, $\forall r \in R_w, w \in W$, itself according to Eq. (4). Therefore, the proposed model can be formulated as a fixed-point problem with regard to the path flows.

Proposition 1 *The transit network equilibrium conditions* (1)–(6) *for a given strategy set of transit operators are equivalent to finding a vector* **h** *such that the following fixed-point problem holds.*

$$\mathbf{h}^* - \mathbf{F}\left(\mathbf{h}^*\right) - \mathbf{0}, \quad \mathbf{h}^* \in \Omega \tag{7}$$

with $\mathbf{F}(\mathbf{h}) = \left(g_w P_r^w, \forall r \in R_w, w \in W\right)$ *and* $\Omega = \left\{\mathbf{h} \mid \sum_{r \in R_w} h_r = g_w, \forall w \in W\right\}.$

It should be pointed out that the link travel time functions and the elastic demand functions that are defined in this paper are assumed to be continuous. As a result, the feasible set Ω is closed because the OD demand is bounded, thus there exists at least one solution to the fixed-point problem (7) according to the Brouwer's fixed-point theory. Moreover, it is easily proved that the uniqueness of the model solution can be guaranteed if all of the link travel time functions and the elastic demand functions are strictly monotone (Patriksson 1994; Zhou et al. 2005). The fixed-point problem (7) can be solved effectively by using a solution algorithm recently proposed by Huang and Li (2007), which is based on the method of successive averages in conjunction with the logit-type assignment process.

3.2 Transit operator optimal fare sub-model

3.2.1 Profit function of transit operators

The fare structures adopted by transit operators would have significant impacts on the route choice decisions of passengers and subsequent effects on the passenger flow distribution in the transit network, and thus, on the profits of transit operators. The net profit of a transit operator is the total revenue that is generated from the passenger fares minus the total transit service operating costs.

Suppose that there are K transit operators and L_k is the set of transit lines that is operated by operator k, then the net profit, Φ_k, of operator k can be expressed as

$$\Phi_k(\mathbf{p}, \mathbf{v}(\mathbf{p})) = \sum_{s \in S} \sum_{l \in A_s \cap L_k} p_s^l v_s^l - \sum_{l \in L_k} \left(C_l^0 + N_l C_l^1\right), \quad \forall k \in K \tag{8}$$

where A_s is the set of attractive lines on link s and $l \in A_s \cap L_k$ means that line l that is provided by operator k passes through link s. p_s^l and v_s^l are the fare and the passenger flow of line l on link s, respectively. $\mathbf{p}$ and $\mathbf{v}$ are the vectors of the fare and the passenger flow, respectively. N_l is the number of transit vehicles on line l, C_l^0 is

the fixed operating cost of line l, and C_l^1 is the operating cost per vehicle-hour on line l. The first term on the right-hand side of Eq. (8) represents the total revenue of operator k. The second term represents the total costs of operator k, which consist of the fixed cost C_l^0 and the variable operating cost $N_l C_l^1$ that is proportional to the total vehicle hours (Chien and Schonfeld 1998; Yang and Woo 2000).

3.2.2 Market equilibrium

With the aforementioned assumptions and model, we now investigate the optimal transit fare structure under the following scenarios: the monopoly market (with profit maximization), the social optimum and the oligopolistic competitive market.

Monopoly market The monopoly market in this paper refers to the situation at which all transit lines are operated by a single profit-driven firm or agency to which the monopoly rights have been conferred by the government. Under this monopoly condition, the profit-maximizing firm will determine the transit fare structure to maximize its net profit that is generated from the operation of the whole transit system. The optimal fare structure under the monopoly market can then be determined by

$$\max_{\mathbf{p}} \sum_{k \in K} \Phi_k(\mathbf{p}, \mathbf{v}(\mathbf{p})) \tag{9}$$

where $\mathbf{v}(\mathbf{p})$ is obtained by solving the following transit passenger flow sub-model (note that $\mathbf{v}$ relates to $\mathbf{h}$ by Eq. (4))

$$\mathbf{h}^* - \mathbf{F}\left(\mathbf{h}^*\right) = \mathbf{0}, \quad \mathbf{h}^* \in \Omega \tag{10}$$

The proposed monopoly model (9)–(10) is actually a mathematical program with equilibrium constraints or a bi-level programming problem (Luo et al. 1996).

Social optimum Similar to the monopoly market with profit maximization, the social optimum case assumes that the transit market is managed by a single operator but with an objective to maximize the total social welfare. The total social welfare (SW) per unit period is defined as the sum of the surplus of consumers and the net profit of the transit operator. Thus, the social optimum solution is derived by

$$\max_{\mathbf{p}} \ \mathrm{SW}(\mathbf{p}, \mathbf{g}(\mathbf{p}), \ \mathbf{v}(\mathbf{p})) = \alpha_1 \sum_{w \in W} \int_0^{g_w(\mathbf{p})} G_w^{-1}(u)du - \alpha_1 \sum_{w \in W} g_w(\mathbf{p})S_w(\mathbf{p})$$

$$+ \sum_{k \in K} \Phi_k(\mathbf{p}, \ \mathbf{v}(\mathbf{p})) \tag{11}$$

where $\mathbf{g}(\mathbf{p})$ and $\mathbf{v}(\mathbf{p})$ solve the following transit passenger flow sub-model (note that $\mathbf{g}$ and $\mathbf{v}$ relate to $\mathbf{h}$ by Eqs. (3) and (4) respectively)

$$\mathbf{h}^* - \mathbf{F}(\mathbf{h}^*) = \mathbf{0}, \quad \mathbf{h}^* \in \Omega \tag{12}$$

The parameter α_1 in Eq. (11) is the value of time for evaluation purposes. The first and second terms on the right-hand side of Eq. (11) are the total utilities of travelers and the expected total disutility (including transit fares) incurred by all passengers in the transit market, respectively. The net of these two terms is the net user benefit. The last term is the profit of the transit operator that is determined by Eq. (8).

Oligopoly market The oligopoly market refers to the situation in which transit services are provided by a number of independent transit operators. Each of the operators seeks to optimize its fares to maximize its own profits. This leads to an oligopolistic competitive equilibrium, or Cournot–Nash game, for which the equilibrium fare structure can be obtained by

$$\max_{\mathbf{p}^k} \Phi_k\big(\mathbf{p}^k, \mathbf{p}^{-k}, \mathbf{v}(\mathbf{p}^k, \mathbf{p}^{-k})\big), \quad \forall k \in K \tag{13}$$

where $\mathbf{v}(\mathbf{p}^k, \mathbf{p}^{-k})$, which relates to $\mathbf{h}$ by Eq. (4), is obtained by solving

$$\mathbf{h}^* - \mathbf{F}(\mathbf{h}^*) = \mathbf{0}, \quad \mathbf{h}^* \in \Omega \tag{14}$$

In the objective function (13), $\mathbf{p}^k$ represents the oligopolistic competitive solution of the transit fares for operator k, and $\mathbf{p}^{-k}$ represents that of the other operators excluding k.

The oligopolistic competitive equilibrium can also be represented as

$$\Phi_k\big(\mathbf{p}^{k*}, \mathbf{p}^{-k*}, \mathbf{v}(\mathbf{p}^{k*}, \mathbf{p}^{-k*})\big) \geq \Phi_k\big(\mathbf{p}^k, \mathbf{p}^{-k*}, \mathbf{v}(\mathbf{p}^k, \mathbf{p}^{-k*})\big), \quad \forall \mathbf{p}^k, k \in K \tag{15}$$

where $\mathbf{v}(\mathbf{p}^k, \mathbf{p}^{-k})$ is given by Eq. (14) in virtue of Eq. (4). Inequality (15) implies that at equilibrium no operator can increase its own profit by unilaterally changing fares. Model (13) or (15) can further be formulated as an equivalent variational inequality problem (Zhou et al. 2005).

Finally, it should be mentioned that all of the above bi-level optimization problems with transit passenger flow sub-model at the lower-level can be solved using some heuristic approaches such as the sensitivity analysis based algorithm (Yang 1997) and simulated annealing algorithm (Li et al. 2007).

4 Numerical studies

To facilitate the presentation of the essential ideas and contributions of this paper, the proposed model is applied to an example transit network. This numerical example is intended to determine the optimal transit fare under different market regimes and to compare the efficiency of different market regimes. It is also used to ascertain the effects of the unreliability of transit services on the optimal market fares and on the market demand shares among the competing transit services.

4.1 Data input

The example transit network is shown in Fig. 2(a). It consists of two OD pairs (A–B and B–A), six transit stops (bus stations C, D, E, and F, and metro stations P and Q), four bus lines (L_1, L_2, L_3, and L_4), one metro line (L_5), and four bi-directional walking links (A–C, A–P, B–F, and B–Q). $L_{1(1)}$, $L_{1(2)}$, and $L_{1(3)}$ are the line segments of L_1 between C and D, between D and E, and between E and F, respectively. $L_{2(1)}$ and $L_{2(2)}$ are the line segments of L_2 between C and D and between D and E, respectively. $L_{3(1)}$ and $L_{3(2)}$ are the line segments of L_3 between D and E and between E and F, respectively. Without loss of generality, it is assumed that the in-vehicle travel times between two consecutive nodes are equal for both directions. The basic input data for the numerical example are given in Tables 1 and 2, respectively. Figure 2(b) shows the modified example transit network.
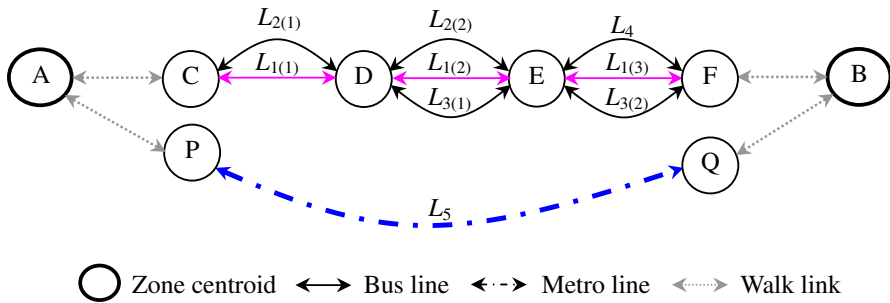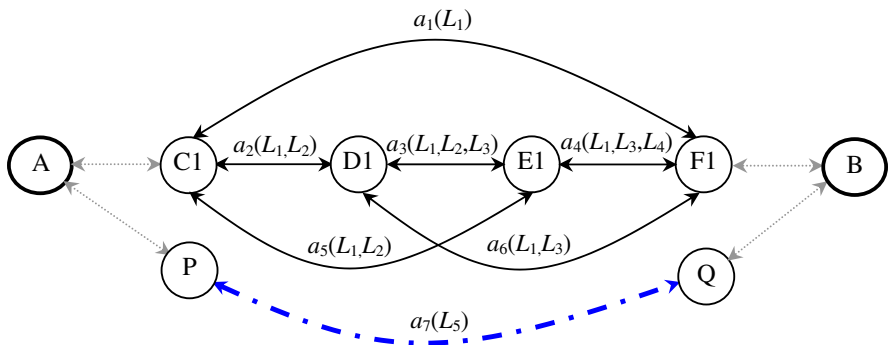
Fig. 2 (a) The example transit network. (b) The modified example network

Table 1 Basic input data for the example transit network

Basic data	Bus line								Metro line
	$L_{1(1)}$	$L_{1(2)}$	$L_{1(3)}$	$L_{2(1)}$	$L_{2(2)}$	$L_{3(1)}$	$L_{3(3)}$	L_4	L_5
Length of line (km)	8.5	8.5	8	7.5	7	7.5	7	7	15
Number of vehicles	14	14	14	11	11	6	6	3	3
Vehicle capacity (pass/veh)	120	120	120	120	120	120	120	120	400

The elastic demand function for each OD pair is specified as

$$g_w = g_w^0 \exp(-\pi_w S_w), \quad \forall w \in W \tag{16}$$

where g_w^0 is the potential demand between OD pair w, and π_w is a parameter that reflects the demand sensitivity to the travel disutility by OD pair.

It is assumed in this numerical example that the potential demand for each OD pair is 5,000 (pass/h). The average speed of travel by bus is 25 km/h and by metro is 60 km/h. It is also assumed that there are two terminals for each of the transit lines, and that the stationary time at each terminal of the bus lines is 0.1 h and the stationary time at each terminal of the metro line is 0.05 h. The free-flow walking time and the capacity for each of the walking links are 0.15 (h) and 5,000 (pass/h), respectively. The baseline discomfort level for all of the in-vehicle links is assumed to be zero. The other model parameters are: α_1=50, β_1=120, α_2=0.5, β_2=120, γ_1=0.15, γ_2=0.15, n_1=3.0, n_2=3.0, B=0.15, x=-0.24, y=5.71, λ_1=1.0, λ_2=1.0, θ=0.9, and π=0.5. In the following analyses, unless specified otherwise, the input data above are considered as the reference or base case.

4.2 Analyses of numerical results

Monopolistic solution in a fully unregulated transit market We first consider the monopolistic solution in which there exists only one profit-maximizing firm or agency that operates the transit services in the market. Figure 3 displays the contours of the total system net profit and social welfare when the bus and metro fare levels are varied. It can be observed in Fig. 3 that different combinations of bus and metro fare levels can lead to three possible outcomes: positive profit, neutral profit, or negative profit. The profit-maximizing solution is given by solving the monopoly problem (9)–(10). The solution occurs at point G with the fare rate of $0.71 and $0.92 for bus and metro rides, respectively, with the maximum profit of $4,923 per hour.

Social optimum solution We now look at the social optimum solution for this numerical example. From the social welfare perspective, efficient allocation is a

Table 2 Basic input data for the model coefficients

	τ_a	φ_a	ρ_a
In-vehicle travel time	1.0	0.9	0.1, Bus; 0.01, Metro
Waiting time	2.0	0.9	0.5, Bus; 0.25, Metro
Walking time	1.8	0.9	0.05

Fig. 3 Total system profit and social welfare for the monopoly and social optimum solutions
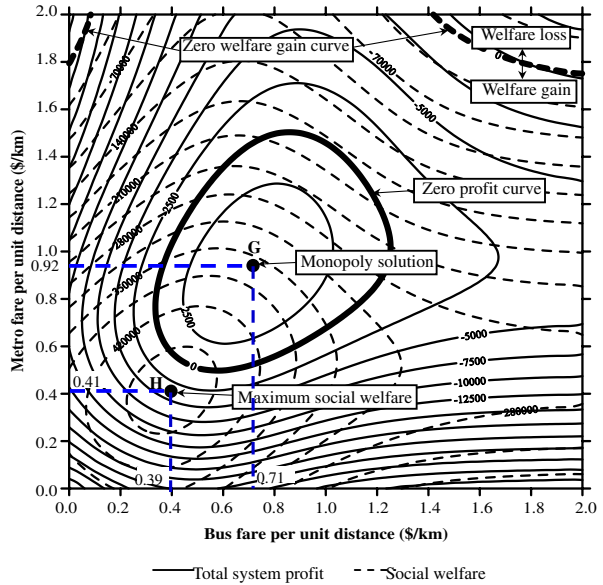

combination of bus and metro fares that would maximize the social welfare as defined by Eq. (11). Figure 3 shows that the social optimum occurs at point H with the fare rate of $0.39 and 0.41 for bus and metro rides, respectively, which results in a social surplus of $470,099 per hour. However, the social optimum solution entails a deficit of $3,979 per hour for the total transit operation, with $456 per hour for bus operation and $3,523 per hour for metro operation. Therefore, in order to achieve the social optimum, a form of subsidy to the transit operator is required. Figure 3 also indicates that welfare loss occurs with the combination of high fare levels for metro and/or bus rides. It is because this fare combination would lead to the result that the total social costs of the transit system cannot be offset by its benefit.

Oligopolistic competitive solution The oligopolistic competitive solution can be determined at the point with the maximum profits for both bus and metro operators in terms of Eq. (13) or (15). Figure 4 plots the profit contours of bus and metro operators in the space of the fare structures of the bus and metro operators. Point M, which is shown in Fig. 4, is the intersection between the two red and dashed curves that pass through the vertices of the profit contours for the bus and metro operators, respectively. The fares of the bus and metro operators at point M are, respectively, $0.57 per unit distance and $0.76 per unit distance, which results in the profits of $2,010 per hour and $2,087 per hour for the bus and metro operators, respectively.

It should be pointed out that point M can be confirmed as an oligopolistic competitive equilibrium solution by checking whether point M satisfies Eq. (15). In fact, the two red and dashed curves that pass through the vertices of the profit contours are the response curves to profit maximization of one operator in response to the fare change of the other operator. Specifically, as the fare rate of the bus operator is fixed at $0.57 per unit distance, the metro operator has to set the fare rate

Fig. 4 Oligopolistic solution with fare competition between bus and metro operators
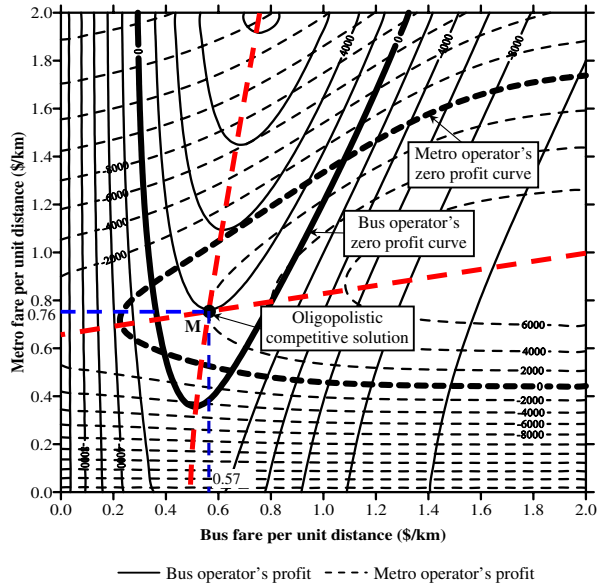
at $0.76 per unit distance to maximize its own profit; any other fare settings for the metro operator would lead to a lower profit, and vice versa. Therefore, point M represents the situation at which neither operator has an incentive to change its fares, which implies that Eq. (15) is satisfied at point M. In addition, it can be seen in Fig. 4 that point M is located within the common positive profit area of the bus and metro operators.

In Figs. 3 and 4, it can be observed that the total profits and the social welfare generated by oligopolistic competition solution lie between those by the monopoly solution and the social optimum solution.

The effects of transit service unreliability on optimal market fares Table 3 illustrates the impacts of the transit service unreliability on the optimal fare structure under different market conditions, including the monopoly, social optimum and oligopo-

Table 3 The impacts of transit service unreliability on the market equilibrium solutions

	Monopoly solution		Oligopoly solution		Social optimum solution	
Degree of transit service unreliability (λ_1)	$\lambda_1=0.0$	$\lambda_1=1.0$	$\lambda_1=0.0$	$\lambda_1=1.0$	$\lambda_1=0.0$	$\lambda_1=1.0$
Bus fare ($/km)	0.67	0.71	0.50	0.57	0.29	0.39
Metro fare ($/km)	0.85	0.92	0.61	0.76	0.26	0.41
Total profit ($/h)	14,005	4,923	11,834	4,097	−1,751	−3,979
Bus profit ($/h)	14,819	2,326	13,349	2,010	7,008	−456
Metro profit ($/h)	−814	2,597	−1515	2,087	−8,759	−3,523
Total demand (pass/h)	3,308	2,576	4,254	3,035	5,883	4,079
Bus demand (pass/h)	1,725	848	2,125	1,011	2,745	1,167
Metro demand (pass/h)	1,582	1,728	2,129	2,024	3,137	2,912
Social welfare ($/h)	553,633	371,218	650,047	420,436	717,852	470,099

**Fig. 5** The supply-demand equilibrium relationship
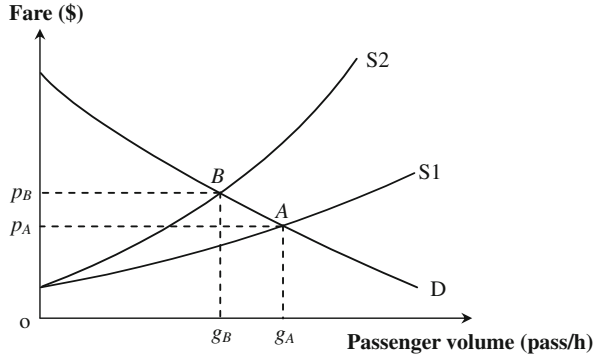
listic competition. It is noted in Table 3 that as the level of transit services degrades from a fully reliable level (i.e. $\lambda_1 = 0.0$) to a less reliable level (i.e. $\lambda_1 = 1.0$), the optimal market fares go up for a given market regime. The interesting phenomenon can be explained by the supply-demand equilibrium curves, as shown in Fig. 5. In fact, the degradation of the transit service level implies that the supply curve for transit services moves from S1 to S2. With the same demand curve D, the market equilibrium point would thus move from A to B, implying that the total transit demand decreases from g_A to g_B, and the transit fare rises from p_A to p_B.

Table 3 also shows that as the level of transit services descends, the total system net profits, total transit demand and total social welfare decrease. Moreover, the degradation of the transit service level would cause that the bus operator's profit decreases but the metro operator's profit increases. This is because with the degraded transit services, the bus demand dramatically decreases, whereas the metro demand only slightly changes. Therefore, with the increasing fare level from p_A to p_B, the metro operator's profit gets an increase instead.

The effects of bus service unreliability on transit market demand share Finally, we examine the effects of the degree of the unreliability of the bus service on the resultant transit demand and market share when the unreliability of the metro service

Fig. 6 The effects of the unreliability of bus service on the resultant transit demand and market share when the unreliability of metro service is fixed at the level of $\lambda_{1,\text{metro}} = 1.0$
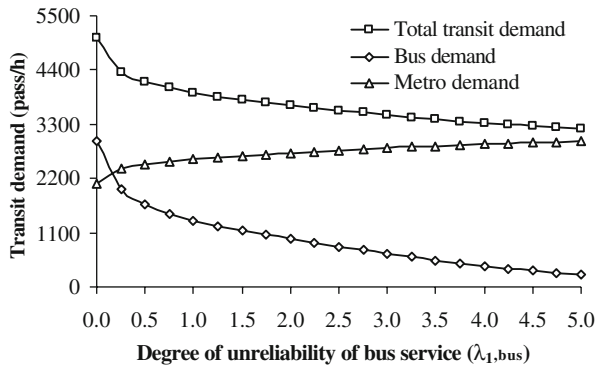
is fixed at the level of $\lambda_{1,\text{metro}}=1.0$. As shown in Fig. 6, when the bus service is fully reliable (i.e. $\lambda_{1,\text{bus}}=0.0$), the passenger demand for the bus service is greater than that for the metro service. As the degree of the unreliability of the bus service increases, the resultant total transit demand and the bus demand decrease and the metro demand increases. Furthermore, when the bus service degrades to a much lower level of reliability (i.e. $\lambda_{1,\text{bus}}=5.0$), the bus operator will be driven out of the transit market, and the metro operator will dominate the whole transit market.

5 Conclusions

In this paper, a network-based model that explicitly incorporates the unreliability of the transit services was proposed to study the optimal transit fare structure under various market regimes. The effects of passenger crowding on the dwelling time of transit vehicles at stops and the walking time on bi-directional walkways were also considered in the proposed model, together with the passenger demand elasticity and the passenger discomfort due to crowding within the transit vehicles. The optimal solutions that correspond to the monopoly, social optimum, and oligopolistic competition were analyzed and compared. The proposed model provides a useful tool for modeling a competitive transit market and evaluating transit policies at the strategic level.

A numerical example was used to examine the efficiency of different transit market regimes and to ascertain the effects of the unreliability of transit services. Some new insights and important findings are obtained. (1) The oligopolistic competitive solution lies between the monopoly solution and the social optimum solution in terms of the total system profits and social welfare. (2) A form of subsidy is required to compensate the deficit of the transit operator under the welfare maximization regime. (3) The degradation of the transit service level would lead to an increasing optimal fare structure for a given market regime. (4) An operator will be driven out of the transit market when the service that is provided by that transit operator degrades to a level that is sufficiently less reliable than that of the competing transit services.

Several directions for future research are suggested as follows: (1) to extend the proposed model to a multimodal transportation network in which the interaction between the auto and transit modes can be considered (Boyce 2007); (2) to consider the risk-taking behavior of passengers towards uncertainty (Bell and Cassir 2002; Szeto et al. 2006; Bell 2007); (3) to assess the performance of the system using the concept of travel demand satisfaction reliability (Heydecker et al. 2007), and (4) to optimize and design transit networks in which both the recurrent and non-recurrent congestion are considered.

Appendix

Appendix A. Link travel disutility

As stated in Section 2.1, there are two types of links in the transit network $G=(N, S)$; namely, transit links and walking links. Let S_1 and S_2 represent the set of transit links and the set of walking links, respectively, and thus $S = S_1 \cup S_2$. In the following, we define in turn the travel disutility functions for the two types of links.

Transit link travel disutility

The travel disutility u_s on transit link s, measured in generalized time units, consists of the following four generalized travel cost components: travel time, transit fare, in-vehicle crowding discomfort, and perceived cost of the unreliability of transit services, i.e.,

$$u_s = E(T_s) + \frac{1}{\alpha_1}p_s + \frac{\beta_1}{\alpha_1}g_s + \frac{\beta_2}{\alpha_2}f(\sigma_s), \quad \forall s \in S_1 \tag{17}$$

where T_s is the random variable of (actual) travel time on transit link s. $E(T_s)$ and σ_s are the mean and the standard deviation of T_s, respectively. p_s and g_s are the transit fare and the in-vehicle crowding discomfort cost on transit link s, respectively. The parameters α_1, β_1 and β_2 are, respectively, the value of time, the value of discomfort, and the value of reliability of passengers, which are all measured in monetary value per unit time. The function $f(\cdot)$ measures the cost of service unreliability, which is a function of the standard deviation, σ_s, of the travel time on transit link s (Noland et al. 1998).The mean travel time $E(T_s)$ on transit link s comprises the mean in-vehicle travel time $E(T_{s1})$ and the mean waiting time $E(T_{s2})$ on link s, i.e.,

$$E(T_s) = \tau_1 E(T_{s1}) + \tau_2 E(T_{s2}), \quad \forall s \in S_1 \tag{18}$$

where T_{s1} and T_{s2} are the random variables of the actual in-vehicle travel time and the actual waiting time on link s, respectively. The parameters (τ) are the reciprocal substitution factors for converting the different time components to the same unit. For the purpose of presentation, let $t_s=E(T_s)$, $t_{s1}=E(T_{s1})$, and $t_{s2}=E(T_{s2})$, then Eq. (18) can be rewritten as

$$t_s = \tau_1 t_{s1} + \tau_2 t_{s2}, \quad \forall s \in S_1 \tag{19}$$

Next, we define the mean in-vehicle link travel time t_{s1}, the mean waiting time t_{s2}, the crowding discomfort cost g_s, and the unreliability cost $f(\sigma_s)$ of transit services.

Mean in-vehicle link travel time

According to A4 in Section 2.2, the actual in-vehicle link travel time is an independent normally distributed random variable with mean t_{s1} and standard deviation σ_{s1}. The mean in-vehicle travel time t_{s1} on link s can be estimated in terms of the mean in-

vehicle travel times of all attractive lines on link s (De Cea and Fernandez 1993; Uchida et al. 2005), i.e.,

$$t_{s1} = \sum_{l \in A_s} t_s^l x_s^l, \quad \forall s \in S_1 \tag{20}$$

where x_s^l is the probability of passengers on link s choosing line l. t_s^l is the mean in-vehicle travel time of passengers using line l on link s, which is assumed to be a constant that is dependent on the length of line l.

We now derive the probability x_s^l of passengers on link s choosing line l for traveling. Note that the transit service frequency f_s on link s can be formulated as the sum of the service frequencies of all attractive lines on link s, i.e.,

$$f_s = \sum_{l \in A_s} f_l, \quad \forall s \in S_1 \tag{21}$$

As the transit line frequency f_l is a random variable that is dependent on the level of the reliability of transit services, the link frequency f_s is also a random variable with

$$E(f_s) = \sum_{l \in A_s} E(f_l), \quad \forall s \in S_1 \tag{22}$$

where $E(f_l)$ and $E(f_s)$ are the mean service frequencies of line l and link s, respectively.

Consequently, according to A3 in Section 2.2, the probability x_s^l of passengers on transit link s choosing line l can be approximated by the proportion of the mean frequency of line l to the mean frequency of link s, i.e.,

$$x_s^l = \frac{E(f_l)}{E(f_s)} = \frac{E(f_l)}{\sum\limits_{l \in A_s} E(f_l)}, \quad \forall l \in A_s, s \in S_1 \tag{23}$$

where $E(f_l)$ can be determined according to Appendix B.

In-vehicle crowding discomfort cost

In general, passenger discomfort is affected by the degree of crowding in transit vehicles. Similar to (20), the crowding discomfort cost on link s can be estimated in terms of the mean discomfort costs of all attractive lines on link s, i.e.,

$$g_s = \sum_{l \in A_s} g_s^l x_s^l, \quad \forall s \in S_1 \tag{24}$$

where g_s^l is the in-vehicle discomfort cost of line l on link s.

According to Spiess and Florian (1989); Wu et al. (1994), and Lo et al. (2003), the in-vehicle crowding discomfort cost, g_s^l, which is measured in terms of generalized time units, on line l passing through link s can be expressed in the

form of the Bureau of Public Roads (BPR) type function with regard to the mean in-vehicle travel time, passenger volume, and vehicle capacity on the line, i.e.,

$$g_s^l = t_s^l \left(g_s^{l0} + \gamma_1 \left(\frac{v_s^l + \overline{v}_s^l}{K_l} \right)^{n_1} \right), \quad \forall l \in A_s, s \in S_1 \tag{25}$$

where g_s^{l0} is the baseline discomfort level or riding quality of line l passing through link s, and γ_1 and n_1 are the positive calibrated parameters of the in-vehicle discomfort function. v_s^l is the passenger flow using line l on link s, and can be estimated by

$$v_s^l = v_s x_s^l, \quad \forall l \in A_s, s \in S_1 \tag{26}$$

where v_s is the passenger flow on link s.

In Eq. (25), the capacity K_l of transit line l can be calculated by

$$K_l = \kappa_l f_l, \quad \forall l \in L \tag{27}$$

where L is the set of transit lines and κ_l is the vehicle capacity on line l.

$\overline{v}_s^l$ is the passenger flow that competes with v_s^l for the same common capacity of line l passing through link s. It consists of two components: (1) the number of passengers boarding at node $i(s)$ (i.e., the tail node of link s), all other links that include line l as an attractive line, i.e., the first term on the right-hand side of Eq. (28) below; and (ii) the passenger volume boarding line l at a node before $i(s)$ and alighting after $i(s)$, i.e., the second term on the right-hand side of (28) below. $\overline{v}_s^l$ can be represented as

$$\overline{v}_s^l = \sum_{e \in A_{i(s)}^{l+}} v_e^l + \sum_{e \in \overline{A}_{i(s)}^{l}} v_e^l, \quad \forall l \in A_s, s \in S_1 \tag{28}$$

where $A_{i(s)}^{l+}$ is the set of links going out from node $i(s)$ on which line l is included as an attractive line but link s is excluded, and $\overline{A}_{i(s)}^{l}$ is the set of links on which line l is included as an attractive line, with an origin node before $i(s)$ and an end node after $i(s)$.

Mean waiting time

The waiting time that is experienced by a transit passenger includes the waiting time for the arrival of the transit vehicle and the overload delay at stops due to the insufficient capacity of the arriving vehicle. The former depends on the arrival distribution of passengers and the average arrival frequency of the vehicles on the wait link, and the latter depends on the passenger volumes boarding the same link and those already in the arriving vehicles. Similar to De Cea and Fernandez (1993) and Lo et al. (2003, 2004), the average waiting time $t_{s2}(v_s)$ on link s can be described as the following volume-delay function,

$$t_{s2}(v_s) = E\left(\frac{\alpha_2}{f_s} \right) + \gamma_2 \left(\frac{v_s + \overline{v}_s}{K_s} \right)^{n_2}, \quad \forall s \in S_1 \tag{29}$$

where α_2, γ_2, and n_2 are positive calibrated parameters. The value of α_2 is dependent on the distributions of transit vehicle headways and passenger arrival times. The

typical value of α_2 adopted in the literature is 0.5 with assumptions of a uniform random arrival distribution of the passengers and of a constant transit vehicle headway (Lam and Morrall 1982). The first term on the right-hand side of Eq. (29) represents the expected waiting time of passenger for the next arriving vehicle, while the second term captures the boarding congestion effect at the transit stops.

In Eq. (29), K_s is the total capacity of the transit vehicles on link s, and

$$K_s = \sum_{l \in A_s} K_l, \quad \forall s \in S_1 \tag{30}$$

where the capacity K_l of line l can be given by Eq. (27).

v_s is the passenger volume waiting to get on link s, and $\overline{v}_s$ is the passenger volume competing with v_s for the same common capacity on link s, which can be calculated by

$$\overline{v}_s = \sum_{l \in A_s} \overline{v}_s^l, \quad \forall s \in S_1 \tag{31}$$

where the passenger flow $\overline{v}_s^l$ can be determined by Eq. (28).

Similar to the derivation of the expected line frequency in Appendix B, the mean $E\left(\frac{\alpha_2}{f_s}\right)$, which is frequency-dependent and used in Eq. (29), can be calculated by

$$E\left(\frac{\alpha_2}{f_s}\right) = \alpha_2 E\left(\frac{1}{f_s}\right) = \frac{\alpha_2}{E(f_s)}\left(1 + \frac{(\sigma(f_s))^2}{(E(f_s))^2}\right), \quad \forall s \in S_1 \tag{32}$$

where the mean $E(f_s)$ and variance $(\sigma(f_s))^2$ of f_s can be given by, respectively,

$$E(f_s) = \sum_{l \in A_s} E(f_l), \quad \forall s \in S_1 \tag{33}$$

$$(\sigma(f_s))^2 = \sum_{l \in A_s} (\sigma(f_l))^2, \quad \forall s \in S_1 \tag{34}$$

where $E(f_l)$ and $\sigma(f_l)$ can be determined according to Appendix B.

Measure of the unreliability of transit services

As stated in Eq. (17), the unreliability cost of transit services can be measured by the function $f(\sigma_s)$ of the standard deviation σ_s. In this paper, for simplicity, we define

$$f(\sigma_s) = \lambda_1 \sigma_s, \quad \forall s \in S_1 \tag{35}$$

where λ_1 measures the degree of the unreliability of transit services. The larger the value of λ_1, the less reliable the transit services, and vice versa.

Similar to Eqs. (18) or (19), the variance σ_s^2 of the travel time on transit link s is the sum of the variances of the in-vehicle link travel time and waiting time on link s, i.e.,

$$\sigma_s^2 = (\tau_1 \sigma_{s1})^2 + (\tau_2 \sigma_{s2})^2, \quad \forall s \in S_1 \tag{36}$$

where σ_{s1} and σ_{s2} are the standard deviations of the in-vehicle travel time and waiting time on link s, respectively. They can be determined according to the relationship between mean and variance, as shown in A4 in Section 2.2.

Walking disutility function

The walking times for access to or egress from transit stops are often assumed to be flow independent in the previous literature (Wu et al. 1994). However, the empirical study of Lam et al. (2003) showed that on a bi-directional walkway with heavy opposing pedestrian flows, both the capacity of the walkway and the pedestrian walking speeds can be reduced significantly, particularly in the minor flow direction.

On the basis of their empirical studies, Lam et al. (2003) proposed a generalized walking time function to account for the bi-directional flow effects on the walkways under different flow conditions, ranging from free-flow to congested situations. Following Lam et al. (2003), the (expected) generalized walking time function that is used in the proposed model is

$$t_{s^{+(-)}}(v_{s^+}, v_{s^-}) = t_s^0 + B(v_{s^{+(-)}}/(v_{s^+} + v_{s^-}))^x ((v_{s^+} + v_{s^-})/C_s)^y, \forall s \in S_2 \qquad (37)$$

where s^+ and s^- are two walking links that represent the physical walkway s, $t_{s^{+(-)}}$ is a unit of the walking time in direction $+ (-)$ on walkway s with bi-directional flows, t_s^0 is the free-flow walking time on walkway s, and C_s is the capacity of the physical walkway s under unidirectional flow conditions. B, x, and y are the parameters to be calibrated with observed data.

Similar to Eq. (35), we can define the uncertainty cost that is caused by the fluctuation of walking time on the congested walkways as below.

$$f(\sigma_s) = \lambda_2 \sigma_s, \quad \forall s \in S_2 \qquad (38)$$

where λ_2 measures the degree of the uncertainty of the walking time. Again, the standard deviation σ_s of the walking time on walking link s can be determined according to the relationship between mean and variance, as shown in A4 in Section 2.2.

Consequently, the walking disutility for access to or egress from transit stops can be formulated as the sum of the walking time and the uncertainty cost that is caused by the fluctuation of the walking time, i.e.,

$$u_s = t_s + \frac{\beta_2}{\alpha_1} f(\sigma_s), \quad \forall s \in S_2 \qquad (39)$$

where the walking time t_s on link s can be calculated by Eq. (37).

Appendix B. Random transit line frequency

As stated in A2 in Section 2.2, the variations of the in-vehicle travel time and the dwelling time of transit vehicles at stops can cause variability of the transit service frequency. Hence, the line frequency f_l is a random variable and its mean and variance are derived as follows.

Let N_l be the number of vehicles on line $l \in L$, and $\Gamma_l(\mathbf{v})$ be the cycle journey time of a transit vehicle on line $l \in L$, and then the line frequency f_l can be obtained by

$$f_l = \frac{N_l}{\Gamma_l(\mathbf{v})}, \quad \forall l \in L \tag{40}$$

where $\mathbf{v}$ is the vector of passenger flows in the transit network.

The cycle journey time $\Gamma_l(\mathbf{v})$ of a transit vehicle on line $l \in L$ is composed of the line-haul travel time, terminal time, and dwelling delays at transit stops (Fernandez and Marcotte 1992; Lam et al. 2002). The uncertainties of the line-haul travel time and the dwelling time of transit vehicles at stops would lead to the variation of the cycle journey time. Hence, $\Gamma_l(\mathbf{v})$ is a random variable, and its mean $E(\Gamma_l(\mathbf{v}))$ and variance $(\sigma(\Gamma_l(\mathbf{v})))^2$ can be calculated by, respectively,

$$E(\Gamma_l(\mathbf{v})) = \zeta t_l^0 + \sum_{m \in l} t_l^m + \sum_{n \in l} \overline{d}t_l^n(\mathbf{v}), \quad \forall l \in L \tag{41}$$

$$(\sigma(\Gamma_l(\mathbf{v})))^2 = \sum_{m \in l} (\sigma_l^m)^2 + \sum_{n \in l} (\sigma(dt_l^n(\mathbf{v})))^2, \quad \forall l \in L \tag{42}$$

where t_l^0 is the constant terminal time on line l and ζ is the number of terminal times on the circular line. $m \in l$ and $n \in l$ imply that line segment m and transfer node n lie on transit line l, respectively. t_l^m and σ_l^m are the mean and standard deviation of the travel time on line segment m on transit line l, respectively. $\overline{d}t_l^n(\mathbf{v})$ and $\sigma(dt_l^n(\mathbf{v}))$ are the mean and standard deviation of the dwelling time $dt_l^n(\mathbf{v})$ at node n on line l, respectively.

According to Lam et al. (1998), the transit vehicle dwelling time at a transit stop is governed by the number of boarding and alighting passengers, i.e., the total interchanging passenger volumes. The expected dwelling time can be expressed as a function with regard to the boarding and alighting volumes (Yin et al. 2004),

$$\overline{d}t_l^n(\mathbf{v}) = max(\overline{d}t_{l0}^n, \eta_0 + \eta_1 Bo_l^n + \eta_2 Al_l^n), \forall l \in L \tag{43}$$

where $\overline{d}t_{l0}^n$ is the minimal (scheduled) dwelling time of line l at stop n. Bo_l^n and Al_l^n are, respectively, the number of passengers boarding and the number of passengers alighting line l at stop n, and they can be determined by the method that is outlined in the study of Lam et al. (2002). The coefficients (η) are the positive parameters, which can be calibrated by the observed data (Lam et al. 1998, 2002).

From (40), for a given value of N_l, the mean $E(f_l)$ and variance $(\sigma(f_l))^2$ of the line frequency f_l can be given by, respectively,

$$E(f_l) = E\left(\frac{N_l}{\Gamma_l(\mathbf{v})}\right) = N_l E\left(\frac{1}{\Gamma_l(\mathbf{v})}\right), \quad \forall l \in L \tag{44}$$

$$(\sigma(f_l))^2 = \left(\sigma\left(\frac{N_l}{\Gamma_l(\mathbf{v})}\right)\right)^2 = N_l^2 \left(\sigma\left(\frac{1}{\Gamma_l(\mathbf{v})}\right)\right)^2, \quad \forall l \in L \tag{45}$$

Applying a quadratic Taylor series approximation to Eqs. (44) and (45), we then have

Proposition B.1 *The mean and variance of the transit line frequency f_l can be calculated by, respectively,*

$$E(f_l) = \frac{N_l}{E(\Gamma_l(\mathbf{v}))}\left(1 + \frac{(\sigma(\Gamma_l(\mathbf{v})))^2}{(E(\Gamma_l(\mathbf{v})))^2}\right), \quad \forall l \in L \tag{46}$$

$$(\sigma(f_l))^2 = N_l^2\,\frac{(\sigma(\Gamma_l(\mathbf{v})))^2}{(E(\Gamma_l(\mathbf{v})))^4}, \quad \forall l \in L \tag{47}$$

where $E(\Gamma_l(\mathbf{v}))$ and $\sigma(\Gamma_l(\mathbf{v}))$ can be calculated by Eqs. (41) and (42), respectively.

Proof For the purpose of presentation, let $\frac{1}{X} = \frac{1}{\Gamma_l(\mathbf{v})}$, and then a quadratic Taylor series approximation of $\frac{1}{X}$ around $X_0 = E(X)$ can be represented as

$$\frac{1}{X} = \frac{1}{X_0} - \frac{1}{1!X_0^2}(X - X_0) + \frac{2}{2!X_0^3}(X - X_0)^2 + \cdots \tag{48}$$

Taking the expectation on both sides of Eq. (48) and ignoring higher order terms yield

$$E\left(\frac{1}{X}\right) = \frac{1}{X_0} - \frac{1}{X_0^2}E(X - X_0) + \frac{1}{X_0^3}E\left((X - X_0)^2\right) \tag{49}$$

Because $E(X - X_0) = 0$ and $E\left((X - X_0)^2\right) = (\sigma(X))^2$, Eq. (49) can be written as

$$E\left(\frac{1}{X}\right) = \frac{1}{E(X)} + \frac{(\sigma(X))^2}{E(X)^3} = \frac{1}{E(X)}\left(1 + \frac{(\sigma(X))^2}{(E(X))^2}\right) \tag{50}$$

Taking the variance on both sides of Eq. (48) and ignoring higher order terms, we obtain

$$\left(\sigma\left(\frac{1}{X}\right)\right)^2 = \frac{(\sigma(X))^2}{(E(X))^4} \tag{51}$$

Substituting $\frac{1}{X} = \frac{1}{\Gamma_l(v)}$ into Eqs. (50) and (51), Eqs. (44) and (45) then become (46) and (47), respectively. This completes the proof of Proposition B.1.

References

Allen WB, Mahmoud MM, McNeil D (1985) The importance of time in transit and reliability of transit time for shippers, receivers, and carriers. Transp Res 19B:447–456

Bell MGH (2007) Mixed routing strategies for hazardous materials: decision-making under complete uncertainty. Int J Sustain Transp 1:133–142

Bell MGH, Cassir C (2002) Risk-averse user equilibrium traffic assignment: an application of game theory. Transp Res 36B:671–681

Bell MGH, Schmoecker JD, Iida Y, Lam WHK (2002) Transit network reliability: an application of absorbing Markov chains. In: Taylor MAP (ed) Transportation and traffic theory. Elsevier, Oxford, pp 43–62

Boyce DE (2007) Forecasting travel on congested urban transportation networks: review and prospects for network equilibrium models. Netw Spat Econ 7:99–128

Carey M (1999) Ex ante heuristic measures of schedule reliability. Transp Res 33B:473–494

Chien S, Schonfeld P (1998) Joint optimization of a rail transit line and its feeder bus system. J Adv Transp 32:253–284

Chien S, Daripally SK, Kim K (2007) Development of a probabilistic model to optimize disseminated real-time bus arrival information for pre-trip passengers. J Adv Transp 41:195–215

De Cea J, Fernandez E (1993) Transit assignment for congested public transport systems: an equilibrium model. Transp Sci 27:133–147

Fernandez E, Marcotte P (1992) Operators-users equilibrium model in a partially regulated transit system. Transp Sci 26:93–105

Frima M, Edvardsson B, Gaerling T (1998) Perceived service quality attributes in public transport: inference from complaints and negative critical incidents. J Public Transp 2:67–89

Hall RW (1983) Travel outcome and performance: the effect of uncertainty on accessibility. Transp Res 17B:275–290

Heydecker BG, Lam WHK, Zhang N (2007) Use of travel demand satisfaction to assess road network reliability. Transportmetrica 3:139–171

Huang HJ, Li ZC (2007) A multiclass multicriteria logit-based traffic equilibrium assignment model under ATIS. Eur J Oper Res 176:1464–1477

Jackson W, Jucker J (1981) An empirical study of travel time variability and travel choice behavior. Transp Sci 16:460–475

Kocur G, Hendrickson C (1982) Design of local bus service with demand equilibration. Transp Sci 16:149–170

Lam WHK, Morrall J (1982) Bus passenger walking distances and waiting times: a summer–winter comparison. Transp Quart 36:407–421

Lam WHK, Zhou J (2000) Optimal fare structure for transit networks with elastic demand. Transp Res Rec 1733:8–14

Lam WHK, Cheung CY, Poon YF (1998) A study of train dwelling time at the Hong Kong mass transit railway system. J Adv Transp 32:285–296

Lam WHK, Zhou J, Sheng Z (2002) A capacity restraint transit assignment with elastic line frequency. Transp Res 36B:919–938

Lam WHK, Lee YS, Chan KS, Goh PK (2003) A generalized function for modeling bi-directional flow effects on indoor walkways in Hong Kong. Transp Res 37A:789–810

Li ZC, Huang HJ, Lam WHK, Wong SC (2007) Time-differential pricing of road tolls and parking charges in a transport network with elastic demand. In: Allsop RE, Bell MGH, Heydecker BG (eds) Transportation and traffic theory. Elsevier, Oxford, pp 55–85

Lo HK, Yip CW, Wan KH (2003) Modeling transfer and non-linear fare structure in multi-modal network. Transp Res 37B:149–170

Lo HK, Yip CW, Wan KH (2004) Modeling competitive multi-modal transit services: a nested logit approach. Transp Res 12C:251–272

Luo ZQ, Pang JS, Ralph D (1996) Mathematical programs with equilibrium constraints. Cambridge University Press, Cambridge, UK

Nguyen S, Pallottino S (1988) Equilibrium traffic assignment for large scale transit networks. Eur J Oper Res 37:176–186

Noland RB, Small KA, Koskenoja PM, Chu X (1998) Simulating travel reliability. Reg Sci Urban Econ 28:535–564

Oppenheim N (1995) Urban travel demand modeling: from individual choices to general equilibrium. Wiley, New York

Patriksson M (1994) The traffic assignment problem—models and methods. VSP, Utrecht, The Netherlands

Richardson AJ, Taylor MAP (1978) Travel time variability on commuter journeys. High Speed Ground Transp J 12:77–99

Sheffi Y (1985) Urban transportation networks: equilibrium analysis with mathematical programming methods. Prentice-Hall, Englewood Cliffs

Spiess H, Florian M (1989) Optimal strategies: a new assignment model for transit networks. Transp Res 23B:83–102

Szeto WY, O'Brien L, O'Mahony M (2006) Risk-averse traffic assignment with elastic demands: NCP formulation and solution method for assessing performance reliability. Netw Spat Econ 6:313–332

Taylor MAP (1982) Travel time variability-the case of two public modes. Transp Sci 16:517–521

Tisato P (1998) Service unreliability and bus subsidy. Transp Res 32B:423–436

Transport Department (2003) Travel characteristics survey 2002—final report, Hong Kong. Transport Department of Hong Kong, Hong Kong

Uchida K, Sumalee A, Watling D, Connors R (2005) Study on optimal frequency design problem for multimodal network using probit-based user equilibrium assignment. Transpn Res Rec 1923:236–245

Wu JH, Florian M, Marcotte P (1994) Transit equilibrium assignment: a model and solution algorithms. Transp Sci 28:193–203

Yang H (1997) Sensitivity analysis for the elastic-demand network equilibrium problem with applications. Transp Res 31B:55–70

Yang H, Woo KK (2000) Modeling bus service under competition and regulation. ASCE J Transp Eng 126:419–425

Yang H, Kong HY, Meng Q (2001) Value-of-time distributions and competitive bus services. Transp Res 37E:411–424

Yin Y, Lam WHK, Miller MA (2004) A simulation-based reliability assessment approach for congested transit network. J Adv Transp 38:27–44

Zhou J, Lam WHK, Heydecker BG (2005) The generalized Nash equilibrium model for oligopolistic transit market with elastic demand. Transp Res 39B:519–544