Optimal Sectional Fare and Frequency Settings for Transit Networks with Elastic Demand

Sujun Sun and W.Y. Szeto[1]
Department of Civil Engineering, The University of Hong Kong, Hong Kong, P. R. China
[1] Phone: (852) 28578552, Fax: (852) 25595337, Email: ceszeto@hku.hk

Abstract

Sectional fares have been used in transit services in practice but are rarely examined analytically and compared with flat and distance-based fares, especially under the considerations of path choice, elastic demand, service frequency, and profitability. This paper proposes a bilevel programming model to jointly determine the fare and frequency setting to maximize transit operator's profit. The preceding three fare structures can be incorporated into the bilevel model. To consider the path choice and elastic demand in the bilevel model, the existing approach-based stochastic user equilibrium transit assignment model for the fixed demand was extended to the elastic demand case and the resultant model was used in the lower level model. To solve the bilevel model, the sensitivity-based descent search method that takes into account the approach-based formulation for the elastic demand transit assignment is proposed, in which the approach-based formulation was solved by the cost-averaging self-regulated averaging method. Numerical studies and mathematical analyses were performed to examine the model properties and compare the three fare structures. The result of the Tin Shui Wai network instance is also provided to illustrate the performance of the solution method.

It is proven that when all passengers' destinations are located at transit terminals, the sectional fare structure is always better than the other two fare structures in terms of profitability. For more general networks, the sectional fare structure is always better than the flat fare structure, but the choice between sectional and distance-based fare structures depends on the geometry of the network (e.g., the route structure and the distance between stops), the demand distribution, and the maximum allowable fares. It is also proven that the optimal profit (total vehicle mileage) is strictly monotonically decreasing (monotonically decreasing) with respect to the unit operating cost. Moreover, it is proven that the lower level approach-based assignment problem with elastic demand has exactly one solution. However, the bi-level problem can have multiple optimal solutions. Interestingly, it is found that from the operator's profitability point of view, providing better information to the passengers may not be good.

Keywords: fare optimization; frequency setting; transit assignment; bilevel optimization; transit network design; sensitivity-based heuristic

1. Introduction

Public transit services are found worldwide. They can also be very profitable and competitive. According to the latest Travel Characteristics Survey 2011 conducted by the Hong Kong Transport Department (Arup Group Limited, 2014), among the 12.6 million daily mechanized trips, 88% of them were made by public transit services including bus, light bus, the Mass Transit Railway, light rail transit, ferry, trams, taxi services, etc. Most of the transit services in Hong Kong are operated by profit-driven private agencies. For instance, there are 5 private bus companies and over 100 private light bus companies in Hong Kong. It is crucial for these transit operators to optimize their services in order to stay competitive in the market.

Transit fares are a basic element in public transit services. They are also considered to be a major factor to attract passenger demand as well as raise operators' revenue. Fleishman et al. (1996) and Nassi and da Costa (2012) provided reviews of and discussions on various fare strategies. Transit fares can be generally classified into two types, namely flat and differentiated fares. A flat fare charges a single and constant price regardless of trip length, time period, and level of service, while differentiated fares vary with one or more factors mentioned previously. Nassi and da Costa (2012) evaluated the transit fare systems of some metropolitan regions worldwide and found out that the distance-based differentiated fare system was ranked the best by operators, professional consultants, and government sectors based on the weighted ranking result of different criteria including complexity, fare, trip attraction, etc.

Vuchic (2007) concluded that some commonly concerned constraints needed to be considered when planning transit fare structures. Firstly, the passenger demand is not constant; it is sensitive to the fare charged and the level of service provided as well as those of the competing transit modes. When the fare of a transit mode increases or the level of service provided by that mode decreases, the willingness of passengers to travel on that mode decreases. Secondly, different individuals may have different perceptions of the value of the service provided versus the fare paid. In order to achieve the operators' objectives under limited resources and other constraints, optimization models were proposed to find out the best decision.

A summary of recently developed fare optimization models is presented in Table 1. Most of the existing fare optimization models consider only one kind of fare structure, including the fare strategy and the pricing level. For example, flat fare optimization models were proposed by Lam and Zhou (2000) and Zhou et al. (2005), who developed bi-level models for the flat fare structure optimization under stochastic user equilibrium with elastic demand (SUEED). Huang (2002), Wang et al. (2014), and Wang et al. (2016) extended the consideration to bi-modal equilibrium models for determining an optimal transit flat fare. Zhang and Yang (2016) also developed a flat fare and frequency optimization model considering passengers' willingness to pay. For distance-based fare optimization, Tsai et al. (2008) proposed a bi-level model under SUEED and trip length differentiation. Li et al. (2009) also developed a bi-level model to optimize the distance-based fare structure under different market regimes with network uncertainty. Borndörfer et al. (2012) considered multiple objectives in their model optimizing the zonal fare structure with monthly tickets. Tsai et al. (2012) developed a fare and frequency optimizing model considering distance-based fares with the zonal partition. Recently, Wang and Qu (2017) developed an optimal sectional fare setting on a *rail corridor* that leads to the system optimal (SO) station choice considering passengers' competition for seats. In the aforementioned studies, no comparison of the effects of adopting different fare structures was made. Moreover, the sectional fare structure (also referred to as the directional fare structure by Chin et al., 2016)—which can retain the advantages of both flat and distance-based fares (e.g., flexibility in price, simplicity, and the low cost of the fare collection process) and is widely adopted by bus companies in Hong Kong (e.g., Kowloon Motor Bus Company Limited, City Bus Limited, and New World First Bus Services Limited)—have not received sufficient attention in the fare optimization literature, especially over transit *networks*.

Table 1. A summary of fare optimization models

References	Behavioral assumption	Bi-level optimization problem	Fare structure		
			Flat fare	Differential fare	
				Distance-based	Sectional
Borndörfer et al. (2012)	Logit-type elastic demand			•	
Chien and Tsai (2007)	Elastic demand		•	•	
Deng et al. (2014)	Elastic demand		•	•	
Huang (2002)	Bi-modal logit with elastic demand		•		
Huang et al. (2016)	Logit SUEED	•	•	•	
Lam and Zhou (2000)	Logit SUEED	•	•		
Li et al. (2009)	Logit SUEED	•		•	
Tsai et al. (2008), (2012)	Elastic demand			•	
Wang et al. (2014)	Bi-modal logit	•	•		
Wang et al. (2016)	Bi-modal logit		•		
Wang and Qu (2017)	UE/SO				•
Zhang and Yang (2016)	Elastic demand		•		
Zhou et al. (2005)	Logit SUEED	•	•		

Limited studies consider two fare structures. These studies often focus on the comparison between the flat and differentiated fare structures or between the flat and distance-based fare structures. Ling (1998) made a comparison between the flat and differentiated fare structures and concluded that the optimal fare structure is affected by the fare elasticity of demand for short and long trips and the numbers of long and short trips. It was found that the distance-based differentiated fare structure is favorable in terms of revenue maximization when the fare elasticity of demand for short trips is greater than that for long trips or the number of short trips is larger than that of long trips. Other analytical models considering more than one kind of fare structures focus on the comparison between flat and distance-based fare structures (e.g., Chien and Tsai, 2007; Deng et al., 2014; Huang et al., 2016). As the time of this writing, there is no study with a comparison of more than two kinds of differentiated fare structures such as the distance-based and sectional fare structures *analytically*. It is unclear whether the sectional fare structure is always better than the distance-based fare structure under profit maximization or vice versa.

Some existing fare optimization models do capture the path choice behavior of passengers, which is formulated into a transit assignment problem (e.g., Lam and Zhou, 2000; Zhou et al., 2005; Li et al., 2009; Huang et al., 2016). Most of the existing transit assignment models are link-based (e.g., Hamdouch and Lawphongpanich, 2008; Hamdouch et al., 2011; Sun et al., 2013; Codina and Rosell, 2017) and path-based (e.g., Cats et al., 2016; Nuzzolo et al., 2016). The path-based models use path flows to present passengers' path choice behavior. Therefore, path-specific information is contained in the formulation and is useful to determine the impacts of path-specific design decisions such as fare discounts for transfers between particular transit lines in the upper level fare setting design problem. However, the path-based formulation requires the full path set to be given, which is known to be time-consuming to find it out completely. On the other hand, passenger link flows are used to formulate the link-based models. They do not require the path choice set information, and the computational efficiency and the convergence of solution methods for the link-based-formulated models for large networks can be ensured. Long et al. (2013) proposed an alternative formulation, namely approach-based formulation, for their dynamic traffic assignment model. This formulation used approach proportions (i.e., the probability of the traffic using each outgoing link from a node) as decision variables to model traffic movements at a node. Like solving for the link flows of the link-based formulation, solving for the equilibrium approach proportions of the equilibrium approach-based formulation does not require the time-consuming path set enumeration. However, the path-based variables such as path flows and path selection probabilities can be traced using approach proportions. In the transit assignment literature, there are still very limited transit assignment models adopting this newly developed formulation method. Table 2 summarizes the existing approach-based transit assignment models in the literature. However, they are all related to fixed demand. *The extensions to SUE with elastic demand (SUEED) have not been found yet*.

Table 2. A summary of approach-based transit assignment models

Reference	Behavioral assumption
Szeto and Jiang (2014a)	User equilibrium (UE)
Jiang and Szeto (2016)	Reliability-based user equilibrium
Sun and Szeto (2018)	Stochastic user equilibrium (SUE)

With a transit assignment model incorporated into bilevel fare optimization models as a lower-level model, gradient-based sensitivity analysis heuristics can be used to solve the whole models. The heuristics are related to the gradient-based sensitivity analysis of network equilibria introduced by Tobin and Friesz (1988), which has been extended for different network equilibrium problems (e.g., Yang, 1995; Yang, 1997; Yang and Bell, 2007). The sensitivity analysis is aimed at calculating the derivatives of equilibrium flows. In the literature, most of the bi-level transit fare optimization models, including those developed by Lam and Zhou (2000), Zhou et al. (2005), Li et al. (2009), and Wang et al. (2014), have been solved by sensitivity analysis-based algorithms, but *the sensitivity analysis of the approach-based SUEED and the corresponding sensitivity analysis-based algorithms have yet been found*.

In this paper, we propose a fare and frequency optimization model that considers the approach-based SUEED over transit networks and the profitability of the operator. This bilevel model can incorporate the sectional, flat, and distance-based fare structures. We compare the sectional fare structure with the two other fare structures in terms of the profitability of the operator under different network topologies, demand patterns, and maximum allowable fares. We also investigate the effects of the passenger perception of travel cost on the profitability of the operator and the relationships between profit, total vehicle mileage, and unit operating cost. Based upon the sensitivity analysis of the approach-based SUEED in our study, we develop a sensitivity-based descent search method to solve the bilevel model, in which the approach-based model is solved by the cost-averaging self-regulated averaging method. To illustrate the performance of the solution methodology, we present a case study of Tin Shui Wai, Hong Kong.

The contributions of this paper are specified as follows:

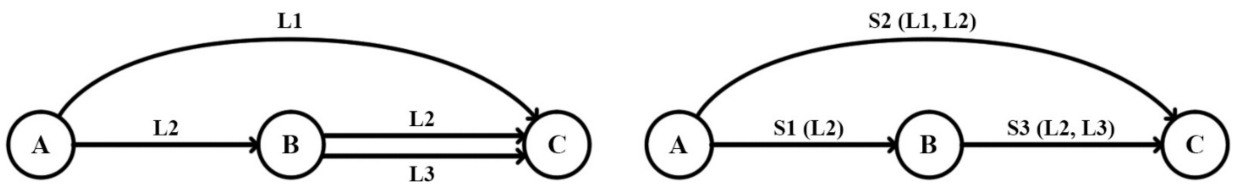
- It proposes a new bi-level model to optimize fare and frequency settings to maximize transit operator's profit under SUEED. To the best of our knowledge, we are the pioneers to consider sectional fares over transit networks.
- It presents the properties of the bilevel model and provides insights on the choice of fare structures. We analytically compare the performance of three different fare structures (i.e., flat, distance-based, and sectional fare structures) in terms of profit maximization. We prove that the flat fare structure is a special case of the sectional fare structure; under a special condition, we find that the distance-based fare structure is a special case of the sectional fare structure. We also prove that the optimal profit and total vehicle mileage are respectively strictly monotonically decreasing and monotonically decreasing with respect to the unit operating cost.
- It proposes an approach-based formulation for the SUEED transit assignment problem, and we prove the existence and uniqueness of the solution to the problem.
- It introduces a sensitivity analysis-based descent search method to determine an optimal fare and frequency setting, *taking into account the approach-based formulation* for the lower level problem. This paper supplements the literature in terms of broadening the concept and theory of approach-based transit assignment to SUEED and developing the corresponding sensitivity analysis methodology that can be used for solving general bi-level transit network design problems.

The remainder of this paper is organized as follows. Section 2 provides the problem formulation and properties. Section 3 gives the details of the solution method. Section 4 presents the numerical studies. Finally, Section 5 concludes the paper.

2. Problem formulation

2.1. Network representations

The node-link representation of a transit network using the concept of attractive lines as described by De Cea and Fernández (1993) was used in this study, which is one of the approaches to handling the common line problem (e.g., Lam et al., 1999; 2002). Using this approach, the problem is transformed into the one similar to the road network assignment problem. In this approach, transit stops are represented as transit nodes in the network, and transit lines traveling between the same pair of transit stops in the network are grouped into one transit link and referred to as the common lines on that link. Passengers can board, alight, or make a transfer only at transit nodes in the network. Figure 1(a) shows an example transit network consisting of a set of transit stops and transit lines, and Figure 1(b) shows its node-link representation using the concept of attractive lines. As shown in the figure, for node pair AB, only transit line L2 travels between node A and node B, and therefore, link S1 is associated with L2 only. For node pair AC, two transit lines (i.e., L1 and L2) connect node A and node C, and therefore S2 is associated with both L1 and L2. Similarly, for node pair BC, link S3 is formed by both L2 and L3.



(a) Transit stops and lines (b) Node-link representation using attractive lines

Figure 1. An example of the network representation using the concept of attractive lines

2.2. Notations

To formulate the studied problem and present the solution method, the following are used:

2.2.1. Sets and indices

a	approach index;
b, b'	route indexes;
d	destination index;
i, i', j	node indexes;
k	iteration number;
r	origin index;
$s, s', s^{\#}$	link indexes;
x	fare structure index. x = 1 to 3 indicates flat, distance-based, and sectional fare structures, respectively;
y, y'	path indexes;
A_i^+	the set of approaches associated with node i or the set of links coming out from that node;
A_i^-	the set of links entering node i;
$\bar{A}_s$	the set of competing links associated with link s;
B	the set of routes;
B_s	the set of attractive routes associated with link s;
D	the set of destinations;
I^b	the set of stops along route b;
I_{i+}^b	the set of subsequent stops of stop i along route b;
N	the set of nodes (i.e., stops) in the transit network;
N^d	the set of nodes in the transit network that are connected to destination d;
N^{id}	the set of nodes in the subnetwork formed by S^{id};
R	the set of origins;
S	the set of links;
S^y	the set of links on path y;
S^{id}	the set of links connecting node i and destination d;
Y^{ij}	the set of paths between nodes i and j;

Y_s^{ij}	the set of paths between nodes *i* and *j* using link *s*;

2.2.2. Functions

c_s	the total expected cost of link *s*;
$\mathbf{c}$	the vector (c_s) with a dimension of $\lvert S \rvert$;
$\overline{C}^{id}$	the perceived minimum expected travel cost between node *i* and destination *d*;
e_s	the exponential function of the total expected travel cost of link *s*;
e_y	the exponential function of the total expected travel cost of path *y*;
h_y^{rd}	passenger flow on path *y* between origin-destination (O-D) pair *rd*;
$m(s)$	an approach using link *s*;
p_s	the fare for passengers using link *s*;
$p_{i,x}^{b}$	the fare charged at stop *i* of route *b* using fare structure *x*;
$u(a)$	the underlying link of approach *a*;
$t(s)$, $h(s)$	the tail and head nodes of link *s*;
v_s	passenger flow on link *s*;
$\overline{v}_s$	passenger flow on the competing links of link *s*;
v_s^d	passenger flow on link *s* to destination *d*;
v_s^{rd}	passenger flow on link *s* traveling from *r* to *d*;
$\mathbf{v}$	the vector of (v_s^d) with a dimension of $\lvert S \rvert \times \lvert D \rvert$;
W_a^d	the weight for passengers who use approach *a* and travel to destination *d*;
α_y^{id}	the probability of passengers using path *y* and traveling from node *i* to destination *d*;

2.2.3. Decision variables

f^b	the frequency of route *b*;
$\mathbf{f}$	the vector (f^b) with a dimension of $\lvert B \rvert$;
q^{rd}	passenger demand from origin *r* to destination *d*;
$\mathbf{q}$	the vector (q^{rd}) with a dimension of $\lvert R \times D \rvert$;
α_a^d	the probability of passengers who use approach *a* and go to destination *d* ;
$\boldsymbol{\alpha}$	the vector (α_a^d) with a dimension of $\left\lvert \bigcup_{i \in N} A_i^+ \right\rvert \times \lvert D \rvert$;
ρ_{flat}^{b}	the fare for passengers using route *b* in the flat fare structure;
ρ_{dist}^{b}	the fare charged per unit distance of route *b* in the distance-based fare structure;
$\rho_{i,\text{sect}}^{b}$	the fare increment at stop *i* along the backward direction of the stop sequence of

	route b in the sectional fare structure;
ρ	the vector (ρ^b_{flat}) with a dimension of $\lvert B \rvert$ in the flat fare structure; the vector (ρ^b_{dist}) with a dimension of $\lvert B \rvert$ in the distance-based fare structure; the vector ($\rho^b_{i,\text{sect}}$) with a dimension of $\lvert I^b \rvert \times \lvert B \rvert$ in the sectional fare structure;

2.2.4. Parameters

$f_{\min}$, $f_{\max}$	the minimum and maximum allowable frequencies;
l_s	the length of link s;
l^b	the total length of route b;
l^b_i	the length of the link connecting node i and the terminal node of route b;
$p_{\min}$, $p_{\max}$	the minimum and maximum allowable fares;
q^{rd}_{base}	the base demand from r to d;
t^b_s	in-vehicle travel time on link s over route b;
σ	the unit conversion parameter in the waiting time function. σ = 60 veh*min/h is used throughout this paper;
ε	the convergence tolerance;
θ	the parameter of the logit model, which reflects the passenger perception of travel cost;
κ^b	the capacity of a single vehicle on route b. The capacity of a double-deck bus in Hong Kong (i.e., κ^b = 150 pass/veh) is used throughout this paper;
λ^k_{DEC}	the step size at the kth iteration in the sensitivity analysis-based descent search method.
μ_{t}, μ_{ω}	the values of in-vehicle travel time and (additional) waiting time. Since the values of in-vehicle travel time and waiting time are not provided in the latest Travel Characteristics Survey 2011 conducted by the Hong Kong Transport Department (Arup Group Limited, 2014), in this paper we used the values adopted in the literature, i.e., $\mu_{\text{t}} = \mu_{\omega}$ = 0.5 HK$/min (Sun and Szeto, 2018);
χ_v, $\chi_{\bar{v}}$	the calibration parameters in the additional waiting time function. χ_v = $\chi_{\bar{v}}$ = 1 is used throughout this paper;
τ	the operating cost per vehicle per unit distance;
ψ^{rd}	the parameter of the linear demand function;
ϖ_s	the calibration parameter in the additional waiting time function associated with link s. ϖ_s = 10 min/h is used throughout this paper;

2.3. Assumptions

In this paper, the model formulation is developed under the following assumptions:

A1) Passengers arrive at transit stops randomly;

A2) Passengers board the first arriving vehicle from a set of attractive transit routes;

A3) The waiting times on links are independent of waiting times on other links;

A4) Transit headways follow exponential distributions;

A5) Passengers select their path choices according to the logit-based stochastic user equilibrium condition.

A6) The passenger demand of each O-D pair is elastic with respect to its minimum expected perceived travel cost. Mathematically, the passenger demand between O-D pair rd can be formulated into

$$q^{rd} = Q^{rd}(\overline{C}^{rd}), \quad \forall r \in R, d \in D, \tag{1}$$

where Q^{rd} is the demand function of O-D pair rd. In this paper, a linear demand function is adopted, as described by

$$q^{rd} = q^{rd}_{\text{base}} - \psi^{rd} \cdot \overline{C}^{rd}, \quad \forall r \in R, d \in D, \tag{2}$$

where q^{rd}_{base} is the base demand between O-D pair rd and ψ^{rd} is a positive parameter;

A7) The transit routes in the network are operated by a single private sector;

A8) The transit operator is allowed to optimize its profit under certain service level requirements;

A9) The fares for various transit routes operated by the single operator can differ from each other even when they share the same stop sequence, assuming that the level of service provided by different transit routes (e.g., the provision/absence of air conditioners) can be different;

A10) For a set of attractive routes with different fares, a weighted mean of fares is used. It is also assumed that the change in fares is marginal and does not change the attractive set of routes;

A11) Similar to the study of Ran and Boyce (1996), passengers between each O-D pair only consider "efficient paths" consisting of links that only take them further away from their origin and closer to their destination. This implies that the sub-network for each O-D pair is acyclic;

A12) All link costs are additive;

A13) The fare structure and level of each transit route are known to all passengers before they decide their paths.

2.4. Link costs

The total expected cost of a link s is formulated as follows.

$$c_s = \mu_{\text{t}} \cdot \frac{\sum_{b \in B_s} f^b t_s^b}{\sum_{b \in B_s} f^b} + \mu_{\omega} \cdot \frac{\sigma}{\sum_{b \in B_s} f^b} + \mu_{\omega} \cdot \varpi_s \left(\frac{\chi_v v_s + \chi_{\bar{v}} \bar{v}_s}{\sum_{b \in B_s} f^b \kappa^b} \right)^n + p_s, \quad \forall s \in S, \tag{3}$$

where μ_{t} and μ_{ω} are the values of in-vehicle travel time and waiting time, respectively. The first term in the equation is the mean in-vehicle travel time cost, calculated by the product of the weighted average of in-vehicle travel times (i.e., t_s^b) of all attractive transit lines associated with link s and the value of in-vehicle travel time. The second term in the equation is the cost of the mean waiting time for the first arriving vehicle in the attractive line set of link s and σ = 60 min/h.

The third term is the perceived congestion time cost, which is a BPR type function calculated using passenger flows on link *s* and its competing links as proposed by Szeto and Jiang (2014a). The competing links of link *s* are defined to be the links associated with passengers who board before/at the tail node of link *s* and alight after the head node of link *s*, using at least one of the attractive lines associated with link *s*. $\chi_v = \chi_{\bar{v}} = n = 1$ and $\varpi_s = 10$ min/h for all *s* are used in this paper. The last term is the mean fare of link *s*, which is calculated by the weighted average of the fares of all the transit lines associated with link *s*. The fare of a transit line depends on the fare structure adopted in the model and will be explained in detail in the following subsection.

2.5. Fare structures

2.5.1. Flat fare

The fare is independent of the transit service provided. Each route in the network charges a fixed price, regardless of the number of stops and distance traveled by the passengers on that route. Let ρ^b_{flat} be the fixed fare of route *b*. Then for a particular link *s*, the expected fare is the weighted sum of the fares of the associated routes as calculated by

$$p_s = \frac{\sum_{b \in B_s} f^b \rho^b_{\text{flat}}}{\sum_{b \in B_s} f^b}, \quad \forall s \in S. \tag{4}$$

2.5.2. Distance-based fare

Following Li et al. (2009), Tsai et al. (2008), and Tsai et al. (2012), the mileage-based distance-based fare is adopted, where the fare charged to each passenger is directly proportional to the distance traveled by him or her. Each route has its fare charging rate, and the fare equals the unit rate multiplied by the distance traveled by the passengers on that route.

Let ρ^b_{dist} be the fare charging rate per unit distance of route *b*. Then for a particular link *s*, the expected fare is

$$p_s = \frac{l_s \sum_{b \in B_s} f^b \rho^b_{\text{dist}}}{\sum_{b \in B_s} f^b}, \quad \forall s \in S, \tag{5}$$

where l_s is the length of link *s*.

2.5.3. Sectional fare

For the sectional fare structure, the transit route is divided into sections. The amount of fare remains unchanged within each section and decreases when crossing sections. Passengers are required to pay the fare when boarding. The fare charged to each passenger depends on which section the passenger's boarding stop belongs to. This kind of fare structure is widely adopted by bus companies in Hong Kong because it is more flexible in price compared with the flat fare structure while retaining simplicity in the fare collection process compared with the distance-based fare structure.

Let $p_{i,\,\text{sect}}^{b}$ be the fare charged at node (stop) *i* of route *b*. The expected fare of link *s* is

$$p_s = \frac{\sum_{b\in B_s} f^b p_{t(s),\,\text{sect}}^{b}}{\sum_{b\in B_s} f^b}, \quad \forall s \in S. \tag{6}$$

In order to handle the constraint that the fare at each stop cannot be higher than that at its previous stops while making the decision variable at each stop nonnegative, the decision variable at that stop is defined as "the fare *increment* at that stop along the *backward* direction of the stop sequence of the route". Therefore, the feasibility constraints for the fare increment $\rho_{i,\,\text{sect}}^{b}$ at stop *i* of route *b* can be defined as

$$\rho_{i,\text{sect}}^{b} \geq 0, \quad \forall i \in I^b, b \in B \ \text{and} \tag{7}$$

$$\sum_{i\in I^b} \rho_{i,\,\text{sect}}^{b} \leq p_{\max}, \quad \forall b \in B, \tag{8}$$

where $p_{\max}$ is the maximum allowable fare.

The fare at node *i* of route *b* is

$$p_{i,\,\text{sect}}^{b} = \rho_{i,\,\text{sect}}^{b} + \sum_{i'\in I_{i+}^{b}} \rho_{i',\,\text{sect}}^{b}, \quad \forall i \in I^b, b \in B. \tag{9}$$

For illustration purposes, Table 3 gives an example of sectional fares of a transit route and the fare increment at each stop.

Table 3. Sectional fares and fare increments of an example transit route

Stop sequence	1	2	3	4 (Destination)
Fare increment $\rho_{i,\,\text{sect}}^{b}$	10	0	5	0
Fare $p_{i,\,\text{sect}}^{b}$	15	5	5	0

Proposition 1 *The flat fare structure is a special case of the sectional fare structure.*

The proof of Proposition 1 is given in Appendix A. According to Proposition 1, the optimal objective value obtained by the sectional fare structure is not worse than that by the flat fare structure, which implies that the sectional fare can outperform the flat fare structure in terms of profit maximization.

Proposition 2 *When the destinations of all passengers are located at the last stop of each transit route, the distance-based fare structure is a special case of the sectional fare structure.*

The proof of Proposition 2 is presented in Appendix A. Proposition 2 implies that the sectional fare structure performs better than the distance-based fare structure for certain types of networks in terms of profit maximization. However, in Section 4.3, examples are given to illustrate that under particular network settings, the distance-based fare structure can achieve a higher profit than the sectional fare structure.

2.6. A bi-level formulation of the problem

2.6.1. The lower level problem

The lower level problem is a logit-based SUEED transit assignment problem. It depicts how passengers select transit lines when the fare and frequency settings are fixed. This problem is formulated using the concept of approach and the theory of the STOCH algorithm (Dial, 1971). Following the definition by Szeto and Jiang (2014a), an approach of a node is a link emanating from that node. The approach does not consider the head node of the link but instead only concerns with the tail node. In other words, an approach is an outgoing link and is solely associated with the tail node. (A link is defined by two nodes but an approach only concerns the tail node.) To illustrate the concept of approaches associated with a node, Figure 2 is given, in which there are two approaches associated with node A in the example network shown in Figure 1.

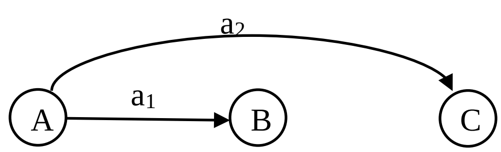


Figure 2. Two approaches associated with node A in the example network shown in Figure 1

The approach probability means that the probability of the passengers who enter the link concerned from the tail node. Under the logit assumption, it is related to the weights of all the approaches coming out from the same node as explained in the following. In contrast to the STOCH algorithm, we calculate the weights of approaches using the exponential functions of the total expected travel cost of links instead. The exponential function of the total expected travel cost of link s is expressed as

$$e_s = \exp\left(-\theta \cdot c_s\right), \quad \forall s \in S. \tag{10}$$

The concept of the backward pass method in Dial's algorithm is applied to formulate the weight of the passengers using approach a and heading towards destination d:

$$W_a^d = e_{u(a)} \cdot \left(\delta_{h(u(a))}^d + \sum_{a' \in A_{h(u(a))}^+} W_{a'}^d \right), \quad \forall a \in A_i^+, i \in N, d \in D, \tag{11}$$

where $\delta_{h(u(a))}^d = 1$ if $h(u(a)) = d$, $\delta_{h(u(a))}^d = 0$ otherwise. The corresponding approach probability is then computed by

$$\alpha_a^d = \frac{W_a^d}{\sum_{a' \in A_{t(u(a))}^+} W_{a'}^d}, \quad \forall a \in A_i^+, i \in N, d \in D. \tag{12}$$

Note that for unused approaches, the corresponding weights and approach probabilities equal zero. Based on Eq. (12), we have $0 \le \alpha_a^d \le 1$ and $\sum_{a \in A_i^+} \alpha_a^d = 1$ for $\forall i \in N, d \in D$.

A path flow probability equals the product of all the approach probabilities of their associated links on the path. Let Y^{ij} be the set of paths connecting nodes *i* and *j* and S^y be the set of links on path *y*. The expected path cost c_y and the probability of path *y* being used, α_y^{id}, can be respectively expressed as

$$c_y = \sum_{s \in S^y} c_s, \quad \forall y \in Y^{ij}; i, j \in N \text{ and} \tag{13}$$

$$\alpha_y^{id} = \prod_{s \in S^y} \alpha_{m(s)}^d, \quad \forall y \in Y^{id}, i \in N, d \in D. \tag{14}$$

Note that $\alpha_a^d = \sum_{y \in Y_{u(a)}^{t(u(a))d}} \alpha_y^{t(u(a))d}, \quad \forall a \in A_i^+, i \in N, d \in D$. The approach probabilities are related to c_y and α_y^{id} as presented in the following.

Proposition 3 *The approach probabilities calculated using Eqs. (10) to (12) satisfy the logit-based SUE condition:* $\alpha_y^{rd} = \dfrac{\exp\left(-\theta \cdot c_y\right)}{\sum_{y' \in Y^{rd}} \left(\exp\left(-\theta \cdot c_{y'}\right)\right)}, \quad \forall y \in Y^{rd}, r \in R, d \in D$.

The proof of Proposition 3 is provided in Appendix A.

The calculation of the weight of an approach requires tracing the weights of all links used between the head node of that approach and its destination. These weights can also be used to calculate the perceived minimum expected travel cost between node *i* and destination *d*, given by

$$\overline{C}^{id} = -\frac{1}{\theta} \cdot \ln\left(\sum_{a \in A_i^+} W_a^d\right), \quad \forall i \in N^d, d \in D. \tag{15}$$

The elastic passenger demand between each O-D pair can, therefore, be calculated using the demand function given by Eq. (1) and expressed in terms of the logsum of the weights as stated below:

Proposition 4 *The passenger demands calculated using Eqs. (1) and (15) satisfy the logit-based elastic demand condition:* $q^{rd} = Q^{rs}\left(-\frac{1}{\theta} \cdot \ln\left(\sum_{y \in Y^{rd}} \left(\exp\left(-\theta \cdot c_y\right)\right)\right)\right), \quad \forall r \in R, d \in D$.

The proof of Proposition 4 is presented in Appendix A.

The passenger flow on link *s* towards destination *d* can be expressed using the concept of the forward pass method. It equals the product of the probability of passengers selecting that link to go to destination *d* and the total inflow rate of its tail node, expressed as

$$v_s^d = \alpha_{m(s)}^d \cdot \left(q^{t(s)d} + \sum_{s' \in A_{t(s)}^-} v_{s'}^d \right), \quad \forall s \in S, d \in D \cdot \tag{16}$$

Note that $q^{t(s)d} = 0$ if $t(s) \notin R$.

The approach-based SUEED assignment problem can be defined by finding **q** and **α** that satisfy Eqs. (1), (3)-(12), (15), and (16) simultaneously. This problem uses passenger demands and approach proportions as the decision variables in which approach proportions take values between zero and one. This problem is different from the standard link-based/path-based SUEED transit assignment problem, which uses both passenger demands and link flows/path flows as the nonnegative decision variables.

With Proposition 3 and Proposition 4, we can conclude that the solution to the approach-based SUEED problem satisfies the logit-based SUEED condition: $\alpha_y^{rd} = \dfrac{\exp(-\theta \cdot c_y)}{\sum_{y' \in Y^{rd}} \left(\exp(-\theta \cdot c_{y'}) \right)}, \quad \forall y \in Y^{rd}, r \in R, d \in D$ and $q^{rd} = Q^{rs} \left(-\dfrac{1}{\theta} \cdot \ln \left(\sum_{y \in Y^{rd}} \left(\exp(-\theta \cdot c_y) \right) \right) \right), \quad \forall r \in R, d \in D$. Based on this result, it is also reasonable and sensible to have Proposition 5 stated below:

Proposition 5 *The approach-based SUEED problem is equivalent to the path-based logit SUEED problem:* $h_y^{rd} = \alpha_y^{rd} \cdot q^{rd} = \dfrac{\exp(-\theta \cdot c_y)}{\sum_{y' \in Y^{rd}} \left(\exp(-\theta \cdot c_{y'}) \right)} \cdot q^{rd}, \quad \forall y \in Y^{rd}, r \in R, d \in D$ *and* $q^{rd} = Q^{rs} \left(-\dfrac{1}{\theta} \cdot \ln \left(\sum_{y \in Y^{rd}} \left(\exp(-\theta \cdot c_y) \right) \right) \right), \quad \forall r \in R, d \in D$.

The proof of Proposition 5 is given in Appendix A. Proposition 5 shows that path flows can be calculated using approach proportions and passenger demands. The path-based logit SUEED problem has a unique passenger demand pattern and a unique flow pattern as discussed by Lam and Zhou (1999). Based on Proposition 5, we can also state the following:

Proposition 6 *There exists a solution to the approach-based SUEED problem in terms of approach probabilities and passenger demands. Moreover, the solution is unique.*

The proof of Proposition 6 is presented in Appendix A.

The approach probabilities defined by Eqs. (10)-(12) and the passenger demands defined by Eqs. (1) and (15) are functions of link travel costs. As a result, the passenger link flows unilaterally determined by approach probabilities and passenger demands as described in Eq. (16) are also functions of link travel costs. On the other hand, the link travel costs defined by Eqs. (3)-(9) are functions of passenger link flows. Therefore, link travel costs are functions of themselves, and the approach-based SUEED problem can be formulated as a fixed point (FP) problem: to find **c** such that

$$\mathbf{c} = \mathbf{F}(\mathbf{c}), \tag{17}$$

where **F** is the mapping function defined by Eqs. (1), (3)-(12), (15), and (16).

The lower level approach-based SUEED problem can be solved by the cost-averaging self-regulated averaging method (c-SRAM) proposed by Long et al. (2014). Similar to the problem studied by Sun and Szeto (2018), the proposed approach-based SUEED problem also satisfies the convergence criterion of the c-SRAM.

2.6.2. The upper level problem

In this paper, our problem of interest is to maximize the transit operator's profit. The transit network design problem can, therefore, be formulated as follows:

$$\max Z(\boldsymbol{\rho}, \mathbf{f}) = \sum_{s \in S} v_s p_s - \tau \cdot \sum_{b \in B} f^b l^b \tag{18}$$

subject to

$$f_{\min} \le f^b \le f_{\max}, \quad \forall b \in B \text{ and} \tag{19}$$

$$p_{\min} \le p^b \le p_{\max}, \quad \forall b \in B. \tag{20}$$

Objective (18) is to maximize the operator's profit calculated by total revenue minus total operating cost. Conditions (19) and (20) are the service level requirement constraints.

Proposition 7 *The optimal profit increases as the unit operating cost τ decreases.*

The proof of Proposition 7 is provided in Appendix A.

Proposition 8 *The optimal total vehicle mileage is monotonically decreasing over* $\tau > 0$.

The proof of Proposition 8 is presented in Appendix A.

The existence and uniqueness of the solution to the lower level problem as stated in Proposition 6 ensures that Z has a well-defined value for any input $(\boldsymbol{\rho}, \mathbf{f})$ to these functions. Moreover, the objective function in (18) is continuous. Furthermore, the solution set formed by constraints (19) and (20) is obviously nonempty and compact. By Weierstrass' Theorem, an optimal solution to the upper level problem exists. However, an example is given in Section 4.1 to illustrate that the solution to the upper level problem may not be unique.

3. Solution method: Sensitivity analysis-based descent search method

Sensitivity analysis has been known to be an effective approach for solving the bi-level programming problems with continuous decision variables. Ying and Miyagi (2001), Ying and Yang (2005), and Ying et al. (2007) have developed the sensitivity analysis of link-based SUE assignment problems based on Dial's (1971) algorithm, which does not require path set enumeration. This approach can be applied to solve an equivalent problem to our proposed bi-level programming problem, in which the lower level problem is formulated as a link-based SUEED instead. However, this paper proposes a more efficient sensitivity analysis-based method to solve our proposed problem.

In this paper, the derivatives of passenger link flows in the lower level problem with respect to the decision variables in the upper level problem, $\nabla_{\mathbf{f}}\mathbf{v}(\boldsymbol{\alpha},\mathbf{q})$ and $\nabla_{\boldsymbol{\rho}}\mathbf{v}(\boldsymbol{\alpha},\mathbf{q})$, are derived to find out the steepest descent direction at any point of the upper level problem. Together with a properly designed step size, some nice properties of the upper level objective function, and an assumption, a stationary point can be obtained by a gradient descent method.

The steepest descent direction at any point of the upper level problem is defined by $\begin{bmatrix} -\nabla_{\mathbf{f}}Z \\ -\nabla_{\boldsymbol{\rho}}Z \end{bmatrix}$, in which $\nabla_{\mathbf{f}}Z$ and $\nabla_{\boldsymbol{\rho}}Z$ are calculated by

$$\nabla_{\mathbf{f}}Z = \frac{\partial Z}{\partial \mathbf{f}} + \frac{\partial Z}{\partial \mathbf{v}}\nabla_{\mathbf{f}}\mathbf{v} \text{ and} \tag{21}$$

$$\nabla_{\boldsymbol{\rho}}Z = \frac{\partial Z}{\partial \boldsymbol{\rho}} + \frac{\partial Z}{\partial \mathbf{v}}\nabla_{\boldsymbol{\rho}}\mathbf{v}. \tag{22}$$

To determine $\nabla_{\mathbf{f}}\mathbf{v}$ and $\nabla_{\boldsymbol{\rho}}\mathbf{v}$ in the above two equations, the lower level approach-based SUEED problem is re-formulated as an FP problem in terms of a passenger link flow vector, expressed as

$$\mathbf{v} = \mathbf{G}(\mathbf{v}), \tag{23}$$

where $\mathbf{G}$ is the mapping function formed by Eqs. (1), (3)-(12), (15), and (16). $\nabla_{\mathbf{f}}\mathbf{v}$ and $\nabla_{\boldsymbol{\rho}}\mathbf{v}$ can then be obtained from the FP problem (23) using the implicit function theorem:

$$\nabla_{\mathbf{f}}\mathbf{v} = [\mathbf{I} - \nabla_{\mathbf{v}}\mathbf{G}]^{-1} \cdot \nabla_{\mathbf{f}}\mathbf{G} \text{ and} \tag{24}$$

$$\nabla_{\boldsymbol{\rho}}\mathbf{v} = [\mathbf{I} - \nabla_{\mathbf{v}}\mathbf{G}]^{-1} \cdot \nabla_{\boldsymbol{\rho}}\mathbf{G}, \tag{25}$$

where $\mathbf{I}$ is the identity matrix with the dimension of $|S \cdot D| \times |S \cdot D|$. As shown above, the inverse operation of the matrix $[\mathbf{I} - \nabla_{\mathbf{v}}\mathbf{G}]$ with the dimension of $|S \cdot D| \times |S \cdot D|$ is required, which is time-consuming.

$\nabla_{\mathbf{v}}\mathbf{G}$ in the preceding two equations can be derived using the product rule of matrix derivatives:

$$\nabla_{\mathbf{v}}\mathbf{G} = \nabla_{\mathbf{c}}\mathbf{G} \cdot \nabla_{\mathbf{v}}\mathbf{c}. \tag{26}$$

Based on Eqs. (16) and (23), the elements in $\nabla_{\mathbf{c}}\mathbf{G}$ in the preceding equation can be calculated using the forward pass method:

$$\frac{\partial G_s^d}{\partial c_{s^\#}} = \frac{\partial \alpha_{m(s)}^d}{\partial c_{s^\#}} \cdot \left(q^{t(s)d} + \sum_{s' \in A_{t(s)}^-} v_{s'}^d \right) + \alpha_{m(s)}^d \cdot \left(\frac{\partial q^{t(s)d}}{\partial c_{s^\#}} + \sum_{s' \in A_{t(s)}^-} \frac{\partial G_{s'}^d}{\partial c_{s^\#}} \right), \quad \forall s, s^\# \in S, d \in D. \tag{27}$$

Note that $q^{t(s)d} = 0$ if $t(s) \notin R$; $\frac{\partial q^{t(s)d}}{\partial c_{s^\#}} = 0$ if $q^{t(s)d} = 0$.

$\frac{\partial \alpha_{m(s)}^d}{\partial c_{s'}}$ and $\frac{\partial q^{t(s)d}}{\partial c_{s'}}$ in the preceding equation can also be obtained by the forward pass method:

$$\frac{\partial \alpha_a^d}{\partial c_s} = \alpha_{m(s)}^d \cdot \left(\left(\delta_s^a - \alpha_a^d \right) \cdot (-\theta) \cdot \delta_s^{t(u(a))} + \sum_{s' \in A_{t(s)}^-} \left(\frac{\partial \alpha_a^d}{\partial c_{s'}} \right) \right), \quad \forall s \in S^{t(u(a))d}, a \in A_i^+, i \in N, d \in D, \quad (28)$$

where $\delta_s^a = 1$ if $s = u(a)$, $\delta_s^a = 0$ otherwise; $\delta_s^{t(u(a))} = 1$ if $t(s) = t(u(a))$, $\delta_s^{t(u(a))} = 0$ otherwise. Note that $\frac{\partial \alpha_a^d}{\partial c_s} = 0$ if $s \notin S^{t(u(a))d}$. The derivation of Eq. (28) is shown in Appendix B.

The expression of $\frac{\partial q^{rd}}{\partial c_s}$ depends on the form of the demand function used. In this paper, we adopt the linear demand function as presented in Eq. (2). The derivative can be obtained by the forward pass method:

$$\frac{\partial q^{rd}}{\partial c_s} = \alpha_{m(s)}^d \cdot \left(\left(-\psi^{rd} \right) \delta_s^r + \sum_{s' \in A_{t(s)}^-} \frac{\partial q^{rd}}{\partial c_{s'}} \right), \quad \forall s \in S^{rd}, r \in R, d \in D, \quad (29)$$

where $\delta_s^r = 1$ if $t(s) = r$, $\delta_s^r = 0$ otherwise. Note that $\frac{\partial q^{rd}}{\partial c_s} = 0$ if $s \notin S^{rd}$. The derivation of Eq. (29) is given in Appendix B.

$\frac{\partial c_s}{\partial v_{s^\#}^d}$, an element of $\nabla_{\mathbf{v}} \mathbf{c}$, can be calculated by

$$\frac{\partial c_s}{\partial v_{s^\#}^d} = \varpi_s \cdot n \cdot \left(\frac{\chi_v v_s + \chi_{\bar{v}} \bar{v}_s}{\sum_{b \in B_s} f^b \kappa^b} \right)^{n-1} \cdot \frac{\left(\chi_v \cdot \delta_{s^\#}^s + \chi_{\bar{v}} \cdot \bar{\delta}_{s^\#}^s \right)}{\sum_{b \in B_s} f^b \kappa^b}, \quad \forall s, s^\# \in S, \forall d \in D, \quad (30)$$

where $\delta_{s^\#}^s = 1$ if $s^\# = s$, $\delta_{s^\#}^s = 0$ otherwise; $\bar{\delta}_{s^\#}^s = 1$ if $s^\# \in \bar{A}_s$, $\bar{\delta}_{s^\#}^s = 0$ otherwise.

Using the chain rule, $\nabla_{\mathbf{f}} \mathbf{G}$ and $\nabla_{\boldsymbol{\rho}} \mathbf{G}$ in Eqs. (24) and (25) can be expressed as follows:

$$\nabla_{\mathbf{f}} \mathbf{G} = \nabla_{\mathbf{c}} \mathbf{G} \cdot \nabla_{\mathbf{f}} \mathbf{c} \text{ and} \quad (31)$$

$$\nabla_{\boldsymbol{\rho}} \mathbf{G} = \nabla_{\mathbf{c}} \mathbf{G} \cdot \nabla_{\boldsymbol{\rho}} \mathbf{c}. \quad (32)$$

The expression of elements in $\nabla_{\mathbf{f}} \mathbf{c}$ in Eq. (31) can be calculated by

$$\frac{\partial c_s}{\partial f^b} = \frac{\mu_t t_s^b + p_s^b - \left(\mu_t t_s + \mu_\omega \omega_s + p_s \right)}{\sum_{b' \in B_s} f^{b'}} - \frac{n \cdot \kappa^b \cdot \mu_\omega \phi_s}{\sum_{b' \in B_s} f^{b'} \kappa^{b'}}, \quad \forall s \in S, b \in B_s. \quad (33)$$

The expression of elements in $\nabla_{\boldsymbol{\rho}} \mathbf{c}$ in (32) depends on the fare structure selected. For the three different fare structures, we have

$$\frac{\partial c_s}{\partial \rho^b_{\text{flat}}} = \frac{f^b}{\sum_{b' \in B_s} f^{b'}}, \quad \forall s \in S, b \in B_s, \tag{34}$$

$$\frac{\partial c_s}{\partial \rho^b_{\text{dist}}} = \frac{l_s f^b}{\sum_{b' \in B_s} f^{b'}}, \quad \forall s \in S, b \in B_s, \text{ and} \tag{35}$$

$$\frac{\partial c_s}{\partial \rho^b_{i,\text{ sect}}} = \frac{\delta^s_i f^b}{\sum_{b' \in B_s} f^{b'}}, \quad \forall s \in S, i \in I^b, b \in B_s, \tag{36}$$

where $\delta^s_i = 1$ if $i = t(s)$ or $i \in I^b_{t(s)+}$, $\delta^s_i = 0$ otherwise.

Note that the calculation of the aforementioned derivative information does not require path set enumeration.

The procedure of the heuristic is given as follows:

1) Set the initial frequency and fare vectors $\mathbf{f}^0$ and $\boldsymbol{\rho}^0$, and the convergence tolerance ε. Set $k = 0$;
2) Solve the lower-level problem defined by Eqs. (1), (3)-(12), (15), and (16) for given $\mathbf{f}^k$ and $\boldsymbol{\rho}^k$ using c-SRAM to obtain $\boldsymbol{\alpha}^k$, $\mathbf{v}^k$, and $\mathbf{q}^k$;
3) Calculate the sensitivity information $\nabla_{\mathbf{f}}\mathbf{v}$ and $\nabla_{\boldsymbol{\rho}}\mathbf{v}$ using $\boldsymbol{\alpha}^k$, $\mathbf{v}^k$, and $\mathbf{q}^k$ and Eqs. (24)-(36);
4) Calculate $\nabla_{\mathbf{f}^k} Z$ and $\nabla_{\boldsymbol{\rho}^k} Z$ using Eqs. (21) and (22);
5) Update $\mathbf{f}^{k+1} := \mathbf{f}^k - \lambda^k_{\text{DEC}} \nabla_{\mathbf{f}^k} Z$; $\boldsymbol{\rho}^{k+1} := \boldsymbol{\rho}^k - \lambda^k_{\text{DEC}} \nabla_{\boldsymbol{\rho}^k} Z$, where the step size λ^k_{DEC} is determined by an inexact line-search method so that the upper boundary constraints (19) and (20) hold;
6) If $\left\| \nabla_{\boldsymbol{\rho},\mathbf{f}^k} Z \right\| < \varepsilon$, then stop. Otherwise, set $k = k + 1$ and return to Step 2).

As discussed in Section 2.6.2, the objective function in (18) is continuous and the solution set formed by constraints (19) and (20) is nonempty and compact. The proposed heuristic converges to a stationary point to the studied problem provided that the upper level objective function is continuously differentiable (i.e., the first-order derivatives exist at any point of the feasible region and are continuous) and $[\mathbf{I} - \nabla_{\mathbf{v}}\mathbf{G}]^{-1}$ always exists. If these convergence criteria are not satisfied, the algorithm may stop before a stationary point is found. If the non-convergence is only due to the non-existence of derivatives at some points, the first-order derivatives can be computed by approximation instead to obtain a stationary point. If $[\mathbf{I} - \nabla_{\mathbf{v}}\mathbf{G}]^{-1}$ and the first-order derivatives always exist, the heuristic becomes a classical gradient descent algorithm.

In the proposed solution methodology, the network loading process as described in Eq. (16) can be done for one origin to all destinations simultaneously using the approach probabilities (defined by the probability of using a specific link towards a specific destination). In contrast, when using the link-based sensitivity analysis algorithms, equilibrium link flows are obtained for each O-D pair at a time as in the STOCH algorithm (Dial, 1971). Each link common to the sub-networks for different O-D pairs but with the same origin is only required to be handled once in our solution methodology, whereas each common link is required to be handled more than once in the STOCH algorithm (The number of times required to handle one common link equals the number of sub-networks with the same origin including this link). Therefore, our proposed algorithm is more computationally efficient.

4. Numerical studies

4.1. Multiple solutions to the bi-level problem

To illustrate that optimal solutions to the upper level problem (18)-(20) may not be unique, an example transit network as shown in Figure 3 was developed with $\theta = 0.1$ and $\tau = 6$. The feasible range of frequency was set to be from 4.5 veh/h to 20 veh/h. The feasible range of the fares was set to be from HK$ 0 to HK$ 25. The parameters of the demand function are $q^{rd}_{\text{base}} = 500$ and $\psi^{rd} = 1.2$.

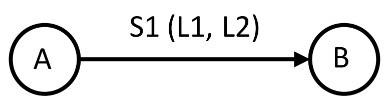


Figure 3. Example transit network I

Due to the simple network topology, the flat, sectional, and distance based fares are identical. The example was solved using the Excel Solver. Three different optimal fare and frequency settings are shown in Table 4. As revealed in Figure 3, L1 and L2 are the two common lines of link S1. Since their optimal fares at node A have reached their upper bound (i.e., 25 HK$), the total expected link cost of S1 calculated by Eqs. (3)-(9) and the total profit calculated by Eq. (18) do not change if the sum of the frequencies of L1 and L2 remains unchanged. Therefore, the solution stays optimal when $f_{\text{L1}} + f_{\text{L2}} = 9.96$ veh/h. This example shows that under certain network settings, multiple optimal solutions to the bi-level fare and frequency design problem exist.

Table 4. Multiple optimal solutions

Route	Decision variable	Optimal solutions		
		1	2	3
L1	Fare charged at node A (HK$)	25	25	25
	Frequency (veh/h)	**4.5**	**4.98**	**5.46**
L2	Fare charged at node A (HK$)	25	25	25
	Frequency (veh/h)	**5.46**	**4.98**	**4.5**

4.2. The effect of the unit operating cost

Figure 4 shows a small example transit network with in-vehicle travel time (in min) indicated next to each link. Nodes 1 to 5 are origin nodes and nodes 6 and 7 are destination nodes. There are 10 O-D pairs in total. There is a transfer node T where passengers can make transfers. This example network is used for the numerical examples in Section 4 unless specified otherwise.

The base demand of each O-D pair is given in Table 5. The linear demand function as presented in Eq. (2) with ψ = -1.07 was used for all O-D pairs. This demand setting is used for the numerical examples in Section 4 unless specified otherwise. The transit network presented in Figure 4, its fixed routes, and their flat fares shown in Table 6 are used in this example. In this example, the maximum and minimum allowable frequencies of each route were assumed to be 40 veh/h and 4.5 veh/h, respectively. Unless specified otherwise, the heuristic for the numerical examples in Section 4 was coded and compiled using Dev C++ 5.11 and ran on a personal computer with a 3.4-GHz Core processor and 16 GB RAM.

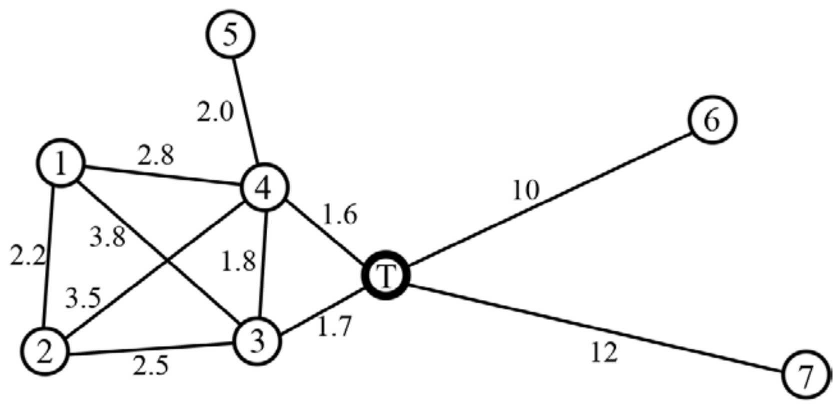


Figure 4. Example transit network II

Table 5. The base demand (pass/h) of each O-D pair in the network

D \ O	1	2	3	4	5
6	576	300	218	108	177
7	782	268	288	145	178

Table 6. Fixed routes and flat fares

Route	Stop sequence	Flat fare (HK$)
1	5-4-T-7	10
2	5-4-T-6	10
3	2-3-4-T-7	10
4	1-3-4-T-6	10

When θ = 0.1, the total (optimal) vehicle mileage of all routes against the unit operating cost τ for each route is plotted in Figure 5. Since the operating cost of each route is directly related to the frequency and length of the route, when the value of τ increases, the total vehicle mileage tends to be lowered to reduce the operating cost. However, the total vehicle mileage obtained is not always decreasing with the increase in the value of τ. When τ is sufficiently large, the total vehicle mileage obtained is bounded by the minimum frequency constraint. This result is consistent with the finding of Proposition 8.

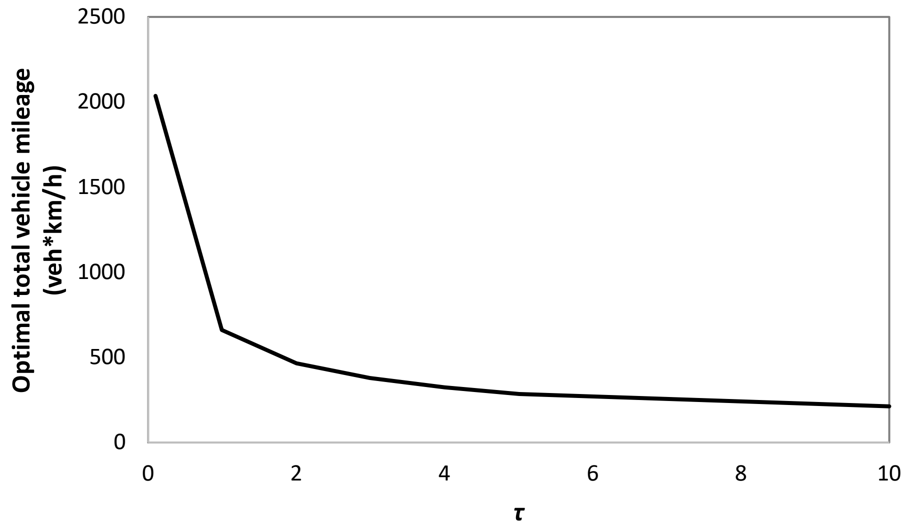


Figure 5. Optimal total vehicle mileage against τ

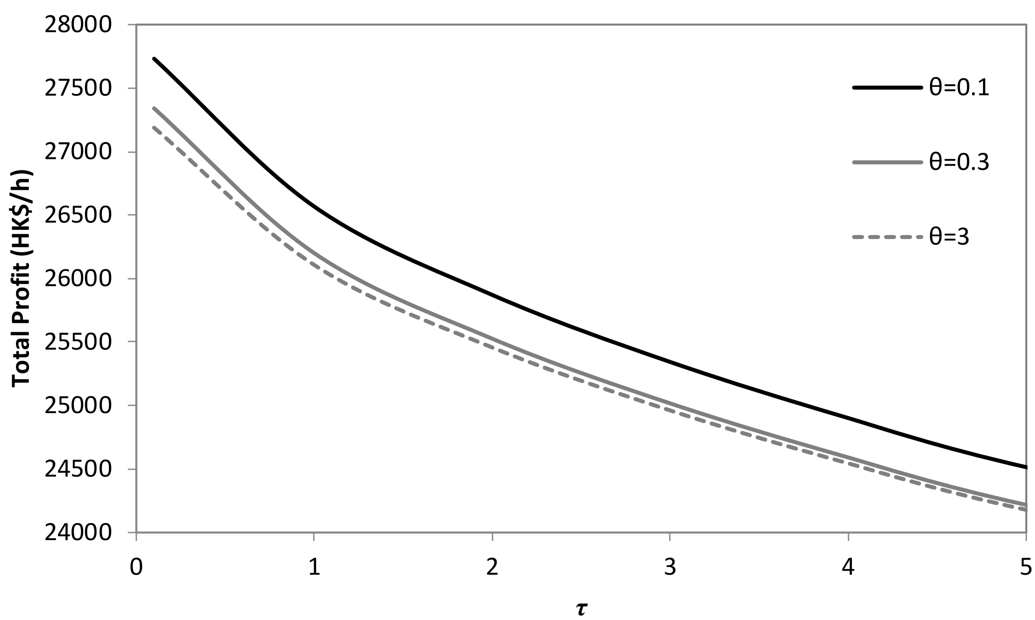


Figure 6. Total profits obtained under different values of θ against τ

Figure 6 displays the total profits obtained under different values of θ against τ. As shown in the figure, for any given value of θ, the total profit obtained decreases as the value of τ increases. It is because for a given set of frequencies, the revenue is fixed under a given set of fares but a higher value of τ gives a larger total operating cost. When τ is sufficiently large, the optimal frequencies are bounded by the minimum frequency constraint (which means that the optimal frequencies remain unchanged) and therefore the total profit obtained decreases linearly against τ. However, when τ is not sufficiently large, the optimal frequency setting and hence the revenue change with τ in addition to the operating cost. Consequently, the total profit decreases nonlinearly against τ. This result is consistent with the finding of Proposition 7.

Figure 6 also reveals that the total profit obtained decreases as the value of θ increases. It is because when passengers have little information about their actual travel cost (i.e., when the value of θ is small), they may choose unreasonably expensive routes, which can increase the total profit. The difference in the total profits obtained under different values of θ can be significant. For example, when $\tau = 3$ and $\theta = 0.1$, the profit is HK\$ 3376.73; when θ increases to 3, the profit is HK\$ 2583.52, in which the difference is about 23%. It is therefore important to estimate the value of θ accurately to determine the total profit in reality.

4.3. A comparison of the three fare structures

To compare the three different fare structures, their optimal solutions were obtained based on the transit network presented in Figure 4, the demand setting and fixed routes in Section 4.2.

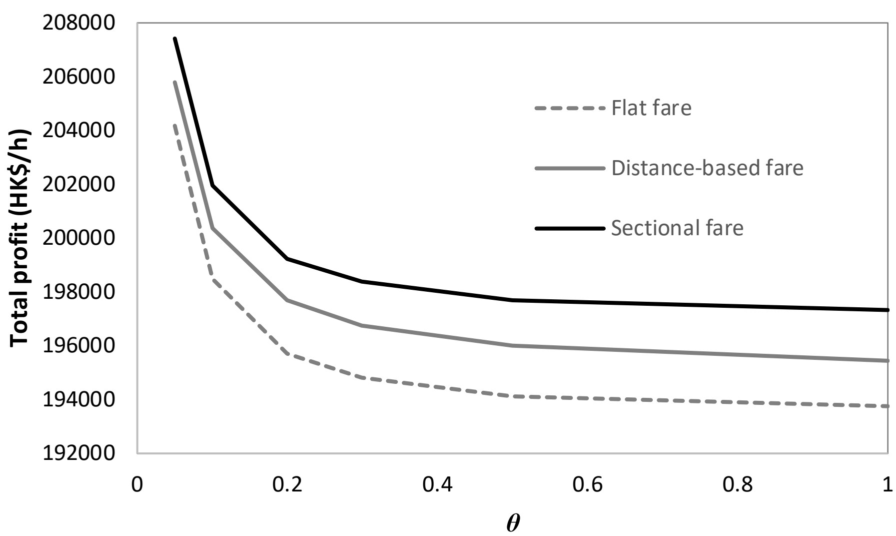


Figure 7. Total profits obtained by different fare structures against θ

The total profits obtained by different fare structures against θ is shown in Figure 7. As reflected by the figure, the total profit obtained by the sectional fare structure is the highest while the total profit obtained by the flat fare structure is the lowest.

For the comparison between the sectional and flat fare structures, the result is in line with expectation: the optimal objective value obtained by the sectional fare structure is not worse than that by the flat fare structure as stated in Proposition 1.

For the comparison between the sectional and distance-based fare structures, since the transit network instance satisfies the condition stated in Proposition 2 where all destinations of passengers are located at the two transit terminals, the obtained objective value using the sectional fare structure is higher as expected.

For more general transit networks with multiple destinations along a transit route, the flat fare structure is still a special case of the sectional fare counterpart. Therefore, the optimal objective value obtained by the sectional fare structure is always not worse than that by the flat fare structure. However, since the same sectional fare is charged to the passengers boarding at the same stop regardless of their trip lengths, the distance-based fare structure is no longer a special structure of the sectional fare counterpart.

In order to make a further comparison between the distance-based and sectional fare structures without the assumption that all passengers alight at the same destination, a small example network was developed as presented in Figure 8. There are two O-D pairs in the network (i.e., O-D pairs (A,B) and (A,C)) with different demand levels and one transit line serving the network (i.e., L1) with a fixed frequency of 5 veh/h. The setting of line L1 is given in Table 7. Note that there is no demand from node B to node C, and hence the sectional fare at node B does not affect the flow pattern and is not a decision variable in this example. It is also noted that the sectional fares on using the two links are the same because both links start from the same node.

Let l_1 and l_2 be the lengths of S1 and S2, respectively. The following three cases are considered: in case 1, the lengths of S1 and S2 and the maximum allowable fare are fixed while the ratio of the demand between O-D pair (A,B) to that between O-D pair (A,C) is varying; in case 2, the demands between O-D pairs (A,B) and (A,C) and their maximum allowable fare are fixed while the ratio of the length of S1 and to that of S2 is varying; in case 3, the demands between O-D pairs (A,B) and (A,C), and the lengths of S1 and S2 are fixed while the maximum allowable fare is varying. In each case, there are different numbers of subcases as shown in Table 8. The detailed characteristics of each subcase are also presented in Table 8. The optimal fare settings under both the sectional and distance-based fare structures were obtained in each subcase, assuming $\theta = 0.5$ and $\psi = 0.5$.

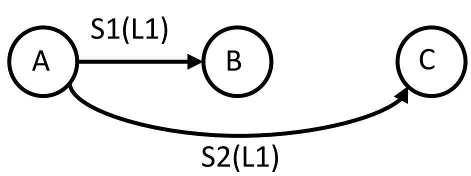


Figure 8. Example transit network III

Table 7. The setting of line L1

Stop sequence	A – B – C
Links	A-B (S1, l_1); A-C (S2, l_2)
Distance-based fares	$\rho_{\text{distance}} \cdot l_1$; $\rho_{\text{distance}} \cdot l_2$
Sectional fares	$p_{\text{A,sectional}}$; $p_{\text{A,sectional}}$

Table 8. The cases' characteristics

	Demand (pass/h)	The in-vehicle travel time of each link (min)	Maximum allowable fare (HK$)
Case 1 (varying passenger demand)	(A,B): 300; (A,C): 100 ($\frac{q^{AC}}{q^{AB}+q^{AC}}=$ 0.25) (A,B): 200; (A,C): 200 ($\frac{q^{AC}}{q^{AB}+q^{AC}}=0.5$) (A,B): 100; (A,C): 300 ($\frac{q^{AC}}{q^{AB}+q^{AC}}=$ 0.75)	S1: 40; S2: 80	50
Case 2 (varying the lengths of links)	(A,B): 100; (A,C): 300	S1: 10; S2: 190 ($\frac{l_1}{l_2}=0.05$) S1: 50; S2: 150 ($\frac{l_1}{l_2}=0.25$) S1: 100; S2: 100 ($\frac{l_1}{l_2}=0.5$) S1: 150; S2: 50 ($\frac{l_1}{l_2}=0.75$) S1: 190; S2: 10 ($\frac{l_1}{l_2}=0.95$)	50
Case 3 (varying the maximum allowable fare)	(A,B): 100; (A,C): 300	S1: 40; S2: 80	25 50 75

The optimal profit obtained in each case using each of the two fare structures is shown in Figure 9. The corresponding fare setting of each case is shown in Table 9. This table also gives the fare settings for all the subcases if they are different. As shown in Table 9, in all cases, the optimal fares reach the upper bound. As a result, the distance-based fare for the short-distance trip (i.e., O-D pair (A,B)) is lower than the sectional fare for that trip in all cases. Note that for the example network presented in Figure 8, all passengers board L1 at node A, and hence the sectional fare is equivalent to the flat fare. This implies that the conclusion is also applicable to the comparison between the flat and distance-based fare structures.

Table 9. The optimal fare setting of each case and subcase

	Sectional fare (HK$)	Distance-based fare (HK$)
Case 1	S1: 50; S2: 50	S1: 25; S2: 50
Case 2	S1: 50; S2: 50	S1: 2.5; S2: 50
		S1: 12.5; S2: 50
		S1: 25; S2: 50
		S1: 37.5; S2: 50
		S1: 47.5; S2: 50
Case 3	S1: 25; S2: 25	S1: 12.5; S2: 25
	S1: 50; S2: 50	S1: 25; S2: 50
	S1: 75; S2: 75	S1: 37.5; S2: 75

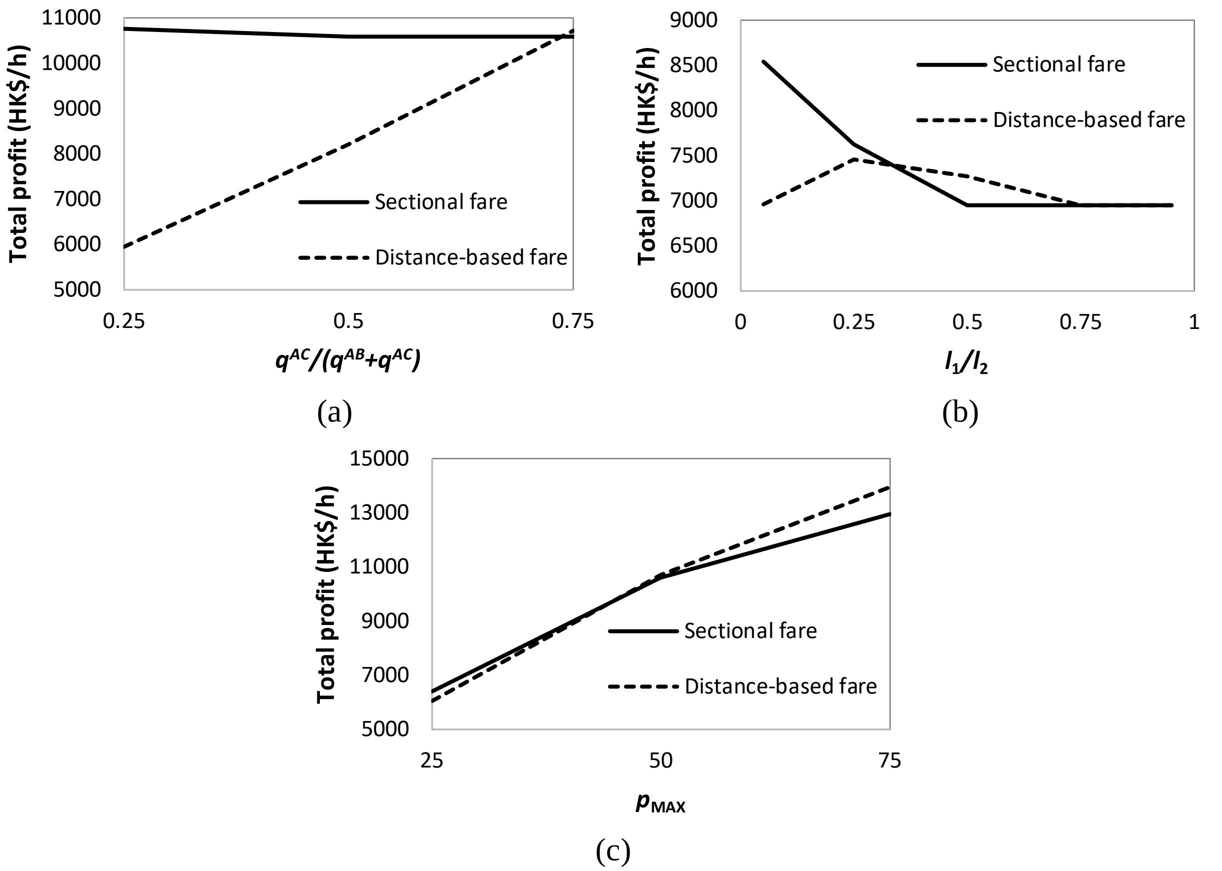


Figure 9. Optimal profits obtained by both fare structures in (a) case 1, (b) case 2, and (c) case 3

In case 1, as shown in Figure 9(a), as the proportion of long-distance-demand (i.e., q^{AC}) increases, the optimal profit obtained by the distance-based fare structure increases significantly. At $q^{AC}/(q^{AB}+q^{AC})$ = 0.75, the optimal profit obtained by the distance-based fare structure is higher than that by the sectional fare structure. This result is consistent with the observation by Ling (1998), by which the flat fare structure (the sectional fare structure in our case) is not favorable in terms of revenue maximization when the number of long trips is significantly greater than that of short trips under a fixed operating cost.

In case 2, as shown in Figure 9(b), the optimal profit obtained by the sectional fare structure monotonically decreases as l_1/l_2 increases; the optimal profit obtained by the distance-based fare structure first increases and then decreases as l_1/l_2 increases; at l_1/l_2 = 0.5, the optimal profit obtained by the distance-based fare structure is higher than that by the sectional fare structure. This result implies that the performance of the two fare structures is also affected by the geometry of the network.

In case 3, as shown in Figure 9(c), the optimal profits obtained by both fare structures monotonically increase as $p_{\max}$ increases. The distance-based fare structure outperforms the sectional fare counterpart in terms of profitability at $p_{\max}$ = 50 and $p_{\max}$ = 75, but is worse off at $p_{\max}$ = 25. As shown in Table 9, the fares charged to long-distance passengers (i.e., q^{AC}) using both fare structures are the same and equal to the upper bound while the distance-based fare charged to passengers on link S1 (i.e., short-distance demand) is fixed to half of the sectional fare. This implies that the difference in the resultant profit comes mainly from the fare charged to the short-distance demand (i.e., q^{AB}). It is because $p_{\max}$ affects the fare charged to passengers on Link S1 due to the fare structure requirement, which in turn affects the demand level on S1 due to demand elasticity.

Concluding from the three cases, the choice between sectional and distance-based fare structures is affected by the geometry of the network (e.g., route structure and distance between stops), the demand distribution, and the maximum allowable fares. While using the sectional fare structure can yield a better profit compared to using the other two fare structures in certain kinds of networks, the number of decision variables in the sectional fare structure (i.e., roughly equal to $\left|I^b\right|\times\left|B\right|$) is significantly larger compared to that in each of the other two fare structures (i.e., $\left|B\right|$). Therefore, a longer computational time is expected for the optimization process of the sectional fare structure.

4.4. The influences of the passenger perception of travel cost

We consider two cases of the effects of the passenger perception of travel cost θ. Firstly, the transit network presented in Figure 4 and the fixed routes shown in Table 10 are used for the sensitivity analysis of θ in the lower level problem. We set τ = 3. The example was solved using the Excel Solver.

Table 10. Fixed routes

Route	Stop sequence	Fare (HK$)	Frequency (veh/h)
1	1-4-T-7	10	5
2	5-4-T-6	10	5
3	2-3-T-6	10	5
4	2-4-T-7	10	5

The plot of total ridership in the network against θ is shown in Figure 10. As demonstrated in the figure, the total ridership decreases as the value of θ increases. When the value of θ is sufficiently large, the SUE condition is equivalent to the UE condition and therefore the total ridership approaches the UE limit.

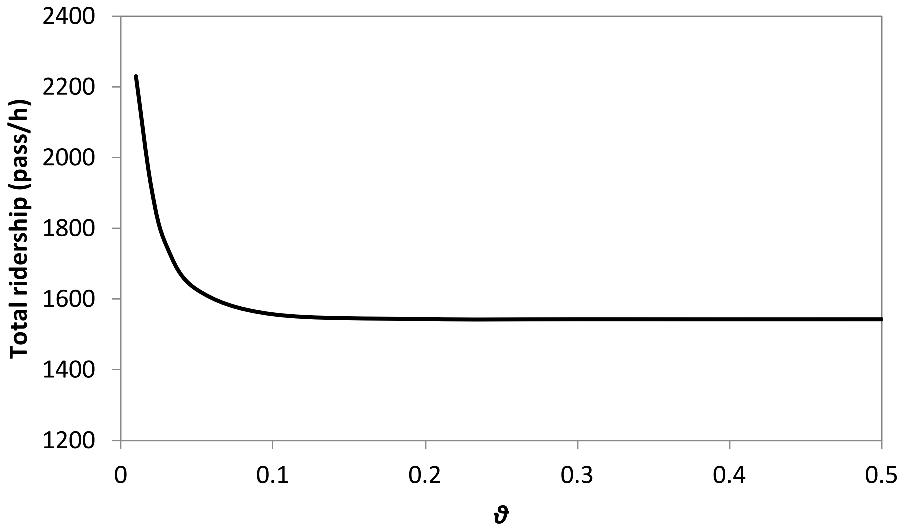
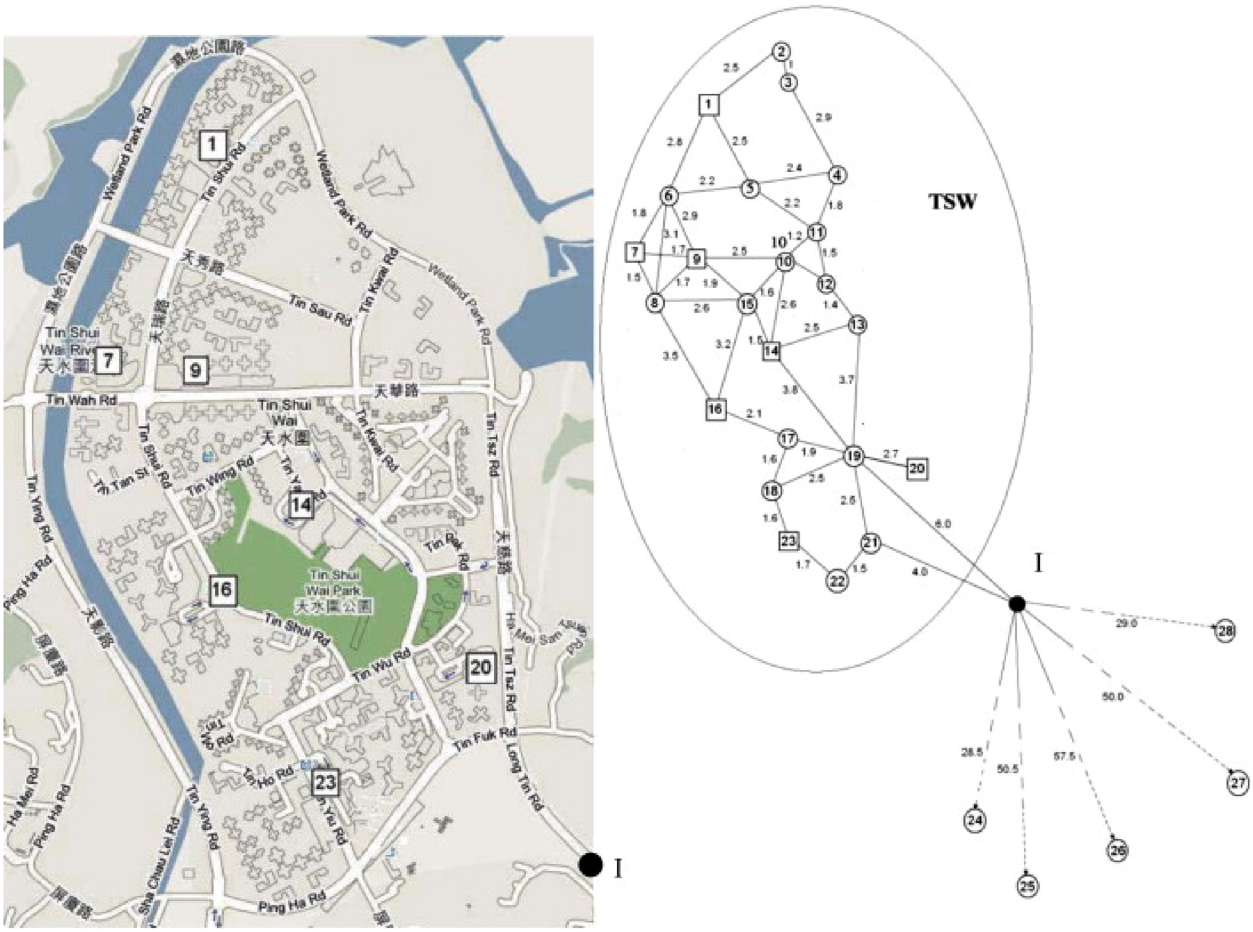


Figure 10. Total ridership under different values of θ

Secondly, for the bi-level optimization problem, we can observe from Figure 6 and Figure 7 that the optimal profit obtained decreases as the value of θ increases. This result implies that from the transit operator's point of view, providing better information to the passengers may not be good in terms of profitability.

4.5. An optimal solution of the TSW instance

In this example, a bilevel profit maximization model was developed and solved based on the bus network of Tin Shui Wai (TSW), a suburban residential area in Hong Kong, as shown in Figure 11(a). Trunk buses serve as the main transportation mode for the residents in TSW working in the urban areas of Hong Kong. Currently, all the trunk bus routes pass through the Tai Lam Tunnel (TLT) from the TSW area to the urban areas. Free transfers can be made at the bus interchange at TLT (marked as "I" in Figure 11). A simplified transit network of the TSW area is presented in Figure 11(b). The TSW network consists of 23 origin nodes (i.e., nodes 1-23), 5 destination nodes (i.e., nodes 24-28), and a transfer node "I". The in-vehicle travel times (in minutes) between nodes are also shown in the figure. There are 115 O-D pairs in the network and the maximum demand of each O-D pair is shown in Table 11. The maximum demands are adopted from the fixed passenger demand data used by Szeto and Jiang (2014b). The O-D demand function is linear as described by Eq. (2).



(a) Map of Tin Shui Wai (b) The Tin Shui Wai bus network

Figure 11. The TSW network (Szeto and Jiang, 2014b)

Table 11. The base demand (pass/h) and the slope of demand function of each O-D pair in the network

D / O	24	25	26	27	28
1	384	296	204	188	298
2	108	78	76	44	108
3	94	80	76	54	110
4	66	44	42	28	60
5	200	148	156	112	204
6	174	154	142	92	226
7	226	152	142	92	206
8	200	152	142	94	234
9	192	126	98	68	170
10	66	48	38	30	68
11	38	28	28	18	46
12	312	268	228	138	330
13	354	210	180	156	286
14	126	96	72	58	118
15	204	162	126	78	186
16	506	340	300	254	426
17	56	40	40	28	54
18	152	126	116	76	142
19	68	50	44	28	60
20	118	78	60	52	98
21	72	46	44	30	56
22	66	50	40	32	56
23	412	368	294	192	418
ψ^d	-0.6	-0.5	-0.6	-0.5	-0.6

Table 12. Fixed routes

Route	Stop sequence
1	20, 19, I, 25
2	1, 6, 8, 16, 17, 18, 23, 22, 21, I, 28
3	9, 10, 11, 5, 6, 8, 16, 17, 18, 23, 22, I, 27
4	7, 6, 1, 2, 3, 4, 11, 12, 13, 19, I, 24
5	14, 15, 8, 9, 10, 12, 13, 19, I, 26

The TSW network instance satisfies the condition stated in Proposition 2. As pointed out in Proposition 1 and Proposition 2, the sectional fare structure is always better than the flat and distance-based fare structures. Therefore, the optimization of the fare and frequency setting was performed based on the sectional fare structure only. The total base demand of the instance is 15924 pass/hr. In the TSW network instance, the numbers of links and approaches equal 206 and 783, respectively. The fixed routes in the instance are presented in Table 12. The feasible range of frequency of each route was set to be from 1 veh/h to 60 veh/h. The feasible range of fare of each route was set to be from 0 HK$ to 25 HK$. The initial fare and frequency settings are shown in Table 13. An optimal solution was obtained based on the parameters of $\theta = 0.5$ and $\tau = 50$.

Table 13. The initial fare and frequency settings of the fixed routes

Route	The fare at each stop (HK$)	Frequency (veh/h)
1	0	1
2	0	1
3	0	1
4	0	1
5	0	1

The objective value (i.e., total profit) as well as $\left\|\nabla_{\mathbf{p},\mathbf{f}^k} Z\right\|$ over iterations are plotted in Figure 12 and Figure 13, respectively. As we can see, the proposed solution scheme converges quickly (i.e., within 9 iterations). It took 104.2 seconds to obtain a convergent solution. The optimal fare and frequency settings are presented in Table 14. The optimal profit obtained under these fare and frequency settings is 214650.5 HK$/h.

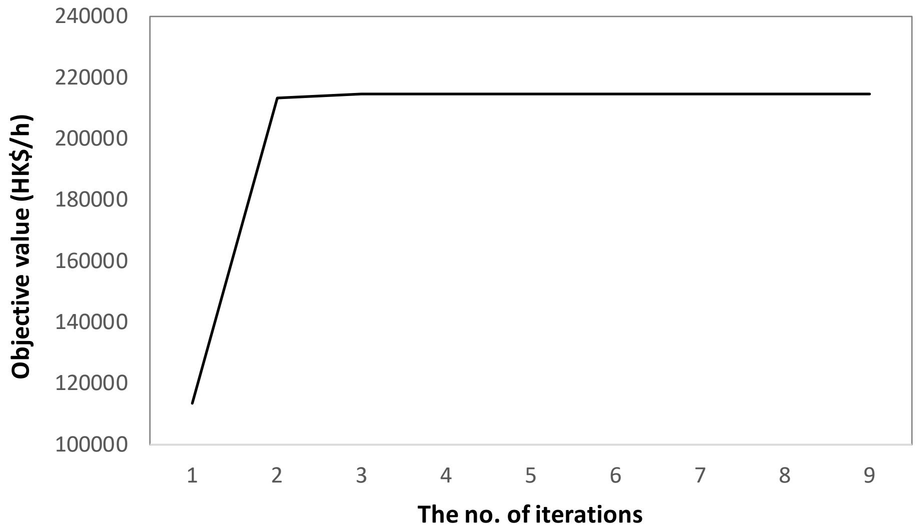


Figure 12. The objective value at each iteration

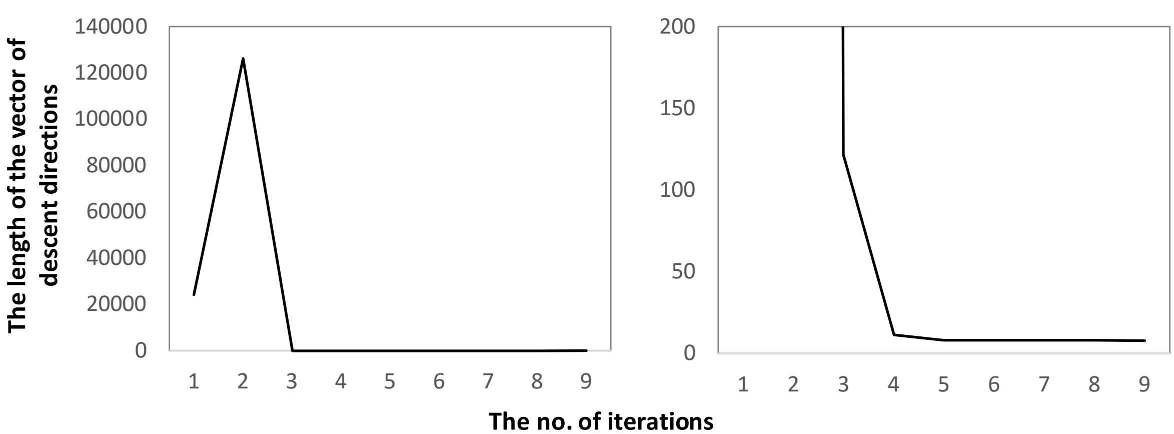


Figure 13. The length of the vector of descent directions $\left\|\nabla_{\mathbf{p},\mathbf{f}^k} Z\right\|$

Table 14. The optimal fare and frequency settings of the fixed routes

Route	The fare at each stop (HK$)	Frequency (veh/h)
1	25 – 24.52 – 15.68	43.48
2	25 – 24.22 – 23.03 – 21.51 – 18.99 – 16.47 – 13.93 – 10.53 – 7.14 – 3.74	60
3	25 – 24.56 – 24.08 – 23.59 – 22.53 – 21.12 – 19.54 – 16.76 – 13.98 – 11.18 – 7.50 – 3.82	20.41
4	25 – 24.44 – 23.26 – 21.30 – 19.32 – 17.34 – 15.36 – 13.38 – 10.72 – 7.47 – 4.19	38.98
5	25 – 24.95 – 23.94 – 21.81 – 19.54 – 17.27 – 13.79 – 9.58 – 5.32	60

5. Conclusion

In this paper, a bi-level profit maximization model is developed to determine the optimal fare and frequency settings under approach-based SUEED. The sensitivity analysis-based descent search method that takes into account approach-based SUEED is proposed to solve the model. The derivatives of the passenger link flows in the approach-based SUEED assignment problem with respect to the frequency and fare variables were derived and used in the proposed solution method. The effectiveness of the proposed solution method is also illustrated based on the TSW transit network instance.

Mathematical analysis and numerical tests were carried out to investigate the model properties and compare flat, distance-based, and sectional fare structures in terms of profitability. It is proven that the lower level approach-based SUEED assignment problem has exactly one solution. However, it is shown by an example that the bi-level problem can have multiple optimal solutions. It is proven that when all destinations are located at transit terminals, the sectional fare structure is always more profitable than the other two fare structures. For more general networks, the sectional fare structure is always better than the flat fare structure, but the choice between the sectional and distance-based fare structures depends on the geometry of the network (e.g., route structure and the distance between stops), the demand distribution, and the maximum allowable fares. The results also show that from the operator's point of view, providing better information to the passengers may not be good in terms of profitability. The results further demonstrate that the optimal profit increases as the unit operating cost decreases, and that the optimal total vehicle mileage is monotonically decreasing with respect to the unit operating cost. These relationships are also confirmed by mathematical proofs.

Our findings should be interpreted with the following limitations. Our model is frequency-based and is suitable for long-term planning purposes from the perspective of private operators. It omits some detailed operation issues such as vehicle scheduling and fleet size determination for simplicity. Moreover, in the proposed model, the congestion and overcrowding issues are captured using the congestion cost function approach. This approach cannot explicitly restrict the maximum number of passengers boarding/sitting in a transit vehicle. As a result, it is only suitable for planning purposes at the design stage of the transit service when the service decisions are still subject to changes. Furthermore, this study assumes that all passengers are homogenous but they are not in reality. In addition, this paper only presents three typical fare schemes but other special fare schemes (e.g., daily/monthly pass tickets, student/elderly tickets) are not considered by our study. Several future research directions are therefore suggested as follows:

- Consider the schedule-based extension of the transit assignment problem with hard capacity constraints to evaluate short-term transit network design strategies and determine the optimal fleet size and vehicle schedules;
- Consider multi-class passenger demand to capture different age groups of passengers and their preferences in transit network design;
- Develop formulations of special fare schemes (e.g., daily/monthly pass tickets, student/elderly tickets) for multi-class transit network design problems.

Acknowledgments

The study was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region of China (HKU17218916). The authors are grateful to the four reviewers for their constructive comments.

References

Arup Group Limited, 2014. Travel Characteristics Survey 2011 Final Report. Transport Department, Hong Kong Government, Hong Kong.

Borndörfer, R., Karbstein, M. and Pfetsch, M. E., 2012. Models for fare planning in public transport. *Discrete Applied Mathematics, 160*(18), pp. 2591–2605.

Cantarella, G. E., 1997. A general fixed-point approach to multimode multi-user equilibrium assignment with elastic demand. *Transportation Science, 31*(2), pp. 107–128.

Cats, O., West, J. and Eliasson, J., 2016. A dynamic stochastic model for evaluating congestion and crowding effects in transit systems. *Transportation Research Part B: Methodological, 89*, pp. 43–57.

Chien, S. I. J. Y. and Tsai, C. F. M., 2007. Optimization of fare structure and service frequency for maximum profitability of transit systems. *Transportation Planning and Technology, 30*(5), pp. 477–500.

Chin, A., Lai, A. and Chow, J., 2016. Nonadditive public transit fare pricing under congestion with policy lessons from a case study in Toronto, Ontario, Canada. *Transportation Research Record, 2544*, pp. 28–37.

Codina, E. and Rosell, F., 2017. A heuristic method for a congested capacitated transit assignment model with strategies. *Transportation Research Part B: Methodological, 106*, pp. 293–320.

De Cea, J. and Fernández, E., 1993. Transit assignment for congested public transport systems: an equilibrium model. *Transportation Science, 27*(2), pp. 133–147.

Deng, L., Zhang, Z., Liu, K., Zhou, W. and Ma, J., 2014. Fare optimality analysis of urban rail transit under various objective functions. *Discrete Dynamics in Nature and Society*, pp.

Dial, R. B., 1971. A probabilistic multipath traffic assignment model which obviates path enumeration. *Transportation Research, 5*(2), pp. 83–111.

Fleishman, D., Shaw, N., Joshi, A. and Freeze, R., 1996. Fare Policies, Structures, and Technologies, TCRP Report 10. Transportation Research Board, National Research Council, Washington, DC.

Hamdouch, Y., Ho, H. W., Sumalee, A. and Wang, G., 2011. Schedule-based transit assignment model with vehicle capacity and seat availability. *Transportation Research Part B: Methodological, 45*(10), pp. 1805–1830.

Hamdouch, Y. and Lawphongpanich, S., 2008. Schedule-based transit assignment model with travel strategies and capacity constraints. *Transportation Research Part B: Methodological, 42*(7), pp. 663–684.

Huang, D., Liu, Z., Liu, P. and Chen, J., 2016. Optimal transit fare and service frequency of a nonlinear origin-destination based fare structure. *Transportation Research Part E: Logistics and Transportation Review, 96*, pp. 1–19.

Huang, H. J., 2002. Pricing and logit-based mode choice models of a transit and highway system with elastic demand. *European Journal of Operational Research, 140*(3), pp. 562–570.

Jiang, Y. and Szeto, W. Y., 2016. Reliability-based stochastic transit assignment: Formulations and capacity paradox. *Transportation Research Part B: Methodological, 93,* pp. 181–206.

Lam, W. H. K., Gao, Z. Y., Chan, K. S., and Yang, H., 1999. A stochastic user equilibrium assignment model for congested transit networks. *Transportation Research Part B: Methodological, 33*(5), pp. 351–368.

Lam, W. H. K., Zhou, J., and Sheng, Z., 2002. A capacity restraint transit assignment with elastic line frequency. *Transportation Research Part B: Methodological, 36*(10), pp. 919–938.

Lam, W. H. K. and Zhou, J., 1999. Stochastic transit assignment with elastic demand. *Journal of Eastern Asia Society for Transportation Studies, 3*(2), pp. 75–87.

Lam, W. H. K. and Zhou, J., 2000. Optimal fare structure for transit networks with elastic demand. *Transportation Research Record: Journal of the Transportation Research Board, 1733*, pp. 8–14.

Li, Z. C., Lam, W. H. K. and Wong, S. C., 2009. The optimal transit fare structure under different market regimes with uncertainty in the network. *Networks and Spatial Economics, 9*(2), pp. 191–216.

Ling, J. H., 1998. Transit fare differentials: a theoretical analysis. *Journal of Advanced Transportation, 32*(3), pp. 297–314.

Long, J. C., Huang, H. J., Gao, Z. Y. and Szeto, W. Y., 2013. An intersection-movement-based dynamic user optimal route choice problem. *Operations Research, 61*(5), pp. 1134–1147.

Long, J. C., Szeto, W. Y. and Huang, H. J., 2014. A bi-objective turning restriction design problem in urban road networks. *European Journal of Operational Research, 237*(2), pp. 426–439.

Nassi, C. D. and da Costa, F. C. D. C., 2012. Use of the analytic hierarchy process to evaluate transit fare system. *Research in Transportation Economics, 36*(1), pp. 50–62.

Nuzzolo, A., Crisalli, U., Comi, A. and Rosati, L., 2016. A mesoscopic transit assignment model including real-time predictive information on crowding. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations, 20*(4), pp. 316–333.

Ran, B. and Boyce, D., 1996. Ideal stochastic dynamic route choice models. In: *Modeling Dynamic Transportation Networks* (pp. 181–210). Springer.

Sun, L., Meng, Q. and Liu, Z., 2013. Transit assignment model incorporating bus dwell time. *Transportation Research Record, 2352*, pp. 76–83.

Sun, S. and Szeto, W. Y., 2018. Logit-based transit assignment: approach-based formulation and paradox revisit. *Transportation Research Part B: Methodological, 112*, pp. 191–215.

Szeto, W. Y. and Jiang, Y., 2014a. Transit assignment: approach-based formulation, extragradient method, and paradox. *Transportation Research Part B: Methodological*, *62*, pp. 51–76.

Szeto, W. Y. and Jiang, Y., 2014b. Transit route and frequency design: bi-level modeling and hybrid artificial bee colony algorithm approach. *Transportation Research Part B: Methodological*, *67*, pp. 235–263.

Tobin, R. L. and Friesz, T. L., 1988. Sensitivity analysis for equilibrium network flow. *Transportation Science, 22*(4), pp. 242–250.

Tsai, F. M., Chien, S. and Spasovic, L., 2008. Optimizing distance-based fares and headway of an intercity transportation system with elastic demand and trip length differentiation. *Transportation Research Record: Journal of the Transportation Research Board, 2089*, pp. 101–109.

Tsai, F. M., Chien, S. and Wei, C. H., 2012. Joint optimization of temporal headway and differential fare for transit systems considering heterogeneous demand elasticity. *Journal of Transportation Engineering, 139*(1), pp. 30–39.

Vuchic, V. R., 2007. *Urban Transit: Operations, Planning, and Economics*, John Wiley & Sons.

Wang, S. and Qu, X., 2017. Station choice for Australian commuter rail lines: equilibrium and optimal fare design. *European Journal of Operational Research, 258*(1), pp. 144–154.

Wang, W., Sun, H., Wang, D. Z. W. and Wu, J., 2014. Optimal transit fare in a bimodal network under demand uncertainty and bounded rationality. *Journal of Advanced Transportation, 48*(8), pp. 957–973.

Wang, W. W., Wang, D. Z. W., Sun, H. and Wu, J., 2016. Public transit service operation strategy under indifference thresholds-based bi-modal equilibrium. *Journal of Advanced Transportation, 50*(6), pp. 1124–1138.

Yang, H., 1995. Sensitivity analysis for queuing equilibrium network flow and its application to traffic control. *Mathematical and Computer Modelling, 22*(4-7), pp. 247–258.

Yang, H., 1997. Sensitivity analysis for the elastic-demand network equilibrium problem with applications. *Transportation Research Part B: Methodological, 31*(1), pp. 55–70.

Yang, H. and Bell, M. G. H. Sensitivity analysis of network traffic equilibrium revisited: the corrected approach. The 4th IMA International Conference on Mathematics in TransportInstitute of Mathematics and its Applications, 2007.

Ying, J. Q., Lu, H. and Shi, J. 2007. An algorithm for local continuous optimization of traffic signals. *European Journal of Operational Research, 181*(3), pp. 1189–1197.

Ying, J. Q. and Miyagi, T., 2001. Sensitivity analysis for stochastic user equilibrium network flows—a dual approach. *Transportation Science*, *35*(2), pp. 124–133.

Ying, J. Q. and Yang, H., 2005. Sensitivity analysis of stochastic user equilibrium flows in a bi-modal network with application to optimal pricing. *Transportation Research Part B: Methodological*, *39*(9), pp. 769–795.

Zhang, J. and Yang, H., 2016. Transit operations with deterministic optimal fare and frequency control. *Transportation Research Procedia, 14*, pp. 313–322.

Zhou, J., Lam, W. H. K. and Heydecker, B. G., 2005. The generalized Nash equilibrium model for oligopolistic transit market with elastic demand. *Transportation Research Part B: Methodological, 39*(6), pp. 519–544.

Appendix A

This appendix gives the proofs of Propositions 1-7.

Proposition 1 *The flat fare structure is a special case of the sectional fare structure.*

Proof: According to the definition of flat fares, let $\boldsymbol{\rho}^*_{\text{flat}} = \left(\rho^{*b}_{\text{flat}}\right)$ with $0 \le \rho^{*b}_{\text{flat}} \le p_{\max}$ be a feasible fare setting under the flat fare structure. Then the fare charged at stop i of route b $\left(p^{*b}_{i,1}\right)$ can be calculated by

$$p^{*b}_{i,1} = \rho^{*b}_{\text{flat}}, \quad \forall i \in I^b, b \in B. \tag{37}$$

Then we can always find a vector $\boldsymbol{\rho}^*_{\text{sectional}} = \left(\rho^{*b}_{i,\text{ sect}}\right)$ calculated by

$$\rho^{*b}_{i,\text{ sect}} = \delta^b_i \cdot \rho^{*b}_{\text{flat}}, \quad \forall i \in I^b, b \in B, \tag{38}$$

where $\delta^b_i = 1$ if i is the last stop of route b, δ^b_i = 0 otherwise. Note that $\rho^{*b}_{i,\text{ sect}}$ satisfies constraints (7) and (8).

Based on Eqs. (9) and (38), the fare charged at stop i of route b under the sectional fare structure can be expressed as

$$p^{*b}_{i,3} = \rho^{*b}_{i,\text{ sect}} + \sum_{i' \in I^b_{i+}} \rho^{*b}_{i',\text{ sect}} = \rho^{*b}_{\text{flat}} = p^{*b}_{i,1}, \quad \forall i \in I^b, b \in B, \tag{39}$$

which means that the sectional fare structure can be expressed as a flat fare structure as a special case. This completes the proof. □

Proposition 2 *When the destinations of all passengers are located at the last stop of each transit route, the distance-based fare structure is a special case of the sectional fare structure.*

Proof: Let l^b_i be the length of the link connecting node i and its next stop along route b. Note that $l^b_i = 0$ if i is the last stop of route b, $l^b_i > 0$ otherwise. According to the definition of distance-based fares, let $\boldsymbol{\rho}^*_{\text{dist}} = \left(\rho^{*b}_{\text{dist}}\right)$ with $0 \le \rho^{*b}_{\text{dist}} \le \dfrac{p_{\max}}{\sum_{i \in I^b} l^b_i}$ be a feasible fare setting under the distance-based fare structure. Then the fare charged at stop i of route b can be calculated by

$$p^{*b}_{i,2} = \rho^{*b}_{\text{dist}} \cdot \left(l^b_i + \sum_{i' \in I^b_{i+}} l^b_{i'} \right), \quad \forall i \in I^b, b \in B. \tag{40}$$

Then we can always find a vector $\boldsymbol{\rho}^{\#}_{\text{sect}} = \left(\rho^{\#b}_{i,\text{ sect}}\right)$ calculated by

$$\rho^{\#b}_{i,\text{ sect}} = \rho^{*b}_{\text{dist}} \cdot l^b_i, \quad \forall i \in I^b, b \in B, \tag{41}$$

and $\rho^{\#b}_{i,\text{ sect}}$ satisfies constraints (7) and (8).

Based on Eqs. (41) and (9), the sectional fare charged at stop i of route b can be formulated as

$$p^{\#b}_{i,3} = \rho^{\#b}_{i,\text{ sect}} + \sum_{i' \in I^b_{i+}} \rho^{\#b}_{i',\text{ sect}} = \rho^{*b}_{\text{dist}} \cdot l^b_i + \sum_{i' \in I^b_{i+}} \left(\rho^{*b}_{\text{dist}} \cdot l^b_{i'} \right)$$

$$= \rho^{*b}_{\text{dist}} \cdot \left(l^b_i + \sum_{i' \in I^b_{i+}} l^b_{i'} \right) = p^{*b}_{i,2}, \quad \forall i \in I^b, b \in B, \tag{42}$$

which means that when the destinations of all passengers are located at the last stop of each transit route, the sectional fare structure can be expressed as a distance-based fare structure as a special case. This completes the proof. □

Proposition 3 *The approach probabilities calculated using Eqs. (10) to (12) satisfy the logit-based SUE condition:* $\alpha^{rd}_y = \dfrac{\exp\left(-\theta \cdot c_y\right)}{\sum_{y' \in Y^{rd}} \left(\exp\left(-\theta \cdot c_{y'}\right)\right)}, \quad \forall y \in Y^{rd}, r \in R, d \in D$.

<u>Proof</u>: Let $e_y = \exp\left(-\theta \cdot c_y\right)$ be the exponential function of the expected path cost of path y (c_y). From Eqs. (10) and (13), we have

$$e_y = \exp\left(-\theta \cdot \sum_{s \in S^y} c_s \right) = \exp\left(\sum_{s \in S^y} \left(-\theta \cdot c_s \right) \right) = \prod_{s \in S^y} \exp\left(-\theta \cdot c_s \right) = \prod_{s \in S^y} e_s, \quad \forall y \in Y^{ij}, i, j \in N. \tag{43}$$

The following equation can be proved using mathematical induction:

$$W^d_a = e_{u(a)} \cdot \left(\delta^d_{h(u(a))} + \sum_{y \in Y^{h(u(a))d}} e_y \right), \quad \forall a \in A^+_i, i \in N, d \in D. \tag{44}$$

Let $\{i_n, i_{n-1}, \ldots, i_k, \ldots, i_2, i_1, d\}$ be the sequence of the nodes of sub-networks formed by S^{id}, where the subscript of node i is the topological order of that node. We assume that $W^d_a = e_{u(a)} \cdot \left(\delta^d_{h(u(a))} + \sum_{y \in Y^{h(u(a))d}} e_y \right), \quad \forall a \in A^+_{i_k}$ is true for the kth node and its succeeding nodes (i.e., $i_{k'}, k' \le k$) in the sequence. Then for the (k+1)th node in the sequence, based on Eqs. (11) and (43), we have

$$W^d_{a'} = \exp\left(-\theta \cdot c_{u(a')}\right) \cdot \left(\delta^d_{h(u(a'))} + \sum_{a'' \in A^+_{h(u(a'))}} W^d_{a''} \right) = e_{u(a')} \cdot \left(\sum_{a'' \in A^+_{h(u(a'))}} W^d_{a''} \right)$$

$$= e_{u(a')} \cdot \left(\sum_{a'' \in A^+_{h(u(a'))}} \left(e_{u(a'')} \cdot \left(\delta^d_{h(u(a''))} + \sum_{y \in Y^{h(u(a''))d}} e_y \right) \right) \right)$$

$$= e_{u(a')} \cdot \left(\sum_{a'' \in A^+_{h(u(a'))}} \left(\sum_{y \in Y^{h(u(a''))d}} \exp\left(-\theta \cdot \left(c_{u(a'')} + c_y \right) \right) \right) \right)$$

$$= e_{u(a')} \cdot \sum_{y \in Y^{h(u(a'))d}} e_y ,$$

where $h(u(a'))=i_{k'}, k'\le k$ and $t(s')=i_{k+1}$. Therefore, it is also true for the (k+1)th node and its succeeding nodes in the sequence. When $k=1$, we have $W_{a*}^{d}=\exp\left(-\theta\cdot c_{u(a*)}\right)=e_{u(a*)}$, where $t(u(a*))=i_1, h(u(a*))=d$. Therefore, $W_a^d=e_{u(a)}\cdot\left(\delta_{h(u(a))}^{j}+\sum_{y\in Y^{h(u(a))d}} e_y\right)$, $\forall a\in A_i^+, i\in N, d\in D$ is true.

From Eqs. (43) and (44), we have

$$\sum_{a\in A_i^+} W_a^d=\sum_{a\in A_i^+}\left(e_{u(a)}\cdot\left(\sum_{y\in Y^{h(u(a))d}} e_y\right)\right)=\sum_{a\in A_i^+}\left(\sum_{y\in Y^{h(u(a))d}}\exp\left(-\theta\cdot\left(c_{u(a)}+c_y\right)\right)\right)$$
$$=\sum_{y\in Y^{id}}\exp\left(-\theta\cdot c_y\right),\quad \forall i\in N, d\in D. \tag{45}$$

Let S^y be the set of links on path y, $y\in Y^{id}, i\in N, d\in D$. We consider three cases when combining Eqs. (10) to (12), (43), and (45): (i) When $s^*\in S^y$ and $t(s^*)=i$, we have

$$\alpha_{m(s^*)}^{d}=\frac{W_{m(s^*)}^{d}}{\sum_{a\in A_i^+} W_a^d}=\frac{e_{s*}\cdot\left(\sum_{y\in Y^{h(s^*)d}} e_y\right)}{\sum_{y\in Y^{id}} e_y}, \tag{46}$$

where $\delta_{h(s^*)}^{d}=0$. (ii) When $s^{\#}\in S^y$ and $h(s^{\#})=d$, we have

$$\alpha_{m(s^{\#})}^{d}=\frac{W_{m(s^{\#})}^{d}}{\sum_{a\in A_{t(s^{\#})}^{+}} W_a^d}=\frac{e_{s^{\#}}}{\sum_{y\in Y^{t(s^{\#})d}} e_y}, \tag{47}$$

where $\delta_{h(s^{\#})}^{d}=1$. (iii) For all the other intermediate links $s\in S^y$, we have

$$\alpha_{m(s)}^{d}=\frac{W_{m(s)}^{d}}{\sum_{a\in A_{t(s)}^{+}} W_a^d}=\frac{e_s\cdot\left(\sum_{y\in Y^{h(s)d}} e_y\right)}{\sum_{y\in Y^{t(s)d}} e_y}, \tag{48}$$

where $\delta_{h(s)}^{d}=0$.

By combining Eqs. (14), (43), and (46) to (48) and canceling out the repeated terms, we have

$$\alpha_y^{id}=\alpha_{m(s^*)}^{d}\cdot\ldots\cdot\alpha_{m(s')}^{d}\cdot\alpha_{m(s)}^{d}\cdot\ldots\cdot\alpha_{m(s^{\#}{}')}^{d}\cdot\alpha_{m(s^{\#})}^{d}$$
$$=\frac{e_{s^*}\cdot\left(\sum_{y'\in Y^{h(s^*)d}} e_{y'}\right)}{\sum_{y'\in Y^{id}} e_{y'}}\cdot\ldots\cdot\frac{e_{s'}\cdot\left(\sum_{y'\in Y^{h(s')d}} e_{y'}\right)}{\sum_{y'\in Y^{t(s')d}} e_{y'}}\cdot\frac{e_s\cdot\left(\sum_{y'\in Y^{h(s)d}} e_{y'}\right)}{\sum_{y'\in Y^{t(s)d}} e_{y'}}\cdot\ldots\cdot\frac{e_{s^{\#}{}'}\cdot\left(\sum_{y'\in Y^{h(s^{\#}{}')d}} e_{y'}\right)}{\sum_{y'\in Y^{t(s^{\#}{}')d}} e_{y'}}\cdot\frac{e_{s^{\#}}}{\sum_{y'\in Y^{t(s^{\#})d}} e_{y'}}$$

$$= \frac{e_{s^*} \cdot \ldots \cdot e_{s'} \cdot e_s \cdot \ldots \cdot e_{s^{\#}{}'} \cdot e_{s^{\#}}}{\sum_{y' \in Y^{id}} e_{y'}} = \frac{\prod_{s \in S^y} e_s}{\sum_{y' \in Y^{id}} e_{y'}} = \frac{e_y}{\sum_{y' \in Y^{id}} e_{y'}}$$

$$= \frac{\exp\left(-\theta \cdot c_y\right)}{\sum_{y' \in Y^{id}} \left(\exp\left(-\theta \cdot c_{y'}\right)\right)}, \quad \forall y \in Y^{id}, i \in N, d \in D, \tag{49}$$

where s' is the preceding link of link s and $h(s') = t(s)$; $s^{\#}{}'$ is the preceding link of link $s^{\#}$ and $h(s^{\#}{}') = t(s^{\#})$. Therefore, when $i = r \in R$, we have

$$\alpha_y^{rd} = \frac{\exp\left(-\theta \cdot c_y\right)}{\sum_{y' \in Y^{rd}} \left(\exp\left(-\theta \cdot c_{y'}\right)\right)}, \quad \forall y \in Y^{rd}, r \in R, d \in D. \tag{50}$$

This completes the proof. □

Proposition 4 *The passenger demands calculated using Eqs. (1) and (15) satisfy the logit-based elastic demand condition:* $q^{rd} = Q^{rs}\left(-\frac{1}{\theta} \cdot \ln\left(\sum_{y \in Y^{rd}} \left(\exp\left(-\theta \cdot c_y\right)\right)\right)\right), \ \forall r \in R, d \in D$.

Proof: From Eq. (45), when $i = r \in R$, Eq. (15) becomes

$$\overline{C}^{rd} = -\frac{1}{\theta} \cdot \ln\left(\sum_{a \in A_r^+} W_a^d\right) = -\frac{1}{\theta} \cdot \ln\left(\sum_{y \in Y^{rd}} \exp\left(-\theta \cdot c_y\right)\right), \quad \forall r \in R, d \in D. \tag{51}$$

By substituting Eq. (51) into Eq. (1), we have

$$q^{rd} = Q^{rd}\left(-\frac{1}{\theta} \cdot \ln\left(\sum_{y \in Y^{rd}} \left(\exp\left(-\theta \cdot c_y\right)\right)\right)\right), \quad \forall r \in R, d \in D. \tag{52}$$

This completes the proof. □

Proposition 5 *The approach-based SUEED problem is equivalent to the path-based logit SUEED problem:* $h_y^{rd} = \alpha_y^{rd} \cdot q^{rd} = \frac{\exp\left(-\theta \cdot c_y\right)}{\sum_{y' \in Y^{rd}} \left(\exp\left(-\theta \cdot c_{y'}\right)\right)} \cdot q^{rd}, \ \forall y \in Y^{rd}, r \in R, d \in D$ *and*

$$q^{rd} = Q^{rs}\left(-\frac{1}{\theta} \cdot \ln\left(\sum_{y \in Y^{rd}} \left(\exp\left(-\theta \cdot c_y\right)\right)\right)\right), \quad \forall r \in R, d \in D.$$

Proof: Consider the subnetwork formed by the links used between O-D pair *rd*, $s \in S^{rd}$. By definition, for an intermediate node *i* in the sub-network formed by S^{rd}, the probability that sub-path $y \in Y^{id}$ is used is equal to the sum of the probabilities of the paths using sub-path *y* over the sum of the probabilities of paths passing through node *i*, written as

$$\alpha_y^{id} = \frac{\sum_{y' \in Y^{rd|y}} \alpha_{y'}^{rd}}{\sum_{y' \in Y^{rd,i}} \alpha_{y'}^{rd}}, \quad \forall y \in Y^{id}, i \in N^{rd}, r \in R, d \in D, \tag{53}$$

where $Y^{rd|y}$ is the set of paths connecting origin r and destination d using sub-path y; $Y^{rd,i}$ is the set of paths connecting origin r and destination d passing through node i.

Based on the logit assumption and Eq. (43) and canceling out the repeated terms, Eq. (53) can become

$$\alpha_y^{id} = \frac{\sum_{y' \in Y^{rd|y}} \alpha_{y'}^{rd}}{\sum_{y' \in Y^{rd,i}} \alpha_{y'}^{rd}} = \sum_{y' \in Y^{rd|y}} \left(\frac{e_{y'}}{\sum_{y'' \in Y^{rd}} e_{y''}} \right) / \sum_{y' \in Y^{rd,i}} \left(\frac{e_{y'}}{\sum_{y'' \in Y^{rd}} e_{y''}} \right) = \frac{\sum_{y' \in Y^{rd|y}} e_{y'}}{\sum_{y' \in Y^{rd,i}} e_{y'}} = \frac{\left(\sum_{y' \in Y^{ri}} e_{y'} \right) \cdot e_y}{\left(\sum_{y' \in Y^{ri}} e_{y'} \right) \cdot \left(\sum_{y' \in Y^{id}} e_{y'} \right)}$$

$$= \frac{e_y}{\sum_{y' \in Y^{id}} e_{y'}}, \quad \forall y \in Y^{id}, i \in N^{rd}, r \in R, d \in D. \tag{54}$$

Let v_s^{rd} be the flow from origin r to destination d using link s. Based on Eq. (14), the path-based logit-based SUEED solution satisfies

$$v_s^{rd} = \sum_{y \in Y_s^{rd}} h_y^{rd} = \sum_{y \in Y_s^{rd}} \left(\alpha_y^{rd} \cdot q^{rd} \right) = q^{rd} \cdot \sum_{y \in Y_s^{rd}} \alpha_y^{rd}$$

$$= q^{rd} \cdot \left(\sum_{y \in Y^{rt(s)}} \left(\prod_{s' \in S^y} \alpha_{m(s')}^d \right) \right) \cdot \alpha_{m(s)}^d \cdot \left(\sum_{y \in Y^{h(s)d}} \alpha_y^{h(s)d} \right)$$

$$= q^{rd} \cdot \left(\sum_{y \in Y^{rt(s)}} \left(\prod_{s' \in S^y} \alpha_{m(s')}^d \right) \right) \cdot \alpha_{m(s)}^d, \tag{55}$$

where $\left(\sum_{y \in Y^{h(s)d}} \alpha_y^{h(s)d} \right)$ equals 1 by definition.

When $s \in A_r^+$, we have

$$v_s^{rd} = q^{rd} \cdot \alpha_{m(s)}^d. \tag{56}$$

When $s \in S^{rd}$ and $t(s) \neq r$, we have

$$v_s^{rd} = q^{rd} \cdot \left(\sum_{y \in Y^{rt(s)}} \left(\prod_{s' \in S^y} \alpha_{m(s')}^d \right) \right) \cdot \alpha_{m(s)}^d$$

$$= q^{rd} \cdot \sum_{s' \in A_{t(s)}^-} \left(\left(\sum_{y \in Y^{rt(s')}} \left(\prod_{s'' \in S^y} \alpha_{m(s'')}^d \right) \right) \cdot \alpha_{m(s')}^d \right) \cdot \alpha_{m(s)}^d$$

$$= \sum_{s' \in A^-_{t(s)}} \left(v^{rd}_{s'} \right) \cdot \alpha^d_{m(s)} . \tag{57}$$

Substituting Eqs. (56) and (57) into Eq. (55), we obtain

$$v^{rd}_s = \left(q^{rd} \cdot \delta^r_s + \sum_{s' \in A^-_{t(s)}} \left(v^{rd}_{s'} \right) \right) \cdot \alpha^d_{m(s)}, \quad \forall s \in S^{rd}, r \in R, d \in D , \tag{58}$$

where $\delta^r_s = 1$ if $t(s) = r$, $\delta^r_s = 0$ otherwise. Therefore, we have

$$v^d_s = \sum_{r \in R} v^{rd}_s = \left(\sum_{r \in R} \left(q^{rd} \cdot \delta^r_s \right) + \sum_{r \in R} \sum_{s' \in A^-_{t(s)}} \left(v^{rd}_{s'} \right) \right) \cdot \alpha^d_{m(s)}$$

$$= \left(q^{t(s)d} + \sum_{s' \in A^-_{t(s)}} \left(v^d_{s'} \right) \right) \cdot \alpha^d_{m(s)}, \quad \forall s \in S, d \in D . \tag{59}$$

Note that $q^{t(s)d} = 0$ if $t(s) \notin R$.

According to the definition of approach probabilities and Eq. (54), we have

$$\alpha^d_a = \sum_{y \in Y^{t(u(a))d}_{u(a)}} \alpha^{t(u(a))d}_y = \frac{\sum_{y \in Y^{t(u(a))d}_{u(a)}} e_y}{\sum_{y \in Y^{t(u(a))d}} e_y}, \quad \forall a \in A^+_i, i \in N, d \in D . \tag{60}$$

The following equation can be proved using mathematical induction:

$$\sum_{y \in Y^{t(u(a))d}_{u(a)}} e_y = W^d_a , \quad \forall a \in A^+_i, i \in N, d \in D . \tag{61}$$

Let $\{i_n, i_{n-1}, \ldots, i_k, \ldots, i_2, i_1, d\}$ be the sequence of the nodes of sub-networks formed by S^{id}, where the subscript of node i is the topological order of that node. We assume that $\sum_{y \in Y^{t(u(a))d}_{u(a)}} e_y = W^d_a , \quad \forall a \in A^+_{i_k}$ is true for the kth node and its succeeding nodes (i.e., $i_{k'}, k' \le k$) in the sequence. Then for the (k+1)th node in the sequence, we have

$$\sum_{y \in Y^{t(u(a'))d}_{u(a')}} e_y = e_{u(a')} \cdot \left(\sum_{y \in Y^{h(u(a'))d}} e_y \right) = e_{u(a')} \cdot \left(\sum_{a'' \in A^+_{h(u(a'))}} \left(\sum_{y \in Y^{h(u(a'))}_{u(a'')}} e_y \right) \right) = e_{u(a')} \cdot \left(\sum_{a'' \in A^+_{i_{k'}}} \left(\sum_{y \in Y^{i_{k'}}_{u(a'')}} e_y \right) \right)$$

$$= e_{u(a')} \cdot \left(\sum_{a'' \in A^+_{i_{k'}}} \left(W^d_{a''} \right) \right) = W^d_{a'} ,$$

where $h(u(a')) = i_{k'}, k' \le k$ and $t(u(a')) = i_{k+1}$. Therefore, it is also true for the (k+1)th node and its succeeding nodes in the sequence. When $k = 1$, we have $\sum_{y \in Y^{t(u(a^*))d}_{u(a^*)}} e_y = e_{u(a^*)} = W^d_{a^*}$, where $t(u(a^*)) = i_1, h(u(a^*)) = d$. Therefore, $\sum_{y \in Y^{t(u(a))d}_{u(a)}} e_y = W^d_a \quad \forall a \in A^+_i, i \in N, d \in D$ is true.

Based on Eq. (61), Eq. (60) can be rewritten as

$$\alpha_a^d = \frac{\sum_{y \in Y_{u(a)}^{t(u(a))d}} e_y}{\sum_{y \in Y^{t(u(a))d}} e_y} = \frac{W_a^d}{\sum_{a' \in A_{t(u(a))}^+} \left(\sum_{y \in Y_{u(a')}^{t(u(a))d}} e_y \right)} = \frac{W_a^d}{\sum_{a' \in A_{t(u(a))}^+} W_{a'}^d}, \quad \forall a \in A_i^+, i \in N, d \in D. \tag{62}$$

From Eq. (61), we can also derive the following:

$$q^{rd} = Q^{rd}\left(-\frac{1}{\theta} \cdot \ln\left(\sum_{y \in Y^{rd}} \left(\exp\left(-\theta \cdot c_y\right)\right)\right)\right)$$
$$= Q^{rd}\left(-\frac{1}{\theta} \cdot \ln\left(\sum_{a \in A_r^+} \sum_{y \in Y_{u(a)}^{rd}} \left(\exp\left(-\theta \cdot c_y\right)\right)\right)\right)$$
$$= Q^{rd}\left(-\frac{1}{\theta} \cdot \ln\left(\sum_{a \in A_r^+} W_a^d\right)\right), \quad \forall r \in R, d \in D. \tag{63}$$

According to Eqs. (62) and (63), we can conclude that the solution to the path-based SUEED problem also satisfies the approach-based SUEED problem defined by Eqs. (1), (3)-(12), (15), and (16). By this conclusion together with the fact that the solution to the approach-based SUEED problem satisfies the logit-based (path-based) SUEED condition (implied by Propositions 3 and 4), we conclude that the approach-based SUEED problem is equivalent to the path-based SUEED problem. This completes the proof. □

Proposition 6 *There exists a solution to the approach-based SUEED problem in terms of approach probabilities and passenger demands. Moreover, the solution is unique.*

Proof:
The finding by Cantarella (1997) implies that a solution exists to the path-based SUEED problem. Let $\left(\mathbf{a}^*(\mathbf{c}^*), \mathbf{q}^*(\mathbf{c}^*)\right)$ be a solution to the path-based SUEED problem with the elements $\mathbf{a}^* = \left(\alpha_y^{*rd}\right)$, $\mathbf{q}^* = \left(q^{*rd}\right)$, and $\mathbf{c}^* = \left(c_y^*\right)$; let $\mathbf{a}^{*id} = \left(\alpha_y^{*id}\right)$ be the vector of the probabilities of sub-paths in the subnetwork formed by the links connecting node *i* and destination *d*, $s \in S^{id} \subseteq S^{rd}$, and the elements of $\mathbf{a}^{*id}$ are calculated using Eq. (53). The existence of $\mathbf{a}^* = \left(\alpha_y^{*rd}\right)$ ensures the existence of $\mathbf{a}^{*id} = \left(\alpha_y^{*id}\right)$. Let $\boldsymbol{\alpha}^*$ be the approach probability vector with the elements α_a^{*d} calculated by $\alpha_a^{*d} = \sum_{y \in Y^{t(u(a))d}} \alpha_y^{*t(u(a))d}$. Then, based on the findings in Proposition 5, $\left(\boldsymbol{\alpha}^*, \mathbf{q}^*\right)$ is a solution to the approach-based SUEED problem defined by Eqs. (1), (10)-(12), (15), and (16). Moreover, the existence of $\mathbf{a}^{*id} = \left(\alpha_y^{*id}\right)$ ensures the existence of $\boldsymbol{\alpha}^*$. Therefore, we can conclude that the solution existence of the path-based SUEED problem ensures that a solution to the approach-based SUEED problem also exists.

We assume that there is more than one solution to the approach-based SUEED problem. Let $(\boldsymbol{\alpha}^{\#},\mathbf{q}^{\#})$ with the elements $\boldsymbol{\alpha}^{\#}=(\alpha_a^{\#d})$ and $\mathbf{q}^{\#}=(q^{\#rd})$ be another solution to the approach-based SUEED problem and $(\boldsymbol{\alpha}^{\#},\mathbf{q}^{\#})\neq(\boldsymbol{\alpha}^{*},\mathbf{q}^{*})$; let $\mathbf{a}^{\#}$ be the corresponding path choice probability vector with the elements $\alpha_y^{\#rd}$ such that $\alpha_a^{\#d}=\sum_{y\in Y^{t(u(a))d}}\alpha_y^{\#t(u(a))d}$ and $\alpha_y^{\#id}$ satisfies Eq. (53). Then by Propositions 3 and 4, $(\mathbf{a}^{\#},\mathbf{q}^{\#})$ is a solution to the path-based SUEED problem. According to Eq. (60), the mapping from path choice probability to approach probability is surjective. Hence, we have $(\mathbf{a}^{\#},\mathbf{q}^{\#})\neq(\mathbf{a}^{*},\mathbf{q}^{*})$. However, $(\mathbf{a}^{\#},\mathbf{q}^{\#})\neq(\mathbf{a}^{*},\mathbf{q}^{*})$ contradicts the solution uniqueness of the path-based SUEED problem (see Cantarella, 1997). As a result, the approach-based SUEED problem has exactly one solution in terms of approach probabilities and passenger demands. This completes the proof. □

Proposition 7 *The optimal profit increases as the unit operating cost τ decreases.*

Proof: Let $(\boldsymbol{\rho}^{*},\mathbf{f}^{*})$ and $Z^{*}(\boldsymbol{\rho}^{*},\mathbf{f}^{*})$ be an optimal fare and frequency setting and the optimal objective value when $\tau=\tau^{*}>0$; let $\tau^{\#}=\tau^{*}-\vartriangle\tau$ and $\tau^{\#},\vartriangle\tau>0$. Obviously, $(\boldsymbol{\rho}^{*},\mathbf{f}^{*})$ is a feasible solution to the maximization problem when $\tau=\tau^{\#}$. When $\tau=\tau^{\#}$, the objective value $Z^{*'}$ at $(\boldsymbol{\rho}^{*},\mathbf{f}^{*})$ is

$$
\begin{aligned}
Z^{*'} &= \sum_{s\in S} v_s^{*} p_s^{*} - \tau^{\#}\cdot\sum_{b\in B} f^{*b} l^{b} = \sum_{s\in S} v_s^{*} p_s^{*} - \left(\tau^{*}-\vartriangle\tau\right)\cdot\sum_{b\in B} f^{*b} l^{b} \\
&= \sum_{s\in S} v_s^{*} p_s^{*} - \tau^{*}\cdot\sum_{b\in B} f^{*b} l^{b} + \vartriangle\tau\cdot\sum_{b\in B} f^{*b} l^{b} \\
&= Z^{*} + \vartriangle\tau\cdot\sum_{b\in B} f^{*b} l^{b} > Z^{*}.
\end{aligned}
\tag{64}
$$

Let $(\boldsymbol{\rho}^{\#},\mathbf{f}^{\#})$ and $Z^{\#}(\boldsymbol{\rho}^{\#},\mathbf{f}^{\#})$ be an optimal fare and frequency setting and the optimal objective value when $\tau=\tau^{\#}$. Then we have $Z^{\#}\geq Z^{*'}>Z^{*}$. Therefore, the objective value increases when the unit operating cost τ decreases. This completes the proof. □

Proposition 8 *The optimal total vehicle mileage is monotonically decreasing over $\tau>0$.*

Proof: Let $(\boldsymbol{\rho}^{*},\mathbf{f}^{*})$ and $Z^{*}(\boldsymbol{\rho}^{*},\mathbf{f}^{*})$ be an optimal fare and frequency setting and the optimal objective value when $\tau=\tau^{*}>0$; let $\tau^{\#}=\tau^{*}+\vartriangle\tau$ and $\tau^{\#},\vartriangle\tau>0$. When $\tau=\tau^{\#}$, the objective (18) can be rewritten as

$$
\max Z^{'}(\boldsymbol{\rho},\mathbf{f}) = \sum_{s\in S} v_s p_s - \tau^{*}\cdot\sum_{b\in B} f^{b} l^{b} - \vartriangle\tau\cdot\sum_{b\in B} f^{b} l^{b}\,. \tag{65}
$$

Let $(\boldsymbol{\rho}^{\#},\mathbf{f}^{\#})$ and $Z^{'\#}(\boldsymbol{\rho}^{\#},\mathbf{f}^{\#})$ be an optimal fare and frequency setting and the optimal objective value to the network design problem (19), (20), and (65). Obviously, $(\boldsymbol{\rho}^{\#},\mathbf{f}^{\#})$ is a feasible solution to the network design problem (18)-(20) when $\tau=\tau^{*}$, and the objective value is denoted as $Z^{\#}(\boldsymbol{\rho}^{\#},\mathbf{f}^{\#})$. Since $(\boldsymbol{\rho}^{*},\mathbf{f}^{*})$ optimizes the problem (18)-(20) when $\tau=\tau^{*}$, we have

$$Z^{\#}(\boldsymbol{\rho}^{\#},\mathbf{f}^{\#}) \le Z^{*}(\boldsymbol{\rho}^{*},\mathbf{f}^{*}) . \tag{66}$$

$(\boldsymbol{\rho}^{*},\mathbf{f}^{*})$ is also a feasible solution to the network design problem (19), (20), and (65). The corresponding objective value $Z^{'*}$ at $(\boldsymbol{\rho}^{*},\mathbf{f}^{*})$ is given by

$$Z^{'*}(\boldsymbol{\rho}^{*},\mathbf{f}^{*}) = \sum_{s\in S} v_s^* p_s^* - \tau^* \cdot \sum_{b\in B} f^{*b} l^b - \triangle\tau \cdot \sum_{b\in B} f^{*b} l^b$$
$$= Z^{*} - \triangle\tau \cdot \sum_{b\in B} f^{*b} l^b . \tag{67}$$

Since $(\boldsymbol{\rho}^{\#},\mathbf{f}^{\#})$ optimizes the problem (19), (20), and (65), we have

$$Z^{'*}(\boldsymbol{\rho}^{*},\mathbf{f}^{*}) \le Z^{'\#}(\boldsymbol{\rho}^{\#},\mathbf{f}^{\#})$$
$$Z^{*} - \triangle\tau \cdot \sum_{b\in B} f^{*b} l^b \le Z^{\#} - \triangle\tau \cdot \sum_{b\in B} f^{\#b} l^b$$
$$\triangle\tau \cdot \left(\sum_{b\in B} f^{\#b} l^b - \sum_{b\in B} f^{*b} l^b \right) \le Z^{\#} - Z^{*} . \tag{68}$$

Combining (66) and (68), we have

$$\triangle\tau \cdot \left(\sum_{b\in B} f^{\#b} l^b - \sum_{b\in B} f^{*b} l^b \right) \le Z^{\#} - Z^{*} \le 0$$
$$\sum_{b\in B} f^{\#b} l^b - \sum_{b\in B} f^{*b} l^b \le 0$$
$$\sum_{b\in B} f^{\#b} l^b \le \sum_{b\in B} f^{*b} l^b . \tag{69}$$

Therefore, the optimal total vehicle mileage is monotonically decreasing over $\tau > 0$. This completes the proof. □

Appendix B

This appendix gives the derivations of Eqs. (28) and (29).

<u>Derivation of Eq. (28):</u>

Based on Eq. (60) and the quotient rule, we have

$$\frac{\partial \alpha_a^d}{\partial c_s}=\frac{\partial\left(\dfrac{\sum\limits_{y\in Y_{u(a)}^{t(u(a))d}}\exp\left(-\theta\cdot c_y\right)}{\sum\limits_{y'\in Y^{t(u(a))d}}\left(\exp\left(-\theta\cdot c_y\right)\right)}\right)}{\partial c_s}$$

$$=\frac{\dfrac{\partial\left(\sum\limits_{y\in Y_{u(a)}^{t(u(a))d}}\exp\left(-\theta\cdot c_y\right)\right)}{\partial c_s}\cdot\cancel{\left(\sum\limits_{y\in Y^{t(u(a))d}}\left(\exp\left(-\theta\cdot c_y\right)\right)\right)}}{\left(\sum\limits_{y\in Y^{t(u(a))d}}\left(\exp\left(-\theta\cdot c_y\right)\right)\right)^{\cancel{2}}}$$

$$-\frac{\cancel{\left(\sum\limits_{y\in Y_{u(a)}^{t(u(a))d}}\exp\left(-\theta\cdot c_y\right)\right)}\cdot\dfrac{\partial\left(\sum\limits_{y\in Y^{t(u(a))d}}\left(\exp\left(-\theta\cdot c_y\right)\right)\right)}{\partial c_s}}{\left(\sum\limits_{y\in Y^{t(u(a))d}}\left(\exp\left(-\theta\cdot c_y\right)\right)\right)^{\cancel{2}}}$$

$$=\frac{\dfrac{\partial\left(\sum\limits_{y\in Y_{u(a)}^{t(u(a))d}}\exp\left(-\theta\cdot c_y\right)\right)}{\partial c_s}-\alpha_a^d\cdot\dfrac{\partial\left(\sum\limits_{y\in Y_s^{t(u(a))d}}\left(\exp\left(-\theta\cdot c_y\right)\right)\right)}{\partial c_s}}{\sum\limits_{y\in Y^{t(u(a))d}}\left(\exp\left(-\theta\cdot c_y\right)\right)}. \qquad (70)$$

When $s\in A_{t(u(a))}^{+}$ and $s\neq u(a)$, we have

$$\frac{\partial \alpha_a^d}{\partial c_s} = \frac{\cancel{\dfrac{\partial\left(\sum_{y\in Y_{u(a)}^{t(u(a))d}} \exp\left(-\theta\cdot c_y\right)\right)}{\partial c_s}} - \alpha_a^d \cdot \dfrac{\partial\left(\sum_{y\in Y_s^{t(u(a))d}} \left(\exp\left(-\theta\cdot c_y\right)\right)\right)}{\partial c_s}}{\sum_{y\in Y^{t(u(a))d}} \left(\exp\left(-\theta\cdot c_y\right)\right)}$$

$$= \frac{-\alpha_a^d}{\sum_{y\in Y^{t(u(a))d}} \left(\exp\left(-\theta\cdot c_y\right)\right)} \cdot \frac{\partial\left(\left(\sum_{y\in Y^{t(u(a))t(s)}} e_y\right)\cdot e_s \cdot \left(\sum_{y\in Y^{h(s)d}} e_y\right)\right)}{\partial c_s}$$

$$= \frac{\left(-\alpha_a^d\right)\cdot\left(\sum_{y\in Y^{t(u(a))t(s)}} e_y\right)\cdot\left(\sum_{y\in Y^{h(s)d}} e_y\right)}{\sum_{y\in Y^{t(u(a))d}} \left(\exp\left(-\theta\cdot c_y\right)\right)} \cdot \frac{\partial e_s}{\partial c_s}$$

$$= \frac{\left(-\alpha_a^d\right)\cdot\left(\sum_{y\in Y^{t(u(a))t(s)}} e_y\right)\cdot\left(\sum_{y\in Y^{h(s)d}} e_y\right)\cdot e_s \cdot (-\theta)}{\sum_{y\in Y^{t(u(a))d}} \left(\exp\left(-\theta\cdot c_y\right)\right)} = \left(-\alpha_a^d\right)\cdot\frac{\sum_{y\in Y_s^{t(u(a))d}} \left(\exp\left(-\theta\cdot c_y\right)\right)}{\sum_{y\in Y^{t(u(a))d}} \left(\exp\left(-\theta\cdot c_y\right)\right)}\cdot(-\theta) \tag{71}$$

$$= -\alpha_a^d \cdot \alpha_{m(s)}^d \cdot \left(\cancel{\sum_{y\in Y^{h(s)d}} \alpha_y^{h(s)d}}\right)\cdot(-\theta) = -\alpha_a^d \cdot \alpha_{m(s)}^d \cdot (-\theta).$$

When $s \in A_{t(u(a))}^{+}$ and $s = u(a)$, we have

$$\frac{\partial \alpha_a^d}{\partial c_s} = \frac{\dfrac{\partial\left(\sum_{y\in Y_s^{t(u(a))d}} \exp\left(-\theta\cdot c_y\right)\right)}{\partial c_s} - \alpha_a^d \cdot \dfrac{\partial\left(\sum_{y\in Y_s^{t(u(a))d}} \left(\exp\left(-\theta\cdot c_y\right)\right)\right)}{\partial c_s}}{\sum_{y\in Y^{t(u(a))d}} \left(\exp\left(-\theta\cdot c_y\right)\right)}$$

$$= \frac{1-\alpha_a^d}{\sum_{y\in Y^{t(u(a))d}} \left(\exp\left(-\theta\cdot c_y\right)\right)} \cdot \frac{\partial\left(\left(\sum_{y\in Y^{t(u(a))t(s)}} e_y\right)\cdot e_s \cdot \left(\sum_{y\in Y^{h(s)d}} e_y\right)\right)}{\partial c_s}$$

$$= \left(1-\alpha_a^d\right)\cdot \alpha_{m(s)}^d \cdot \left(\cancel{\sum_{y\in Y^{h(s)d}} \alpha_y^{h(s)d}}\right)\cdot(-\theta)$$

$$= \left(1-\alpha_a^d\right)\cdot \alpha_{m(s)}^d \cdot (-\theta). \tag{72}$$

When $s \in S^{h(u(a))d}$, we have

$$\frac{\partial \alpha_a^d}{\partial c_s} = \frac{\dfrac{\partial\left(\sum_{y \in Y_{u(a)}^{t(u(a))d}} \exp\left(-\theta \cdot c_y\right)\right)}{\partial c_s} - \alpha_a^d \cdot \dfrac{\partial\left(\sum_{y \in Y_s^{t(u(a))d}} \left(\exp\left(-\theta \cdot c_y\right)\right)\right)}{\partial c_s}}{\sum_{y \in Y^{t(u(a))d}} \left(\exp\left(-\theta \cdot c_y\right)\right)}$$

$$= \frac{e_{u(a)} \cdot \dfrac{\partial\left(\sum_{y \in Y_s^{h(u(a))d}} \exp\left(-\theta \cdot c_y\right)\right)}{\partial c_s}}{\sum_{y \in Y^{t(u(a))d}} \left(\exp\left(-\theta \cdot c_y\right)\right)} - \alpha_a^d \cdot (-\theta) \cdot \sum_{y \in Y_s^{t(u(a))d}} \alpha_y^{t(u(a))d}$$

$$= \frac{e_{u(a)} \cdot \left(\sum_{y \in Y_s^{h(u(a))d}} \exp\left(-\theta \cdot c_y\right)\right) \cdot (-\theta)}{\sum_{y \in Y^{t(u(a))d}} \left(\exp\left(-\theta \cdot c_y\right)\right)} - \alpha_a^d \cdot (-\theta) \cdot \left(\sum_{y \in Y^{t(u(a))t(s)}} \left(\prod_{s' \in S^y} \alpha_{m(s')}^d\right)\right) \cdot \alpha_{m(s)}^d$$

$$= \frac{\left(\sum_{y \in Y_{u(a)}^{t(u(a))d} \cap Y_s^{t(u(a))d}} \exp\left(-\theta \cdot c_y\right)\right) \cdot (-\theta)}{\sum_{y \in Y^{t(u(a))d}} \left(\exp\left(-\theta \cdot c_y\right)\right)} - \alpha_a^d \cdot (-\theta) \cdot \left(\sum_{y \in Y^{t(u(a))t(s)}} \left(\prod_{s' \in S^y} \alpha_{m(s')}^d\right)\right) \cdot \alpha_{m(s)}^d$$

$$= \left(\sum_{y \in Y_{u(a)}^{t(u(a))d} \cap Y_s^{t(u(a))d}} \alpha_y^{t(u(a))d}\right) \cdot (-\theta) - \alpha_a^d \cdot (-\theta) \cdot \left(\sum_{y \in Y^{t(u(a))t(s)}} \left(\prod_{s' \in S^y} \alpha_{m(s')}^d\right)\right) \cdot \alpha_{m(s)}^d$$

$$= \alpha_a^d \cdot \left(\sum_{y \in Y^{h(u(a))t(s)}} \left(\prod_{s' \in S^y} \alpha_{m(s')}^d\right)\right) \cdot \alpha_{m(s)}^d \cdot \cancel{\left(\sum_{y \in Y^{h(s)d}} \alpha_y^{h(s)d}\right)} \cdot (-\theta)$$

$$-\alpha_a^d \cdot (-\theta) \cdot \left(\sum_{y \in Y^{t(u(a))t(s)}} \left(\prod_{s' \in S^y} \alpha_{m(s')}^d\right)\right) \cdot \alpha_{m(s)}^d$$

$$= \alpha_a^d \cdot (-\theta) \cdot \alpha_{m(s)}^d \cdot \left(\sum_{y \in Y^{h(u(a))t(s)}} \left(\prod_{s' \in S^y} \alpha_{m(s')}^d\right) - \sum_{y \in Y^{t(u(a))t(s)}} \left(\prod_{s' \in S^y} \alpha_{m(s')}^d\right)\right)$$

$$= \alpha_{m(s)}^d \cdot \alpha_a^d \cdot (-\theta) \cdot \sum_{s' \in A_{t(s)}^-} \left(\left(\sum_{y \in Y^{h(u(a))t(s')}} \left(\prod_{s'' \in S^y} \alpha_{m(s'')}^d\right)\right) \cdot \alpha_{m(s')}^d - \left(\sum_{y \in Y^{t(u(a))t(s')}} \left(\prod_{s'' \in S^y} \alpha_{m(s'')}^d\right)\right) \cdot \alpha_{m(s')}^d\right)$$

$$= \alpha_{m(s)}^d \cdot \sum_{s' \in A_{t(s)}^-} \left(\frac{\partial \alpha_a^d}{\partial c_{s'}}\right). \tag{73}$$

Using Eqs. (71)-(73), Eq. (70) can be rewritten as

$$\frac{\partial \alpha_a^d}{\partial c_s} = \alpha_{m(s)}^d \cdot \left(\left(\delta_s^a - \alpha_a^d \right) \cdot (-\theta) \cdot \delta_s^{t(u(a))} + \sum_{s' \in A_{t(s)}^-} \left(\frac{\partial \alpha_a^d}{\partial c_{s'}} \right) \right), \quad \forall s \in S^{t(u(a))d}, a \in A_i^+, i \in N, d \in D, \quad (74)$$

where $\delta_s^a = 1$ if $s = u(a)$, $\delta_s^a = 0$ otherwise; $\delta_s^{t(u(a))} = 1$ if $t(s) = t(u(a))$, $\delta_s^{t(u(a))} = 0$ otherwise.

<u>Derivation of Eq. (29):</u>

According to the chain rule, we have

$$\frac{\partial q^{rd}}{\partial c_s} = \frac{\partial q^{rd}}{\partial \overline{C}^{rd}} \cdot \frac{\partial \overline{C}^{rd}}{\partial c_s}, \quad \forall s \in S^{rd}, r \in R, d \in D. \quad (75)$$

From Eq. (51), $\frac{\partial \overline{C}^{rd}}{\partial c_s}$ in Eq. (75) can be calculated by

$$\frac{\partial \overline{C}^{rd}}{\partial c_s} = \frac{\partial \left(-\frac{1}{\theta} \cdot \ln \left(\sum_{y \in Y^{rd}} \exp\left(-\theta \cdot c_y\right) \right) \right)}{\partial c_s} = \frac{-\frac{1}{\theta}}{\sum_{y \in Y^{rd}} \exp\left(-\theta \cdot c_y\right)} \cdot \frac{\partial \left(\sum_{y \in Y^{rd}} \exp\left(-\theta \cdot c_y\right) \right)}{\partial c_s}$$

$$= \frac{-\frac{1}{\theta}}{\sum_{y \in Y^{rd}} \exp\left(-\theta \cdot c_y\right)} \cdot \frac{\partial \left(\sum_{y \in Y_s^{rd}} \exp\left(-\theta \cdot c_y\right) \right)}{\partial c_s}$$

$$= \frac{-\frac{1}{\theta}}{\sum_{y \in Y^{rd}} \exp\left(-\theta \cdot c_y\right)} \cdot \frac{\partial \left(\left(\sum_{y \in Y^{rt(s)}} e_y \right) \cdot e_s \cdot \left(\sum_{y \in Y^{h(s)d}} e_y \right) \right)}{\partial c_s} = \frac{-\frac{1}{\theta} \cdot \left(\sum_{y \in Y^{rt(s)}} e_y \right) \cdot \left(\sum_{y \in Y^{h(s)d}} e_y \right)}{\sum_{y \in Y^{rd}} \exp\left(-\theta \cdot c_y\right)} \cdot \frac{\partial e_s}{\partial c_s}$$

$$= \frac{\cancel{-\frac{1}{\theta}} \cdot \left(\sum_{y \in Y^{rt(s)}} e_y \right) \cdot \left(\sum_{y \in Y^{h(s)d}} e_y \right)}{\sum_{y \in Y^{rd}} \exp\left(-\theta \cdot c_y\right)} \cdot e_s \cdot \cancel{(-\theta)}$$

$$= \frac{\sum_{y \in Y_s^{rd}} \exp\left(-\theta \cdot c_y\right)}{\sum_{y \in Y^{rd}} \exp\left(-\theta \cdot c_y\right)}, \quad \forall s \in S^{rd}, r \in R, d \in D. \quad (76)$$

Based on Eq. (50), Eq. (76) can be rewritten as

$$\frac{\partial \overline{C}^{rd}}{\partial c_s} = \frac{\sum_{y \in Y_s^{rd}} \exp\left(-\theta \cdot c_y\right)}{\sum_{y \in Y^{rd}} \exp\left(-\theta \cdot c_y\right)} = \sum_{y \in Y_s^{rd}} \alpha_y^{rd} = \left(\sum_{y \in Y^{rt(s)}} \left(\prod_{s' \in S^y} \alpha_{m(s')}^{d} \right) \right) \cdot \alpha_{m(s)}^{d} \cdot \cancel{\left(\sum_{y \in Y^{h(s)d}} \alpha_y^{h(s)d} \right)}$$

$$= \left(\sum_{y \in Y^{rt(s)}} \left(\prod_{s' \in S^y} \alpha_{m(s')}^{d} \right) \right) \cdot \alpha_{m(s)}^{d} . \tag{77}$$

For $s^0 \in A_r^+$ (i.e., $t(s^0) = r$), we have

$$\frac{\partial \overline{C}^{rd}}{\partial c_{s^0}} = \alpha_{m(s^0)}^{d} . \tag{78}$$

For all the other links $s \in S^{rd}$ and $t(s) \neq r$, we have

$$\frac{\partial \overline{C}^{rd}}{\partial c_s} = \left(\sum_{y \in Y^{rt(s)}} \left(\prod_{s' \in S^y} \alpha_{m(s')}^{d} \right) \right) \cdot \alpha_{m(s)}^{d} = \sum_{s' \in A_{t(s)}^-} \left(\left(\sum_{y \in Y^{rt(s')}} \left(\prod_{s'' \in S^y} \alpha_{m(s'')}^{d} \right) \right) \cdot \alpha_{m(s')}^{d} \right) \cdot \alpha_{m(s)}^{d}$$

$$= \sum_{s' \in A_{t(s^*)}^-} \left(\frac{\partial \overline{C}^{rd}}{\partial c_{s'}} \right) \cdot \alpha_{m(s)}^{d} . \tag{79}$$

Therefore, Eq. (77) can be expressed as

$$\frac{\partial \overline{C}^{rd}}{\partial c_s} = \alpha_{m(s)}^{d} \cdot \left(\delta_s^r + \sum_{s' \in A_{t(s)}^-} \frac{\partial \overline{C}^{rd}}{\partial c_{s'}} \right), \quad \forall s \in S^{rd}, r \in R, d \in D , \tag{80}$$

where $\delta_s^r = 1$ if $t(s) = r$, $\delta_s^r = 0$ otherwise.

Using the linear demand function as stated in Eq. (2) and the above equation, Eq. (75) can be rewritten as

$$\frac{\partial q^{rd}}{\partial c_s} = \left(-\psi^{rd}\right) \cdot \frac{\partial \overline{C}^{rd}}{\partial c_s} = \alpha_{m(s)}^{d} \cdot \left(\left(-\psi^{rd}\right) \cdot \delta_s^r + \sum_{s' \in A_{t(s)}^-} \frac{\partial q^{rd}}{\partial c_{s'}} \right), \quad \forall s \in S^{rd}, r \in R, d \in D , \tag{81}$$

where $\delta_s^r = 1$ if $t(s) = r$, $\delta_s^r = 0$ otherwise.