<table>
<thead>
<tr>
<th>Title</th>
<th>Optimization of Number of Operators and Allocation of New Lines in an Oligopolistic Transit Market</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Li, ZC; Lam, WHK; Wong, SC</td>
</tr>
<tr>
<td>Citation</td>
<td>Networks And Spatial Economics, 2012, v. 12 n. 1, p. 1-20</td>
</tr>
<tr>
<td>Issued Date</td>
<td>2012</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/10722/150662">http://hdl.handle.net/10722/150662</a></td>
</tr>
<tr>
<td>Rights</td>
<td>The original publication is available at <a href="http://www.springerlink.com">www.springerlink.com</a>; This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.</td>
</tr>
</tbody>
</table>
Optimization of Number of Operators and Allocation of New Lines in an Oligopolistic Transit Market

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

Abstract This paper proposes a novel model for determining the optimal number of transit operators and the allocation of new lines in an oligopolistic transit market. The proposed model consists of three interrelated sub-models that are associated with three types of players; namely, transit authority, transit operators, and transit passengers. In practice, the operating cost per unit of transit line of each operator is decreasing in the number of lines that it operates. These effects which are referred to as the scale economies of transit operations are explicitly incorporated in the proposed model. On the basis of a logit-type transit passenger travel choice sub-model with elastic demand, the fares and frequencies of transit services are determined by an oligopolistic competitive equilibrium model (i.e. transit operator sub-model). The transit authority sub-model for optimization of the number of operators and the allocation of new lines is expressed as a 0-1 integer programming problem. It can be solved by an implicit enumeration heuristic solution algorithm. Numerical results show that both the scale economies and the market demand level have significant impacts on the optimal number of operators and the allocation schemes of new lines. Ignoring the effects of scale economies on transit operations may lead transit authorities to make biased decisions.

Keywords Transit market, Oligopolistic competition, Number of operators, Allocation of new lines, Scale economies, Implicit enumeration

Z.-C. Li, William H. K. Lam
Department of Civil & Structural Engineering, Hong Kong Polytechnic University
Hung Hom, Kowloon, Hong Kong, China
email: cehklam@polyu.edu.hk

Z.-C. Li
School of Management, Huazhong University of Science & Technology
Wuhan 430074, China
email: smzcli@gmail.com

S. C. Wong
Department of Civil Engineering, University of Hong Kong
Pokfulam Road, Hong Kong, China
email: hhecwsc@hkucc.hku.hk
1 Introduction

In recent years, there has been an accelerating trend towards the deregulation of transit systems, particularly in developing countries. This is attributed to the brief that the privatization and deregulation of urban transit services can offer the possibility of significant improvements in efficiency and reliability of the transit system (Lo and Yip, 2001). However, empirical studies by Estache and Gomez-Lobo (2005) and Ardila (2008) recently showed that transit service competition may to some extent lead to market failure, including oversupply of transit vehicles, inflated fares and low service quality. They also showed that transit systems that were initially privately owned were eventually regulated, when cities became increasingly reliant on transit modes as a mass people mover. The phenomenon was called the cycle of private and public involvement (Estache and Gomez-Lobo, 2005).

Similar phenomenon has also recently occurred in Hong Kong. Specifically, the transit industry in Hong Kong has experienced several significant mergers in the past several years. Two (i.e., the New World First Bus Company and the Citybus Company) of five bus companies (wholly-owned by the private sector) were merged in 2003. More recently, the two railway corporations (i.e., the MTR and the KCR) in Hong Kong have also been merged. These mergers raise some intriguing and important questions. If this tendency persists, then will the number of operators in the transit market continue to reduce and finally arrive at a monopoly? How many operators are optimal in a transit market? How can the transit authority (e.g., the government) create a “win-win-win” regulatory or incentive environment in which private investors can procure more profits while providing high-quality services, passengers’ travel needs can be satisfied with affordable services, and society as a whole is better off than it was before? The solutions to these problems would have a long-term implication on sustainable transportation in which a sustainable public transportation system can be maintained and improved to support the urban development and growth, especially for cities that are characterized by low car ownership and compact urban structure, such as Hong Kong (Tong and Wong, 1997; Loo and Chow, 2008).

To address these interesting issues, advanced models need to be developed to help us understand and assess the effects of competition in transit market with various numbers of transit operators. In general, the number of operators in a transit system will influence the strategies of competing operators in relation to service frequencies and fares, the travel choice
behavior of passengers, and the performance of the system itself. For example, the monopoly regime (i.e. a single operator in a market) and oligopoly regime (i.e. multiple operators in a market) can lead to a significant difference in the efficiency and reliability of the transit system. It is therefore very important that transit authorities optimize the number of operators, particularly in large Asian cities, such as Hong Kong.

Significant progress has been made in transit system modeling over the past decades, particularly in modeling competition between transit operators (Harker, 1988; Fernandez and Marcotte, 1992; Zubieta, 1998; Lo et al., 2000, 2004; Zhou et al., 2005; Wang and Yang, 2005). However, most of these existing studies have assumed that the number of operators in the transit market was given and fixed exogenously. Little attention has been paid to exploring the impacts of the number of operators on a transit system. Some exceptions are Williams and Abdulaal (1993) and Williams and Martin (1993), who studied the market behavior of an arbitrary number of operators under oligopolistic competition. However, their studies did not provide a methodology for the determination of the optimal number of operators in a transit market. Moreover, their models implicitly assumed that each of the operators in the market only operated a single transit line (i.e., the one-to-one assumption). In reality, the number of transit companies (or operators) in a city is far less than the number of transit lines. It implies that each transit company operates multiple transit lines. Thus, these models may not be able to address the problems with unequal numbers of transit lines and operators, such as the allocation problem of new transit lines, which is a timely problem in large Asian cities.

For example, the transit markets in Hong Kong are partially regulated, and the bus services can be divided into three categories: franchised bus, public light bus or minibus (including red minibuses and green minibuses) and residential bus. In Hong Kong, some new transit lines (e.g., bus and/or railway lines) are proposed regularly and put into operation to accommodate the growing travel demands particularly in the newly developed outlying areas. The operating rights of these new lines are usually conferred by the Transport Department on one or several transit companies through competitive bidding exercise. Specifically, a transit company wishing to invest capital in an additional transit line has to bid for it in advance from the Transport Department, and if it is successful, the selected transit company is then granted a franchise for operating that line. This franchise for regular bus or green minibus routes is usually given to the existing transit companies or transit operators. However, for the third category of residential bus services, new transit companies would have the right to apply for
granting the license to provide the commuting services. Naturally, this raises another important issue, i.e., how should the authority allocate these new lines to bidders to create a “win-win-win” situation?

In addition, the one-to-one assumption adopted in the models of Williams and Abdulaal (1993) and Williams and Martin (1993) also gives rise to another major limitation, i.e., the effects of the scale economies of transit operations cannot be considered properly. “Scale economies” means that the operating cost per unit of transit line of each operator decreases with the number of lines that it operates. A number of empirical studies have shown that these economies are commonplace in transit operations (Berechman and Giuliano, 1985; Savage, 1997). Thus, it is necessary to incorporate the effects of scale economies into the allocation problem of new transit lines particularly in an oligopolistic transit market with competition.

In view of the above, this paper proposes (1) to develop a novel model for simultaneously determining the optimal number of operators and the allocation of new lines in an oligopolistic transit market, (2) to incorporate the effects of scale economies on the transit system in the proposed model, (3) to develop a heuristic solution algorithm to solve the simultaneous optimization problem, and (4) to explore the effects of the allocation of additional lines, the scale economies and the transit market demand level.

In order to facilitate the presentation of the essential ideas, this paper focuses on a transit market that consists of commuting trips between a residential area and an urban central area. In the proposed model, three types of players are involved. They are transit authority, transit operators, and transit passengers. The interactions between these three players result in a multi-level hierarchical system, as shown in Figure 1. In this multi-level system, the transit authority aims to determine simultaneously the optimal number of operators and to allocate the new transit lines to bidders in a way that maximizes the total social welfare of the transit system. However, the authority is not willing to subsidize the transit operators, but rather wants them to be able to survive in a competitive market. This is often the case in many large Asian cities with high density populated development, such as Hong Kong. The transit operators in the market seek to maximize their own profits by adjusting their service fares and frequencies, while accounting for the transit passengers’ travel choice decisions that minimize their own perceived disutility of travel.
The remainder of this paper is organized as follows. In Section 2, some basic assumptions and notation are described. Section 3 presents the model formulation and solution algorithm. Section 4 provides a numerical example to illustrate the application of the proposed model. Finally, conclusions are given in Section 5 together with recommendations for further studies.

2 Basic considerations

2.1 Assumptions

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

**A1** In the transit market, there are \( L \) lines, including new and existing lines, and \( K \) alternative companies or operators that are bidding for the operating rights of new lines. Suppose that all the bidders act independently to determine their fares and frequencies. Collusive behavior between the bidders is not considered in this paper, but will be explored in a future study.

**A2** Each transit operator in the market can operate more than one transit line, but each transit line can be allocated to one operator only. This relaxes the strict assumption in Williams and Abdulaal (1993) and Williams and Martin (1993) that each operator or company can operate one transit line only.

**A3** All transit passengers are assumed to be homogenous concerning their values of time (VOT) and disutility perception. However, this can easily be relaxed to incorporate passenger heterogeneity by introducing a distribution of passenger's VOT, as done in Yang et al. (2001) and Huang and Li (2007). Transit passengers make their travel choice decisions in a stochastic manner based on the tradeoff between the service qualities or travel disutilities of different transit services. The disutility is measured by the weighted sum of the walking time for access to and egress from transit stations, the in-vehicle travel time, the waiting time at transit stations, the in-vehicle crowding discomfort, and the fare (De Cea and Fernandez, 1993; Wu et al., 1994; Lo et al., 2004; Uchida et al., 2007; Li et al., 2008).

**A4** The effects of the scale economies of transit operations are explicitly incorporated into the definition of transit operating cost and thus the profit function of transit operators. This
relaxes the assumption in many previous related studies, such as that of Uchida et al. (2007) and Yang and Woo (2000). They assumed that the transit operating cost was a linear function of the service frequency or fleet size.

An elastic demand function is used to capture the responses of passengers to the level of transit services and fares. The responses include switching to alternative modes (e.g. automobile) and not making the journey at all (Zhou et al., 2005; Li et al., 2007).

2.2 Notation

Sets

$L$ set of all transit lines in the transit market
$|L^k|$ total number of lines operated by operator $k$
$\bar{L}$ set of new or additional transit lines
$K$ set of transit operators in the transit market

Variables associated with transit passengers

$D$ total resultant passenger demand in the market (passengers/h)
$D_l$ passenger demand selecting line $l$ (passengers/h)
$D$ vector of resultant passenger demand; $D = (D_l, \forall l \in L)$
$\lambda_l$ proportion of passengers selecting line $l$
$u_l$ average travel disutility or generalized travel cost on line $l$ (h)
$w_l$ average waiting time of passengers for using line $l$ (h)
$g_l$ crowding discomfort cost of passengers in transit vehicles on line $l$ (h)

Variables associated with transit operators

$N_l$ transit fleet size or number of transit vehicles on line $l$ (vehicles)
$f_l$ service frequency of line $l$ (veh/h)
$p_l$ fare of line $l$ ($)
$\Phi_k$ profit of operator $k$ ($/h$)
$\Delta\Phi_k$ change of operator $k$’s profit due to an additional line
$y$ vector of fares and frequencies; $y = (p,f)$
$p$ vector of fare; $p = (p_l, \forall l \in L)$
\( f \) vector of service frequency; \( f = (\{ f_l \}, \forall l \in L) \)

**Variables associated with transit authority**

\( x_{lk} \) 0-1 variable; equal to 1 if line \( l \) is allocated to operator \( k \) and 0 otherwise

\( x \) vector of the 0-1 variable \( x_{lk} \); \( x = (x_{lk}, \forall l \in L, k \in K) \)

**Constants**

\( C_l \) capacity of transit vehicle (including seats and standees) on line \( l \) (passengers/veh)

\( D \) potential (latent) passenger demand in the transit market (passengers/h)

\( d_l \) total walking distance for access to and egress from transit stations on line \( l \) (km)

\( \kappa_0, \kappa_1 \) parameters in the in-vehicle crowding discomfort cost function

\( L_l \) length of line \( l \) (km)

\( t_l \) average in-vehicle travel time on line \( l \) (h)

\( T_l \) cycle journey time of a transit vehicle on line \( l \) (h)

\( \tau_l \) average walking time for access to and egress from transit stations on line \( l \) (h)

\( V_l \) average operating speed of transit vehicles on line \( l \) (km/h)

\( \omega \) average walking speed of passengers (km/h)

\( \varphi \) expected disutility of using the transit services (h)

\( \mu_k^0 \) fixed cost per unit of transit line for operator \( k \) ($/h)

\( \eta_l \) (variable) operating cost per vehicle-hour on line \( l \) ($/veh-h)

\( \alpha_0 \) value of in-vehicle travel time ($/h)

\( \alpha_1 \) value of walking time ($/h)

\( \alpha_2 \) value of waiting time ($/h)

\( \alpha_3 \) value of in-vehicle crowding discomfort ($/h)

\( \beta \) demand dispersion parameter in the elastic demand function (1/h)

\( \theta \) parameter for representing the perception variation of passengers on travel disutility

\( \rho \) parameter for capturing the effects of the scale economies of transit operations

### 3 Model formulation

Consider a transit market that consists of the commuting trips between an origin-destination
(OD) pair that connects \( H \) (a residential area) and \( W \) (a workplace) in an urban central area. 

As previously stated, there are three types of players in the transit market, i.e., the transit passengers, the transit operators, and the transit authority. They constitute a multi-level hierarchical decision system. The decision variables associated with them are, respectively, transit service choice, service fares and frequencies, and number of operators and allocation of new transit lines. These can be formulated as three interrelated sub-models, i.e., the transit passenger travel choice sub-model, the transit operator optimal fare and frequency sub-model, and the transit authority optimal number of operators and allocation of lines sub-model. In the following, these three sub-models are presented in order.

3.1 Transit passenger travel choice sub-model

According to A3, transit passengers choose transit services based on their own perceptions of the service quality or the disutility of the transit services. A multinomial logit formulation has often been adopted in the previous studies to model transit passenger travel choice behavior (Lam et al. 1999, 2002; Lo et al., 2000; Uchida et al., 2007). With the use of a logit-type formulation, the passenger demand, \( D_l \), selecting line \( l \) can be determined by

\[
D_l = D \lambda_l = D \frac{\exp(-u_l)}{\sum_l \exp(-u_l)}, \quad \forall l \in L,
\]

where \( L \) is the set of all transit lines in the market, including new and existing lines, \( D \) is the total resultant passenger demand in the market, \( \lambda_l \) is the proportion of passengers selecting line \( l \), and \( u_l \) is the average disutility or generalized travel cost on line \( l \). The parameter \( \theta \) represents the variation in passenger perceptions of travel disutility. The higher the value of \( \theta \), the smaller the variation in passenger perceptions, and vice versa.

The average travel disutility \( u_l \), measured in terms of equivalent time units, on line \( l \) is composed of the average walking time \( \tau_l \) for access to and egress from transit stations, the average in-vehicle travel time \( t_l \), the average passenger waiting time at stations \( w_l \), the in-vehicle crowding discomfort cost \( g_l \), and fare \( p_l \), i.e.,

\[
u_l = t_l + \frac{\alpha_1}{\alpha_0} \tau_l + \frac{\alpha_2}{\alpha_0} w_l + \frac{\alpha_3}{\alpha_0} g_l + \frac{1}{\alpha_0} p_l, \quad \forall l \in L,
\]

where the parameters \( \alpha_0, \alpha_1, \alpha_2 \) and \( \alpha_3 \) are, respectively, the value of in-vehicle travel
time, the value of walking time, the value of waiting time and the value of crowding discomfort of passengers, which are all measured in monetary value per unit time.

The average in-vehicle travel time \( t_l \) depends on the length \( L_l \) of line \( l \) and the average operating speed \( V_l \) of transit vehicles on line \( l \), expressed as

\[
 t_l = \frac{L_l}{V_l}, \quad \forall l \in L, \tag{3}
\]

where the average vehicle operating speed \( V_l \) is calculated on the basis of the whole journey time including boarding and alighting times at transit stations (Yang and Woo, 2000).

The average walking time \( \tau_l \) of passengers using line \( l \) can be determined by the total walking distance \( d_l \) for access to and egress from transit stations on line \( l \) divided by the average walking speed \( \omega \) of passengers, i.e.,

\[
 \tau_l = \frac{d_l}{\omega}, \quad \forall l \in L. \tag{4}
\]

The average passenger waiting time \( w_l \) on line \( l \) can be estimated by

\[
 w_l = \frac{\gamma}{f_l}, \quad \forall l \in L, \tag{5}
\]

where \( f_l \) is the service frequency of line \( l \). The value of \( \gamma \) is dependent on the distributions of transit vehicle headways and passenger arrival times. The typical value of \( \gamma \) adopted in the literature is 0.5, with an assumption of a uniform random arrival distribution of passengers and a constant transit vehicle headway (Lam and Morrall, 1982).

Similar to Wu et al. (1994), Lo et al. (2004), and Uchida et al. (2007), the in-vehicle crowding discomfort cost, \( g_l \), measured in terms of generalized time units, can be expressed as a function of passenger volume on line \( l \) and the capacity of vehicle on line \( l \), formulated as

\[
 g_l = \kappa_0 t_l \left( \frac{D_l}{f_lC_l} \right)^{\kappa_1}, \quad \forall l \in L, \tag{6}
\]

where \( \kappa_0 \) and \( \kappa_1 \) are constants, \( C_l \) is the vehicle capacity (including seats and standees) on line \( l \), and \( f_lC_l \) is the total capacity of line \( l \). \( t_l \) can be calculated by Equation (3).
To capture the responses of passengers to the level of transit services and fares, an exponential form of elastic demand function, which is widely used in transport models (Zhou et al., 2005; Li et al., 2007), is adopted and specified as

\[ D = \bar{D} \exp(-\beta \varphi), \]  

(7)

where \( \bar{D} \) is the potential (latent) passenger demand in the transit market and \( \varphi \) is the expected disutility of using the transit services. \( \beta \) is the demand dispersion parameter that reflects the demand sensitivity to the expected disutility of the transit services. According to the random utility theory, \( \varphi \) can be measured by the following log-sum formula (Oppenheim, 1995)

\[ \varphi = -\frac{1}{\theta} \ln \sum_{l \in L} \exp(-\theta u_l). \]  

(8)

It should be pointed out that for given transit fares and frequencies, according to Equations (1)-(8), the passenger demand \( D_l, \forall l \in L \) on line \( l \) is a function of the travel disutility \( u_l, \forall l \in L \), which would be affected by the crowding discomfort cost \( g_l \), which is a function of the passenger demand \( D_l, \forall l \in L \) itself in terms of Equation (6). Therefore, the passenger travel choice sub-model can be formulated as a fixed-point problem with regard to the passenger demand \( D_l, \forall l \in L \).

**Proposition 1** (The passenger travel choice sub-model). For a given transit fare and frequency pattern, the logit-based passenger travel choice sub-model is equivalent to finding a vector \( \mathbf{D}^* \), such that the following fixed-point problem holds.

\[ \mathbf{D}^* - \mathbf{F}(\mathbf{D}^*) = \mathbf{0}, \quad \mathbf{D}^* \in \Omega \]  

(9)

with \( \mathbf{D} = (D_l, \forall l \in L), \quad \mathbf{F}(\mathbf{D}) = (D\lambda_l, \forall l \in L) \) and \( \Omega = \{ \mathbf{D} | \sum_l D_l = D, \forall l \in L \} \).

**Remark 1.** The crowding discomfort cost function (6) and the elastic demand function (7) in this paper are assumed to be continuous. Consequently, the feasible set \( \Omega \) is closed because the OD demand is bounded, thus there exists at least one solution to the fixed-point problem (9) according to the Brouwer’s fixed-point theory. Moreover, it is easily proved that the uniqueness of the model solution can be guaranteed if the discomfort cost function and the elastic demand function are strictly monotone (Patriksson, 1994). The fixed-point problem (9)
can be solved effectively by using the solution algorithm recently proposed by Huang and Li (2007), which is based on the method of successive averages (MSA) in conjunction with the logit-type assignment process.

3.2 Transit operator optimal fare and frequency sub-model

In Hong Kong, over 90% of the 11 million daily person trips are served by privately operated transit services (Transport Department, 2003). The transit industry in Hong Kong operates profitably without government subsidy. This is different from the situation in most Western cities in which the transit industry is often subsidized by the local government due to insufficient ridership. Under such a profitable operating environment in Hong Kong, the principal objective of the private transit operators is neither welfare gain nor the efficient utilization of road space, but rather profit maximization. Consequently, these privately owned transit operators compete against each other for maximizing their patronage and revenues (Zhou et al., 2005).

The net profit of a transit operator is the total revenue that is generated from the passenger fares minus the total transit service operating costs. For formulating the net profit obtained by a transit operator, we let

$$ x_{lk} = \begin{cases} 
1 & \text{if line } l \text{ is operated by operator } k, \\
0 & \text{otherwise.} 
\end{cases} $$

(10)

The net profit, \( \Phi_k \), of operator \( k \) can then be expressed as

$$ \Phi_k \left( y^k, y^{-k} \right) = \sum_{l \in L} \left( p_l D_l(y) \right) x_{lk} - \mu_k \left( \sum_{l \in L} x_{lk} \right)^p + \sum_{l \in L} \left( \eta_l N_l \right) x_{lk}, \quad \forall k \in K, $$

(11)

where \( K \) is the set of all operators in the transit market. \( y^k = (p^k, f^k) \) is the vector of the fare and frequency of operator \( k \), and \( y^{-k} = (p^{-k}, f^{-k}) \) is that of the other operators excluding \( k \). \( \mu_k \) is the fixed cost per unit of transit line for operator \( k \), \( \eta_l N_l \) is the variable operating cost on line \( l \), where \( \eta_l \) is the operating cost per vehicle-hour on line \( l \) and \( N_l \) is the transit fleet size or number of transit vehicles on line \( l \). \( N_l \) can be determined by \( N_l = T_l f_l \), where \( T_l \) is the cycle journey time of a transit vehicle on line \( l \) and \( T_l = 2L/V_l \).
Remark 2. The first term on the right-hand side of Equation (11) represents the total revenue of operator \( k \). The square brackets represent the total cost of operator \( k \), which consists of the fixed cost (i.e. the first term in the bracket) and the variable operating cost (i.e. the second term in the bracket) that is proportional to the total vehicle-hours (Yang and Woo, 2000). According to White (2002), the fixed cost consists of the capital costs for building the depot and providing other fixed assets, overhead and administration costs of the offices etc. It is related to the total number of lines (i.e. \( \sum_{l \in L} x_{lk} \)) that are operated by an operator. The variable operating cost includes the fuel costs, vehicle maintenance costs, crew wages, tires, and insurance etc.

Remark 3. According to Berechman and Giuliano (1985), the exponent \( \rho \ (0.0 \leq \rho \leq 1.0) \) in Equation (11) can be introduced to capture the effects of the scale economics of transit operations, as shown in Figure 2. In this figure, \( \mu^0 \) is the fixed cost per unit of transit line. It can be seen in Figure 2 that when \( \rho = 1.0 \), the total fixed cost is a linear function of the total number of lines that are operated by an operator (i.e. \( \sum_{l \in L} x_{lk} \)), implying that the effects of the scale economies of transit operations are not considered. When \( \rho = 0.0 \), the total fixed cost is a constant independent of the total number of lines operated. When \( 0.0 < \rho < 1.0 \), the fixed cost that results from an additional transit line service (i.e. the marginal cost) decreases as the total number of lines operated increases. Hence, the parameter \( \rho \ (0.0 < \rho < 1.0) \) can indeed capture the effects of the scale economies of transit services. In particular, for a given number of lines, the smaller the value of \( \rho \), the larger the scale economies, and vice versa.

In a competitive transit market, profit-driven operators have to decide their service fares and frequencies carefully to compete with other operators so as to maximize their own profits. Clearly, the profit obtained by an operator depends significantly on the strategies adopted by the other competitors and the responses of passengers in the transit market. At the same time, the strategies adopted by one operator have a significant impact on the strategies of the competitors and on the travel choice behavior of passengers. In turn, the subsequent effects on the passenger demand share over transit lines would affect the profits of the operator itself. This leads to an oligopolistic competitive equilibrium or a Cournot-Nash game between operators.
Proposition 2 (The transit operator optimal fare and frequency sub-model). The equilibrium fares and frequencies of operator $k$ in an oligopolistic competitive market can be obtained by the following maximization problem.

$$\max_{\{y^k\}} \Phi_k \left( y^k, y^{-k}, D \left( y^k, y^{-k} \right) \right), \quad \forall k \in K,$$

where the passenger demand vector $D \left( y^k, y^{-k} \right)$ can be determined by the passenger travel choice sub-model (9).

To search the oligopoly solution resulting from the competition between $K$ operators, a diagonalization method here is developed. The basic idea of this method is to solve the individual maximization problem (12) of the operators separately and sequentially, holding the decision variables of the other operators fixed in turn until the sequence converges. This method is heuristic, and thus cannot guarantee that the solution is a global optimum. The step-by-step procedure is given as follows.

**Step 0. Initialization.** Choose an initial fare and frequency pattern $y^{(1)} = \{y^{(1)}_k, k = 1, 2, \ldots, K\}$ for all $K$ operators. Set iteration counter $n = 1$.

**Step 1. Determination of passenger demand pattern.** Determine the passenger demand pattern $D^{(n)} = \{D^{(n)}_l\}$ by the passenger travel choice sub-model (9).

**Step 2. Diagonalization.** Solve the transit operator sub-model (12) independently for all $K$ operators, one then obtains the fare and frequency pattern $y^{(n+1)} = \{y^{(n+1)}_k, k = 1, 2, \ldots, K\}$.

**Step 3. Convergence check.** If $\|y^{(n+1)}_k - y^{(n)}_k\| < \varepsilon$, $k = 1, 2, \ldots, K$, where $\varepsilon$ is a pre-specified tolerance, then stop. Otherwise, set $n = n + 1$, and go to Step 1.

3.3 Transit authority sub-model

As previously stated, the transit authority aims to determine the optimal number of operators and to allocate new transit lines to bidders in a way that maximizes the total social welfare of the transit system. At the same time, it is not willing to subsidize the transit operators, but rather expects them to be able to survive in the competitive market. Transit operators, as profit maximizing entities, are not obliged to serve the lines the transit authority allows them to bid.
That is, a transit operator may not serve that line if its (marginal) profit decreases by serving this extra line. Consequently, the objective of the authority is, in fact, to determine an optimal line-operator matrix (referred to as the decision matrix in this paper), as shown in Equation (13), that maximizes the total social welfare, subject to the constraint that the profit of each operator is non-decreasing with the introduction of an additional line.

The decision matrix is an \( L \times K \) matrix, where \( L \) is the set of new transit lines to be allocated. Each row of the matrix is associated with a new line \((L_l)\), and each column is associated with a bidder \((O_k)\). Each element \( x_{lk} \) in the matrix is a binary variable, i.e.,

\[
x_{lk} = 0 \text{ or } 1, \quad \forall l \in L, k \in K.
\]

According to \( A2 \), each line can be allocated to one bidder only, thus implying that the sum of all the elements in each row of the decision matrix is equal to one, i.e.,

\[
\sum_k x_{lk} = 1, \quad \forall l \in L.
\]

In addition, an operator expects to bid an extra line if and only if the marginal profit due to the additional line is nonnegative, i.e.,

\[
\Delta \Phi_k \left( y \left( |L^k + 1|, |L^k| \right) \right) \geq 0, \quad \forall k \in K,
\]

where \( \Delta \Phi_k (\cdot) \) is the change of operator \( k \)'s profit caused by an additional line, and \(|L^k|\) is the total number of lines that are being operated by operator \( k \).

The total social welfare of the transit system is the sum of the consumer surplus and the producer surplus. Following Williams and Abdulaal (1993) and Evans (1987), the consumer surplus can be represented by \((D/\beta)\), measured in time units. It expresses the perceived benefits experienced by the potential passenger demand. The producer surplus is the total...
profits of all transit operators in the network, i.e., \( \sum_k \Phi_k \), measured in monetary units. Consequently, we obtain the following transit authority sub-model.

**Proposition 3** (The transit authority sub-model). The optimal number of operators and the allocation of new transit lines can be obtained by solving the following maximization problem.

\[
\max_{\{x_{lk}\}} \left( D\left( u(y(x)) \right), \Phi(y(x)) \right) = \frac{\alpha_0 D}{\beta} + \sum_{k \in K} \Phi_k(y(x)),
\quad (17)
\]

subject to \((14)-(16),\) \((18)\),

where \(x\) is the vector that consists of the elements in the decision matrix \(x_{lk}\). Vectors \(D\) and \(\Phi\), which are the functions in \(x\) through \(y\), can be determined by the passenger sub-model \((9)\) and the operator sub-model \((12)\), respectively. \(\alpha_0\) (see Equation \((2)\)) is the passenger’s value of time that is used to convert time into monetary cost.

The transit authority sub-model \((17)-(18)\) is a 0-1 integer programming problem with the binary variable \(x_{lk}\) as the decision variable. Note that only one element in each row of the decision matrix \(x_{lk}\) is 1, and all other elements in the same row are zero (see Equation \((15)\)). This characteristic of the decision matrix significantly reduces the size of the feasible solution space. In view of this, an implicit enumeration solution algorithm can be developed to solve the transit authority problem \((17)-(18)\). The basic idea behind this algorithm is to execute repeatedly a step-by-step procedure for maximizing the objective function \((17)\), by examining the local conditions but considering one input at a time. At each step, a decision is made about whether a particular input is an optimal solution. The step-by-step procedure of the solution algorithm is given as follows.

**Step 0. Initialization.** Choose an initial feasible solution, for example, allocate all new lines to a single operator (say, operator \(k\)). Set \(Z = -\infty\) (i.e., the lower bound of the objective function \((17)\)).

**Step 1. First loop operation.** Set iteration counter \(j = 1\).

**Step 2. Second loop operation.** Set the level (or new line to be allocated) counter \(l = 1\).

**Step 3. Third loop operation.** At level \(l\), check all \(K\) allocations sequentially in each of which line \(l\) is allocated to one operator. Set the operator counter \(k = 1\).
Step 3.1. If \( k > K \), then go to Step 4.

Step 3.2. Use the operator sub-model (12) and the passenger sub-model (9) to determine the fare and frequency pattern \( y^{(k)} = \{p^{(k)}, f^{(k)}\} \), the corresponding passenger demand pattern \( D^{(k)} = \{D_{ij}^{(k)}\} \), and the travel disutility \( u^{(k)} = \{u_{ij}^{(k)}\} \). Then, calculate the value \( Z^{(k)} \) of the objective function (17).

Step 3.3. Termination check for the third loop operation. If \( Z^{(k)} > Z \) and \( \Delta \Phi_{k}() \geq 0 \), \( \forall k \in K \), then put \( y^{*} = y^{(k)} \), \( D^{*} = D^{(k)} \), \( Z^{*} = Z^{(k)} \), \( k = k + 1 \), and go to Step 3.1. Otherwise, set \( k = k + 1 \), and go to Step 3.1.

Step 4. Termination check for the second loop operation. If \( l > L \), then go to Step 5. Otherwise, set \( l = l + 1 \), and go to Step 3.

Step 5. Termination check for the first loop operation. Repeat Steps 2-4. If no better solution can be found for a complete search across the \( L \) new lines, then terminate the algorithm and output the optimal solution \( \{x^{*}, y^{*}, D^{*}\} \) and the corresponding objective function value \( Z^{*} \). Otherwise, set \( j = j + 1 \), and go to Step 2.

It should be pointed out that the proposed implicit enumeration algorithm determines a solution by a sequence of decisions, each of which searches for the best solution at that stage. That is, each decision is locally optimal. Therefore, the proposed solution algorithm cannot guarantee the solution is a global optimum, but offers a heuristic method of finding an approximate but acceptable solution to the line allocation problem.

3.4 Summary of the model and solution algorithms

In the previous sections, we have formulated three interrelated sub-models that are associated with three types of players in the transit market, respectively. The solution algorithms for these three sub-models are also discussed and presented. For the convenience of readers, the interactions among the three sub-models together with the solution algorithms are further summarized in Table 1. It can be seen in Table 1 that the input of a sub-model is the outputs of the other two sub-models, and vice versa. The solution algorithm consists of three stages, each of which is corresponding to solving one of the three sub-models. The solution algorithm
proposed in our paper is a heuristic algorithm, thus a global optimum solution is not always guaranteed.

4 Numerical studies

To facilitate the presentation of the essential ideas and illustrate the contributions of this paper, we apply the proposed model and solution algorithm to a simple transit network. We explore the effects of the allocation of additional lines on the equilibrium fares and frequencies of the operators and on the transit system performance. The effects of the scale economies of transit operations and the market demand levels are also explored.

4.1 Problem setting

The example transit network is shown in Figure 3. There are four alternative transit lines that connect residential area $H$ with workplace $W$. The dashed lines represent the walk links for access to or egress from transit stations. Lines L1 and L2 are two existing lines, and Lines L3 and L4 are two newly constructed lines. The numbers shown beside the lines in Figure 3 are the lengths of the corresponding lines. The numbers shown beside the lines in Figure 3 are the lengths of the corresponding lines. The numbers shown beside the lines in Figure 3 are the lengths of the corresponding lines. Suppose that the two existing lines L1 and L2 are being operated by Operator A. Two operators from outside of the market (Operators B and C) and Operator A inside the market (a total of three profit-driven operators) are bidding for the operating rights of the new lines L3 and L4.

According to the discussion in the previous sections, the simultaneous optimization problem for the number of operators and the allocation of new lines is equivalent to the determination of the following decision matrix.

$$
\begin{pmatrix}
O_A & O_B & O_C \\
L_3 & x_{3A} & x_{3B} & x_{3C} \\
L_4 & x_{4A} & x_{4B} & x_{4C}
\end{pmatrix}
$$

(19)

where each element in the matrix is 0 or 1, and the sum of all the elements in each row is 1.

Different bidders or operators can be characterized to some extent by their sizes or market shares (in terms of number of lines operated). Without loss of generality, it is assumed that Operator A is the largest company, followed by Operator B and then Operator C. In this example, the numbers of lines that are being operated by the three operators (excluding the
two new lines L3 and L4) are assumed to be 300, 150, and 15, respectively. Consequently, for a given value of $\rho$ (0.0 < $\rho$ < 1.0), the scale economies of Operators A, B, and C decrease in order of their sizes, that is, the scale economy of Operator A is largest, that of Operator C is smallest, and that of Operator B somewhere in between.

Suppose that the average walking speed of passengers and the length for each of the walk links are 5 km/h and 0.1 km, respectively. The average operating speed of the vehicles and the vehicle capacity of each operator are 18 km/h and 120 passengers/veh, respectively. The fixed cost, $\mu^0$, per unit of transit line for each operator is $2,000/h. The variable operating cost, $\eta$, per vehicle-hour is $50/veh-hour. The coefficients of the disutility function in Equation (2) are: $\alpha_0 = 5.0$, $\alpha_1 = 10.0$, $\alpha_2 = 10.0$, and $\alpha_3 = 7.5$ ($/h). The other model parameters are: $\bar{D} = 3000$, $\rho = 0.5$, $\theta = 1.0$, $\beta = 0.5$, $\kappa_0 = 0.02$ and $\kappa_1 = 2.0$. In the following analysis, unless specifically stated otherwise, these input data and the model parameter values are considered as the reference case.

4.2 Analyses of results

The proposed solution algorithm was coded in programming language C and run on a personal computer with an Intel Pentium 1.4-GHz CPU and 256 MB of random-access memory (RAM). The iterative process of the solution algorithm is as follows: for a given allocation scheme for the new line L3 (e.g. L3 is allocated to Operator A), the proposed algorithm sequentially checks the alternative allocation schemes for the new line L4, considering one input at a time (i.e. L4 is allocated to one alternative operator at a time). Then, the allocation scheme for L3 is updated (e.g. L3 is allocated to the next operator), and the alternative schemes for L4 are again sequentially checked one input at a time. The process is repeated until the value of objective function (17) cannot be further increased. The iterative process for the reference case takes about 52 seconds of CPU time.

Tables 2 and 3 show the impacts of the introduction of the new transit lines L3 and L4 on the transit system when $\bar{D}$ and $\rho$ are fixed as 3000 pass/h and 0.5, respectively. Note that before the introduction of lines L3 and L4, the transit market is a monopolistic one in which there is only one operator (i.e. Operator A), which operates the two existing lines: L1 and L2. After the introduction of lines L3 and L4, Operators B and C join the market for bidding the
additional lines, thus leading to the oligopolistic competition between operators. As a result of the competition, the additional allocation leads to a sharp decrease in the fares of the existing transit services, as shown in Table 2. Specifically, the fares of lines L1 and L2 decrease from $9.9 to $6.3 and from $10.0 to $6.4, respectively, thus causing a slight decrease in their corresponding service frequencies.

Table 2 also shows that the new line L3 (10 km length) is allocated to Operator B (a medium-scale company), and the new line L4 (15 km length) is allocated to Operator C (a small-scale company). That is, short-haul lines are allocated to larger companies, whereas long-haul lines are allocated to smaller companies. This means that the optimal allocation scheme of the additional lines depends on the company sizes of bidders. An investigation of this impact would make an interesting and important research topic, but is outside the scope of this paper and thus left to a future study. In addition, the fares (or frequencies) of lines L1 and L2 are higher (or lower) than those of lines L3 and L4, respectively. This is because among the three bidders, Operator A has the largest market share (almost 50%) to an extent that gives it significant market power in setting fares.

Table 3 shows the effects of the addition of the new lines L3 and L4 on the transit system performance, in terms of the total resultant market demand, total system profit and the social welfare of the system. It can be seen in Table 3 that after the introduction of lines L3 and L4, all these performances concerned dramatically increase. Specifically, the resultant market demand increases to 2,076 from 998 (pass/h), the total system profit increases to 10,116 from 8,918 ($/h), and the social welfare increases to 30,857 from 18,895 ($/h).

Table 4 shows the impacts of the scale economies of transit operations and the market demand levels. Two different values are specified for the scale economy parameter $\rho$ and the potential market demand level $D$. The two values of the parameter $\rho$ are 0.5 and 1.0, respectively. $\rho = 1.0$ is actually associated with a case that does not consider the effects of the scale economies. The two potential market demand levels are 3,000 and 6,000 pass/h, respectively. It can be seen in Table 4 that when $\rho = 1.0$ at the low potential demand level (i.e., 3,000 pass/h), the two new lines L3 and L4 are simultaneously allocated to Operator B, and Operator C, which is a small-scale company, has no incentive to join the market, thus leaving just two operators (i.e. Operators A and B). This is because if the effects of the scale
economies are ignored at such a low demand level, then the profit of Operator C decreases from the additional line services due to the overestimated operating cost. However, when the effects of the scale economies are considered (i.e., $\rho = 0.5$), then Operator C can achieve a positive profit increase by serving an extra line. This indicates that ignoring the effects of the scale economies can lead an authority to make a biased decision.

Table 4 also shows that when the potential demand level is doubled (i.e., $D$ increases from 3,000 to 6,000 pass/h), Operator C has the incentive of positive profit to join the market, regardless of whether the effects of scale economies are considered or not. This means that the market demand levels have an important impact on the optimal number of operators and the allocation of new lines as well. In particular, when $\rho = 1.0$, allocating either one of the new lines L3 and L4 to Operator B and the other to Operator C yields the same system performance, both in terms of the total resultant market demand, total system profit and social welfare. This is because when $\rho = 1.0$, the operating costs of Operator B and C are linear with regard to the number of lines that they operate and the service frequencies that they set. However, when $\rho = 0.5$, the short-haul line L3 is allocated to the larger company (Operator B) and the long-haul line L4 is allocated to the smaller company (Operator C). Table 4 also shows that for a given potential demand level, the total system profit and the social welfare when $\rho = 1.0$ is smaller than when $\rho = 0.5$. This is because ignoring the effects of scale economies leads to the overestimation of the transit operating costs and thus the underestimation of the benefits or profits of transit operations and the social welfare of the transit system.

5 Conclusion and future studies

This paper has proposed a new model for simultaneous optimization of the number of operators and the allocation of new transit lines in an oligopolistic transit market. The proposed model addressed the interaction among the transit authority, transit operators, and transit passengers and also explicitly incorporated the impacts of the scale economics of the transit operations. On the basis of a multi-level hierarchical system framework, the proposed model was formulated as a 0-1 integer programming problem which can be solved by an implicit enumeration algorithm. The proposed model provides a useful tool for modeling competitive transit services and evaluating transit policies at the strategic level.
A numerical example was used to illustrate the application of the proposed model and solution algorithm and to assess the impacts of the additional allocation of new transit lines, the scale economies and the market demand levels. Some new insights and important findings have been obtained, as follows. (1) The introduction of new transit lines significantly influences the equilibrium fares and frequencies of the existing transit services and the transit system performances, in terms of the total resultant market demand, total system profit and social welfare. For example, after the introduction of new lines, both the fares and frequencies of the existing transit services are reduced. (2) The optimal allocation scheme of new transit lines is related to the company sizes of bidders or operators. (3) Both the scale economies and the market demand level have significant impacts on the optimal number of operators and the allocation schemes of new lines. In particular, ignoring the effects of scale economies would underestimate the benefits of transit operations and the social welfare of the system. Hence, the transit authority may make a biased decision in allocating new transit lines to competitive bidders.

It should be pointed out that, although the numerical results that are presented in this paper are sensible, case studies on large and realistic transit networks are necessary to further justify the findings of this paper and the performance of the proposed model. Other relevant issues that we are currently investigating include a consideration of collusive behavior between different operators (Williams and Martin, 1993; Ibeas et al., 2006), the explicit incorporation of the effects of transit service unreliability (Bell et al., 2002; Li et al., 2008) and passenger heterogeneity (Yang et al., 2001), the optimization of transit fleet size (Ceder, 2005), and the development of efficient solution algorithms for solving the simultaneous optimization problem in large-scale transit networks.

Acknowledgements

The authors would like to thank two anonymous referees for their helpful comments and suggestions on an earlier draft of the paper. The work that is described in this paper was supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region (PolyU 5202/06E), the Research Committee of the Hong Kong Polytechnic University (G-YX1V), the National Natural Science Foundation of China (70971045), the Research Foundation for the Author of National Excellent Doctoral
Dissertation of PR China (200963), and the University Research Committee of the University of Hong Kong (10400582).

References

Li ZC, Lam WHK, Wong SC, Zhu DL, Huang HJ (2007) Modeling park-and-ride services in...
Loo BPY, Chow ASY (2008) Changing urban form in Hong Kong: what are the challenges on sustainable transportation? Inter J Sustainable Transp 2: 177-193
Transport Department (2003) Travel characteristics survey 2002—final report, Hong Kong. Transport Department of Hong Kong
Williams HCWL, Martin D (1993) Public transport services under market arrangements II: a model of competition between groups of services. Transp Res 27B: 389-399
Zubieta L (1998) A network equilibrium model for oligopolistic competition in city bus services. Transp Res 32B: 413-422
Table 1 Summary of three interrelated sub-models and associated solution algorithms

<table>
<thead>
<tr>
<th>Transit authority sub-model (Proposition 3)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Output:</strong> allocation of lines (x)</td>
<td></td>
</tr>
<tr>
<td><strong>Input:</strong> service fare (P), frequency (f), and passenger demand on each line (D)</td>
<td></td>
</tr>
<tr>
<td><strong>Solution algorithm:</strong> see Section 3.3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Transit operator optimal fare and frequency sub-model (Proposition 2)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Output:</strong> transit service fare (P) and frequency (f)</td>
<td></td>
</tr>
<tr>
<td><strong>Input:</strong> allocation of lines (x), and passenger demand on each line (D)</td>
<td></td>
</tr>
<tr>
<td><strong>Solution algorithm:</strong> see Section 3.2</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Transit passenger travel choice sub-model (Proposition 1)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Output:</strong> passenger demand on each line (D)</td>
<td></td>
</tr>
<tr>
<td><strong>Input:</strong> allocation of lines (x), service fare (P) and frequency (f)</td>
<td></td>
</tr>
<tr>
<td><strong>Solution algorithm:</strong> see Section 3.1 and Huang and Li (2007)</td>
<td></td>
</tr>
</tbody>
</table>

Table 2 Optimal allocation schemes of lines, fares and frequencies before and after introduction of new lines

<table>
<thead>
<tr>
<th></th>
<th>Operator</th>
<th>Optimal allocation scheme of lines</th>
<th>Fare ($)</th>
<th>Frequency (veh/h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before the introduction of new transit lines (monopoly market)</td>
<td>A</td>
<td>L1</td>
<td>9.9</td>
<td>7.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L2</td>
<td>10.0</td>
<td>5.3</td>
</tr>
<tr>
<td>After the introduction of new transit lines (oligopoly market)</td>
<td>A</td>
<td>L1</td>
<td>6.3</td>
<td>7.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L2</td>
<td>6.4</td>
<td>5.1</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>L3</td>
<td>5.8</td>
<td>7.8</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>L4</td>
<td>5.6</td>
<td>5.6</td>
</tr>
</tbody>
</table>

Table 3 Resultant system performances before and after introduction of new lines

<table>
<thead>
<tr>
<th></th>
<th>Resultant market demand (pass/h)</th>
<th>Profit ($/h)</th>
<th>Social welfare ($/h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before the introduction of new transit lines (monopoly market)</td>
<td>998</td>
<td>8,918</td>
<td>18,895</td>
</tr>
<tr>
<td>After the introduction of new transit lines (oligopoly market)</td>
<td>A: 959</td>
<td>A: 5,170</td>
<td>30,857</td>
</tr>
<tr>
<td></td>
<td>B: 637</td>
<td>B: 3,179</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C: 480</td>
<td>C: 1,767</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total: 2,076</td>
<td>Total: 10,116</td>
<td></td>
</tr>
</tbody>
</table>
Table 4  Effects of scale economies and market demand levels on transit system

<table>
<thead>
<tr>
<th>Potential demand level (pass/h)</th>
<th>Optimal number of operators</th>
<th>Resultant market demand (pass/h)</th>
<th>Total profit ($/h)</th>
<th>Social welfare ($/h)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \rho = 1.0 )</td>
<td>( \rho = 0.5 )</td>
<td>( \rho = 1.0 )</td>
<td>( \rho = 0.5 )</td>
</tr>
<tr>
<td>( \bar{D} = 3,000 )</td>
<td>A: L1, L2</td>
<td>2</td>
<td>1,971</td>
<td>2,076</td>
</tr>
<tr>
<td></td>
<td>B: L3, L4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>C: L4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \bar{D} = 6,000 )</td>
<td>A: L1, L2</td>
<td>3</td>
<td>4,222</td>
<td>4,222</td>
</tr>
<tr>
<td></td>
<td>B: L3 (or L4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>C: L4 (or L3)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 1  The hierarchical structure of three types of players in the transit market
Fig. 2 Change in the fixed cost of operator $k$ with the number of lines operated

Fig. 3 The example transit network