Resolving the issue of frequency undersupply or oversupply in transit operations under monopoly

Junlin Zhang, Hai Yang
Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, P.R. China

Abstract
The conflicting arguments on the issue of frequency undersupply or oversupply implies an inadequate understanding in the connection between the transit market and the commodity market. To resolve the frequency supply issue and elucidate the correspondences between the two markets, we present a general modeling framework for transit operations under monopoly.

In a monopoly transit market, there exists an interplay among passengers, the transit operator and the regulator. The demand structure of passengers is based on the definition of passengers’ willingness to pay. Price and quality are the two major instruments that the operator manages to achieve its targets. The full price in transit is a passenger’s generalized cost per trip. While transit frequency is an important measure of service quality that can be flexibly adjusted during operations. Transit operator’s objective is a weighted combination of profit and consumer surplus with a balancing parameter that indicates the operator’s nature. Propositions of independence of frequency rule, marginal cost pricing, free public transport and price elasticity at optimum are revealed under general assumptions of cost structures. Model variations shown in the mainstream of literature are discussed. Regulation policies and influences are analyzed. Passengers’ heterogeneity in value of time is investigated. A decision tree of frequency undersupply or oversupply is generated.

Keywords:
Monopoly transit market; frequency supply; regulation policies; willingness to pay; value of time

1. Introduction

There is a research debate recently (van Reeven, 2008; Basso and Jara-Díaz, 2010; Savage and Small, 2010; Karamychev and van Reeven, 2010) about whether or not the Morhing effect (Mohring, 1972) is relevant in support of transit subsidies and brings the frequency supply issue in transit operations. It is shown in these papers that whether the profit-maximizing operator will undersupply or oversupply frequency relative to the socially optimal operator depends on specifications of demand and cost structures. Gómez-Lobo (2014) pointed out that all the above discussions are special cases of Spence (1975) when the provision of frequency is treated as a
quality attribute. Spence (1975) and Sheshinski (1976) articulated that the monopoly’s profit-maximizing decision is likely to be biased away from the social optimum due to possible differences between average and marginal valuation of quality. If the transit market, of which value of time (VOT) is a critical indicator of the valuation of service quality, is indeed a special case of the commodity market, passengers’ heterogeneity in VOT then must play a role in frequency undersupply or oversupply. Yet this is not explained by Gómez-Lobo (2014). Without clarifying passengers’ heterogeneity in VOT, the connection between the transit market and the commodity market remains unclear.

Another problem is that Spence-Sheshinski model (Spence, 1975; Sheshinski, 1976) has its limitations when applied to transit operations directly. Product quantity and quality are taken as the decision variables in Spence-Sheshinski model because they are under full control of the monopolist in a commodity market. Unfortunately, the quantity of rides or passenger demand in a transit market is not fully controlled by the operator; instead it is the service price and quality that the operator manages to accommodate passengers and achieve its desired goals.

To handle these research questions, we present a general modeling framework for transit operations under monopoly. In a monopoly transit market, there exists an interplay among passengers, the transit operator and the regulator. A passenger’s selection on transit is affected by many factors such as transit fare, travel time, access convenience and comfort levels. Passengers perceive differently on the same transit service due to their diverse backgrounds such as income, habits and knowledge. As a result, the aggregate passenger behaviors are hard to predict and the estimation of passenger demand becomes complex mechanisms (Ceder, 2015). Savage (2010) expressed passenger demand as a function of passengers’ generalized cost and a set of exogenous demand variables. To incorporate passengers’ characteristics and their heterogeneity, we further present a form of demand structure of passengers based on the definition of passengers’ willingness to pay (WTP). Various demand structures in the literature can be derived as model variations.

As mentioned, service price and quality are the two major instruments that the operator manages to approach its objective. The full price in transit is a passenger’s generalized cost per trip. While transit frequency is an important measure of service quality that can be flexibly adjusted during operations. Therefore, we take transit fare and frequency as our decision variables. A large number of papers (e.g. Newell, 1971; Jansson, 1980; Jansson, 1993; Pedersen, 2003; Basso and Jara-Díaz, 2010; Moccia and Laporte, 2016) have studied fare and frequency determinations over the past decades. We adopt transit operator’s objective given by Jørgensen and Pedersen (2004). The objective is a weighted combination of operator’s profit and consumer surplus. We introduce
a balancing parameter that indicates the operator’s nature. The objectives of profit maximization and social welfare maximization are thus included with different values of the balancing parameter. We find that optimal solutions crucially depend on the cost structures of passengers and the operator. Significant implications such as independence of frequency rule, marginal cost pricing, free public transport and price elasticity at optimum are revealed under general assumptions of cost structures.

For the regulator, it has its own concerns and is responsible to supervise the transit market to perform healthily. We explore the regulation policies that can be utilized by the regulator. Four specific policy tools, namely objective regulation, fare regulation, frequency regulation, and fiscal regulation (called “OFFF” regulations), and their influences are analyzed. The equivalent influences of fiscal regulation and objective regulation are discussed.

After establishing the modeling framework with passengers’ homogeneity in VOT, we continue to investigate passengers’ heterogeneity in VOT. The existence of both heterogeneous WTP and heterogeneous VOT alters the demand structure of passengers. We show that the main results found with homogeneous VOT still apply to generally heterogeneous passengers. The proposition of frequency undersupply with given demand is consistent with the proposition of quality undersupply given by Spence (1975). It shows that the profit-maximizing operator deviates from the socially optimal operator in frequency supply if the expected VOT differs between realized passengers and marginal passengers. With this proposition, the connection between the transit market and the commodity market becomes evident. And the issue of frequency undersupply or oversupply can be fully resolved with all of our findings. A decision tree is generated and the cases considered in previous studies correspond to only specific branches respectively.

The remainder of the paper comprises six sections. In Section 2, the modeling framework with passengers’ homogeneity in VOT is established. Model variations shown in the mainstream of literature are described in Section 3. Regulations policies and influences are analyzed in Section 4. Passengers’ heterogeneity in VOT is investigated in Section 5. The issue of frequency undersupply or oversupply is discussed in Section 6. We give concluding remarks in Section 7. The primary notations used in the paper are listed in Table 1.
<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P )</td>
<td>Transit fare per trip</td>
</tr>
<tr>
<td>( F )</td>
<td>Transit frequency (reciprocal of transit headway)</td>
</tr>
<tr>
<td>( X )</td>
<td>A vector of exogenous demand variables</td>
</tr>
</tbody>
</table>

\( Q(X) \)  
Potential passenger demand

\( D = q(C,X) = \tilde{q}(P,F,X) \)  
Realized passenger demand

\( K = k(F,D) = \tilde{k}(P,F,X) \)  
Transit operating cost

\( C = c(P,F,D) = \tilde{c}(P,F,X) \)  
Passengers’ generalized cost per trip

\( S = s(P,F,D) = \tilde{s}(P,F,X) \)  
A subsidy scheme proposed by the regulator

\( G(\omega) \)  
Cumulative distribution function of WTP for the group of transit passengers

\( g(\omega) \)  
Probability density function of WTP for the group of transit passengers

\( \omega \)  
Passengers’ WTP

\( \overline{\omega} \)  
The supremum of passengers’ WTP

\( \beta \)  
Passengers’ VOT

\( \pi \)  
Transit operator’s profit

\( \psi \)  
Consumer surplus of transit passengers

\( \gamma \in [0,1] \)  
A balancing parameter in transit operator’s objective

\( h(\omega, \beta) \)  
The bivariate joint probability density function for passengers’ WTP and VOT

\( g_M \)  
Probability density of marginal passengers

\( D_M \)  
Marginal passenger demand

\( E_\Omega(\beta) \)  
Expected VOT of realized passengers

\( E_M(\beta) \)  
Expected VOT of marginal passengers


2. Modeling framework with passengers’ homogeneity in VOT

2.1. Demand structure of passengers
In general, realized passenger demand or the quantity of transit rides \( D \), is expressed as a function of passengers’ generalized cost per trip \( C \) and a vector of exogenous demand variables \( X \) (Savage, 2010):

\[
D = q(C, X) \tag{1}
\]

Passengers’ generalized cost stands for the full price in transit and is written as

\[
C = c(P, F, D) \tag{2}
\]

to include widespread situations. In this expression, \( P \) is the transit fare per trip; while \( F \) is the transit frequency, which is an important measure of service quality. As implied by Eqs. (1) and (2), passenger demand \( D \) is actually a function of transit fare and frequency and we write

\[
D = \tilde{q}(P, F, X) \tag{3}
\]

We take fare and frequency as two decision variables. To avoid ambiguity when calculating partial derivatives, we denote \( C = c(P, F, D) = \tilde{c}(P, F, X) \) when passenger demand is substituted. We suppose \( \partial \tilde{c} / \partial P > 0 \) and \( \partial \tilde{c} / \partial F < 0 \). No more information about passenger demand can be obtained from Eqs. (1)-(3). To further examine the relationship between fare, frequency and passenger demand, we need to make assumptions on passengers’ characteristics and their heterogeneity. We give the following definition of passengers’ WTP.

**Definition.** Each passenger has a reserved maximum amount of the generalized cost, which is the passenger’s willingness to pay (WTP, denoted by \( \omega \)) for the transit trip. A passenger will not use the transit service if his or her perceived generalized cost per trip exceeds the person’s WTP.

Passengers may have different levels of WTP for the transit due to their dissimilar backgrounds. Besides, passengers’ perception of the generalized cost may vary due to their heterogeneity in VOT (Verhoef and Small, 2004). In accordance with the major literature, we assume passengers’ homogeneity in VOT for the time being until Section 5. With identical VOT, passengers’ perceived generalized cost is identical.

Let \( G(\omega) \) be the cumulative distribution function of WTP for the group of transit passengers, which is defined over the interval \([0, \bar{\omega}]\). The symbol \( \bar{\omega} \) denotes the supremum of WTP among all potential passengers. We have \( G(0) = 0 \) and \( G(\bar{\omega}) = 1 \). When the number of the entire passenger group is large enough, \( G(\omega) \) can be approximately supposed to be differentiable, and
we assume a corresponding continuous probability density function for the group of transit passengers, which is denoted by \( g(\omega) \). Accordingly, passenger demand can be rewritten as

\[
D = q(C, X) = Q(X) \cdot [1 - G(C)]
\]

(4)

where \( Q(X) \) is the potential passenger demand that solely depends on the exogenous demand variables.

Eq. (4) adds a ceiling to passenger demand compared with Eq. (1). This restriction is reasonable for the number of passengers is always limited. For example, an apparent constraint is the population in the studied area. With such kind of demand structure of passengers, endogenous and exogenous factors that influence passenger demand are separated, and passengers’ characteristics are incorporated in the distribution function. Various demand structures are derived from this expression, which will be elaborated in Section 3.3.

2.2. Objective and fare-frequency settings of transit operator

We continue to look at the problem of transit operations from the perspective of the operator. Jørgensen and Pedersen (2004) suggested an objective function held by the operator. Their objective function is a weighted combination of the operator’s profit \( \pi \) and consumer surplus \( \psi \):

\[
\max_z = \alpha \pi + (1 - \alpha) \psi
\]

(5)

where \( \alpha \in [0.5, 1] \) is a weighting parameter. Transit operator’s profit is given by the difference between the operator’s farebox revenue \( P \cdot D \) and transit operating cost \( K \). Generally, transit operating cost is a function of transit frequency and passenger demand:

\[
K = k(F, D)
\]

(6)

Likewise, we also denote \( K = k(F, D) = \tilde{k}(P, F, X) \) when passenger demand is substituted. The corresponding partial derivatives are given by

\[
\frac{\partial \tilde{k}}{\partial P} = \frac{\partial k}{\partial D} \cdot \frac{\partial \tilde{q}}{\partial P}
\]

(7)

\[
\frac{\partial \tilde{k}}{\partial F} = \frac{\partial k}{\partial D} + \frac{\partial k}{\partial D} \cdot \frac{\partial \tilde{q}}{\partial F}
\]

(8)

From Eq. (4), consumer surplus is given by

\[
\psi = \int_c^\infty (\omega - C)Q(X)g(\omega)d\omega
\]

(9)

For convenience of subsequent analysis, we introduce a balancing parameter \( \gamma = 1/\alpha - 1 \in [0, 1] \). The detailed form of transit operator’s objective in our model is then written as
\[
\max z = \pi + \gamma \psi = (P \cdot D - K) + \gamma \int_C (\omega - C) Q(X) g(\omega) d\omega 
\] (10)

The balancing parameter indicates the operator’s nature. When \(\gamma = 0 \ (\alpha = 1)\), the operator is a private sector and pursues profit maximization; when \(\gamma = 1 \ (\alpha = 0.5)\), it is a public sector and pursues social welfare maximization; and when \(0 < \gamma < 1 \ (0.5 < \alpha < 1)\), it is a hybrid sector and pursues weighted social optimum.

We present the necessary first order conditions of Eq. (10) with respect to fare and frequency for the unconstrained optimality (Appendix A.1), which then yield

\[
-D = \frac{\partial c}{\partial P^*} \left[ D_M \left( P^* - \frac{\partial k}{\partial P} \right) - \gamma D \right] 
\] (11)

\[
\frac{\partial k}{\partial F^*} = \frac{\partial c}{\partial F^*} \left[ D_M \left( P - \frac{\partial k}{\partial D} \right) - \gamma D \right] 
\] (12)

where \(D_M = \partial q/\partial C = -Qg\) from Eq. (4) is the marginal passenger demand. This definition of marginal demand is a bit different from the one in economics as the full price in transit is passengers’ generalized cost. For given frequency \(F^*\), \(P^*\) determined by Eq. (11) denotes the attained optimal transit fare. For given fare \(P^*\), \(F^*\) determined by Eq. (12) denotes the attained optimal transit frequency. If fare is also optimal in Eq. (12), substituting Eq. (11) into (12) yields

\[
\frac{\partial k}{\partial F^*} \cdot \frac{\partial c}{\partial P^*} = -D \cdot \frac{\partial c}{\partial F^*} 
\] (13)

When the operator is free to adjust both fare and frequency, \(P^*\) and \(F^*\) that are determined by Eqs. (11) and (13) denote the concurrently attained optimal transit fare and frequency. Eqs. (11)-(13) are possible to produce multiple solutions. To simplify discussion, we may impose strictly concave conditions on transit operator’s objective. This requires the Hessian matrix of Eq. (10) with respect to fare and frequency to be negative definite. Additional restrictions on the distribution of passengers’ WTP are thus needed.

2.3. Cost structures and their implications

As we can see from Eqs. (11)-(13), the optimal solutions crucially depend on the cost structures of passengers and the operator. We impose some general restrictions on cost structures of passengers and the operator, and explore corresponding implications. First we give an additional assumption on cost structure of passengers.

**Assumption 1.** The fare effect on passengers’ perceived generalized cost is normalized to unity: \(\partial \tilde{c}/\partial P = 1\).
An additively separable form of passengers’ generalized cost where passengers’ senses of money are not scaled (i.e. cost structure like $C = 2P$ is not accepted) and externality does not incur (i.e. $c(P, F, D) = c(P, F)$ only) satisfies this assumption. Normally, a natural representation of passengers’ generalized cost is a linear combination of collective terms, such as fare, passenger waiting time cost, in-vehicle travel time cost, access time cost, transferring time cost and so on. These terms will be discussed in details in Section 3.2. Assumption 1 excludes cost terms such as boarding and alighting time cost and in-vehicle crowding cost from discussion under this representation of passengers’ generalized cost. We have the following proposition by observing that Eq. (13) is independent of the balancing parameter $\gamma$.

**Proposition 1.** (Independence of frequency rule) Under Assumption 1, if $\partial c/\partial F$ is independent of transit fare $P$, then the optimal frequency rule (optimal frequency and passenger demand relationship) given by Eq. (13) is independent of transit operator’s nature.

Proposition 1 describes the common optimal frequency rule adopted by the operator no matter whether it is a private, public or hybrid sector. For example, several papers (Basso and Jara-Díaz, 2010; Tirachini et al., 2010; Zhang et al., 2016) find that same frequency rule applies under both profit maximization and social welfare maximization with specific demand and cost structures. However, it does not mean that optimal frequency will always be the same irrespective of the operator’s natures. The reason is that optimal fare is dependent on the balancing parameter by Eq. (11), and passenger demand differs when fare changes. Basso and Jara-Díaz (2010) and Tirachini et al. (2010) showed that an approximate ratio of $\sqrt{2}$ is needed for the optimal frequency between social welfare maximization and profit maximization with linear demand structure of passengers. Only when passenger demand is fixed and passengers are homogeneous in VOT, can we claim that optimal frequency yields the same value regardless of operator’s natures. The issue of frequency undersupply or oversupply will be further discussed in Section 6. We have another proposition by observing that Eq. (11) yields $P^* = \partial k/\partial D$ under Assumption 1 and $\gamma = 1$.

**Proposition 2.** (Marginal cost pricing) Under Assumption 1, the socially optimal fare in transit operations should be equal to the marginal cost associated with operating cost and passenger demand.

Pedersen (2003) and Jørgensen and Pedersen (2004) found similar result by using specific demand and cost structures. Assumption 1 guarantees that passengers’ external cost does not
exist or is neutralized\(^1\) so that the optimal price is just equal to the marginal cost, and no Pigovian tax or subsidy is required. If passengers’ senses of money are not scaled and there is a negative externality, we will have \(\frac{\partial c}{\partial P} < 1\) and the optimal price is above marginal cost from Eq. (11). The optimal price is below marginal cost when the externality is positive. This is consistent with the fundamental theory in welfare economics.

Proposition 2 can be treated as a price reference to achieve social optimum in regulation. Fare regulation with marginal cost pricing is necessary to objective regulation with social welfare maximization. Regulation policies and influences will be discussed in Section 4. Another implication of this proposition is transit subsidization. If the operator has economies of scale, transit operating cost per passenger is decreasing with passenger demand. The marginal cost of transit operating cost is lower than the average cost. Marginal cost pricing then results in deficit of the operator. To sustain financial balance of transit service and the objective of social optimum simultaneously, transit subsidy is justified.

We then give an additional assumption on cost structure of the operator.

**Assumption 2.** The impact of passenger demand on transit operating cost is negligible:
\[ \frac{\partial k}{\partial D} = 0. \]

The rationale of this assumption attributes to the distinct feature of the transit market, of which the service is shared with many passengers. Especially when passenger demand is low, the additional costs created by passengers (e.g. fuel consumption, wear and tear, ticket sales, fare enforcement and information provision) can be omitted. With this acceptable assumption, we have an immediate corollary from Proposition 2.

**Corollary. (Free public transport)** Under Assumptions 1 and 2, the socially optimal fare in transit operations should be zero.

Basso and Jara-Díaz (2010), Savage and Small (2010) and Karamychev and van Reeven (2010) mentioned zero fare under social welfare maximization, but based on specific demand and cost structures. For example, Basso and Jara-Díaz (2010) assumed a linear demand structure, a cost structure of passengers only composed of fare and waiting time costs, and a simple cost structure of the operator being proportional to frequency. In contrast, the above corollary provides a solid theoretical support for the promotion of free public transport under general Assumptions 1 and 2.

A number of empirical studies were conducted on the impacts of free public transport. De Witte et al. (2006; 2008) surveyed the impacts of free public transport in Brussels. Thøgersen (2009)

\(^1\)Consider \( C = 2P + 0.01D \) and \( D = Q - 100P \). Such cost and demand structures may not be reasonable but they satisfy Assumption 1 and exhibit externality.
evaluated the policy of free public transport through a field experiment. Cools et al. (2016) examined zero-price effects (an overreaction to a free product) and recommended introducing free fare to increase public transport usage. Indeed, some cities or towns around the world already introduced various forms (either route-specific or area-wide) of free public transport. The Belgian city of Hasselt was a pioneer to introduce free public transport for both inhabitants and visitors in 1997 (van der Vliet, 2009) (this policy was abandoned in 2014 due to financial struggle with rising costs).

Nonetheless, with free public transport, it means that the regulator, or another third party, will pay the cost of public transport, or a subsidy must be introduced in transit operations. Conceivably, the subsidy to sustain free public transport can be substantial and no third parties can bear it in the long run. Yet, free public transport may induce socially undesirable excessive ridership. In the case of a free public transport policy, there might be a sharp increase in passenger demand that may invalidate Assumption 1 or 2, and the corollary may fail.

Under Assumptions 1 and 2, we obtain \( P^* = (\gamma - 1) \cdot D/D_M \) from Eq. (11). By virtue of price elasticity of passenger demand, \( e_p = P/D \cdot \partial q/\partial P = P/D \cdot \partial q/\partial C \cdot \partial c/\partial P = P \cdot D_M / D \), we readily have the following proposition.

**Proposition 3.** *(Price elasticity at optimum)* Under Assumptions 1 and 2, the optimal fare is set so that the price elasticity of passenger demand is equal to the balancing parameter minus one: 
\[
 e_p^* = \gamma - 1.
\]

The price elasticity at optimum equals minus one when the operator pursues profit maximization \( (\gamma = 0) \). In the commodity market, the price elasticity equals minus one under revenue maximization. It is due to Assumption 2 that we have this result under profit maximization in the transit market, in another word, profit maximization and revenue maximization coincide with each other under Assumption 2.

Finally we give a combined assumption on cost structures of passengers and the operator.

**Assumption 3.** \( \frac{\partial \bar{c}}{\partial F} / \frac{\partial k}{\partial F} \) is a single variable function of transit frequency \( F \), and is increasing with \( F \).

This is a relatively strong assumption under which “Mohring effect” (Mohring, 1972) emerges in our study context. As can be verified, this assumption generally holds for typical cost structures of passengers and the operator used in the literature (Section 3). From Eq. (13) and under
Assumption 1 that $\partial \tilde{c}/\partial P = 1$, we have $-D \cdot \left( \frac{\partial \tilde{c}}{\partial F^*} / \frac{\partial k}{\partial F^*} \right) = 1$, which means that a rise in demand increases $\frac{\partial \tilde{c}}{\partial F^*} / \frac{\partial k}{\partial F^*}$, and will increase the concurrently attained optimal frequency $F^*$ as well under Assumption 3. Since $\partial \tilde{c}/\partial F < 0$, a rise in demand then reduces passengers’ generalized cost, which implies increasing returns to scale. This is an observation of the Mohring effect in the sense that average passenger cost decreases with passenger demand. However, Mohring effect can still occur when the operator has scale diseconomies.\(^2\) Therefore, the operator may nevertheless be able to run a surplus under marginal cost pricing, and subsidy may not be required in order to achieve social welfare maximization. This reminds us of recent debates in the literature (van Reeven, 2008) about whether or not Mohring effect is relevant to support transit subsidies. At least the above analysis tells us that Mohring effect does not suffice to justify transit subsidization under marginal cost pricing.

3. Model variations

In this section, we describe the variations of demand and cost structures and show how our general modeling framework can be specified to give rise to the specific structural models employed in the literature.

3.1. Variations of cost structure of the operator
The following linear cost function of frequency is widely adopted in the literature (Jansson, 1980; Jara-Díaz and Gschwender, 2003; Zhang and Yang, 2016):

$$K = \phi F$$  \hspace{1cm} (14)

When the impact of passenger demand on transit operating cost is considered (Assumption 2 not satisfied), the operating cost is usually expressed as (Tirachini et al., 2010; Gómez-Lobo, 2014):

$$K = \phi F + \lambda D$$  \hspace{1cm} (15)

where $\phi$ and $\lambda$ are the relevant parameters.

\(^2\) Scale diseconomies lead to $k/D < \partial k/\partial D$ in our study context. Consider $K = \phi F + \lambda D - \theta$, where $\phi$, $\lambda$ and $\theta$ are the relevant positive parameters. Scale diseconomies and Mohring effect (Assumption 3 is valid) can both happen when $\theta$ is large.
3.2. Variations of cost structure of passengers
The cost structure of passengers is much more complicated than the cost structure of the operator in the literature. A linear representation of passengers’ generalized cost consists of collective terms, such as transit fare and passenger waiting time cost $C_w$ given below.

$$C = P + C_w$$  \hspace{1cm} (16)

By assuming average waiting time equal to half headway (as initially assumed in Mohring, 1972), then $C_w = \beta_w/(2F)$, where $\beta_w$ is passengers’ value of waiting time. This assumption is largely adopted by many researchers afterwards (Jansson, 1980; van Reeven, 2008; Zhang et al., 2016). Oldfield and Bly (1988) attempted to take into account the impacts of transit overflow by assuming that passenger waiting time depends on both service frequency and vehicles occupancy rate (some passengers may be unable to board the first arrival vehicle). The average waiting time is hence revised as $C_w = \varepsilon \beta_w/F$, where the correction factor $\varepsilon \geq 0.5$. Tirachini et al. (2010) tried to take into account the impacts of schedule unreliability by assuming that there is a threshold headway for passengers to “forget the timetable”. Their function of passenger waiting time cost is of the form $C_w = \beta_w (t_0 + t_1 \cdot \varepsilon/F)$, where the values of $t_0$ and $t_1$ depend on the threshold.

Besides waiting time cost, the access time cost $C_a$ and the in-vehicle time cost $C_v$ are usually treated as constants (Mohring, 1972). Passengers’ boarding and alighting time cost $C_b$ depends on queue length. Hence it relates to passenger demand and may make Assumption 1 invalid. Jansson (1980) and Tirachini et al. (2010) expressed boarding and alighting time cost as $C_b = \beta_b \cdot D/F$, where $\beta_b$ is the value of passengers’ boarding and alighting time. Additional terms can be considered, but just complicate the cost structure of passengers and makes the solution intractable. For instance, Evans and Morrison (1997) incorporated accident risk and disruption. Moccia and Laporte (2016) evaluated the stop spacing, train length and crowding penalty. Consequently, the linear representation of passengers’ generalized cost can be written as:

$$C = P + C_w + C_a + C_v + C_b + \cdots$$  \hspace{1cm} (17)

If we assume that passengers’ waiting time cost is the dominating one among the frequency-related cost terms, we immediately have the following well-known “square root formula”.

**Proposition 4.** (Square root formula) Under Assumption 1 and with the cost structure of the operator in Eq. (14) or (15) and the cost structure of passengers in Eq. (16), where waiting time cost $C_w$ is inversely proportional to frequency, then the concurrently attained optimal frequency is proportional to the square root of passenger demand.
Proof. From Assumption 1, $\partial c/\partial P = 1$; by Eq. (14) or (15), $\partial k/\partial F = \phi$; and from Eq. (16) with $C_w \propto 1/F$, we have $\partial c/\partial F = -\varepsilon \beta / F^2$. Then the optimal frequency rule by Eq. (13) yields

$$F^* = \sqrt{\frac{\varepsilon \beta D}{\phi}}$$

(18)

This completes the proof. ■

We note here in passing that alternative frequency rule can be obtained under different specification of cost structures. For example, if waiting time cost is assumed to be a negative logarithmic function of frequency (Luethi et al., 2007), ceteris paribus, the optimal frequency rule by Eq. (13) will lead to the “linear formula” (optimal frequency being proportional to passenger demand).

3.3. Variations of demand structure of passengers

The unrealistic but simple demand structure is a fixed demand of homogeneous passengers in WTP (van Reeven, 2008; Tirachini et al., 2010):

$$G(\omega) = \begin{cases} 0, & \omega \in [0, \bar{\omega}) \\ 1, & \omega = \bar{\omega} \end{cases}$$

(19)

Linear passenger demand function is often used (Chang and Schonfeld, 1991; Basso and Jara-Díaz, 2010; Savage, 2010). In this case passengers’ WTP is uniformly distributed:

$$G(\omega) = \frac{\omega}{\bar{\omega}}, \quad \omega \in [0, \bar{\omega}]$$

(20)

More advanced demand function can be generated by using a more general distribution of power function (Karamychev and van Reeven, 2010):

$$G(\omega) = \left(\frac{\omega}{\bar{\omega}}\right)^\kappa, \quad \omega \in [0, \bar{\omega}] \quad and \quad \kappa \in [1, +\infty)$$

(21)

As $\kappa$ increases, passengers’ heterogeneity in WTP decreases.

Demand function with constant elasticity is also of wide interest (Oldfield and Bly, 1988). It can be obtained from the following WTP distribution:

$$G(\omega) = 1 - \left(\frac{\omega}{\omega_0}\right)^{-\kappa}, \quad \omega \in [\omega_0, \bar{\omega}] \quad and \quad \bar{\omega} \gg \omega_0, \quad \kappa > 0$$

(22)

where $\omega_0$ is an artificial lower bound of passengers’ minimum WTP to avoid unlimited passenger demand. Other distribution functions of passengers’ WTP can be considered as well, such as gamma distribution or log-normal distribution.
4. Regulation policies and influences

In this section we look at transit operations under regulation. At least four kinds of policy tools can be utilized by the regulator in our framework, namely, objective regulation, fare regulation, frequency regulation, and fiscal regulation (OFFF regulations). A regulation policy will influence the operator’s decision and probably change its settings of fare and frequency.

Objective regulation is a policy tool whereby the regulator confines the operator’s nature and objective directly. For a given value of balancing parameter \( \gamma \in [0,1] \), the optimal fare and frequency are determined from Eqs. (11) and (13). Fare regulation is a policy tool whereby the regulator imposes restriction on transit fare. For each given fare, the optimal frequency is given by Eq. (12). Frequency regulation is a policy tool whereby the regulator restricts transit frequency. For each given frequency, the optimal fare is given by Eq. (11).

Fiscal regulation involves either levies or subsidies. Suppose that the regulator proposes a subsidy scheme \( S = s(P, F, D) \). It is a levy scheme if \( S < 0 \). Similarly, we denote \( S = s(P, F, D) = \tilde{s}(P, F, X) \) when passenger demand is substituted. Under a subsidy scheme, transit operator’s objective is then amended as follows.

\[
\max z = (P \cdot D - K + S) + \int_\omega^{\tilde{\omega}} (\omega - C) Q(X) g(\omega) d\omega 
\]

The counterparts of the first order conditions (11)-(13) are given as

\[
-\left( D + \frac{\partial \tilde{s}}{\partial P} \right) = \frac{\partial \tilde{c}}{\partial P} \left[ D_M \left( P^* - \frac{\partial k}{\partial D} \right) - \gamma D \right] 
\]

\[
\left( \frac{\partial k}{\partial F^*} \right) - \left( \frac{\partial \tilde{s}}{\partial F^*} \right) = \frac{\partial \tilde{c}}{\partial P^*} \left[ D_M \left( P - \frac{\partial k}{\partial D} \right) - \gamma D \right] 
\]

\[
\left( \frac{\partial k}{\partial F^*} - \frac{\partial \tilde{s}}{\partial F^*} \right) \cdot \frac{\partial \tilde{c}}{\partial P^*} = - \left( D + \frac{\partial \tilde{s}}{\partial F^*} \right) \cdot \frac{\partial \tilde{c}}{\partial F^*}
\]
Fig. 1 illustrates the above four types of regulation policies in the fare–frequency two dimensional space. From the figure one can see how the optimal fare and frequency jointly change with respective regulation variable under each regulation policy. Of particular interest is the overlapping phenomenon of the curves of objective regulation and fiscal regulation. Its implication can be summarized as the following proposition.

**Proposition 5.** *(Fiscal regulation equivalence)* There exists a subsidy scheme \( S = \mu D \) or \( S = \mu \psi \) such that the fiscal regulation and the objective regulation are equivalent to each other, where \( \mu \) is a regulation variable dictating the level of subsidy.

**Proof.** The proof is relegated to Appendix B.1. \( \blacksquare \)

Proposition 5 is very useful. It means that the regulator is able to use the “indirect” fiscal regulation to achieve the same goal as the “direct” objective regulation. Such fiscal regulation is

---

\(^3\) Settings: potential passenger demand \( Q = 60000 \); demand structure \( \ln \omega \sim \mathcal{N}(2.5, 0.5^2) \); cost structure of the operator \( K = \phi F \), \( \phi = 6000 \) HK$; cost structure of passengers \( C = P + \frac{\beta}{2F} \), \( \beta = 20 \) HK$/hr; subsidy scheme \( S \propto D \).
a subsidy scheme based on passenger demand or consumer surplus. As a special case, the regulator is capable of making a profit-maximizing operator behave as a socially optimal one.

5. Passengers’ heterogeneity in VOT

So far we considered elastic transit demand which implies passengers’ heterogeneity in WTP. Now we move on to examine the impacts of passengers’ heterogeneity in VOT. For a passenger with a VOT $\beta$, we consider a cost structure written as

$$C = P + \beta T$$

(27)

where $T = T(F)$ is the time component of a trip, and $dT/dF < 0$.

![Fig. 2. Characteristics of realized passengers and marginal passengers.](image)

With both distributed WTP $\omega$ and VOT $\beta$, a bivariate distribution $(\omega, \beta)$ is thus used to characterize each passenger. Let $h(\omega, \beta)$ denote the joint probability density function of $(\omega, \beta)$ and assume it to be differentiable when there is a large group of passengers. A passenger will choose the transit if the passenger’s WTP is not less than his or her perceived generalized cost, i.e. $\omega \geq P + \beta T$. We suppose that the supremum of passengers’ VOT $\bar{\beta} \geq (\bar{\omega} - P)/T$ (the domain
of the bivariate distribution can be extended if this is not satisfied). As shown in Fig. 2, the characteristics of realized passengers are defined in the closed region $\Omega$:

$$\Omega = \{(\omega, \beta) | \omega \geq P + \beta T, 0 \leq \omega \leq \bar{\omega}, \beta \geq 0\}$$

(28)

and the characteristics of marginal passengers are defined in the line segment $M$:

$$M = \{(\omega, \beta) | \omega = P + \beta T, P \leq \omega \leq \bar{\omega}\}$$

(29)

Hence passenger demand can be written as

$$D = \bar{q}(P, F, X) = Q(X) \int_{(\omega, \beta) \in \Omega} h(\omega, \beta) d\omega d\beta$$

(30)

Comparing Eq. (30) with Eqs. (4) and (27) reveals that the demand structure of passengers is altered due to passengers’ heterogeneity in VOT. After incorporating the above demand structure into transit operator’s objective, we obtain the following first-order optimality conditions with respect to fare and frequency (Appendix A.2).

$$-D = D_M \left( P^* - \frac{\partial k}{\partial D} \right) - \gamma D$$

(31)

$$\frac{\partial k}{\partial F^*} = \frac{dT}{dF^*} \left[ D_M \left( P - \frac{\partial k}{\partial D} \right) \cdot E_M(\beta) - \gamma D \cdot E_\Omega(\beta) \right]$$

(32)

where $D_M = -Q \cdot g_M$ is the marginal passenger demand; $E_M(\beta)$ is the expected VOT of the marginal passengers; and $E_\Omega(\beta)$ is the expected VOT of realized passengers. Substituting Eq. (31) into (32) yields

$$\frac{\partial k}{\partial F^*} = -\frac{dT}{dF^*} \left[ \gamma E_\Omega(\beta) + (1 - \gamma) E_M(\beta) \right]$$

(33)

Independence of frequency rule by Proposition 1 becomes invalid if the expected VOT differs between realized passengers and marginal passengers. We then have the following proposition.

**Proposition 6.** *(Frequency undersupply with given demand)* Under Assumption 3 and with the cost structure of passengers in Eq. (27), for given passenger demand, the profit-maximizing operator undersupplies frequency relative to the socially optimal operator when $E_\Omega(\beta) > E_M(\beta)$ and conversely.

Given demand does not mean that demand is inelastic to passengers’ generalized cost because it also depends on exogenous demand variables. Proposition 6 is consistent with the proposition of quality undersupply given by Spence (1975) and Sheshinski (1976). It points to the fact that VOT is a critical indicator of the valuation of quality in transit services. With this proposition, the connection between the transit market and the commodity market becomes evident.
One can easily show that the aforementioned propositions of marginal cost pricing, free public transport, price elasticity at optimum and square root formula remain valid in the presence of heterogeneous VOT. The fiscal regulation equivalence with heterogeneous VOT is also valid for a subsidy scheme based on passenger demand $S = \mu D$ and its proof is provided in Appendix B.2. The propositions of the independence of frequency rule and fiscal regulation equivalence for a subsidy scheme based on consumer surplus may become invalid (Appendix B.2) due to the difference in expected VOT between realized and marginal passengers. This extends the scope on the implications of these propositions.

6. Frequency undersupply or oversupply

Being consistent with the observation in Spence-Sheshinski model (Spence, 1975; Sheshinski, 1976), Proposition 6 establishes a firm connection between the transit market and the commodity market. As transit fare and frequency are our two decision variables, a counterpart of quality undersupply in Spence-Sheshinski model is presented for our model.

**Proposition 7.** *(Frequency undersupply with given fare)* For given transit fare, the profit-maximizing operator always undersupplies frequency relative to the socially optimal operator.

**Proof.** As shown by Spence (1975), it is equivalent to show that the consumer surplus derivative $\partial \psi / \partial F > 0$. When passengers are homogeneous in VOT, $\partial \psi / \partial F = -D \cdot \partial \tilde{c} / \partial F$; when passengers are heterogeneous in VOT, $\partial \psi / \partial F = -D \cdot \mathbb{E}_{\tilde{\alpha}} (\beta) \cdot dT / dF$ (see Appendix A). As $\partial \tilde{c} / \partial F < 0$ and $dT / dF < 0$, the proposition is valid. ■

The proposition matches the fact in microeconomics that the firm will undersupply service unless the price is set at the efficient point where it equals marginal cost. Unlike transit fare that is always overcharged by the profit-maximizing operator, frequency undersupply or oversupply depends on specific assumptions and conditions, which are made clear in Assumptions 1 and 3, Eqs. (13) and (33), and Propositions 6 and 7. Fig. 3 depicts the decision tree that leads to the frequency undersupply or oversupply. This decision tree fully resolves the recent debates on the frequency supply issue in the literature. As elaborated as follows, the cases considered in previous studies correspond to only specific branches respectively.

van Reeven (2008). Passengers are homogeneous in VOT. The passenger demand is fixed and the WTP distribution is given by Eq. (19). So the demand for both profit and welfare
maximization is identical: \( D^\pi = D^o \). The cost structures of the operator and passengers are given by Eq. (14) and Eq. (16), respectively. The authors obtained the square root formula. The profit and welfare maximizing frequencies are the same when consumers ignore the schedule timetable. For consumers who are aware of the timetable, they showed that the average waiting time reduces by half and affects the optimal frequency under social welfare maximization. The frequency for profit maximization is then higher.

Basso and Jara-Díaz (2010). Passengers are homogeneous in VOT, and a linear passenger demand structure is assumed and the WTP distribution is given by Eq. (20). The cost structures of the operator and passengers are given by Eq. (14) and Eq. (16), respectively. The authors also obtained the square root formula. The stable passenger demand solution for profit maximization is less than that for welfare maximization: \( D^\pi < D^o \). If allowed to choose fare freely, the profit-maximizing operator will undersupply frequency.

Savage and Small (2010). The demand structure and cost structure of passengers are in general forms, and the cost structure of the operator is given by Eq. (14). The authors found that the profit-maximizing operator will undersupply frequency for a given fare. Without specification of the demand and cost structures, they claimed that it is theoretically indeterminate which situation

---

**Fig. 3.** Decision tree of frequency undersupply or oversupply.
produces a higher frequency. This is contrasted with our analysis in Fig. 3, where undersupply or oversupply of frequency can be ascertained in each branch of the general decision tree.

Karamychev and van Reeven (2010). Passengers are homogeneous in VOT. The demand structure is such that the WTP distribution is of a power function form given by Eq. (21). The cost structure of the operator and passengers are given by Eq. (14) and Eq. (16), respectively. Whether undersupply or oversupply of frequency occurs depends on the parameter $\kappa$ in the demand structure. As $\kappa$ increases, passenger demand becomes less elastic, in the extreme case when $\kappa$ goes to infinity, it reduces to the fixed demand considered by van Reeven (2008); when $\kappa = 1$, it becomes the model by Basso and Jara-Díaz (2010). Thus, the authors concluded that either undersupply or oversupply of frequency can occur, depending on the degree of passengers’ heterogeneity in WTP.

7. Concluding remarks

This study investigates transit operations under monopoly using a general modeling framework. The demand structure of passengers is based on the definition of passengers’ WTP. Transit operator’s objective is a weighted combination of operator’s profit and consumer surplus. We introduce a balancing parameter that indicates the operator’s nature. We find that the optimal fare and frequency crucially depend on the cost structures of passengers and the operator. Propositions of independence of frequency rule, marginal cost pricing, free public transport and price elasticity at optimum are shown under general assumptions of cost structures. Specific demand structure of passengers, transit operator’s objective, and cost structures of passengers and the operator are discussed as model variations. OFFF regulations and their influences are analyzed. The fiscal regulation equivalence is established.

Transit service can be regarded as a kind of commodity and general theories of the commodity market shall be applicable to the transit market as well. The quantity of rides is passenger demand. The full price in transit is the generalized cost per trip. While transit frequency is an important measure of service quality that can be flexibly adjusted during operations. And VOT is a critical indicator of the valuation of quality in transit service. With these correspondences, we establish the connection between the transit market and the commodity market under monopoly. Passengers’ heterogeneity in VOT is investigated. We show that the main results found with a homogeneous VOT still apply to the case with generally heterogeneous passengers. The issue and debates of frequency undersupply or oversupply are fully resolved.
The transit market also has its distinct features compared with the commodity market. Its service is shared with many passengers and its operator is unable to fully control the quantity of rides. Furthermore, transit services generally exhibit Mohring effect and government subsidy is often introduced. Transit subsidies are generally justified on the ground of economies of scale in the literature. This study defends that economies of scale constitute a justification for transit subsidization but finds that Mohring effect alone does not suffice to support transit subsidies. There are two reasons for the latter. First, the operator may nevertheless be able to run a surplus under marginal cost pricing; second, either frequency undersupply or oversupply may occur under different conditions.

Acknowledgements. The work described in this paper is supported by a grant from Hong Kong’s Research Grants Council under project No. HKUST16222916.

Appendix A. First order conditions of transit operator’s objective

A.1. Passengers’ homogeneity in VOT
Let \( D_m = \frac{\partial q}{\partial C} = -Q_g \) from Eq. (4) as marginal passenger demand, the derivatives of transit operator’s objective with respect to fare and frequency are given below.

\[
\frac{\partial z}{\partial P^*} = \left( P^* \frac{\partial \hat{q}}{\partial P^*} + D - \frac{\partial k}{\partial D} \frac{\partial \hat{q}}{\partial P^*} \right) + \gamma \frac{\partial \psi}{\partial P^*}
\]

\[
= \left( P^* \cdot D_m \frac{\partial \hat{c}}{\partial P^*} + D - \frac{\partial k}{\partial D} \cdot D_m \frac{\partial \hat{c}}{\partial P^*} \right) - \gamma D \frac{\partial \hat{c}}{\partial P^*} \tag{34}
\]

\[
= \frac{\partial \hat{c}}{\partial P^*} \left[ D_m \left( P^* - \frac{\partial k}{\partial D} \right) - \gamma D \right] + D = 0
\]

\[
\frac{\partial z}{\partial F^*} = \left( P \cdot \frac{\partial \hat{q}}{\partial F^*} - \frac{\partial \hat{k}}{\partial F^*} \right) + \gamma \frac{\partial \psi}{\partial F^*}
\]

\[
= \left[ P \cdot D_m \frac{\partial \hat{c}}{\partial F^*} - \left( \frac{\partial k}{\partial F^*} + \frac{\partial k}{\partial D} \cdot D_m \frac{\partial \hat{c}}{\partial F^*} \right) \right] - \gamma D \frac{\partial \hat{c}}{\partial F^*} \tag{35}
\]

\[
= \frac{\partial \hat{c}}{\partial F^*} \left[ D_m \left( P - \frac{\partial k}{\partial D} \right) - \gamma D \right] - \frac{\partial k}{\partial F^*} = 0
\]

The derivative term with star means the derivative evaluated at the optimal value, e.g. \( \frac{\partial k}{\partial F^*} \) means \( \frac{\partial k}{\partial F} \left( F^*, D \right) \). We then have the first order conditions of transit operator’s objective with passengers’ homogeneity in VOT.
A.2. Passengers’ heterogeneity in VOT

The following derivatives are calculated by using Leibniz’s rule for differentiation under the integral sign (Flanders, 1973). From Eq. (30), passenger demand is given by

\[
D = \tilde{q}(P, F, X) = Q \cdot P_\Omega = Q \cdot \int_0^{\beta_{-P}} d\beta \int_{P+\beta T}^{\beta} h(\omega, \beta) d\omega
\]

where \( P_\Omega = \int_{(\omega, \beta) \in \Omega} h(\omega, \beta) d\omega d\beta \) is the portion of realized passengers in region \( \Omega \). By the cost structure of Eq. (27), We have

\[
\frac{\partial \tilde{q}}{\partial P} = -Q \cdot \int_0^{\beta_{-P}} h(P + \beta T, \beta) d\beta = -Q \cdot g_M
\]

where \( g_M = \int_0^{\beta_{-P}} h(P + \beta T, \beta) d\beta \) is the probability density of marginal passengers on margin \( M \), and marginal passenger demand \( D_M = \frac{\partial \tilde{q}}{\partial C} = \frac{\partial \tilde{q}}{\partial P} / \frac{\partial c}{\partial P} = \frac{\partial \tilde{q}}{\partial P} = -Q \cdot g_M \). The derivative in frequency is given by

\[
\frac{\partial \tilde{q}}{\partial F} = -\frac{dT}{dF} Q \cdot \int_0^{\beta_{-P}} \beta h(P + \beta T, \beta) d\beta
\]

\[
= -\frac{dT}{dF} Q g_M \cdot \int_0^{\beta_{-P}} \beta h(P + \beta T, \beta) \frac{g_M}{g_M} d\beta = D_M \cdot E_M(\beta) \frac{dT}{dF}
\]

where \( E_M(\beta) = \int_0^{\beta_{-P}} \frac{\beta h(P + \beta T, \beta)}{g_M} d\beta \) is the conditional expectation of VOT on margin \( M \) or it is the expected VOT of marginal passengers. Consumer surplus is given by

\[
\psi = Q \cdot \int_{(\omega, \beta) \in \Omega} \left[ \omega - (P + \beta T) \right] h(\omega, \beta) d\omega d\beta
\]

and the derivatives of consumer surplus in fare and frequency are given by

\[
\frac{\partial \psi}{\partial P} = -Q \cdot \int_0^{\beta_{-P}} d\beta \int_{P+\beta T}^{\beta} h(\omega, \beta) d\omega = -D
\]

\[
\frac{\partial \psi}{\partial F} = -\frac{dT}{dF} Q \cdot \int_{(\omega, \beta) \in \Omega} \beta h(\omega, \beta) d\omega d\beta
\]

\[
= -\frac{dT}{dF} Q \cdot P_\Omega \cdot \int_{(\omega, \beta) \in \Omega} \beta \frac{h(\omega, \beta)}{P_\Omega} d\omega d\beta = -D \cdot E_\Omega(\beta) \frac{dT}{dF}
\]

where \( E_\Omega(\beta) = \int_{(\omega, \beta) \in \Omega} \frac{\beta h(\omega, \beta)}{P_\Omega} d\omega d\beta \) is the conditional expectation of VOT in region \( \Omega \) or it is the expected VOT of realized passengers.
With these derivatives, we are able to derive the first order conditions of transit operator’s objective with passengers’ heterogeneity in VOT, which are given by Eqs. (31) and (32). When passengers’ heterogeneity in VOT disappears, we will have the same first order conditions with passengers’ homogeneity in VOT under the specific cost structure of Eq. (27).

Appendix B. Proof of Proposition 5

Suppose the objective regulation requires that the operator’s nature moves from $\gamma_0$ to $\gamma_1$, $\gamma_0, \gamma_1 \in [0,1]$, the corresponding optimal fare and frequency will shift from $(P_0, F_0)$ to $(P_1, F_1)$, as shown in Fig. 4. And passenger demand also shifts from $D_0$ to $D_1$, with corresponding shift of marginal passenger demand from $D_{M_0}$ to $D_{M_1}$.

B.1. Passengers’ homogeneity in VOT

For transit subsidization based on passenger demand, construct a subsidy scheme $S = \mu D$, where $\mu = - (\gamma_1 - \gamma_0) \cdot D_1 / D_{M_1}$ with $\mu > 0$ if $\gamma_1 > \gamma_0$. After incorporating this subsidy scheme into Eq. (26), the subsidy effect on both sides of Eq. (26) is canceled out:

$$\frac{\partial \delta}{\partial F^*} \cdot \frac{\partial \hat{c}}{\partial P^*} = \mu D_{M_1} \cdot \frac{\partial \hat{c}}{\partial F^*} \cdot \frac{\partial \hat{c}}{\partial P^*} = \frac{\partial \delta}{\partial F^*} \cdot \frac{\partial \hat{c}}{\partial P^*}$$

(42)

As already mentioned, passenger demand is a function of fare and frequency given by Eq. (3). Therefore, Eq. (13) or (26) is actually an implicit function of the optimal fare and frequency. The cancellation of the subsidy effect in Eq. (26) means the same curve on the fare-frequency two-dimensional space in the presence and absence of such a subsidy scheme. This explains the overlapping between the fiscal regulation and objective regulation in Fig. 1.

We now move on to check whether optimal fare and frequency will shift to $(P_1, F_1)$ with the operator’s nature $\gamma_0$ and subsidy scheme $S$. We note that

$$- \left( D_1 + \frac{\partial \hat{\delta}}{\partial P_1} \right) = - \left( D_1 + \mu D_{M_1} \frac{\partial \hat{c}}{\partial P_1} \right) = -D_1 + (\gamma_1 - \gamma_0) \cdot D_1 \frac{\partial \hat{c}}{\partial P_1}$$

$$= \frac{\partial \hat{c}}{\partial P_1} \left[ D_{M_1} \left( P_1 - \frac{\partial k}{\partial D_1} \right) - \gamma_1 D_1 \right] + (\gamma_1 - \gamma_0) \cdot D_1 \frac{\partial \hat{c}}{\partial P_1}$$

$$= \frac{\partial \hat{c}}{\partial P_1} \left[ D_{M_1} \left( P_1 - \frac{\partial k}{\partial D_1} \right) - \gamma_0 D_1 \right]$$

(43)
which also satisfies Eq. (24). Thus \((P_1, F_1)\) is attainable under operator’s nature \(\gamma_0\) and subsidy scheme \(S = \mu D\).

For transit subsidization based on consumer surplus, we construct a subsidy scheme \(S = \mu \psi\), where \(\mu = \gamma_1 - \gamma_0\). Similar to the above process, the subsidy effect on both sides of Eq. (26) is canceled out and \((P_1, F_1)\) is also attainable under operator’s nature \(\gamma_0\) and subsidy scheme \(S = \mu \psi\).

![Fig. 4. Equivalent influences of fiscal regulation and objective regulation.](image)

B.2. Passengers’ heterogeneity in VOT

With a subsidy scheme \(S = s(P, F, D) = \tilde{s}(P, F, X)\), Eqs. (31) and (33) are amended to:

\[
-D + \frac{\partial \tilde{s}}{\partial P^*} = D_M \left( P^* - \frac{\partial k}{\partial D} \right) - \gamma D
\]

\[
\frac{\partial k}{\partial F^*} - \frac{\partial \tilde{s}}{\partial F^*} = -\frac{dT}{dF^*} D \left[ \gamma E_M (\beta) + (1 - \gamma) E_M (\beta) \right] - \frac{\partial \tilde{s}}{\partial P^*} E_M (\beta) \cdot \frac{dT}{dF^*}
\]

Passenger demand \(D\) is still a function of fare and frequency, which is now detailed as Eq. (30). For transit subsidization based on passenger demand, construct the same kind of subsidy scheme \(S = \mu D, \mu = -(\gamma_1 - \gamma_0) \cdot D_1 / D_{M1}\), a proof similar to Appendix B.1 applies.

However, for a subsidy scheme \(S = \mu \psi\), the subsidy effect on both sides on Eq. (45) cannot be canceled out:
\[
\frac{\partial \tilde{s}}{\partial F^*} = -\mu D \cdot E_\alpha (\beta) \cdot \frac{dT}{dF^*} = -\frac{\partial \tilde{s}}{\partial P^*} \cdot E_M (\beta) \cdot \frac{dT}{dF^*}
\] (46)

Therefore, the fiscal regulation equivalence for a subsidy scheme based on consumer surplus may fail due to the possible difference in expected VOT between realized and marginal passengers. ■

References


