Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Transportation Research Part A 39 (2005) 17–40
TRANSPORTATION RESEARCH PART A
www.elsevier.com/locate/tra

# Regulating taxi services in the presence of congestion externality

Hai Yang [a,*], Min Ye [a], Wilson H. Tang [a], S.C. Wong [b]

[a] *Department of Civil Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, PR China*

[b] *Department of Civil Engineering, The University of Hong Kong, Hong Kong, PR China*

Received 21 October 2003; received in revised form 28 April 2004; accepted 4 May 2004

**Abstract**

In most large cities, the taxi industry is subject to various types of regulation, such as entry restrictions and price controls, and economists have examined the economic consequences of such regulation extensively. Unfortunately, in conventional economic analyses of competition and regulation in the taxi industry little attention has been paid to one important issue: congestion externalities due to both occupied and vacant taxi movements together with normal vehicular traffic. This study investigates the nature of equilibrium and regulation in the taxi market by taking account of congestion externalities and adopting a realistic distance-based and delay-based taxi fare structure. The monopoly, the social optimum and the stable competitive solutions are examined and illustrated with a numerical example.


*Keywords:* Taxi industry; Competition; Equilibrium; Regulation; First-best solution; Congestion externality

* Corresponding author. Tel.: +852 2358 7178; fax: +852 2358 1534.
*E-mail address:* cehyang@ust.hk (H. Yang).

0965-8564/$ - see front matter 
doi:10.1016/j.tra.2004.05.004

## 1. Introduction

Because of their convenience, efficiency, flexibility and 24-h availability, taxi service becomes an important and indivisible part of the urban public transport in most large cities. As reviewed and summarized recently by Yang et al. (2003), a substantial number of studies are available in the literature concerning the models and economics of taxi services under various types of regulation such as entry restriction and price control. A few notable contributions include the early studies by Douglas (1972), De vany (1975), and recent studies by Cairns and Liston-Heyes (1996), Arnott (1996) and Yang et al. (1998, 2002). These studies have overwhelmingly emphasized the role of customer waiting time and the complex intervening relationship between users (customers) and suppliers (firms) of the taxi service. The waiting time is generally considered as an important value or quality of the services received by customers. This variable affects customers' decision as to whether or not to take a taxi, and thus plays a crucial role in the determination of the price level and the resulting equilibrium of the market.

In most large cities, taxis make considerable demands on limited road resources and contribute significantly to traffic congestion. In the urban area of Hong Kong (Hong Kong Island and Kowloon), the total taxis currently form about 25% of the overall traffic mix. In some critical locations, taxis even account for as much as 50% to 60% of the traffic mix (Hong Kong Transport Department, 1986–2000). Thus, effective regulation of the taxi market needs appropriate consideration of the congestion externalities due to both vacant and occupied taxi movements. Unfortunately, previous economic analyses of taxi services are generally based on a constant average taxi ride time or distance and the effect of traffic congestion is neglected in regulating price and setting optimal service standards. It is true that the real taxi fare includes both a distance-based and delay-based charge in nearly every large city. Even with a constant average taxi ride distance as assumed in previous analytical studies, the delay-based charge varies with the congestion level, and hence with the changes of taxi fleet size and/or normal traffic demand.

Table 1 summarizes the fare structure implemented in a few major cities around the world. Although the fare structure varies slightly across cities, it is a common feature that the price consists of initial flag-fall charge, time delay-based charge and travel distance-based charge. Taxi fare control, together with fleet size restriction, is generally introduced in most cities for efficient regulation of the taxi industry. The empirical evidence shown in the table clearly point to the necessity of incorporating the congestion effect in taxi service analysis, which will help precise understanding of the nature of equilibrium and provide reliable information on the price and fleet size regulation of the market.

This study investigates the monopoly, the social optimum and the stable competitive solutions of cruising taxi services in the presence of congestion externalities by adopting a realistic distance-based and delay-based fare structure. The paper is organized as follows: In Section 2, we introduce the essential elements and the basic analytical model required to characterize taxi services with congestion effects. In Section 3, we examine the first-best solution and show that the price charged at the social optimum with congestion externalities should exceed the marginal cost, and as a result the first-best solution may lie in the positive profit region. In Section 4, we address the competitive solution with emphasis on two special cases: the stable competitive solution and the second-best solution. In Section 5, we show in a monopoly market how a profit-maximizing taxi firm will choose markup pricing when congestion effect is built into the model. A numerical

Table 1
Taxi fare structure in major cities around the world

| | Hong Kong | Singapore | New York | London | Tokyo | Shanghai |
|---|---|---|---|---|---|---|
| Initial charge | HK\$15.0/first 2 km | S\$2.4/first 1 km | \$2.0 for getting in | £2.00/first 361.0 m or 77.6 s | 660$^{\text{type1}}$ 340$^{\text{type2}}$ yen/first 2 km | ¥10/first 3 km |
| Distance-based charge | HK\$1.4/0.2 km thereafter | S\$0.1 for every 240 m thereafter up to 10 km, S\$0.1 for every 225 m thereafter from 10 km onwards | \$0.3/0.25 mile | 20p/each additional 180.5 m or 38.8 s if the fare displayed is less than £12.40, 20p/each 128.9 m or 27.7 s thereafter | 80 yen/274 m$^{\text{type1}}$; 80 yen for the next 250 m and 80 yen/274 m$^{\text{type2}}$ | ¥2/3–10 km; ¥3/thereafter |
| Delay-based charge | HK\$1.4/per min | S\$0.1 every 30 s | \$0.3/30 s | | | ¥0.4/per min from 3 to 10 km, ¥0.6/per min thereafter |
| Comments | HK\$5.0/per baggage | All items charged 50% more from 12:00pm to 6:00 am; CBD and holiday surcharge | Additional night charge \$0.50, tip in the range of 15%∼20% | The price charge applies for Monday to Friday between 6:00 am and 8:00 pm | The fare goes up by 80 yen per 211 m from 11:00 pm to 5:00 am | All items charged 30% more after 11 pm |

example is provided in Section 6 to highlight the major theoretical findings, and finally, general conclusions are given in Section 7.

## 2. The basic model

### 2.1. Basic assumptions

Consider a realistic taxi fare structure consisting of three components: a flag-fall or a constant initial flat charge, a distance-based charge, and a delay-based charge. Let

$P^0$ = initial flag-fall charge per ride (HKD)
$\beta^d$ = fare charge per occupied unit distance (HKD/km)
$\beta^t$ = fare charge per delay hour (HKD/h)
$L$ = average taxi ride length (km) (assumed to be a constant)
$T$ = average taxi ride time (h) (congestion dependent)
$T^0$ = average taxi ride time without congestion (h) (assumed to be a constant)

Then, the total trip fare $P$ (HKD) is given by

$$P = P^0 + \beta^d L + \beta^t (T - T^0) \tag{1}$$

Clearly, we have $T - T^0 \geqslant 0$.

Customer demand, $Q$ (trip/h), for taxis is assumed to be a decreasing function of trip fare, $P$, in-vehicle travel time, $T$, and waiting time, $W$ (h), respectively, and takes the following form:

$$Q = f(P, T, W) \tag{2}$$

Customer waiting time, $W$, a measure of the service quality of the taxi market, is determined by the total vacant taxi-hours available. Let $N$ be the taxi fleet size and $N^v$, $N^o$ and $N^n$ be the numbers of vacant taxis, occupied taxis and normal vehicles (private cars) running in the network at a given instant. Clearly,

$$N^o = QT \tag{3}$$

$$N^v = N - QT \tag{4}$$

Then, the average customer waiting time for taxis is a decreasing function of the available vacant taxi-hours:

$$W = w(N^v) = w(N - QT) \tag{5}$$

where $w' = \mathrm{d}w/\mathrm{d}N^v < 0$.

Naturally, the in-vehicle travel time for a given average length of taxi ride is given as a function of the vehicle density on the road,

$$T = t(N^v, N^o, N^n) \tag{6}$$

Suppose the same average length, $L$, for trips by normal vehicles, then for a given normal traffic demand, $Q^{\mathrm{n}}$ (veh/h), the number of normal vehicles running throughout the whole network at a given instant is given by

$$N^{\mathrm{n}} = Q^{\mathrm{n}} \frac{L}{v^{\mathrm{n}}} \tag{7}$$

where $v^{\mathrm{n}}$ is the travel speed of normal vehicles (km/h). Suppose all vehicles (taxis and normal vehicles) run on the roads at the same speed, then, we have,

$$N^{\mathrm{n}} = Q^{\mathrm{n}} T \tag{8}$$

Without sacrificing the realism of our model we further assume that the normal vehicle demand $Q^{\mathrm{n}}$ is fixed, and its impact on the taxi market equilibrium will be treated through parametric sensitivity analysis. [1] As seen from Eq. (8), the number of normal vehicles, $N^{\mathrm{n}}$, running in the network at a given instant is an endogenous variable for fixed normal traffic demand, $Q^{\mathrm{n}}$, in the presence of the congestion externalities considered here.

From Eqs. (6), (4) and (8), we can see that the in-vehicle travel time is a function of taxi fleet size, customer demand for taxis, normal vehicle demand and in-vehicle travel time itself:

$$T = t(N^{\mathrm{v}}, N^{\mathrm{o}}, N^{\mathrm{n}}) = t(N - QT, QT, N^{\mathrm{n}}(T, Q^{\mathrm{n}})) \tag{9}$$

Note that here $N^{\mathrm{n}}$ is an endogenous variable, whereas $N$ is a control or regulation variable. For a given normal traffic demand $Q^{\mathrm{n}}$, $N^{\mathrm{n}}$ changes with in-vehicle travel time $T$, which in turn is a function of taxi fleet size $N$ and taxi fare $P$, as the latter variables influence customer demand $Q$ and hence occupied taxis $N^{\mathrm{o}}$. Such a complex intertwining relationship makes the choice of the regulatory variables $N$ and $P$ rather less than straightforward, but we have to consider their interplay under congested traffic conditions.

To explore the equilibrium relationship in a convenient manner, we define the partial derivatives of $f$ in function (2) with respect to $P$, $T$ and $W$ as $f_1$, $f_2$ and $f_3$, respectively. Clearly, $f_1 < 0$, $f_2 < 0$ and $f_3 < 0$. In view of $P^0 + \beta^{\mathrm{d}} L$ in Eq. (1) being constant for a given average ride distance $L$, we combine and denote the two terms as the fixed component of taxi fare $P^{\mathrm{f}}$: $P^{\mathrm{f}} = P^0 + \beta^{\mathrm{d}} L$. Hence, the total taxi fare per ride $P$ is a function of fixed fare charge $P^{\mathrm{f}}$, hourly delay charge $\beta^{\mathrm{t}}$ and average taxi ride time $T$. Namely,

$$P = p(P^{\mathrm{f}}, \beta^{\mathrm{t}}, T) \tag{10}$$

where $T$ in turn depends on $P^{\mathrm{f}}$ and $\beta^{\mathrm{t}}$ ($T$ depends on the vacant and occupied taxis and hence the customer demand; customer demand depends on the fare charged or on $P^{\mathrm{f}}$ and $\beta^{\mathrm{t}}$). Similarly, let $p_1$, $p_2$ and $p_3$ denote the partial derivatives of $p$ in $P^{\mathrm{f}}$, $\beta^{\mathrm{t}}$ and $T$ in function (10), respectively. Then we have $p_1 > 0$ (in fact, $p_1 = 1.0$ with fare structure given by Eq. (1)), $p_2 > 0$ and $p_3 > 0$. Finally, let $t_1$, $t_2$ and $t_3$ denote the partial derivatives of $t$ in function (9) in $N^{\mathrm{v}}$, $N^{\mathrm{o}}$ and $N^{\mathrm{n}}$, respectively. We also have $t_1 > 0$, $t_2 > 0$ and $t_3 > 0$.

[1] This assumption will greatly simplify our analysis and facilitate interpretation of our results. One should acknowledge the fact that normal vehicle demand responds to congestion (and thus taxi fleet size and fare). Nevertheless, our analysis focuses on the characteristics of taxi services under congested traffic conditions; congestion effects are examined by varying the level of normal vehicle demand and taxi fleet size.

### 2.2. Comparative static effects of regulatory variables

With the distance and delay-based taxi fare structure, we now have three market regulatory variables of constant fare charge $P^{\mathrm{f}}$, delay-based charge rate $\beta^{\mathrm{t}}$ and taxi fleet size $N$. Their comparative static effects on the market are investigated below.

We first look at the effects of the three regulatory variables on the total taxi fare per ride (in this case, the regulatory variables $(P^{\mathrm{f}}, \beta^{\mathrm{t}}, N)$ are regarded as independent variables; the average taxi ride time $T$ is an endogenous or intermediate variable). Taking the partial derivative of $P$ given in (10) with respect to $P^{\mathrm{f}}$,

$$\frac{\partial P}{\partial P^{\mathrm{f}}} = p_1 + \beta^{\mathrm{t}} \frac{\partial T}{\partial P^{\mathrm{f}}} \tag{11}$$

Similarly we have,

$$\frac{\partial P}{\partial \beta^{\mathrm{t}}} = p_2 + \beta^{\mathrm{t}} \frac{\partial T}{\partial \beta^{\mathrm{t}}} \tag{12}$$

$$\frac{\partial P}{\partial N} = p_3 \frac{\partial T}{\partial N} = \beta^{\mathrm{t}} \frac{\partial T}{\partial N} \tag{13}$$

We now look at the effects of the regulatory variables on the average taxi ride time. Taking the partial derivative of $T$ given in (6) with respect to $P^{\mathrm{f}}$ and utilizing Eqs. (4), (3) and (8), we have

$$\frac{\partial T}{\partial P^{\mathrm{f}}} = t_1 \frac{\partial N^{\mathrm{v}}}{\partial P^{\mathrm{f}}} + t_2 \frac{\partial N^{\mathrm{o}}}{\partial P^{\mathrm{f}}} + t_3 \frac{\partial N^{\mathrm{n}}}{\partial P^{\mathrm{f}}} = t_1 \left( -T \frac{\partial Q}{\partial P^{\mathrm{f}}} - Q \frac{\partial T}{\partial P^{\mathrm{f}}} \right) + t_2 \left( T \frac{\partial Q}{\partial P^{\mathrm{f}}} + Q \frac{\partial T}{\partial P^{\mathrm{f}}} \right) + t_3 Q^{\mathrm{n}} \frac{\partial T}{\partial P^{\mathrm{f}}}$$

This gives rise to

$$\frac{\partial T}{\partial P^{\mathrm{f}}} = \frac{(t_2 - T_1)T}{1 + t_1 Q - t_2 Q - t_3 Q^{\mathrm{n}}} \frac{\partial Q}{\partial P^{\mathrm{f}}} \tag{14}$$

Similarly,

$$\frac{\partial T}{\partial \beta^{\mathrm{t}}} = t_1 \frac{\partial N^{\mathrm{v}}}{\partial \beta^{\mathrm{t}}} + t_2 \frac{\partial N^{\mathrm{o}}}{\partial \beta^{\mathrm{t}}} + t_3 \frac{\partial N^{\mathrm{n}}}{\partial \beta^{\mathrm{t}}} = t_1 \left( -T \frac{\partial Q}{\partial \beta^{\mathrm{t}}} - Q \frac{\partial T}{\partial \beta^{\mathrm{t}}} \right) + t_2 \left( T \frac{\partial Q}{\partial \beta^{\mathrm{t}}} + Q \frac{\partial T}{\partial \beta^{\mathrm{t}}} \right) + t_3 Q^{\mathrm{n}} \frac{\partial T}{\partial \beta^{\mathrm{t}}}$$

or

$$\frac{\partial T}{\partial \beta^{\mathrm{t}}} = \frac{(t_2 - t_1)T}{1 + t_1 Q - t_2 Q - t_3 Q^{\mathrm{n}}} \frac{\partial Q}{\partial \beta^{\mathrm{t}}} \tag{15}$$

and

$$\frac{\partial T}{\partial N} = t_1 \frac{\partial N^{\mathrm{v}}}{\partial N} + t_2 \frac{\partial N^{\mathrm{o}}}{\partial N} + t_3 \frac{\partial N^{\mathrm{n}}}{\partial N} = t_1 \left( 1 - T \frac{\partial Q}{\partial N} - Q \frac{\partial T}{\partial N} \right) + t_2 \left( T \frac{\partial Q}{\partial N} + Q \frac{\partial T}{\partial N} \right) + t_3 Q^{\mathrm{n}} \frac{\partial T}{\partial N}$$

or

$$\frac{\partial T}{\partial N} = \frac{t_1 + (t_2 - t_1) T \partial Q / \partial N}{1 + t_1 Q - t_2 Q - t_3 Q^{\mathrm{n}}} \tag{16}$$

We next look at the effects of the regulatory variables on customer demands. Taking the partial derivative of $Q$ in (2) with respect to $P^{\mathrm{f}}$ and utilizing Eqs. (10) and (5),

$$\frac{\partial Q}{\partial P^{\mathrm{f}}} = f_1 \frac{\partial P}{\partial P^{\mathrm{f}}} + f_2 \frac{\partial T}{\partial P^{\mathrm{f}}} + f_3 \frac{\partial W}{\partial P^{\mathrm{f}}} = f_1 \left( p_1 + p_3 \frac{\partial T}{\partial P^{\mathrm{f}}} \right) + f_2 \frac{\partial T}{\partial P^{\mathrm{f}}} + f_3 w' \left( -T \frac{\partial Q}{\partial P^{\mathrm{f}}} - Q \frac{\partial T}{\partial P^{\mathrm{f}}} \right) \tag{17}$$

This leads to

$$\frac{\partial Q}{\partial P^{\mathrm{f}}} = \frac{f_1 p_1 + (f_1 p_3 + f_2 - f_3 w' Q) \partial T / \partial P^{\mathrm{f}}}{1 + f_3 w' T} \tag{18}$$

Substitute Eq. (14) into Eq. (18) gives,

$$\frac{\partial Q}{\partial P^{\mathrm{f}}} = \frac{f_1 p_1 (1 + t_1 Q - t_2 Q - t_3 Q^{\mathrm{n}})}{(1 + f_3 w' T)(1 - t_3 Q^{\mathrm{n}}) - (t_2 - t_1)((f_1 p_3 + f_2) T + Q)} \tag{19}$$

Similarly, we obtain

$$\frac{\partial Q}{\partial \beta^{\mathrm{t}}} = \frac{f_1 p_2 (1 + t_1 Q - t_2 Q - t_3 Q^{\mathrm{n}})}{(1 + f_3 w' T)(1 - t_3 Q^{\mathrm{n}}) - (t_2 - t_1)(T(f_1 p_3 + f_2) + Q)} \tag{20}$$

$$\frac{\partial Q}{\partial N} = \frac{f_3 w' (1 - t_2 Q - t_3 Q^{\mathrm{n}}) + (f_2 + f_1 p_3) t_1}{(1 + f_3 w' T)(1 - t_3 Q^{\mathrm{n}}) - (t_2 - t_1)((f_1 p_3 + f_2) T + Q)} \tag{21}$$

Finally, we consider how the regulatory variables $(P^{\mathrm{f}}, \beta^{\mathrm{t}}, N)$ affect the customer waiting time, which is an important service quality measure. Taking the partial derivatives of $W$ in (5) with respect to $P^{\mathrm{f}}$, $\beta^{\mathrm{t}}$ and $N$ respectively, we get,

$$\frac{\partial W}{\partial P^{\mathrm{f}}} = -w' \left( T \frac{\partial Q}{\partial P^{\mathrm{f}}} + Q \frac{\partial T}{\partial P^{\mathrm{f}}} \right) \tag{22}$$

$$\frac{\partial W}{\partial \beta^{\mathrm{t}}} = -w' \left( T \frac{\partial Q}{\partial \beta^{\mathrm{t}}} + Q \frac{\partial T}{\partial \beta^{\mathrm{t}}} \right) \tag{23}$$

$$\frac{\partial W}{\partial N} = w' \left( 1 - T \frac{\partial Q}{\partial N} - Q \frac{\partial T}{\partial N} \right) \tag{24}$$

where $\partial Q / \partial P^{\mathrm{f}}$, $\partial Q / \partial \beta^{\mathrm{t}}$ and $\partial Q / \partial N$ are given by Eqs. (19)–(21) and $\partial T / \partial P^{\mathrm{f}}$, $\partial T / \partial \beta^{\mathrm{t}}$ and $\partial T / \partial N$ are given by Eqs. (14)–(16).

We note that all the above expressions obtained so far can be greatly simplified if we assume that both vacant and occupied taxis have about the same marginal impact on traffic flow (or running speed) in the considered cruising taxi market, namely $\partial t / \partial N^{\mathrm{v}} = \partial t / \partial N^{\mathrm{o}}$ in Eq. (6) or $t_1 = t_2$. With this assumption, from Eqs. (14) and (15) we have the following simplified qualitative relationships:

$$\frac{\partial T}{\partial P^{\mathrm{f}}} = \frac{\partial T}{\partial \beta^{\mathrm{t}}} = 0 \tag{25}$$

and from Eq. (16)

$$\frac{\partial T}{\partial N} = \frac{t_1}{1 - t_3 Q^{\mathrm{n}}} > 0 \tag{26}$$

where $\partial T/\partial N > 0$ is always true for the traffic flow operating in the normal flow regime in a linear traffic flow model. [2] Also, Eqs. (11)–(13) turn out to be,

$$\frac{\partial P}{\partial P^{\mathrm{f}}} = p_1 > 0 \tag{27}$$

$$\frac{\partial P}{\partial \beta^{\mathrm{t}}} = p_2 > 0 \tag{28}$$

$$\frac{\partial P}{\partial N} = \frac{t_1 p_3}{1 - t_3 Q^{\mathrm{n}}} > 0 \text{ (from Eq. (26))} \tag{29}$$

Moreover, in view of $f_1 < 0$, $f_3 < 0$, $p_1 > 0$, $p_2 > 0$, $w' < 0$ and $T > 0$, we have

$$\frac{\partial Q}{\partial P^{\mathrm{f}}} = \frac{f_1 p_1}{1 + f_3 w' T} < 0 \tag{30}$$

$$\frac{\partial Q}{\partial \beta^{\mathrm{t}}} = \frac{f_1 p_2}{1 + f_3 w' T} < 0 \tag{31}$$

$$\frac{\partial Q}{\partial N} = \frac{f_3 w'(1 - t_2 Q - t_3 Q) + (f_2 + f_1 p_3) t_1}{(1 + f_3 w' T)(1 - t_3 Q^{\mathrm{n}})} = \frac{f_3 w'}{1 + f_3 w' T} + \frac{(f_2 + f_1 p_3 - f_3 w' Q) t_1}{(1 + f_3 w' T)(1 - t_3 Q^{\mathrm{n}})} \tag{32}$$

Eq. (32) describes the aggregate impact of taxi fleet size on customer demand in two opposite manners. The first term of the right-hand side shows that, everything else being equal, increase in taxi fleet size will reduce customer waiting time ($w' < 0$) and thus increase customer demand ($f_3 w' > 0$). The first term is always positive (note that $1 + f_3 w' T > 0$). The second term of the right-hand side represents the negative impact of congestion on customer demand arising from the entry of one additional taxi into the market. Whenever a new taxi enters the market, the average taxi ride time will increase. This will lead to direct decrease in customer demand ($f_2 t_1 < 0$), decrease in customer demand due to increased delay-based charge ($f_1 p_3 t_1 < 0$) and also due to increased cus-

[2] We consider a linear traffic flow model: $v = a - bk$, where speed $v$ is a linear function of traffic density $k$ with parameter $a > 0$ and $b > 0$. Suppose traffic is flowing in the normal flow regime, which means that $0 < a/2 < v \leqslant a$ with $a$ being the maximal free-flow speed (May, 1990). Substitute $k = (N + N^{\mathrm{n}})/\overline{L}$ into the linear flow model where $\overline{L}$ is the total road length in the network and since $T = L/v$, we obtain $t_3 = \partial T/\partial N^{\mathrm{n}} = bL/\overline{L}v^2$. Using $L/v = N^{\mathrm{n}}/Q^{\mathrm{n}}$ from Eq. (7), we have:

$$t_3 = \frac{1}{Q^{\mathrm{n}}} \frac{bN^{\mathrm{n}}}{v\overline{L}} < \frac{1}{Q^{\mathrm{n}}} \frac{b(N^{\mathrm{n}} + N)}{v\overline{L}} = \frac{1}{Q^{\mathrm{n}}} \frac{(a - v)}{v}$$

From $0 < a/2 < v$ we have $0 < (a - v)/v < 1$ and thus $t_3 < 1/Q^{\mathrm{n}}$ in the above equation. As a result, $t_3 Q^{\mathrm{n}} < 1$ or Eq. (26) is always positive as long as the traffic flow is operating in the normal flow regime.

tomer waiting time ($-f_3 w' Q t_1 < 0$) as a result of more occupied and less vacant taxi-hours. One can thus readily understand that $\partial Q/\partial N$ could be positive or negative, depending the relative magnitudes of the two impacts. The latter negative impact on demand, however, depends on the current level of congestion. In fact, suppose the level of traffic congestion depends only the total number of vehicles on the roads regardless of occupied, vacant and normal vehicles, namely, $t_1 = t_2 = t_3$ in Eq. (6). Then from Eq. (32), we have $\partial Q/\partial N > 0$ if $t_1 < 1/(Q + Q^{\mathrm{n}} - ((f_2 + f_1 p_3)/f_3 w'))$. This means that, if the current market is less (or moderately) congested, entry of additional taxis will improve service quality and thus increase customer demand. If, however, the market is already very highly congested due to a high level of normal vehicle demand or from an already large taxi fleet size, the improvement in service quality cannot offset the negative impacts due to extra congestion after entry of one additional taxi into the market, and as a result, new taxi entry will only diminish customer demand. Of course, the latter situation rarely occurs in reality; it is typical in large urban commuting areas that an increase in taxis leads to a greater demand for taxis.

Similarly, substituting Eqs. (30) and (25) into (22), Eqs. (31) and (25) into Eq. (23), and Eq. (32) and Eq. (26) into (24), respectively, we obtain

$$\frac{\partial W}{\partial P^{\mathrm{f}}} = -w' T \frac{f_1 p_1}{1 + f_3 w' T} < 0 \tag{33}$$

$$\frac{\partial W}{\partial \beta^{\mathrm{t}}} = -w' T \frac{f_1 p_2}{1 + f_3 w' T} < 0 \tag{34}$$

$$\frac{\partial W}{\partial N} = w' \left( 1 - T \frac{f_3 w' (1 - t_3 Q^{\mathrm{n}}) + (f_1 p_3 + f_2 - f_3 w' Q) t_1}{(1 + f_3 w' T)(1 - t_3 Q^{\mathrm{n}})} - Q \frac{t_1}{1 - t_3 Q^{\mathrm{n}}} \right) \tag{35}$$

Like Eq. (32), $\partial W/\partial N$ in Eq. (35) could be positive or negative, depending on the current degree of congestion. Similarly, we can find that $\partial W/\partial N < 0$ if $t_1 < 1/(Q + Q^{\mathrm{n}} + T(f_2 + f_1 p_3))$. In particular, if there is no congestion effect or $t_1 = t_2 = t_3 = 0$ in Eq. (6), then we have $\partial W/\partial N = w'/(1 + f_3 w' T^0) < 0$ in Eq. (35).

### 2.3. Effects of normal traffic demand

To gain insight into the effect of normal vehicle demand on customer demand and taxi service quality, we have the following relationships. From Eq. (6),

$$\frac{\partial T}{\partial Q^{\mathrm{n}}} = \frac{\partial t}{\partial N^{\mathrm{n}}} \frac{\partial N^{\mathrm{n}}}{\partial Q^{\mathrm{n}}} = t_3 T > 0 \tag{36}$$

From Eq. (2) and then using Eq. (10), we obtain

$$\frac{\partial Q}{\partial Q^{\mathrm{n}}} = f_1 \frac{\partial P}{\partial Q^{\mathrm{n}}} + f_2 \frac{\partial T}{\partial Q^{\mathrm{n}}} + f_3 \frac{\partial W}{\partial Q^{\mathrm{n}}} = f_1 p_3 \frac{\partial T}{\partial Q^{\mathrm{n}}} + f_2 \frac{\partial T}{\partial Q^{\mathrm{n}}} + f_3 \frac{\partial W}{\partial Q^{\mathrm{n}}} \tag{37}$$

From Eq. (5) we have

$$\frac{\partial W}{\partial Q^{\mathrm{n}}} = w' \left( -T \frac{\partial Q}{\partial Q^{\mathrm{n}}} - Q \frac{\partial T}{\partial Q^{\mathrm{n}}} \right) \tag{38}$$

Substituting Eq. (38) into Eq. (37) and solving for $\partial Q/\partial Q^{\mathrm{n}}$ give rise to:

$$\frac{\partial Q}{\partial Q^{\mathrm{n}}} = \frac{(f_1 p_3 + f_2 - f_3 w' Q)}{(1 + f_3 w' T)} \frac{\partial T}{\partial Q^{\mathrm{n}}} < 0 \tag{39}$$

where $\partial Q/\partial Q^{\mathrm{n}} < 0$ comes from the fact that $1 + f_3 w' T > 0$ and $f_1 p_3 + f_2 - f_3 w' Q < 0$ from $f_1 < 0$, $f_2 < 0, f_3 < 0$, $p_3 > 0$ and $w' < 0$ as well as $\partial T/\partial Q^{\mathrm{n}} > 0$ from Eq. (36). This shows that the congestion externality certainly deteriorates the service quality and drops the customer demand for taxi. As a result, even possibly gaining additional revenue from the delay-based fare charge, taxi firms also face the threat of losing customers in the presence of congestion effects, which would drop their revenue in return. Therefore, the presence of congestion externality in the taxi market makes an intriguing issue of regulating price and setting service standard. Both the taxi firms and the regulator should figure out their preferred solutions respectively by balancing the hourly delay charge and taxi fleet size by properly taking into account the congestion level associated with the various normal traffic demand and taxi fleet size.

### *2.4. Synchronous relationship among variables*

Based on the qualitative analysis developed in the previous section, Fig. 1 illustrates the complicated interrelationships among the exogenous and endogenous variables. A few major observations from the figure and the previous analysis are worthwhile to mention.

Firstly, recalling the assumption mentioned above, the increase in taxi fleet size has positive impacts on in-vehicle travel time and total fare per taxi ride charged for a given fare charge structure. The figure also indicates that the increase in taxi fleet size enhances the taxi availability at the same time. Thus the change in customer demand for taxis is a joint consequence of the positive impact of the increased vacant taxi-hour and the negative impact of the increased in-vehicle travel time and increased fare as well. As described in Eq. (32), the customer demand could increase or decrease as the taxi fleet size increases, depending on the net effect of the two opposite forces (or the current level of congestion and taxi availability).

Secondly, it is obvious, if we assume a vacant and an occupied taxi inflicts the same marginal impact on traffic flow, that raising the fixed taxi fare component and/or the delay-based taxi charge rate will only reduce the customer demand as shown in Eqs. (30) and (31). Increase in normal vehicle demand will make congestion more severe and thus definitely result in a decline in customer demand as well, as described in Eq. (39).

Thirdly, keeping all exogenous variables fixed, it is interesting to note that there is an internal cycle among the customer demand, vacant taxi-hour and customer waiting time. It is this internal cycle of interaction or internal market force which drives the market to a stable equilibrium for fixed exogenous variables. For example, suppose there is a small amount of increase in customer demand from the current equilibrium point, then the vacant taxi time will decrease due to increased customer demand, and as a result average customer waiting time will increase. Increase in waiting time will in turn lead to a decrease in customer demand. This means that the market force will pull down the demand to the original level and stabilize the equilibrium.

The above observations show how the regulatory variables of taxi fleet size and fare structure affect the endogenous variables or demand–supply equilibrium in the taxi market, including the customers' in-vehicle travel time, total taxi fare per ride, service quality, demand for taxis and thus

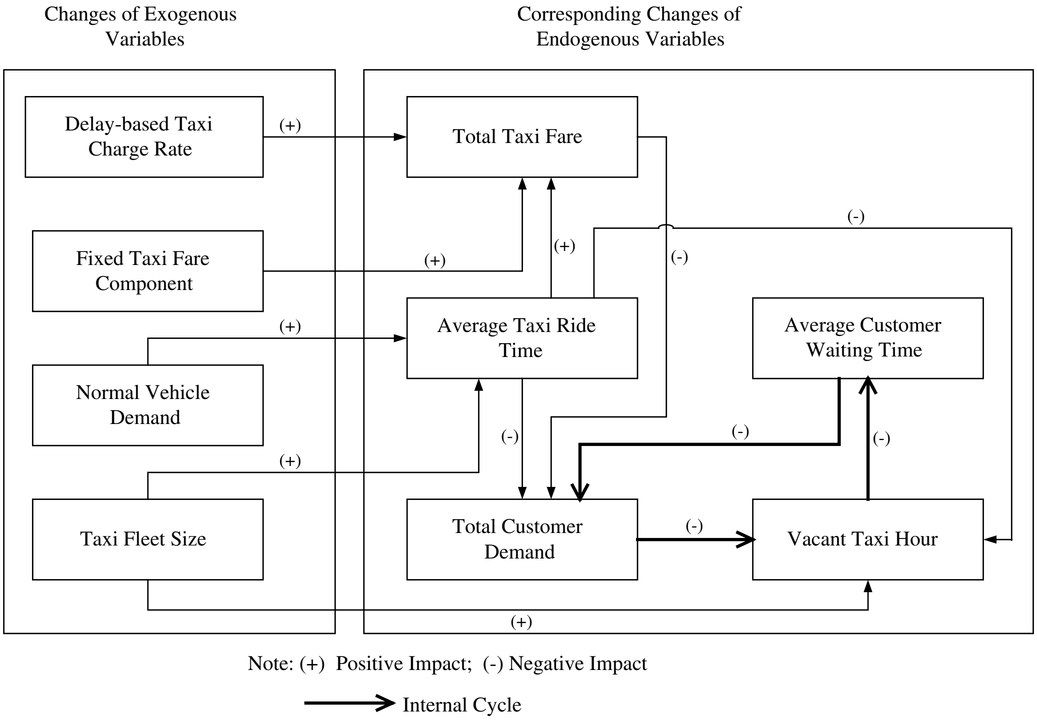


Fig. 1. The synchronous relationships among the exogenous and endogenous variables in the taxi market with congestion effects.

the taxi firms' profit and social welfare. The regulator, therefore, may achieve various objectives by effectively regulating taxi fare structure and fleet size with proper consideration of congestion externality.

Before concluding this section, we note that at the equilibrium points, with or without regulations, the combination of the initial flag-fall charge, the distance-based charge and the hourly delay-charge is indeterminate in giving the same total fare charge per taxi ride. Hence it is reasonable to first fix the sum of the initial flag-fall charge and distance-based charge for the various representative solutions in the subsequent sections, when we ignore the congestion externalities or when there is no congestion effect. The hourly delay-charge rate is then determined when the congestion effect is built into the model, in addition to the predetermined congestion-free initial and distance-based charges. This two-stage fare determination is carried out in the latter sections.

## 3. Social optimum

Now we consider the choice of taxi fleet size and fare structure to maximize the social welfare. Assume that the full price for a taxi trip is given as $\rho = P + \tau T + \kappa W$, where $\tau$ and $\kappa$ are the values of customers' in-vehicle travel time and waiting time, respectively. The demand function is assumed to be a decreasing function of the full price given as $Q = f(\rho)$ where $f' < 0$. As mentioned previously, $P^{\mathrm{f}}$ is predetermined as the corresponding price charged in the absence of congestion effects.

Assuming a constant hourly operation cost $c$ (HKD per taxi-hour) for both vacant and occupied taxis and defining the social welfare as the sum of the consumer surplus and the producer surplus (profit of taxi firms) minus the additional congestion delay cost of normal traffic, then the social welfare maximization problem is given below:

$$\max \ S(\beta^{\mathrm{t}}, N) = \int_{\rho}^{\infty} f(\omega)\,\mathrm{d}\omega + PQ - cN - \tau Q^{\mathrm{n}}(T - T^{0}) \tag{40}$$

where, as assumed before, the normal traffic demand, $Q^{\mathrm{n}}$, is treated with the same trip length as those made by taxis, for simplicity. The same value of in-vehicle travel time is assumed for both taxi and normal vehicle users.

Consider maximizing $S$ with respect to $\beta^{\mathrm{t}}$ and $N$. From $\partial S/\partial\beta^{\mathrm{t}} = 0$, we have

$$\tau(Q + Q^{\mathrm{n}})\frac{\partial T}{\partial\beta^{\mathrm{t}}} + \kappa Q\frac{\partial W}{\partial\beta^{\mathrm{t}}} = P\frac{\partial Q}{\partial\beta^{\mathrm{t}}}$$

From $\partial S/\partial N = 0$, we obtain

$$c + \tau(Q + Q^{\mathrm{n}})\frac{\partial T}{\partial N} + \kappa Q\frac{\partial W}{\partial N} = P\frac{\partial Q}{\partial N}$$

From Eqs. (23) and (24), we obtain:

$$(\tau(Q + Q^{\mathrm{n}}) - \kappa w' Q^{2})\frac{\partial T}{\partial\beta^{\mathrm{t}}} = (P + \kappa w' QT)\frac{\partial Q}{\partial\beta^{\mathrm{t}}} \tag{41}$$

$$(P + \kappa w' QT)\frac{\partial Q}{\partial N} = c + \kappa w' Q + (\tau(Q + Q^{\mathrm{n}}) - \kappa w' Q^{2})\frac{\partial T}{\partial N} \tag{42}$$

From Eqs. (25), (31) and (41), we have

$$P = -\kappa w' QT \tag{43}$$

This in turn means that the left-hand side of Eq. (42) is zero. In view of $\partial T/\partial N > 0$ from Eq. (26), Eq. (42) leads to:

$$c = (\kappa w' Q^{2} - \tau(Q + Q^{\mathrm{n}}))\frac{\partial T}{\partial N} - \kappa w' Q \tag{44}$$

Therefore,

$$P - cT = (\tau(Q + Q^{\mathrm{n}}) - \kappa w' Q^{2})T\frac{\partial T}{\partial N} \tag{45}$$

Eq. (45) is the central result of the social optimum regulation of the taxi market with congestion externalities. The equation implies that if $\partial T/\partial N = 0$, namely, if there is no congestion effect at the social optimum, we in effect have $P = cT$. In this case $PQ = cTQ$, implying that the total revenue just equals (or just covers) only the total cost of occupied taxi-hours and, in the aggregate, the taxis operate at a loss equal to the cost of vacant taxi-hours. We thus arrive at the same results examined by Arnott (1996), who concluded that the first-best solution is located in the negative profit region and taxi services at the social optimum should be subsidized. If the congestion effect cannot be ignored, ($\partial T/\partial N > 0$ from Eq. (26)), then $P > cT$ comes about as a result of $T(\tau(Q + Q^{\mathrm{n}}) - Q^2\kappa w')\partial T/\partial N > 0$ (note that $w' < 0$). This means that the taxi firms' loss is relaxed; the aggregate loss is less than the cost of vacant taxi-hours due to $P > cT > cT^0 = P^{\mathrm{f}}$, where $P^{\mathrm{f}}$ is the corresponding optimal price charged in the absence of congestion effects at the first-best solution. As a result, we conclude that the price charged is higher at the first-best solution [3] in the presence of congestion externality.

Naturally, one question arises from the above observation. Can the taxi firms' profit be enhanced sufficiently so that the first-best solution is located in the non-negative profit region? To answer this question, we now look at the total profit of the taxi firms at the social optimum. From Eqs. (43) and (44) we have

$$PQ - cN = \kappa w' Q(N - QT) + N(\tau(Q + Q^{\mathrm{n}}) - \kappa w' Q^2)\frac{\partial T}{\partial N} \tag{46}$$

In view of $w' < 0$ and $N - QT$ being the total vacant taxi-hour, we must have $N(\tau(Q + Q^{\mathrm{n}}) - \kappa w' Q^2) > 0$ and $\kappa w' Q(N - QT) < 0$. Therefore, a sufficient condition for obtaining positive profit at the social optimum is

$$\frac{\partial T}{\partial N} \geqslant -\frac{\kappa w' Q(N - QT)}{N(\tau(Q + Q^{\mathrm{n}}) - \kappa w' Q^2)} (> 0) \tag{47}$$

or alternatively, with the assumption of $t_1 = t_2 = t_3$ in the congestion characterization equation (6), the following condition is met:

$$t_1 \geqslant \left(Q^{\mathrm{n}} + \frac{NQ}{N - QT} + \frac{\tau N(Q + Q^{\mathrm{n}})}{kw' Q(N - QT)}\right)^{-1} \tag{48}$$

Eq. (47) or (48) implies that a profitable first-best social optimum emerges in a severely congested taxi market, where the entry of additional taxis into the market makes a large marginal congestion effect (and thus entry should be highly controlled at the social optimum).

## 4. Monopoly solution

In a monopoly market, a single taxi firm operates all taxis to maximize its profit. The profit per unit time is given by

$$\Pi = PQ - cN = \left(P^{\mathrm{f}} + \beta^{\mathrm{t}}\left(T - T^0\right)\right)Q - c(N^{\mathrm{v}} + N^{\mathrm{o}}) \tag{49}$$

Maximizing profit $\Pi$ with respect to hourly delay charge rate $\beta^{\mathrm{t}}$ and taxi fleet size $N$ yields the following first-order conditions:

$$\frac{\partial \Pi}{\partial \beta^{\mathrm{t}}} = Q\frac{\partial P}{\partial \beta^{\mathrm{t}}} + P\frac{\partial Q}{\partial \beta^{\mathrm{t}}} = 0 \tag{50}$$

[3] Strictly speaking, here it is more appropriate to use the term "quasi-first-best solution" due to the *unpriced* external congestion effect in the taxi market.

$$\frac{\partial \Pi}{\partial N} = Q\frac{\partial P}{\partial N} + P\frac{\partial Q}{\partial N} - c = 0 \tag{51}$$

Combining Eqs. (28), (31) and (50) gives,

$$P = -\frac{(1 + f_3 w' T)Q}{f_1} \tag{52}$$

Substituting Eqs. (29), (32) into (51) leads to

$$c = -\frac{f_3 w' Q}{f_1} - \frac{t_1(f_2 - f_3 w' Q)Q}{f_1(1 - t_3 Q^{\mathrm{n}})} \tag{53}$$

In view of Eq. (26), $f_k < 0$, $t_k > 0$ $(k = 1, 2, 3)$ and $w' < 0$, we obtain

$$P - cT = -\frac{Q}{f_1} + \frac{t_1(f_2 - f_3 w' Q)TQ}{f_1(1 - t_3 Q^{\mathrm{n}})} = -\frac{Q}{f_1} + \frac{(f_2 - f_3 w' Q)TQ}{f_1}\frac{\partial T}{\partial N} > -\frac{Q}{f_1} > 0 \tag{54}$$

Eq. (54) is the monopoly price markup formula in the presence of congestion externalities. It is interesting to note that when $\partial T/\partial N$ equals zero or there is no congestion effect, Eq. (54) reduces to

$$P - cT = -Q/f_1 \tag{55}$$

This is exactly the same result obtained by Cairns and Liston-Heyes (1996) when the congestion effect is ignored. The price markup formula (55) shows the monopoly power of the single taxi firm to set the taxi fare above its marginal cost in equilibrium. Moreover, from Eq. (54) the taxi fare will exceed marginal cost per trip by a greater amount when there is congestion effect in the monopoly market.

We now further examine the monopoly markup price. In view of the full price $\rho = P + \tau T + \kappa W$, we essentially have $f_1 = \partial Q/\partial \rho$. With $Q = -\frac{\partial}{\partial \rho}\int_\rho^\infty f(\omega)\,\mathrm{d}\omega$, we obtain

$$-\frac{Q}{f_1} = \left(\frac{\partial}{\partial \rho}\int_\rho^\infty f(\omega)\,\mathrm{d}\omega\right)\left(\frac{\partial Q}{\partial \rho}\right)^{-1} = \frac{\partial}{\partial Q}\int_\rho^\infty f(\omega)\,\mathrm{d}\omega \tag{56}$$

where $\frac{\partial}{\partial Q}\int_\rho^\infty f(\omega)\,\mathrm{d}\omega$ is the marginal consumer surplus (net willingness-to-pay) for a taxi ride. Thus, Eq. (54) can be rewritten as

$$P - cT = \left(1 - (f_2 - f_3 w' Q)T\frac{\partial T}{\partial N}\right)\frac{\partial}{\partial Q}\int_\rho^\infty f(\omega)\,\mathrm{d}\omega > \frac{\partial}{\partial Q}\int_\rho^\infty f(\omega)\,\mathrm{d}\omega \tag{57}$$

In the absence of congestion effect, Eq. (57) simplifies to:

$$P - cT = \frac{\partial}{\partial Q}\int_\rho^\infty f(\omega)\,\mathrm{d}\omega \tag{58}$$

Eq. (58) implies that, in an effort to extract as much profit as possible, the monopolist would charge a price in excess of marginal cost per ride by an amount equal to the consumer's marginal net willingness-to-pay for a ride. Eq. (57) shows that this markup will be further increased in the presence of congestion effect. This observation is consistent with the results of the private road pricing examined in Lindsey and Verhoef (2001). Therefore, we can conclude that previous studies ignored congestion effects and as a result underestimated the price charged at the monopoly solution.

## 5. Competitive solution

Assuming that the market is comprised of owner-operated taxis, with one taxi per owner or firm, then in this competitive free entry market, the resultant supply will satisfy the market equilibrium, where the marginal revenue obtained by the last unit of taxi service just covers its cost (profit is nil). It is at this point that the individual incentive to join the taxi industry disappears. Hence equilibrium occurs at

$$PQ - cN = 0 \tag{59}$$

Note that the solutions to the above non-linear equation form a closed zero-profit curve (roughly smoothed right-angled triangular shape to be demonstrated later in a numerical example) in the two dimensional space of taxi fare and fleet size. The equilibrium, at smaller fleet size, is unstable and inferior from a welfare perspective, and thus the lower fleet size equilibrium can seldom emerge in reality. Two special cases of the competitive solution along the zero-profit curve deserve our attention: the stable competitive solution in a fully unregulated taxi market and the second-best social optimum solution when the first-best solution is located in the negative profit region.

First, as pointed out in Yang et al. (2003), under a free entry taxi market, the taxi fleet size will self-adjust to an equilibrium level for a given taxi fare, and it is sufficient to regulate taxi fare alone to obtain a specific competitive solution. When the taxi market is fully uncontrolled or unregulated, there exists a stable equilibrium point at which taxi fleet size reaches a maximum, and the incentive for all individual firms to change fare and/or for the number of taxis in service to change disappears. The total market revenue also reaches a maximum as a result of $PQ - cN = 0$ and $N = N_{\max}$. The necessary condition for a stable maximum fleet size and thus maximum revenue requires $Q + P\partial Q/\partial P = 0$ or the price elasticity of customer demand $e_P = (\partial Q/\partial P)P/Q = -1.0$. This simply means that the maximum competitive taxi fleet size and hence the maximum total market revenue occur at the unit elasticity of the customer demand at which the increase in the revenue with a higher ride fare is cancelled out with the loss due to the reduced customer demand.

As observed in the previous section, the "quasi-first-best solution" can be located in the positive or negative profit domain depending on the congestion level in the market. The second-best solution in a competitive market is of particular interest from a sustainable market regulatory perspective when the first-best solution does entail an aggregate loss. Namely, the first-best solution lies in the negative profit region when the inequality (47) is reversed to:

$$\frac{\partial T}{\partial N} < -\frac{\kappa w' Q(N - QT)}{N(\tau(Q + Q^{\mathrm{n}}) - \kappa w' Q^2)} \tag{60}$$

In this case the second-best solution can be obtained from the following welfare maximization problem with a zero-profit constraint:

$$\max\ S(\beta^{\mathrm{t}}, N) = \int_{\rho}^{\infty} f(\omega)\,\mathrm{d}\omega + PQ - cN - \tau Q^{\mathrm{n}}(T - T^0) \tag{61}$$

subject to the zero-profit constraint (59), where, as mentioned before, $P = P^{\mathrm{f}} + \beta^{\mathrm{d}}L + \beta^{\mathrm{t}}(T - T^0)$ with $P^{\mathrm{f}}$ being the zero-profit fare associated with $N$, in the absence of congestion effects.

To obtain the efficient regulated taxi fare, we use the Lagrange multiplier $\lambda$ to incorporate the zero-profit constraint (59) into the objective function (61) to form the following Lagrangian:

$$\max\ S(\beta^{\mathrm{t}}, N) = \int_{P+\tau T+\kappa W}^{\infty} f(\omega)\,\mathrm{d}\omega - \tau Q^{\mathrm{n}}(T - T^0) + (\lambda + 1)(PQ - cN) \tag{62}$$

Then,

$$\frac{\partial S}{\partial \beta^{\mathrm{t}}} = \tau(Q + Q^{\mathrm{n}})\frac{\partial T}{\partial \beta^{\mathrm{t}}} + \kappa Q \frac{\partial W}{\partial \beta^{\mathrm{t}}} - \lambda Q \frac{\partial P}{\partial \beta^{\mathrm{t}}} - (\lambda + 1) P \frac{\partial Q}{\partial \beta^{\mathrm{t}}} = 0 \tag{63}$$

$$\frac{\partial S}{\partial N} = \tau(Q + Q^{\mathrm{n}})\frac{\partial T}{\partial N} + \kappa Q \frac{\partial W}{\partial N} - \lambda Q \frac{\partial P}{\partial N} - (\lambda + 1)\left(P \frac{\partial Q}{\partial N} - c\right) = 0 \tag{64}$$

Substituting Eqs. (25), (34) and (28) into Eq. (63) gives rise to:

$$P = \frac{-\lambda Q(1 + f_3 w' T) - f_1 \kappa w' Q T}{(\lambda + 1) f_1} \tag{65}$$

Similarly, from Eq. (64) and Eqs. (24), (26) and (32), we obtain

$$c = -\frac{\lambda f_3 w' Q + \kappa w' f_1 Q}{(\lambda + 1) f_1} - \frac{t_1(\lambda(f_2 - f_3 w' Q)Q + f_1(\tau(Q + Q^{\mathrm{n}}) - \kappa w' Q^2))}{(\lambda + 1) f_1 (1 - t_3 Q^{\mathrm{n}})} \tag{66}$$

Therefore,

$$P - cT = -\frac{\lambda Q}{(\lambda + 1) f_1} + \frac{\lambda(f_2 - f_3 w' Q) T Q}{(\lambda + 1) f_1}\frac{\partial T}{\partial N} + \frac{(\tau(Q + Q^{\mathrm{n}}) - \kappa w' Q^2) T}{(\lambda + 1)}\frac{\partial T}{\partial N} \tag{67}$$

From Eq. (56), Eq. (67) then becomes

$$P - cT = \frac{\lambda}{\lambda + 1}\left(1 - (f_2 - f_3 w' Q) T \frac{\partial T}{\partial N}\right)\frac{\partial}{\partial Q}\int_{\rho}^{\infty} f(\omega)\,\mathrm{d}\omega + \frac{1}{\lambda + 1}(\tau(Q + Q^{\mathrm{n}}) - \kappa w' Q^2) T \frac{\partial T}{\partial N} \tag{68}$$

where the Lagrange multiplier $\lambda > 0$ when the first-best solution is associated with a negative profit or when inequality (47) is met.

By comparison of Eq. (68) with Eqs. (45) and (57), it is interesting to observe that, in the presence of congestion externality with at least a level of marginal congestion effect given by inequality (47), the markup price above the marginal cost at the second-best solution is in effect a weighted average of the first-best markup price (the second term of the right-hand side of Eq. (68)) and the monopoly markup price (the first term of the right-hand side of Eq. (68)), with a weighting factors of $1/(\lambda + 1)$ and $\lambda(\lambda + 1)$ ascertained by the Lagrange multiplier. This observation is consistent with the Ramsey pricing result in the presence of external costs (Oum and Tretheway, 1988; Arnott and Kraus, 1993).

When there is no congestion effect, Eq. (68) reduces to:

$$P - cT = \frac{\lambda}{\lambda + 1}\frac{\partial}{\partial Q}\int_{\rho}^{\infty} f(\omega)\,\mathrm{d}\omega \tag{69}$$

By comparing Eqs. (69) and (58), we can further observe that the second-best markup price is a portion of the marginal consumer surplus in contrast with that the monopoly markup price is exactly equal to the marginal consumer surplus in the absence of congestion externality. This discrepancy reflects the use of different decision rules by the market regulator and the monopolist: the objective of the regulator is to balance the benefits to both the consumers and the taxi firms, whereas the monopolist is to simply maximize the total profit by fully exploiting the consumers' marginal willingness to pay for taxi rides.

Finally, two comments on the competitive solution are deduced. First, the stable competitive solution with a maximum taxi fleet size is associated with a marginal cost of taxi-hour in excess of marginal social value, resulting in wasteful competition among firms, due to the competitive fare above the efficient level. In contrast, the second-best solution leads to a more efficient utilization (higher time occupancy rate) of taxis, with a higher demand served by a smaller taxi fleet size and a lower fare. Second, as observed later, the presence of congestion effects will result in a shrunken competitive taxi market with a smaller taxi fleet size and less customer demand. The positive profit domain demarcated by the zero-profit curve becomes smaller.

## 6. A numerical example

To elucidate the discussion and results obtained so far, we now present a numerical example with the following specific negative exponential demand function used in Yang et al. (2002):

$$\begin{aligned} Q &= \overline{Q}\exp(-\alpha(P + \tau T + \kappa w)) \\ &= \overline{Q}\exp\left(-\alpha\left(P^0 + \beta^{\mathrm{d}}L + \beta^{\mathrm{t}}(T - T^0) + \tau T + \kappa\frac{\gamma}{N - QT}\right)\right) \end{aligned} \tag{70}$$

With reference to the calibrated data of Hong Kong (Yang et al., 2002), we take $L = 20$ (km), $\tau = 35$ (HKD/h), $\kappa = 60$ (HKD/h), $\alpha = 0.03$ (1/HKD), $\gamma = 400.0$ (veh h). The potential customer demand is assumed to be $\overline{Q} = 1.0 \times 10^4$ (trip/h). To characterize the congestion effects, we employ a linear speed–density function: $v = a - bk$ where $k = (N^{\mathrm{n}} + N)/\overline{L} = Q^{\mathrm{n}}L/\overline{L}v + N/\overline{L}$ from Eq. (7) and $\overline{L}$ (km) is the total length of the roads in the network. The parameter values are assumed to be $a = 100$ (km/h), $b = 0.67$ (km$^2$/veh h), $\overline{L} = 500$ (km). In addition, we have $T = L/v$ and $T^0 = L/a$. With these inputs and considering only the normal flow regime, the average running speed of all vehicles (taxis and normal vehicles) on the network is given by

$$v = \frac{1}{2}\left(a - \frac{bN}{\overline{L}} + \sqrt{\left(a - \frac{bN}{\overline{L}}\right)^2 - 4\frac{bQ^{\mathrm{n}}L}{\overline{L}}}\right) \tag{71}$$

Our central concern in the numerical example is about the impact of traffic congestion on the market equilibrium. Figs. 2–8 depict the changes of the first-best solution, the *stable* competitive

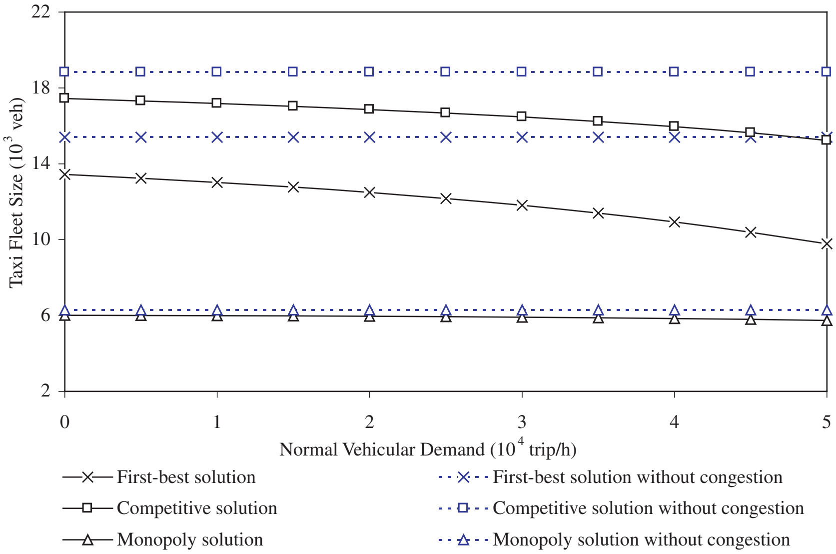


Fig. 2. Variation of hourly delay charge with respect to normal vehicle demand.

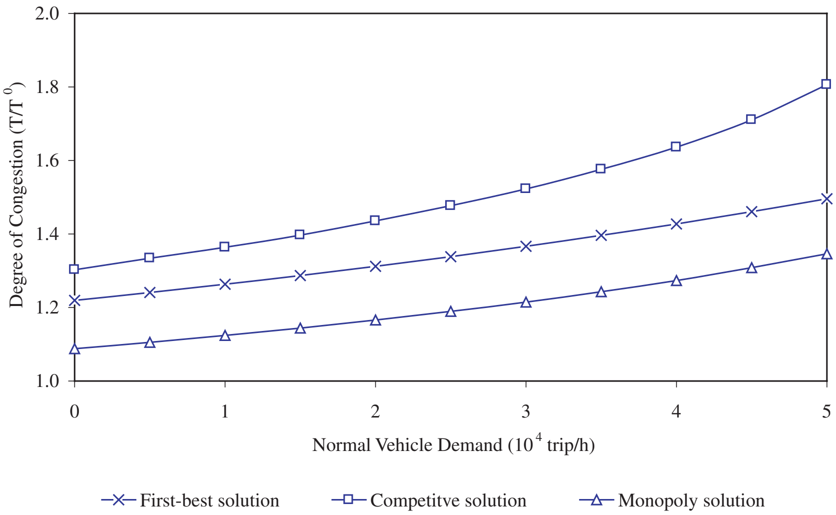


Fig. 3. Variation of the degree of congestion with respect to normal vehicle demand.

solution and the monopoly solution with respect to the level of normal traffic demand (level of congestion) respectively, in terms of taxi fleet size, hourly delay charge rate, total price charged per taxi ride, customer waiting time, customer demand, total taxi profit and social welfare.

As seen from Fig. 2, taxi fleet size for all three representative solutions declines in response to the level of congestion. In particular, the fleet size for the first-best solution declines more quickly, because the regulator, with a concern about the total community benefit including the normal

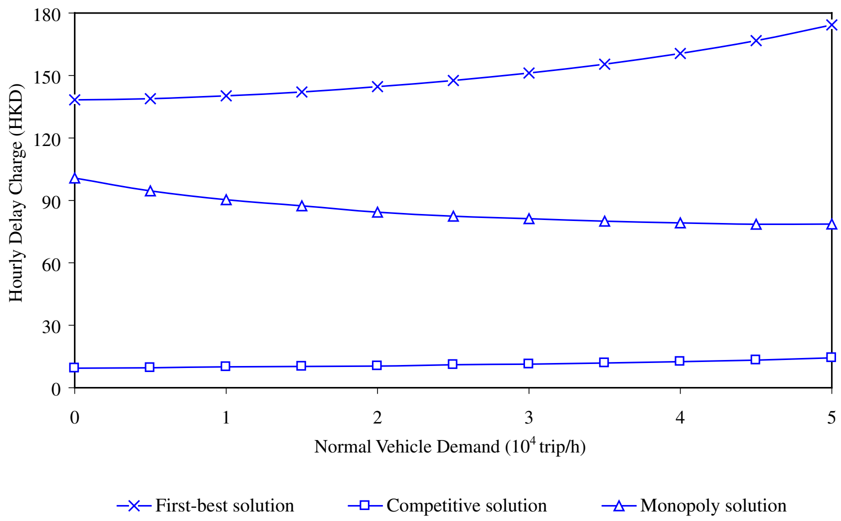


Fig. 4. Variation of total price charge per ride with respect to normal vehicle demand.

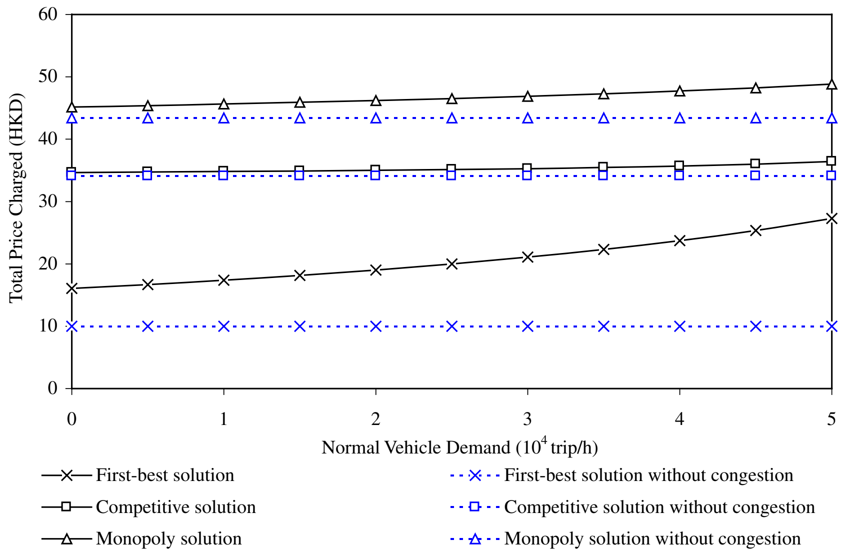


Fig. 5. Variation of taxi fleet size with respect to normal vehicle demand.

vehicle users, would like to reduce traffic congestion caused by both occupied and vacant taxi movements as well as normal vehicular traffic.

Fig. 3 plots the change of the degree of congestion with normal vehicle demand in terms of the ratio of the congested to the free-flow average taxi ride times. Because of the larger taxi fleet size in comparison with those associated with the first-best and monopoly solutions, the average ride time at the unregulated competitive stable equilibrium is longer, its sharper increase with normal

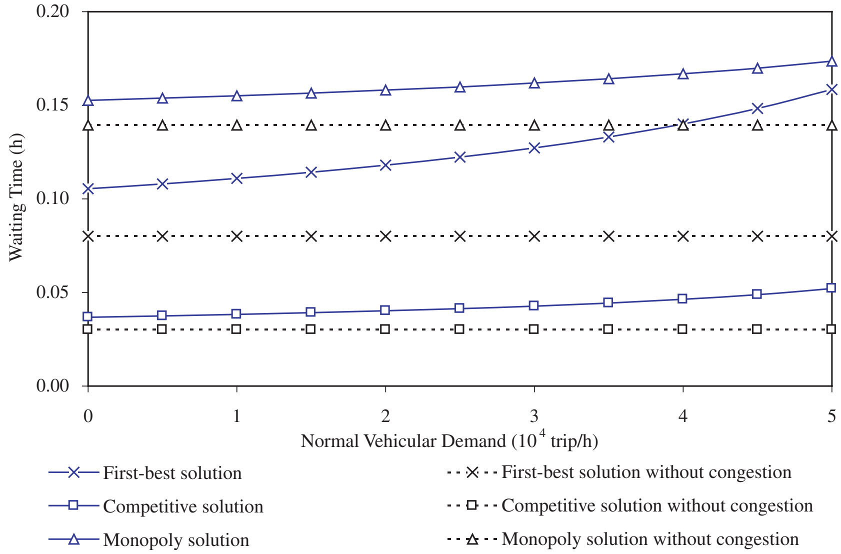


Fig. 6. Variation of customer waiting time with respect to normal vehicle demand.

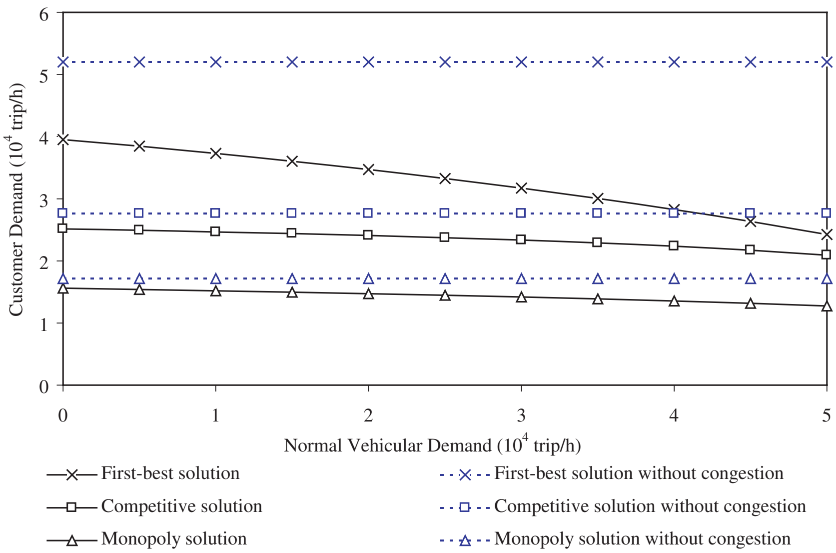


Fig. 7. Variation of customer demand with respect to normal vehicle demand.

vehicle demand is largely due to the fact that the marginal travel time is greater at high flows than at low flow levels.

Fig. 4 shows that the hourly delay charge rate increases gradually with the level of normal vehicle demand for the first-best and competitive solutions, but decreases for the monopoly solution. The two opposite changes reflect the different behaviors of the monopolist and the social regula-

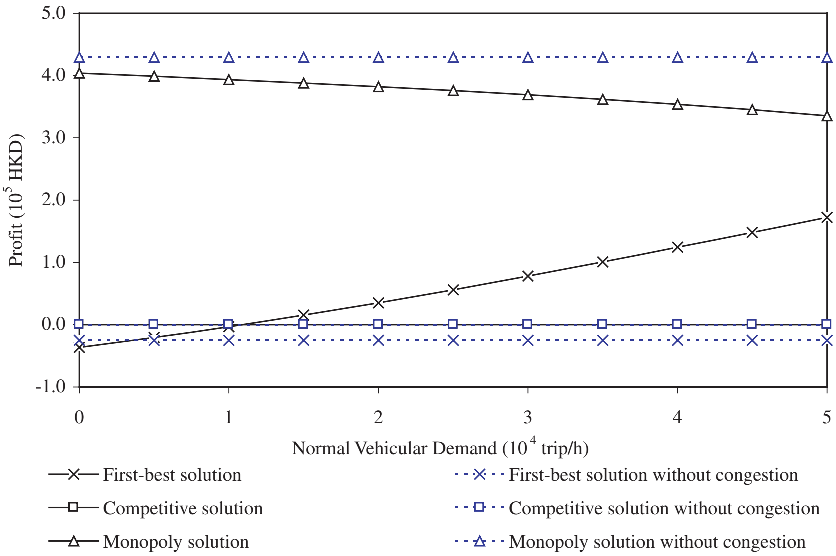


Fig. 8. Variation of social welfare with respect to normal vehicle demand.

tor. The single monopolist tries to attract customers by lowering the hourly delay charge due to congestion and he is able to seize the profit effect of such action in the whole market. Whereas the social regulator takes account of the benefit effect of his price decision on the whole society including the customers, taxi firms and the normal vehicle users. The total price charge per taxi ride for all three representation solutions, as displayed in Fig. 5, becomes higher in the presence of congestion externality and increases with the level of congestion. The sharper increase in the total price charge associated with the first-best solution comes out of the double effect of increased ride time and increased hourly delay charge rate. The increase in total monopoly price reflects the fact that the price effect of increased ride time outweighs that of reduced hourly delay charge rate.

Both the service quality in terms of customer waiting time shown in Fig. 6 and the customer demand shown in Fig. 7 decline, as traffic becomes more congested. A sharper decline of the customer demand under the first-best solution reflects the regulator's preference to diminish the demand for taxis by reducing taxi fleet size (Fig. 2) and increasing taxi fare (Figs. 4 and 5) with a hope of mitigating overall traffic congestion.

As shown in Fig. 8, the profits associated with the first-best and the monopoly solution exhibit an opposite trend of change with the level of normal vehicle demand (note that the competitive solutions with or without congestion are always associated with zero-profit as per its definition, and hence their corresponding (zero) profit curves coincide with each other in the figure). As found in previous studies, taxi services at the first-best social optimum are associated with a negative profit and should thus be subsidized. This is indeed true in the absence of congestion externality or when the congestion effect is mild. Nevertheless, as the level of normal vehicle demand increases, the total profit at the social optimum increases and eventually becomes positive. This is because, as proved earlier, the socially optimal price charged per taxi ride exceeds the corresponding marginal cost and goes up with increasing level of congestion. Note that, in spite of higher markup pricing over marginal cost in the monopoly market, as shown earlier, the total monopoly profit decreases with increasing level of congestion due to reduced customer demand. The social welfare, as depicted in Fig. 9, decreases with increasing level of congestion in all three cases. This

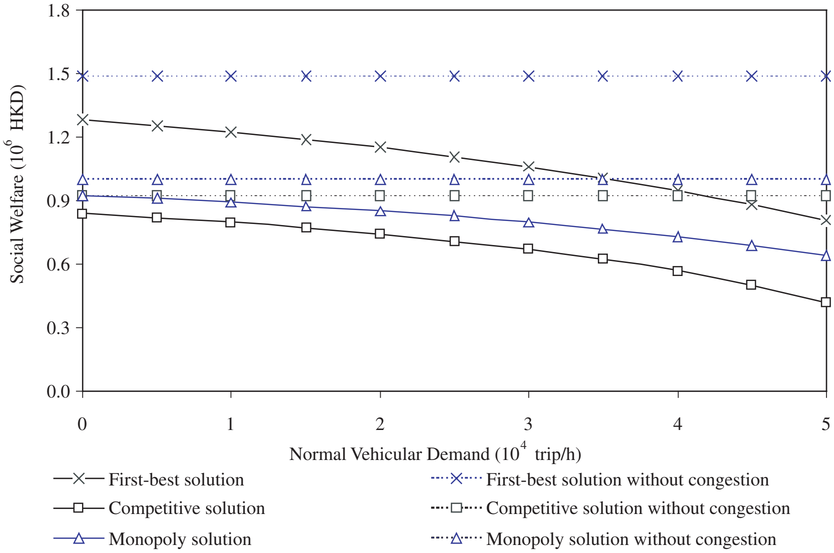


Fig. 9. Variations of taxi firm profit with respect to normal vehicle demand.

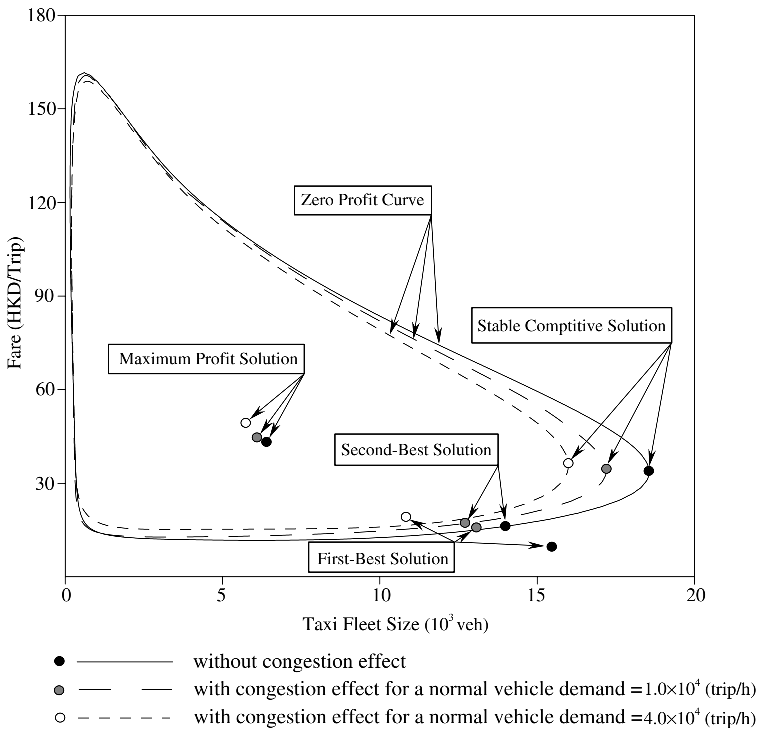


Fig. 10. Impact of traffic congestion on the zero-profit curve and representative solutions.

decrease can mainly be attributed to both the increased journey time of normal vehicle users and the reduced customer surplus (reduced customer demand for taxi services shown in Fig. 7). It is also observed that the monopoly yields a higher social surplus than the competitive market, with or without congestion. Without congestion, the difference in social welfare between the two markets is small. As congestion level (normal vehicle demand) increases, social welfare in the competitive market declines faster, because there are too many competing taxis in the market that make congestion even worse.

Finally, as we expected, Fig. 10 demonstrates that the positive profitable domain (the area within the closed zero-profit curve) shrinks in the presence of congestion effects. In particular, the stable competitive taxi fleet size shrinks most. Without including congestion externality, we actually overestimate the feasible size of the taxi market. The figure also points out that the first-best solution gives rise to a negative profit when there is no congestion or when there is light congestion associated with a normal vehicle demand of $1.0 \times 10^4$ (veh/h). In this case, a sustainable second-best solution can be identified along the zero-profit curve as shown in the figure. When the normal vehicle demand rises to $4.0 \times 10^4$ (veh/h), resulting in a high level of congestion, the first-best solution entails a positive profit which is made by a smaller taxi fleet size. In this case, the second-best solution becomes meaningless and thus does not appear in the figure.

## 7. Conclusions

We have examined the equilibrium mechanisms of the taxi market under various market settings with congestion externalities. Previous analytical results were generalized by adopting a realistic distanced-based and delay-based taxi fare structure and explicitly taking into account the negative impacts of traffic congestion on the taxi firms, the consumers and the whole society. Table 2 summarizes the major theoretical findings made in our study together with a comparison

Table 2
Comparison of the representative solutions with and without congestion externalities

| | Without congestion effects | With congestion effects |
|---|---|---|
| Monopoly solution | $\widetilde{P} - cT^0 = -\frac{\widetilde{Q}}{\widetilde{f}_1} = \frac{\partial}{\partial Q}\int_{\widetilde{\rho}}^{\infty} f(\omega)\,\mathrm{d}\omega$ (Cairns and Liston-Heyes, 1996) | $P - cT = \left(1 - (f_2 - f_3 w' Q)T\frac{\partial T}{\partial Q}\right)\int_{\rho}^{\infty} Q(\omega)\,\mathrm{d}\omega,\ P > \widetilde{P}$ |
| Social optimum | $\widetilde{P} - cT^0 = 0$ (Arnott, 1996) | $P - cT = \left(\tau(Q + Q^{\mathrm{n}}) - \kappa w' Q^2\right)T\frac{\partial T}{\partial N},\ P > \widetilde{P}$ |
| Stable competitive solution | $\widetilde{P}\widetilde{Q} - c\widetilde{N} = 0,\ \varepsilon_P = \frac{\widetilde{P}}{\widetilde{Q}}\frac{\partial Q}{\partial P} = -1.0$ | $PQ - cN = 0,\ \varepsilon_p = \frac{P}{Q}\frac{\partial Q}{\partial P} = -1.0,\ N < \widetilde{N}$ |
| Second-best solution | $\widetilde{P}\widetilde{Q} - c\widetilde{N} = 0$ | $PQ - cN = 0$ |
| | $\widetilde{P} - c\widetilde{T} = \frac{\tilde{\lambda}}{\tilde{\lambda}+1}\frac{\partial}{\partial Q}\int_{\widetilde{\rho}}^{\infty} f(\omega)\,\mathrm{d}\omega > 0$ | $P - cT = \frac{\lambda}{\lambda+1}\left(1 - (f_2 - f_3 w' Q)T\frac{\partial T}{\partial N}\right)\frac{\partial}{\partial Q}\int_{\rho}^{\infty} Q(\omega)\,\mathrm{d}\omega + \frac{1}{\lambda+1}\left(\tau(Q + Q^{\mathrm{n}}) - \kappa w' Q^2\right)T\frac{\partial T}{\partial N}$ |
| | $(\tilde{\lambda} > 0)$ | $= \frac{\lambda}{\lambda+1}(P - cT)_{\text{at monopoly}} + \frac{1}{\lambda+1}(P - cT)_{\text{at social optimum}}$ |
| | | for $\frac{\partial T}{\partial N} < -\frac{\kappa w' Q(N - QT)}{N(\tau(Q + Q^{\mathrm{n}}) - \kappa w' Q^2)}$ $(\lambda > 0)$ |

with existing results. In contrast with previous observations, the most interesting finding here has been that the loss to the taxi firm at the first-best solution is less than the cost of vacant taxi-hour. In fact, as proved theoretically and confirmed numerically the profit at the first-best solution can become positive when traffic congestion rises to a certain level.

## Acknowledgement

The authors wish to thank Richard Arnott of Boston College, Robin Lindsey of University of Alberta and three anonymous referees for their helpful comments on an earlier version of the paper. This research was substantially supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. HKUST6107/03E).

## References

Arnott, R., 1996. Taxi travel should be subsidized. Journal of Urban Economics 40, 316–333.

Arnott, R., Kraus, M., 1993. The Ramsey problem for congestible facilities. Journal of Public Economics 50, 371–396.

Cairns, R.D., Liston-Heyes, C., 1996. Competition and regulation in the taxi industry. Journal of Public Economics 59, 1–15.

De vany, A.S., 1975. Capacity utilization under alternative regulatory constraints: An analysis of taxi markets. Journal of Political Economy 83, 83–94.

Douglas, G.W., 1972. Price regulation and optimal service standards: The taxicab industry. Journal of Transport Economics and Policy 20, 116–127.

Lindsey, R., Verhoef, E.T., 2001. Traffic congestion and congestion pricing. In: Button, K.J., Hensher, D.A. (Eds.), Handbook of Transport Systems and Traffic Control. Elsevier Science, Oxford, pp. 77–105.

May, A.D., 1990. Traffic Flow Fundamentals. Prentice Hall, Englewood Cliffs, NJ.

Oum, T.H., Tretheway, M.W., 1988. Ramsey pricing in the presence of external costs. Journal of Transport Economics and Policy 22, 307–317.

Transport Department, 1986–2000. The Level of Taxi Services. TTSD Publication Series, Hong Kong SAR Government, Hong Kong.

Yang, H., Lau, Y.W., Wong, S.C., Lo, H.K., 1998. A macroscopic taxi model for passenger demand, taxi utilization and level of services. Transportation 27, 317–340.

Yang, H., Wong, S.C., Wong, K.I., 2002. Demand–supply equilibrium of taxi services in a network under competition and regulation. Transportation Research 36B, 799–819.

Yang, H., Ye, M., Tang, W.H., Wong, S.C., 2003. Modeling urban taxi services: a literature survey and an analytical example. In: Lam, W.H.K., Bell, M.G.H. (Eds.), Advanced Modeling for Transit Operations and Service Planning. Pergamon, pp. 257–286.